Georgia State University

ScholarWorks @ Georgia State University

Risk Management and Insurance Dissertations Department of Risk Management and Insurance

8-9-2022

Risk Analysis and Uncertainty Quantification in Insurance Ratemaking

Seul Ki Kang

Follow this and additional works at: https://scholarworks.gsu.edu/rmi_diss

Recommended Citation

Kang, Seul Ki, "Risk Analysis and Uncertainty Quantification in Insurance Ratemaking." Dissertation, Georgia State University, 2022.
doi: https://doi.org/10.57709/BD2G-V577

This Dissertation is brought to you for free and open access by the Department of Risk Management and Insurance at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Risk Management and Insurance Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

RISK ANALYSIS AND UNCERTAINTY QUANTIFICATION IN INSURANCE RATEMAKING

By

SEUL KI KANG

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree
of
Doctor of Philosophy
In the Robinson College of Business
of
Georgia State University

GEORGIA STATE UNIVERSITY
ROBINSON COLLEGE OF BUSINESS

2022

Copyright by
Seul ki Kang
2022

ACCEPTANCE

This dissertation was prepared under the direction of the *Seul ki Kang* Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Business Administration in the J. Mack Robinson College of Business of Georgia State University.

Richard Phillips, Dean

Dissertation Chair: Dr. Liang Peng

Committee: Dr. Ajay Subramanian
Dr. Stephen H. Shore
Dr. G. Peter Zhang

ABSTRACT

RISK ANALYSIS AND UNCERTAINTY QUANTIFICATION IN INSURANCE RATEMAKING

By

SEUL KI KANG

July 2022

Committee Chair: *Dr. Liang Peng*

Major Academic Unit: *Department of Risk Management and Insurance*

Insurance ratemaking, which is the process of setting an adequate amount of premium for an insured entity, is an essential role of insurance actuaries. For the success of this process, they need to perform a delicate and sound statistical analysis of insurance data, considering all the information it contains. Recently, several works of literature that explore the Value-at-Risk (VaR) for premium calculation have been reported, such as Heras, Moreno, and Vilar-Zanón (2018). Motivated by the importance of risk forecast in insurance ratemaking, this dissertation proposes diverse approaches to making inferences about risk measures and quantifying uncertainty. Specifically, I start by disputing the argument in Heras, Moreno, and Vilar-Zanón (2018) that their two-step inference method with quantile regression at the second stage with categorical variables can make a better forecast of VaR of aggregate losses than usual simple nonparametric estimates. By constructing a confidence interval using a novel empirical likelihood method, I provide sound evidence of my disputing argument. I further expand the risk analysis in more general settings to make an inference about VaR using both categorical and continuous explanatory variables and to quantify uncertainty using a random weighted bootstrap method. Lastly, I propose a three-step inference method for forecasting quantile risk measures, such as VaR and Expected Shortfall (ES), at a high-risk level. I adopt a Generalized Pareto Distribution (GPD) with a dynamic threshold for modeling excess losses and prove that I have made an efficient and robust risk forecast. Empirically, I use a well-known Australian automobile insurance dataset to illustrate the developed methods.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. Liang Peng, who made this journey possible. Without his persistent support and advice, I could not have gone through the process successfully. I also want to thank my committee members, Dr. Ajay Subramanian, Dr. Stephen H. Shore, and Dr. G. Peter Zhang, for sharing their invaluable comments and feedback which improved my thesis tremendously. Additionally, I would like to give special thanks to Department of Risk Management and Insurance as a whole for generous scholarships and fellowships.

Last but not least, I want to convey my heartiest appreciation to my husband and daughter, Jason and Christine, who show unstinting support and understanding for my work. The completion of this study could not have been possible without their dedication.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Insurance companies in current competitive markets are experiencing numerous challenges stemming from changes in regulatory environments, unstable market conditions, or transitions in distribution channels. In such complicated environments, there is a high demand for actuaries to make sound decisions about each individual policy in terms of premiums, underwriting, and more. For this purpose, it is mandatory to perform accurate statistical analyses of insurance data containing all detailed information about a policy and formulate a delicate actuarial evaluation of the risk considering all the provided information.

The insurance premium, which is an amount that the insured pays to the insurer for relocating their risk, carries out two important tasks for both the insured and the insurer. First, the premium needs to be fair regarding each individual's risk. The premium of policyholders with disparate risk profiles should be arranged differently considering the cost that they are expected to generate. In other words, the premium should be able to reflect the fact that cost is fairly and equitably allocated among insured risks so that insurance can act as an instrument of pooling and sharing risks of the insured. Second, the premium should be sufficient to cover the total costs and expenses, which are indicated as an insurer's liability in the insurance contract. To put it another way, the premium should be adequately collected to cover expected future costs, which can prevent the insurer from being insolvent. This is undoubtedly one of the crucial roles of the premium because being in a solvent condition at any given time is an essential obligation of the insurance company. Namely, the premium needs to be equitable to make individual agents willing to pay for hedging their risks, while it also should be charged with an adequate amount to guarantee that insurers avoid ruin with certainty and continue their business.

1.1 Premium Principle

In this section, I explore the general definition of the *premium calculation principle*, discuss the desired properties of a good premium principle, and describe some of the risk measures that are appropriate for a premium calculation principle.

The effective analysis to find an appropriate price for an insurance policy is a critical role of actuarial professionals. However, determining a fair price is not an easy task because both low and high premiums can result in unfavorable conditions for the insurer. For example, premiums cannot be too low because it can lead to insolvency conditions for the insurer, but they also cannot be too high because this can lead to a loss of customers to competitors. For that reason, finding out the best strategy to set a premium is an essential element of an insurer maintaining a healthy operational environment.

An actuarial premium principle is a mechanism that allocates an adequate price to each insurance policy, defined as a functional P that assigns a numerical value to $P(S)$, which is the premium that an insurer charges to cover a risk S. Generally, an insurance premium relies on certain statistical characteristics of the distribution of risk, S, such as the expectation of S, $\mathbb{E}(S)$, or the variance of S, Var(S). For example, one of the basic premium principles, the *Equivalence premium principle*, takes the expectation of S for setting a premium, that is, $P(S) = \mathbb{E}(S)$.

A premium calculation principle usually consists of a separate assessment of two components: *pure premium* and *risk loading*. In traditional analysis, the pure premium is often evaluated using a central estimate, which is generally assessed as the expected value of the outcome variables. However, such a central estimate often may not be statistically robust, especially with claim data with outliers or extreme losses, which may lead to failure to capture accurate claims liability.

The risk loading is introduced to overcome the limitation of the pure premium for evaluating the claims liability. It is originated to promote the coverage of the premium for an excess loss over the pure premium, which results from the adverse deviation of the claims. By adding this to the pure premium, it is more likely that the claim liability can be accurately evaluated so that the premium

can provide a more sufficient capital level that is required by regulations to prevent the insurer's insolvency. According to this nature of the risk loading, it is expected that a portfolio with a heavier tail or more volatile features tends to have a higher level of the risk loading because of the higher probability of having a larger variation in reserves. Also, it is anticipated to be positive unless the distribution of S is concentrated at a single point, or else insurers could bump into ruin with certainty even if they hold very large initial reserves.

The following are the traditional premium principles, which are frequently considered in actuarial practice.

- *Expected Value Principle:* For some $\alpha \geq 0$,

$$P(S) = (1 + \alpha)\mathbb{E}(S). \tag{1.1}$$

- *Variance Principle:* For some $\alpha \geq 0$,

$$P(S) = \mathbb{E}(S) + \alpha \text{Var}(S). \tag{1.2}$$

- *Standard Deviation Principle:* For some $\alpha \geq 0$,

$$P(S) = \mathbb{E}(S) + \alpha\sqrt{\text{Var}(S)}. \tag{1.3}$$

The *Expected Value Principle* seems sound and reasonable, but it often receives criticism because it does not consider the variability of the underlying risk S, and this may put insurers in danger. To overcome this limitation, other classes of premium principles are introduced, such as *Variance Principle* or *Standard Deviation Principle*, where the risk loadings depend on the variability of S. The principles given above have several good properties as a "good" premium principle. For example, the variance principle is additive for independent risks, and it is consistent. The standard deviation principle is also consistent and subadditive, provided that covariances of risks are zero.

However, the aforementioned principles also violate some of the desired properties of premium calculation principles, such as monotonicity with respect to stochastic ordering. For more detailed information about a premium calculation principle, I refer to Rolski et al. (2009).

As expected, a pertinent evaluation of claim liabilities is depicted as one of the most concerning tasks of an insurance company. Although a wide variety of actuarial literature has reported the prediction method for the pure premium (Smyth and Jørgensen (2002), Henckaerts, Antonio, and Clijsters (2018)), not many studies have focused on the statistical analysis of the risk loading. The importance of reasonable evaluation of such risk loading has been discussed by professionals, both in practical and academical fields. For example, the Solvency II Directive of the European Parliament Article 77(3) (see EU Commission (2009)), *"Calculation of technical provisions"*, provides the following prescriptive characteristics of the risk loading:

"Where insurance and reinsurance undertakings value the best estimate and the risk loading separately, the risk loading shall be calculated by determining the cost of providing an amount of eligible own funds equal to the Solvency Capital Requirement necessary to support the insurance and reinsurance obligations over the lifetime thereof."

In practice, the risk loading should be assessed distinctly from the pure premium using a specified representative methodology, and the methodology that Solvency II chooses is called the *cost-of-capital method*. However, because the actuary is responsible for deciding the risk loading with some reasonable assumptions about the underlying risk based on the nature of the portfolio or business environment, there is no imperative method for determining the risk loading. The following methods are some of the alternative methods for modeling risk loadings suggested by IASB (2007) and Pelkiewicz et al. (2020):

- Value-at-Risk (VaR).
- Expected shortfall (ES).

- Provision for Adverse Deviation.
- Run-off Capitalisation.

Note that the decision process of premium calculation principle $P(S)$ should consider the important fact that the premium requires to be sufficiently collected to compete with incoming losses, and the portfolio should be in the solvent condition at any given time. For example, International Financial Report Standard No.17 (see IASB (2017)) specifies that an insurer should disclose the confidence level of the techniques that are used for the premium calculation. Considering this concern, another premium calculation principle employing an interesting class of measures is introduced, which is called the *quantile premium principle*. Under this premium principle, $P(S)$ can be expressed using the quantile measures of the risk S.

The most commonly applied measure for the quantile premium principle in an insurance sector is the Value at Risk (VaR) measure, where the premium P is defined as follows:

$$P(S) = \text{VaR}_\alpha(S). \tag{1.4}$$

For example, Article 101 of the Solvency II directive (see EU Commission (2009)), which is a document presenting the *"Calculation of the Solvency Capital Requirement",* makes the following statement:

" The Solvency Capital Requirement(SCR) is calibrated to the Value-at-Risk of the basic own funds of an insurance or reinsurance undertaking subject to a confidence level of approximately 99.5% over a one year time horizon. "

Here, the "basic own funds" is the net assets on an insurer's balance sheet. From this directive, it seems sensible to write $P(S) = \text{VaR}_{99.5\%}(S)$.

In this dissertation, I am interested in the family of approaches, that are related to the quantile risk measures, *Value-at-Risk (VaR)* and *Expected shortfall (ES)* for the premium calculation principles

and risk loadings. Assessing risk loadings based on VaR is one of the commonly used methods demonstrating uncertainty for adding the excess values to the expected losses, which controls for the probability that the actual claims will be less than a certain threshold amount that meets the target confidence level. ES is an alternative tail risk measure that can be calculated by averaging all losses in a distribution that are above VaR at a specified risk level.

The use of risk measures such as VaR or ES for the risk loadings has several advantages over the traditional measures, such as standard deviations or variance. If the standard deviation or variance is considered as a risk measure, the number of standard deviations or variance may decrease to obtain a particular confidence level with a more skewed distribution compared with normal distribution. However, using VaR or ES for risk measures requires a larger number of risk loadings than it would with traditional measures, which makes those methods more risk sensitive and robust for dealing with risks with skewness at the extreme end. See more detail about this in IAA (2009).

Recently, many pieces of the research literature have explored quantile measures for the premium principles. For example, Kudryavtsev (2009) first proposes a model framework, which applies the quantile measures for the insurance ratemaking. In particular, the quantile regression method is adopted for individual ratemaking where a set of risk factors are considered. Heras, Moreno, and Vilar-Zanón (2018) refine the risk forecast from Kudryavtsev (2009) and propose a two-stage quantile regression model, that considers the different risk levels for the different policies for estimation of the quantile of losses. Baione and Biancalana (2019) approach the two-stage quantile regression of Heras, Moreno, and Vilar-Zanón (2018) in a different way, where the overparametrization problem of Heras, Moreno, and Vilar-Zanón (2018) is mitigated. Baione and Biancalana (2021) extend the framework introduced by Heras, Moreno, and Vilar-Zanón (2018) to provide a parsimonious model at the second stage to improve computational effectiveness by adopting the parametric model of quantile regression in Frumento and Bottai (2016).

The basic model framework of the aforementioned literature is as follows. Suppose the actuarial dataset $\{\mathbf{X}_i, N_i, \{L_{i,j}\}_{j=1}^{N_i}\}_{i=1}^{n}$ in a given year is observed, where $\mathbf{X}_i$ is the characteristic vector representing the ith policyholder, $N_i \geq 0$ is the number of claims, and $\{L_{i,j} \geq 0\}_{j=1}^{N_i}$ are the

corresponding losses. The aggregate loss is defined as $S_i = \sum_{j=1}^{N_i} L_{i,j}$. The goal is to forecast the risk of the aggregate loss S of a future policyholder with characteristic vector $\mathbf{X}$ by calculating the value of the VaR of the α risk level of S, $\text{VaR}_S(\alpha|\mathbf{X}) = \inf\{s : P(S \leq s|\mathbf{X}) \geq \alpha\}$. It is clear from the conditional probability theorem that

$$P(S_i \leq s|\mathbf{X}_i) = P(N_i = 0|\mathbf{X}_i) + P(N_i > 0, S_i \leq s|\mathbf{X}_i)$$

and

$$P(N_i > 0, S_i \leq s|\mathbf{X}_i) = P(N_i > 0|\mathbf{X}_i)P(S_i \leq s|N_i > 0, \mathbf{X}_i).$$

Thus,

$$\begin{aligned} P(S_i \leq s|\mathbf{X}_i) &= P(N_i = 0|\mathbf{X}_i) + P(N_i > 0|\mathbf{X}_i)P(S_i \leq s|N_i > 0, \mathbf{X}_i) \\ &= P(N_i = 0|\mathbf{X}_i) + P(N_i > 0|\mathbf{X}_i)P(\tilde{S}_i \leq s|\mathbf{X}_i), \end{aligned} \tag{1.5}$$

where $\tilde{S}_i$ represents the conditional loss of S_i given $N_i > 0$.

Let $p_i = P(N_i = 0, |\mathbf{X}_i)$ be the probability that policy i makes zero claims. Define the adjusted probability level for VaR, α_i^*, as follows by:

$$\alpha_i^* = \frac{\alpha - p_i}{1 - p_i}.$$

Then it is clear that:

$$P(S_i \geq s|\mathbf{X}_i) = (1 - p_i)P(\tilde{S}_i \geq s|\mathbf{X}_i).$$

In other terms, the VaR of S_i at level α is equivalent to that of the conditional loss S_i given $N_i > 0$ at the adjusted level $\alpha_i^* = \frac{\alpha - p_i}{1 - p_i}$. That is:

$$\text{VaR}_{S_i}(\alpha_i|\mathbf{X}_i) = \text{VaR}_{\tilde{S}_i}(\alpha_i^*|\mathbf{X}_i). \tag{1.6}$$

Using this model framework, Kudryavtsev (2009) estimates the probability of having positive losses p for all portfolios and computes the adjusted probability $\alpha_i^* = \alpha^*$ for all i, which is the confidence level of the quantile of the positive losses. For example, he sets the quantile probability α as $\alpha = 0.95$ and estimates that the proportion of claims in the insurance portfolio of their data is 20%, i.e., the estimated probability of having positive losses $\hat{p}$ as 0.2. Then his estimated modified quantile probability level α^* becomes

$$\alpha^* = \frac{\alpha - \hat{p}}{1 - \hat{p}} = \frac{0.95 - 0.8}{0.2}$$

for all i.

Following the conventional formulation for the aggregate claims of policies (e.g., De Jong and Heller (2008), Frees (2010)), he defines the random variable $\tilde{\tilde{S}}_i = \log(\tilde{S}_i)$, which is the log of the aggregate claim amount conditioned on the policy having at least one claim. Then, by running quantile regression, he recovers the VaR of the original response S_i by means of:

$$\text{VaR}_{\tilde{\tilde{S}}_i}(\alpha^*|\mathbf{x}_i) = \exp(\mathbf{x}_i^\tau \gamma_{\alpha^*}) = \gamma_{0,\alpha^*} + \gamma_{1,\alpha^*}\mathbf{x}_{i1} + \cdots \gamma_{m,\alpha^*}\mathbf{x}_{im}, \tag{1.7}$$

where γ_{α^*} are the coefficients of the quantile regression, which are determined depending on the selected risk level α^*. In practice, this regression model is often estimated using the nonparametric approach (see Koenker and Bassett Jr (1978), Koenker and Hallock (2001), Koenker (2005)), where the estimator $\hat{\gamma}_{\alpha^*}$ can be attained by solving the following minimization problem:

$$\min_{\gamma_{\alpha^*}} \Big\{ \sum_{i:\tilde{\tilde{S}} \geq \mathbf{x}_i^\tau \gamma_{\alpha^*}} \alpha^* |\tilde{\tilde{S}}_i - \mathbf{x}_i^\tau \gamma_{\alpha^*}| + \sum_{i:\tilde{\tilde{S}} < \mathbf{x}_i^\tau \gamma_{\alpha^*}} (1 - \alpha^*) |\tilde{\tilde{S}} - \mathbf{x}_i^\tau \gamma_{\alpha^*}| \Big\}. \tag{1.8}$$

Heras, Moreno, and Vilar-Zanón (2018) consider an improvement of the model in Kudryavtsev (2009) by proposing to use the different estimated probability of having positive claims for the estimation of the risk levels that are employed in the quantile regression at the second stage. Because the estimation of the severity of the aggregate losses largely depends on the probability of each

insured having positive claims, and each insurance risk profile has heterogenous probabilities of having positive losses, adopting a different probability for the quantile estimation at the second stage is quite significant for the model accuracy. Precisely, their model consists of two steps, where the first step focuses on the estimation of the probability p_i of policy having zero losses. The usual logistic regression model

$$\text{logit}(1 - F_{N_i}(0|\mathbf{x}_i)) = \mathbf{x}_i^\tau \beta \tag{1.9}$$

is applied for this estimation where $\text{logit}(1 - F_{N_i}(0|\mathbf{x}_i)) = \ln((1 - p_i)/p_i)$. Using this estimated probability $\hat{p}_i$ from the first step, the modified quantile probability level can be computed as:

$$\hat{\alpha}_i^* = \frac{\alpha - \hat{p}_i}{1 - \hat{p}_i}.$$

Note that this adjusted probability level is variant depending on the policy's risk characteristics.

In the second step, they employ the quantile regression technique for each risk profile with an adjusted probability level $\hat{\alpha}_i^*$. In other words, they run quantile regression and estimate the quantile of the log of the positive losses using the following formula with the parameter of the quantile regression:

$$\text{VaR}_{\widetilde{\widetilde{S}}_i}(\alpha_i^*|\mathbf{x}_i) = \exp(\mathbf{x}_i^\tau \gamma_{\alpha_i^*}). \tag{1.10}$$

Baione and Biancalana (2021) refine the model of Heras, Moreno, and Vilar-Zanón (2018) and introduce a novel approach to overcome the overparameterization problem of the previous model. Note that the parameters of (1.10) are variant depending on the adjusted risk level $\hat{\alpha}_i^*$. For instance, the coefficients of the quantile regression of the risk profile whose adjusted risk level $\hat{\alpha}_i^* = 0.8$ are generally different from those with adjusted risk level $\hat{\alpha}_i^* = 0.9$. In other words, it is demanded to run the quantile regression multiple times to obtain the quantile of the outcome variable for all considered risk profiles, which requires quite expensive computational time, especially for insurance data with a substantial number of different risk characteristics.

To resolve this limitation in the model of Heras, Moreno, and Vilar-Zanón (2018), Baione and Biancalana (2021) calibrate a parametric model of the quantile regression. In particular, they

consider the parametric regression quantile model of Frumento and Bottai (2016), where the regression coefficients are considered as the parametric functions of the quantile level α. In the model of Baione and Biancalana (2021), the coefficients of the quantile regression take the following functional form:

$$\gamma_j(\alpha^*, \theta_j) = \theta_{j0} + \theta_{j1} \cdot b_1(\alpha^*) + \cdots + \theta_{jl} \cdot b_l(\alpha^*), \quad 0 \leq j \leq m,$$

where $b_1(\alpha^*) \cdots b_l(\alpha^*)$ are known functions of α^* and m represents the number of different levels of risk profiles. Then, the conditional quantile of the aggregated losses can be represented as follows:

$$\text{VaR}_{\widetilde{\approx}_{S_i}}(\alpha_i^*|x_i) = \gamma_0(\alpha^*, \theta_0) + \gamma_1(\alpha^*, \theta_1) \cdot x_{i1} + \cdots + \gamma_m(\alpha^*, \theta_m) \cdot x_{im}. \tag{1.11}$$

In other words, the entire conditional quantile of the aggregate losses can be recovered by obtaining a m $\times$ n matrix of parameters, instead of solving the quantile regression multiple times.

In the following subsection, I will introduce two widely employed quantile risk measures, VaR and ES, and examine some strengths and weaknesses of these two measures. Furthermore, I will describe a widely used statistical model to estimate a relationship between quantile risk measures and predictors, which I refer to as the *quantile regression model* in the subsequent subsection.

1.1.1 Quantile Risk Measures

1.1.1.1 Value-at-Risk(VaR)

Let S be a loss at a certain fixed time horizon Δt, and denote $F_S(s) = P(S \leq s)$ to be the corresponding density function of loss S. The Value-at-Risk(VaR) is a statistic based on the distribution F_S that measures the "maximum loss that is not exceeded with a given high probability over the time period Δt."

Definition 1.1 (VaR)**.** Given some confidence level $\alpha \in (0, 1)$, the VaR of a risk S at the confidence level α is given by the smallest number s such that the probability that the risk S exceeds s is no

larger than $1-\alpha$. Formally,

$$\text{VaR}_\alpha(S) = \inf\{s \in \mathbb{R} : P(S > s) \leq 1-\alpha\} = \inf\{l \in \mathbb{R} : F_S(s) \geq \alpha\}. \tag{1.12}$$

In simplistic probability terminology, VaR can be rephrased as a quantile of the loss distribution. VaR is probably the most widely used risk measure by corporate treasurers and fund managers as well as by financial institutions because of the simplicity of understanding its intuition. It is the measure that is traditionally used for assessing concerns about the capital requirements for risk existing in the market by regulators. For instance, VaR deviation is utilized for measuring the width of daily loss distribution of a portfolio under financial regulations, such as Basel I and Basel II. Although VaR has several good properties, such as easiness of understanding and intuitiveness, it also has some shortcomings, such as numerical difficulties of optimization (Rockafellar and Uryasev (2000) and Rockafellar (2007)).

To assess VaR, two parameters are needed to be chosen: Δt and α. There is no definite choice for the parameters but there are some aspects that actuarial professionals need to consider in arranging these values.

The confidence level α can be arranged differently depending on the purpose of use of the risk measure. For example, a bank needs to hold capital for market risk in the trading book under the Basel risk management system, which requires to estimate 99% VaR during a 10-day horizon. Under Solvency II, an insurer needs to assess VaR at the 99.5% level for their requirement of solvency capital for a one-year period. Compared with such a capital requirement, for the purpose of the backtesting of VaR models, the lower confidence levels at a shorter period may be used to maintain sufficient statistical power.

The time horizon Δt should represent the time duration of the holding portfolio of financial institutions when the risk measure is used for their risk management purposes, which will be influenced by contractual or legal requirements. If the risk measure is used for their operational risk management purpose, the time horizon should be chosen to fit the market convention. For

example, because the liability time horizon of the insurance product is generally one year, insurance companies typically measure the VaR with a time horizon of one year to represent their risk of liability and asset portfolios.

1.1.1.2 Expected shortfall(ES)

Expected shortfall (ES) is an alternative to VaR for managing risk. It is a coherent risk measure introduced to evaluate the likelihood of loss exceeding the VaR.

Definition 1.2 (ES)**.** For a loss S with $E(|S|) < \infty$ and df F_S, the ES at confidence level $\alpha \in (0, 1)$ is defined as

$$ES_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 q_u(F_S) du, \tag{1.13}$$

where $q_u(F_S) = F_S^{\leftarrow}(u)$ is the quantile function of F_S.

The condition $E(|S|) < \infty$ ensures that the integral is well defined. By definition, ES can be reformulated using VaR as

$$\text{ES}_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_u(S) du. \tag{1.14}$$

From this formula, it can be understood that ES_α can be computed by averaging all the losses, that are worse than the VaR at a given risk level α. Because ES considers the entire distribution of loss above VaR, it is more responsive to the entire tail distribution of the loss. For more detailed information about VaR and ES, see McNeil, Frey, and Embrechts (2015).

1.1.2 Quantile Regression

Let $(S_i, X_i), i = 1, \ldots n$ be the sample of size n, where S_i is a variable of interest in the regression equation, and $X_i = (X_{i1}, \ldots, X_{im})$ are explanatory variables (covariates). Then, the linear quantile model of the given observations is defined as follows:

$$Q_\alpha(S_i|X_i) = X_i^\tau \beta_\alpha, \tag{1.15}$$

where $Q_\alpha(S_i|X_i)$ represents the conditional quantile function for probability α of dependent variable S_i with respect to independent variables X_i. Quantile regression problem aims to find the regression coefficients, β_α, of the corresponding linear quantile model.

Quantile regression problem can be often estimated by the distribution free approach in practice (Koenker and Bassett Jr (1978), Koenker and Hallock (2001), Koenker (2005), Hao and Naiman (2007)). Given a sufficiently large number of observations $(S_i, X_i), i = 1, \ldots, n$, the estimator $\hat{\beta}_\alpha$ of coefficient vector β_α of the quantile regression problem could be attained by solving the minimization problem:

$$\min_{\beta_\alpha} \frac{1}{n} \Big\{ \sum_{i:S_i \geq X_i^\tau \beta_\alpha} \alpha |S_i - X_i^\tau \beta_\alpha| + \sum_{i:S_i < X_i^\tau \beta_\alpha} (1-\alpha) |S_i - X_i^\tau \beta_\alpha| \Big\}. \tag{1.16}$$

Quantile regression problem may also be reformulated as a linear program:

$$\min_{(\beta_\alpha, u, v) \in \mathbb{R}^p \times \mathbb{R}_+^{2n}} \{\alpha 1_n^\tau u + (1-\alpha) 1_n^\tau \nu | X^\tau \beta_\alpha + u - \nu = S\}.$$

Here a linear function is minimized with polyhedral constraint sets, where the absolute value of positive and negative parts of the residual vector $S - X^\tau \beta_\alpha$, u and v, reside. The reformulation of quantile regression as a linear program leads to some critical interpretation because most of the essential properties of the solutions $\hat{\beta}_\alpha$ promptly follow conventional properties of solutions of linear programs. First of all, the solution can be obtained by a finite number of iterations. Also, it can easily be shown that the estimation process is robust, which means that the absolute value of residual $|S - X^\tau \beta_\alpha|$ can be extremely large without ruining the estimates of β_α. I refer to Koenker (2005) and Kudryavtsev (2009) for more detailed information about quantile regression.

1.2 Uncertainty Quantification

The inference for the aforementioned two-stage quantile regression models is expected to be complicated in nature because it is a multi-level problem where the estimate of the first step is

included in the second stage. Performing the traditional method for the asymptotic analysis of quantile regression may result in an unreliable outcome because of the uncertainty coming from the first stage. For that reason, the previous literature merely focused on developing models for the prediction and accurate discussion of the asymptotic properties and uncertainty quantification of those models, which are essential cornerstones to investigate the accuracy of those proposed models, have been negligent.

To overcome this difficulty of quantifying the uncertainty of multistep inference of quantile measures, I introduce advanced statistical tools, such as *Empirical likelihood* and *Random weighted bootstrap method*, which allow me to perform the asymptotic analysis and construct confidence intervals of the proposed inferences.

1.2.1 Empirical Likelihood

Empirical likelihood (EL) is a nonparametric method of making inferences about data that does not require strong distributional assumptions while performing the data analysis in a likelihood manner. It was first pioneered by Owen (1988, 1990) and extensively studied and employed in the literature due to its great features, such as efficiency and universality. The generality of nonparametric methods other than parametric methods may be counterbalanced by their cost of reduced power, but EL tests have strong power properties.

Compared to other nonparametric methods, such as the bootstrap method, EL's main advantages are originated from the characteristics of a likelihood function. EL provides data-driven confidence intervals where ancillary information such as parameter constraints or estimating equations can be easily incorporated. Unlike the bootstrap method, EL can improve the coverage accuracy for inference by Barlett correction. Moreover, it makes it easier to combine information from different sources, as the usual parametric likelihood method does. I summarize some of the details about EL in Owen (2001) in this subsection.

I first start by defining the empirical distribution function and nonparametric likelihood.

Definition 1.3. Let $X_1, \ldots, X_n \in \mathbb{R}^d$. The empirical distribution function (EDF) of $X_1 \ldots, X_n$ is

$$F_n = \frac{1}{n} \sum_{i=1}^{n} 1_{X_i}, \tag{1.17}$$

where 1_x denote the distribution under which $X = x$ with probability 1.

Definition 1.4. Given $X_1, \ldots, X_n \in \mathbb{R}^d$, assumed independent with common density function F_0, the nonparametric likelihood of the density function F is

$$L(F) = \prod_{i=1}^{n} (F(\{X_i\}), \tag{1.18}$$

where $F(\{X_i\})$ is the probability of getting the value X_i in a sample from F.

Note that $L(F)$ denotes the probability of obtaining the given sample points $X_1 \ldots X_n$ from the density function F. To obtain a positive nonparametric likelihood, $F(\{X_i\})$ should be a positive value for every observation $X_i, i = 1, \ldots n$.

Now, the following theorem proves that the EDF is the nonparametric maximum likelihood estimate (NPMLE) of F, which means that the nonparametric likelihood attains its maximum at EDF.

Theorem 1.1. *Let $X_1 \ldots, X_n \in \mathbb{R}^d$ be independent random variables with a common density function F_0. Let F_n be their EDF and let F be any density function. If $F \neq F_n$, then $L(F) < L(F_n)$.*

Proof. The proof of this theorem can be found in the proof of Theorem 2.1 in Owen (2001).

In parametric inference, the likelihood ratio can be utilized for constructing confidence intervals or performing hypothesis tests. Let $L(\theta)$ be the likelihood function of the parameter θ. Wilk's theorem proves that $-2\log(L(\theta_0)/L(\hat{\theta}))$ converges to a χ^2-distribution as $n \to \infty$ under mild regularity conditions, where $\hat{\theta}$ is the MLE of θ. Under this fact, one is able to decide the scope of $L(\theta)$ that θ is rejected and the confidence region for θ_0. Moreover, one may reject the hypothesis that $\theta_0 = \theta$ if θ is outside of the constructed confidence interval. In other words, a confidence

interval for θ can be formulated as

$$\{\theta | L(\theta) \geq cL(\hat{\theta})\}.$$

Similarly, the nonparametric likelihood can be used as a likelihood for executing hypothesis tests and constructing confidence intervals in a nonparametric manner. Define the nonparametric likelihood ratio of distribution F as follows:

$$R(F) = \frac{L(F)}{L(F_n)}. \tag{1.19}$$

Note that $L(F_n)$ can be considered as a nonparametric version of the likelihood function of the MLE in parametric inference.

Suppose we try to make inferences about a parameter $\theta = W(F)$ for some functional value W of distribution F, and the true unknown parameter of the common density function F_0 is $\theta_0 = W(F_0)$. Now define the profile likelihood ratio function as

$$\mathcal{R}(\theta) = \sup\{R(F) | W(F) = \theta, F \in \mathcal{F}\}, \tag{1.20}$$

where $\mathcal{F}$ is the set of all distributions on $\mathbb{R}$. Then, the null hypothesis $H_0 : W(F_0) = \theta_0$ is rejected when $\mathcal{R}(\theta_0) < r_0$ for a certain threshold value r_0 in EL setting. Similarly, the EL confidence region can be formulated as follows:

$$\{\theta | \mathcal{R}(\theta) \geq r_0\}.$$

Consider the case that $X_i \neq X_j$ if $i \neq j$, i.e., data has no ties. If $p_i \geq 0$, $\sum_{i=1}^n p_i \leq 1$ is denoted to be the probability that the distribution F places on the observation X_i, then $L(F) = \prod_{i=1}^n p_i$ and so

$$R(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^{n} np_i. \tag{1.21}$$

When there are some ties, p_j could be the the probability where the distinct values y_j arises $n_j \geq 1$ times in the observation. Then, one may find specific weights $w_i \geq 0, i = 1, \ldots, n$ for each

observation, where the weights can be selected to satisfy that sum of w_i over all i with $X_i = y_j$ is p_j. In other words, F can be reproduced with a distribution that assigns weights w_i on observation X_i, and also any functional values of $W(F)$ of F. Then, the likelihood of distribution F can be established with respect to these weights w_i as $L(F) = \prod_{i=1}^{n} w_i$, and so

$$R(F) = \prod_{i=1}^{n} nw_i. \tag{1.22}$$

1.2.1.1 EL for estimating equations

In many circumstances, it may be required to obtain an inference about parameters satisfying certain conditions. Estimating equations present an easy and flexible way to perform such an inference and obtain related statistics. Suppose we are interested in the parameter $\theta \in \mathbb{R}^p$ with a given random variable $X \in \mathbb{R}^d$, which satisfy the following condition,

$$E(m(X, \theta)) = 0, \tag{1.23}$$

where $m(X, \theta) \in \mathbb{R}^q$ is a vector-valued function of θ and X. Under certain conditions of $m(X, \theta)$ and F, and the case when $p = q$, there is a unique solution θ for the above equation, and this true value θ_0 may be estimated by the estimator $\hat{\theta}$, which satisfies

$$\frac{1}{n} \sum_{i=1}^{n} m(X_i, \hat{\theta}) = 0. \tag{1.24}$$

This equation is called the estimating equation and the $m(X, \theta)$ is refered to as the estimating function.

Now consider the EL ratio function for θ satisfying the estimating equation as follows:

$$\mathcal{R}(\theta) = \max\Big\{ \prod_{i=1}^{n} nw_i | \sum_{i=1}^{n} w_i m(X_i, \theta) = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \Big\}. \tag{1.25}$$

If it is able to be shown that there is a uniquely defined θ_0 for the estimating equation and a consistent estimator $\hat{\theta}$ for θ, corresponding confidence intervals can be created, and hypothesis tests can be performed through the below theorem.

Theorem 1.2. *Let $X_1, \dots X_n \in \mathbb{R}^d$ be independent random vectors with common distribution F_0. For $\theta \in \Theta \subset \mathbb{R}^p$, and $X \in \mathbb{R}^d$, let $m(X, \theta) \in \mathbb{R}^q$. Let $\theta_0 \in \Theta$ be such that* $\text{Var}(m(X_i, \theta_0))$ *is finite and has rank $s > 0$. If θ_0 satisfies $E(m(X, \theta_0)) = 0$, then $-2\log\mathcal{R}(\theta_0) \to \chi^2_{(s)}$ in distribution as $n \to \infty$.*

Proof. The proof of this theorem can be found in proof of Theorem 3.4. in Owen (2001).

The use of EL for estimating equations can be broadened widely by presenting the nuisance parameters. The nuisance parameters can be incorporated in the following manner in the estimating equations: let $\theta \in \mathbb{R}^p$ be parameters that are in the interest and $\nu \in \mathbb{R}^q$ be nuisance parameters. Furthermore, there is an estimating function $m(X, \theta, \nu) \in \mathbb{R}^s$ that satisfies the estimating equation $E(m(X, \theta, \nu)) = 0$. Thus, the EL ratio can be defined as,

$$\mathcal{R}(\theta, \nu) = \max\Big\{ \prod_{i=1}^{n} nw_i | \sum_{i=1}^{n} w_i m(X_i, \theta, \nu) = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \Big\}, \tag{1.26}$$

and the profile EL ratio function, which maximizes the EL ratio over the nuisance parameters, as,

$$\mathcal{R}(\theta) = \max_{\nu} \mathcal{R}(\theta, \nu). \tag{1.27}$$

If there exists any value ν which results in a large EL ratio $\mathcal{R}(\theta, \nu)$, then it can be said that θ is in the confidence region. Also, it can be shown that under mild conditions.

$$-2\log\mathcal{R}(\theta) \to \chi^2_{(p)}. \tag{1.28}$$

1.2.1.2 EL for quantiles

EL can provide the confidence interval for quantile estimators. For $0 < \alpha < 1$, we are interested in inference about the α quantile of F, Q^α, which satisfy

$$E(1_{X \leq Q^\alpha} - \alpha) = 0. \tag{1.29}$$

Now define the EL ratio of the quantile as

$$\mathcal{R}(\alpha, q) = \max\Big\{ \prod_{i=1}^{n} n w_i | \sum_{i=1}^{n} w_i Z_i(\alpha, q) = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \Big\}, \tag{1.30}$$

where $Z_i(\alpha, q) = 1_{X_i \leq q} - \alpha$ with $-\infty < q < \infty$ and $0 < \alpha < 1$. Then, using the fact that the profile EL ratio function for a given α,

$$\mathcal{R}(\alpha) = \max_q \mathcal{R}(\alpha, q). \tag{1.31}$$

converges to χ^2_1, we can construct the confidence region for Q^α, where q makes $\mathcal{R}(\alpha, q)$ be sufficiently large.

1.2.2 Random Weighted Bootstrap Method for Quantile Regression

The finite sample distributions of quantile regression are extensively studied (Koenker (2005)), but their use for statistical inference is limited for several reasons. For example, under moderate conditions, the asymptotic distributions of quantile regression are normal, but the asymptotic variance is difficult to obtain because it depends on the conditional densities of observations that are usually not identified. However, if one uses a nonparametric estimate of the asymptotic variance, this requires a choice of a smoothing parameter, which may lead to unstable estimates. Even though it is possible to obtain the estimate of asymptotic variance using finite sample variance, its accuracy may be variant depending on the quantile level or design matrix. Hence, different conclusions can

be easily derived because of different choices of detail in the modeling process, even when the same data is used.

Resampling methods, such as bootstrap, jackknife, and more, are useful tools in dealing with this limitation (Hahn (1995), He and Shao (1996) and others) and provide a sound statistical inference for quantile regression under extensively different settings. Resampling methods generally generate samples from the observations X's, which have similar 'characteristics' to X's, and obtain an estimate of interesting parameters θ by optimizing the corresponding objective functions using generated samples. By repeating this process to obtain a large collection of these estimates, an inference about the true value of θ, θ_0 can be made. Because each resampling method has its own theoretical justification, which is needed to be made for each case, there is no universal theoretical ground for inferences of resampling methods.

The bootstrap is a widely used methodology for making inferences from regression models. Hahn (1995) depicts the paired bootstrap method whose sampling takes place from the original sample with replacement and studies its asymptotic properties. Bickel and Freedman (1981) describe the residual bootstrap method, where the samples are drawn from the estimated residuals and show that the bootstrap variance estimated from the variance of the resamples can be used to approximate the sampling variance. Some studies develop a modified version of the common bootstrap methods for resampling the M estimator. For example, Wu (1986) and Liu (1988) consider the wild bootstrap method, which uses random weights for the residual bootstrap to handle heteroscedasticity, and Rao and Zhao (1992) describe a random weights resampling method that applies to the loss functions.

Several bootstrap methods have been applied and introduced for quantile estimation (Hahn (1995), Chen, Wei, and Parzen (2004), Feng, He, and Hu (2011), Davidson (2012), Hagemann (2017)). Jin, Ying, and Wei (2001) introduce a simple resampling method that disturbs the minimand function directly. This method has the advantage of computational simplicity in that the overall procedure does not have to consider any complex and subjective nonparametric functional estimator.

The resampling scheme for the regression quantile can be formulated as follows: With the observations $\{(X_i, S_i), i = 1, \ldots, n\}$, the regression quantile estimates of the linear quantile model

$Q_{S_i}(\alpha|X_i) = X_i^\tau \beta_\alpha$ can be obtained by the following optimization problem (Koenker and Bassett Jr, 1978),

$$\hat{\beta}_\alpha = \text{argmin}_{\alpha \in \mathbb{R}^p} \Big\{ \sum_{i=1}^{n} \rho_\alpha (S_i - X_i^\tau \beta) \Big\}, \tag{1.32}$$

where $\rho_\alpha(x) = x(\alpha - 1(x < 0))$ is the quantile loss function. Because quantile regression, as the above model portrays, has estimating equations based on signs of the residuals $S_i - X_i^\tau$, it needs other forms of bootstrap besides the classical bootstrap, which is a common resampling method of the regression model.

Chatterjee and Bose (2005) present a generalized bootstrap technique motivated by weighted bootstrap of Barbe and Bertail (1995). In their model setting, the bootstrapped parameter $\hat{\beta}_\alpha$ for the quantile regression can be obtained by solving

$$\hat{\beta}_\alpha = \text{argmin}_{\beta \in \mathbb{R}^p} \Big\{ \sum_{i=1}^{n} w_i \rho_\alpha (S_i - X_i^\tau \beta) \Big\}, \tag{1.33}$$

where $\{w_i\}$ is a random sample drawn from a specific distribution such as an exponential distribution, Multinomial distribution, and so on. Unlike the wild bootstrap, the generalized bootstrap allows a wide range of choices for the distribution of w_i for the asymptotic soundness. Chen et al. (2008) investigate several options for sampling weights w_i for making inferences for the quantile regression functions.

Suppose we are aiming to test the hypothesis $H_0 : \beta_\alpha \in \Omega_0$, where Ω_0 is a subspace of $\mathbb{R}^p$. The following test statistic can be suggested for the hypothesis test,

$$M_n := \text{min}_{\beta \in \Omega_0} \sum_i \rho_\alpha (S_i - X_i^\tau \beta) - \text{min}_{\beta \in \mathbb{R}^p} \sum_i \rho_\alpha (S_i - X_i^\tau \beta). \tag{1.34}$$

However, the asymptotic characteristics of this test statistic are expected to depend on the conditional densities of S given X. To overcome this difficulty, it is shown that the resampling distribution of

$$M_n^* = \min_{\beta\in\Omega_0} \sum_i w_i\rho_\alpha(S_i - X_i^\tau\beta) - \min_{\beta\in\mathbb{R}^p} \sum_i w_i\rho_\alpha(S_i - X_i^\tau\beta)$$
$$- \sum_i w_i\rho_\alpha(S_i - X_i^\tau\hat{\beta}_{H_0}) - \sum_i w_i\rho_\alpha(S_i - X_i^\tau\hat{\beta}) \tag{1.35}$$

can approximate the distribution of M_n, where $\hat{\beta}_{H_0}$ is the quantile estimate under H_0 with the original sample, if a random sample $w_1, \ldots, w_n$ is drawn from a positively valued distribution with mean 1 and variance 1.

For more information about the general summary of bootstrap methods of quantile regression, please refer to Koenker et al. (2017).

1.3 Summary

Acknowledging the limitations of the previous studies, I am aiming to achieve several goals in this dissertation, such as providing newly developed models for forecasting risk measures, suggesting proper inference methodology for such models, and investigating the asymptotic behaviors of the proposed forecasts.

Firstly, I introduce my first study, which refers to the paper Kang, Peng, and Xiao (2020). In this research, I start with the limitations of the two-step inference methods proposed in Heras, Moreno, and Vilar-Zanón (2018). As previously mentioned, Heras, Moreno, and Vilar-Zanón (2018) introduce a two-step inference using logistic regression and quantile regression to better forecast the VaR of aggregate insurance losses. However, it is suspected that the application of quantile regression at the second step of their inference does not yield statistical improvement over the simple sample quantile estimation when explanatory variables are categorical. This is mainly because the adjusted quantile level for the second step is variant depending on the policy's categorical level, and it prevents the quantile regression from pooling information from all levels of the explanatory variables. Moreover, their inference does not provide details about model assumptions, related asymptotic properties, or uncertainty quantification. Hence, I propose an alternative two-step

inference for VaR via logistic regression and the sample quantile rather than quantile regression and develop an EL method to quantify the inference uncertainty, which was not addressed in Heras, Moreno, and Vilar-Zanón (2018). A simulation study confirms the good finite sample performance of the proposed method. A detailed methodology and related results are presented in Chapter 2.

Previous research has its limitations in that proposed inferences and methods of quantifying uncertainty only can be applied to categorical covariates. Because the two-step inference in Heras, Moreno, and Vilar-Zanón (2018) applies to both continuous and categorical covariates, one may wonder how to quantify the inference uncertainty when $\boldsymbol{X}_i$ involves some continuous variables. Motivated by this, I conduct the research, which refers to the paper Kang, Peng, and Golub (2021), based on two objectives. First, I propose a random weighted bootstrap method, which is a generalized uncertainty quantification method for both categorical and continuous explanatory variables, for two-step inference in Heras, Moreno, and Vilar-Zanón (2018). Secondly, I propose an alternative two-step inference, where a weighted quantile regression is employed in the second stage, and its uncertainty is quantified using the weighted bootstrap method. Unlike the quantile regression in Heras, Moreno, and Vilar-Zanón (2018), I do not adjust the risk level for the quantile regression, and the resulting estimation of VaR is not equal to the empirical quantile estimation when $\boldsymbol{X}_i$ is categorical. In Chapter 3, the details regarding the proposed inference and uncertainty quantification methods are listed.

I further extend my proposed methods to accommodate more extreme cases. As catastrophic events happen more and more frequently, accurately forecasting risk at a higher level is vital for the financial stability of the insurance industry. In Chapter 4, I propose an efficient three-step procedure to forecast extreme risk in insurance ratemaking by using logistic regression for estimating the non-zero claim probability, quantile regression for selecting a dynamic threshold, and a generalized Pareto distribution for fitting exceedances over the chosen threshold. I derive the asymptotic limit of the proposed risk forecasts under certain regularity conditions and employ a random weighted bootstrap method to quantify the uncertainty of the final risk forecast. Next, I apply the proposed

method to an automobile insurance dataset for testing the model performance. Finally, in Chapter 5, I conclude the dissertation with some discussions of the results.

Chapter 2

Risk Analysis With Categorical Explanatory Variables

This chapter is based on my publication: Kang, S., Peng, L. and Xiao, H.(2020), Risk analysis with categorical explanatory variables, *Insurance: Mathematics and Economics, 91(2), 238-243*.

Consider a set of insurance policies with n independent policyholders in a given period time. For the i-th policyholder, we observe the aggregate loss S_i and the d-dimensional categorical explanatory vector $\boldsymbol{X}_i$, representing some characteristics such as the driver's age and the age of the vehicle in automobile insurance. Because $\boldsymbol{X}_i$ is categorical, we assume the possible values are $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m$ without loss of generality. Let S denote the aggregate loss of a new policyholder, and let $\boldsymbol{X}$ represent the corresponding characteristics. Then the practical question is to forecast the Value-at-Risk (VaR) of S at a given level $\alpha \in (0,1)$ given $\boldsymbol{X} = \boldsymbol{x}$, which is defined as

$$\mathrm{VaR}_S(\alpha|\boldsymbol{x}) = \inf\{s : P(S > s|\boldsymbol{X} = \boldsymbol{x}) \leq 1 - \alpha\}.$$

Because $\boldsymbol{X}$ is categorical, we can classify these n policies into m categories, defined as $A_{\boldsymbol{x}_j} = \{i : \boldsymbol{X}_i = \boldsymbol{x}_j, 1 \leq i \leq n\}$ for $j = 1, \ldots, m$. So we only need to forecast $\{\mathrm{VaR}_S(\alpha|\boldsymbol{x}_j)\}_{j=1}^m$. A naive estimator for $\mathrm{VaR}_S(\alpha|\boldsymbol{x}_j)$ is the sample quantile of those S_i's in the category $A_{\boldsymbol{x}_j}$ without really using the explanatory variables. Alternatively, one can model the quantile of S_i as a linear function of $\boldsymbol{X}_i$ and employ the quantile regression technique to forecast $\mathrm{VaR}_S(\alpha|\boldsymbol{x}_j)$ for $j = 1, \cdots, m$. Applying the quantile regression technique to insurance ratemaking is not new. For example, Kudryavtsev (2009) employs the quantile regression technique to estimate quantiles of the net premium rate in ratemaking. We refer to Koenker (2005) for an overview of the quantile regression technique.

As a common feature of having many zero insurance losses due to no insurance claims, the above two simple VaR estimators are not efficient and often underestimate the risk because they do

not model the probability of having zero losses. When $P(S > 0|\boldsymbol{X} = \boldsymbol{x}_j) > 1 - \alpha$, we can write

$$\begin{aligned} \mathrm{VaR}_S(\alpha|\boldsymbol{x}_j) &= \inf\{s > 0 : P(S > s|\boldsymbol{X} = \boldsymbol{x}_j) \leq 1 - \alpha\} \\ &= \inf\{s > 0 : P(S > s|\boldsymbol{X} = \boldsymbol{x}_j, S > 0)P(S > 0|\boldsymbol{X} = \boldsymbol{x}_j) \leq 1 - \alpha\} \\ &= \inf\{s > 0 : P(S > s|\boldsymbol{X} = \boldsymbol{x}_j, S > 0) \leq \tfrac{1-\alpha}{P(S>0|\boldsymbol{X}=\boldsymbol{x}_j)}\}, \end{aligned} \tag{2.1}$$

which motivates the two-step inference in Heras, Moreno, and Vilar-Zanón (2018) by separately estimating $\alpha^*_{\boldsymbol{x}_j} = 1 - \frac{1-\alpha}{P(S>0|\boldsymbol{X}=\boldsymbol{x}_j)}$ and $\mathrm{VaR}_S(\alpha^*_{\boldsymbol{x}_j}|\boldsymbol{x}_j, S > 0)$. Hence, when we model and infer $P(S > 0|\boldsymbol{X} = \boldsymbol{x})$ soundly, the two-step inference should be better than the above two simple estimations without modeling $P(S > 0|\boldsymbol{X})$.

More specifically, the first step in Heras, Moreno, and Vilar-Zanón (2018) employs logistic regression to model and estimate $p_i = P(S_i > 0|\boldsymbol{X}_i)$, belonging to the generalized linear models commonly used for estimating the probability of having nonzero claims in actuarial science; see De Jong and Heller (2008) and Goldburd et al. (2016). Note that the number of distinct p_i's is m at most because $\boldsymbol{X}_i$ is categorical with m levels. After obtaining the estimator $\hat{\alpha}^*_{\boldsymbol{x}_j}$ for $\alpha^*_{\boldsymbol{x}_j}$, which pools information from all levels of the explanatory variables, the second step in Heras, Moreno, and Vilar-Zanón (2018) uses the quantile regression technique to estimate $\mathrm{VaR}_S(\hat{\alpha}^*_{\boldsymbol{x}_j}|\boldsymbol{x}_j, S > 0)$. When we model the conditional quantile of S_i at the level p given $S_i > 0$ and $\boldsymbol{X}_i$ by a linear function of $\boldsymbol{X}_i$, we can estimate $\mathrm{VaR}_S(p|\boldsymbol{x}_j, S > 0)$ by the quantile regression technique, which pools information from all levels of the explanatory variables. However, it remains unclear how Heras, Moreno, and Vilar-Zanón (2018) apply the quantile regression technique to estimate $\mathrm{VaR}_S(\hat{\alpha}^*_{\boldsymbol{x}_j}|\boldsymbol{x}_j, S > 0)$ because the risk level $\hat{\alpha}^*_{\boldsymbol{x}_j}$ depends on the category $\boldsymbol{x}_j$. Nevertheless, it is clear that Heras, Moreno, and Vilar-Zanón (2018) in their data analysis apply quantile regression to those positive S_i's and the related $\boldsymbol{X}_i$'s in each category of $\{A_{\boldsymbol{x}_j}\}_{j=1}^m$ rather than all positive S_i's. Hence, we suspect that the quantile regression estimation in Heras, Moreno, and Vilar-Zanón (2018) fails to pool information from all levels of the explanatory variables and is the same as the sample quantile based on those positive S_i's in each category of $\{A_{\boldsymbol{x}_j}\}_{j=1}^m$.

This Chapter of dissertation proposes an alternative two-step inference for the VaR via logistic regression and the sample quantile rather than quantile regression. To quantify the inference uncertainty, which has not been addressed in Heras, Moreno, and Vilar-Zanón (2018), we develop an empirical likelihood method. Applying the proposed empirical likelihood test to the real dataset in Heras, Moreno, and Vilar-Zanón (2018), we find the risk estimates in Heras, Moreno, and Vilar-Zanón (2018) are not significantly different from the true values without modeling the conditional quantile of S given $\boldsymbol{X}$ and $S > 0$. That is, using quantile regression in the second step is not necessary for categorical explanatory variables. We refer to Owen (2001) for an overview of the empirical likelihood method, which has been proved to be quite effective in interval estimations and hypothesis tests.

As we often do not observe all policies in the full cycle, it is of importance to take the exposure rates into account. Heras, Moreno, and Vilar-Zanón (2018) do adjust the exposure rates in the first step but ignore it in the second step. Moreover, Heras, Moreno, and Vilar-Zanón (2018) neither state the model assumptions explicitly nor provide asymptotic results for their two-step inference. In this paper, we provide explicit model assumptions and a rigorous two-step inference with exposure rates adjusted in both steps. Also, we propose an efficient empirical likelihood method for uncertainty quantification.

We organize this Chapter 2 as follows. Section 2.1 presents the explicit model, a two-step inference with asymptotic results, and an empirical likelihood method for uncertainty quantification. Section 2.2 applies the proposed method to the same insurer database in Heras, Moreno, and Vilar-Zanón (2018) for testing whether the risk estimates in Heras, Moreno, and Vilar-Zanón (2018) are significantly different from the true values without modeling the conditional quantile of S given $\boldsymbol{X}$ and $S > 0$. A simulation study is conducted in Section 2.3 to examine the finite sample performance of the proposed empirical likelihood method. Some conclusions are summarized in Section 2.4. All proofs are put in the Appendix.

2.1 Methodologies and Main Results

As argued in the introduction, this paper forecasts the Value-at-Risk of insurance losses given some categorical explanatory variables of a policyholder and develops an efficient empirical likelihood method for uncertainty quantification. To better appreciate the idea, we start with the detailed model description.

For each policyholder $i = 1, \ldots, n$, we observe the aggregate loss S_i in a particular year and some characteristics of this policyholder denoted by the d-dimensional categorical explanatory vector $\boldsymbol{X}_i$ with m categories $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$. Also, we observe the exposure rates $R_1, \ldots, R_n$ because policies may not be observed in a full year. Let S_i^* denote the aggregate loss of the i-th policyholder in a full year, which equals S_i if $R_i = 1$. We use S^* and $\boldsymbol{X}$ to denote the aggregate loss in a full year and the related characteristics of a future policyholder, respectively. The question is to forecast the risk $\mathrm{VaR}_{S^*}(\alpha|\boldsymbol{X} = \boldsymbol{x}_j)$ for $j = 1, \ldots, m$. Throughout, we write

$$p_i = P(S_i > 0|\boldsymbol{X}_i) \text{ and } p_i^* = P(S_i^* > 0|\boldsymbol{X}_i).$$

Assume

$$p_i = p_i^* R_i \text{ and } S_i^* = S_i h(R_i) \text{ for } i = 1, \ldots, n, \text{ and a known positive function } h. \tag{2.2}$$

Like (2.1), when

$$P(S^* > 0|\boldsymbol{X} = \boldsymbol{x}_j) > 1 - \alpha \text{ for } j = 1, \ldots, m, \tag{2.3}$$

we can write

$$\begin{aligned} \mathrm{VaR}_{S^*}(\alpha|\boldsymbol{x}_j) &= \inf\{s > 0 : P(S^* > s|\boldsymbol{X} = \boldsymbol{x}_j) \leq 1 - \alpha\} \\ &= \inf\{s > 0 : P(S^* > s|\boldsymbol{X} = \boldsymbol{x}_j, S^* > 0)P(S^* > 0|\boldsymbol{X} = \boldsymbol{x}_j) \leq 1 - \alpha\} \\ &= \inf\{s > 0 : P(S^* > s|\boldsymbol{X} = \boldsymbol{x}_j, S^* > 0) \leq \tfrac{1-\alpha}{P(S^*>0|\boldsymbol{X}=\boldsymbol{x}_j)}\}, \end{aligned} \tag{2.4}$$

which suggests the following two-step inference by separately estimating $\alpha^*_{\boldsymbol{x}_j} = 1 - \frac{1-\alpha}{P(S^*>0|\boldsymbol{X}=\boldsymbol{x}_j)}$ and $\mathrm{VaR}_{S^*}(\alpha^*_{\boldsymbol{x}_j}|\boldsymbol{x}_j, S^* > 0)$. To efficiently estimate $P(S^* > 0|\boldsymbol{X})$, the first-step uses logistic regression to model p_i:

$$\mathrm{logit}(\frac{p_i/R_i}{1-p_i/R_i}) = \boldsymbol{\beta}^T \bar{\boldsymbol{X}}_i \qquad \text{for} \quad i = 1, \cdots, n, \tag{2.5}$$

which is equivalent to

$$\mathrm{logit}(\frac{p^*_i}{1-p^*_i}) = \boldsymbol{\beta}^T \bar{\boldsymbol{X}}_i \text{ for } i = 1, \ldots, n, \text{ under condition}(2.2),$$

where $\bar{\boldsymbol{X}}_i$ denotes the corresponding dummy variable of $\boldsymbol{X}_i$ and A^T denotes the transpose of the vector or matrix A. Using the standard logistic regression estimation technique, we can obtain estimator $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$, which gives

$$\hat{p}_i = \frac{R_i}{1+\exp\{-\hat{\boldsymbol{\beta}}^T \bar{\boldsymbol{X}}_i\}} \text{ and } \hat{p}^*_i = \frac{1}{1+\exp\{-\hat{\boldsymbol{\beta}}^T \bar{\boldsymbol{X}}_i\}} \text{ for } i = 1, \cdots, n. \tag{2.6}$$

As $\boldsymbol{X}_i$ only takes m values, $\bar{\boldsymbol{X}}_i$ has m different values, implying that $\{\hat{p}^*_i\}_{i=1}^n$ have m distinct values at most. Hence, for $j = 1, \cdots, m$, we write

$$p^*_{\boldsymbol{x}_j} = \frac{1}{1+\exp\{-\boldsymbol{\beta}^T \bar{\boldsymbol{x}}_j\}} \text{ and } \hat{p}^*_{\boldsymbol{x}_j} = \frac{1}{1+\exp\{-\hat{\boldsymbol{\beta}}^T \bar{\boldsymbol{x}}_j\}} \tag{2.7}$$

and estimate $\alpha^*_{\boldsymbol{x}_j}$ by

$$\hat{\alpha}^*_{\boldsymbol{x}_j} = 1 - \frac{1-\alpha}{\hat{p}^*_{\boldsymbol{x}_j}},$$

where $\bar{\boldsymbol{x}}_j$ is the corresponding dummy variable of $\boldsymbol{x}_j$.

Define

$$A^+_{\boldsymbol{x}_j} = \{i : \boldsymbol{X}_i = \boldsymbol{x}_j,\ S_i > 0,\ 1 \le i \le n\} \text{ for } j = 1, \ldots, m.$$

It follows from (2.2) that we can estimate $\mathrm{VaR}_{S^*}(\alpha^*_{\boldsymbol{x}_j}|\boldsymbol{x}_j, S^* > 0)$ by the sample quantile at the level $\hat{\alpha}^*_{\boldsymbol{x}_j}$ based on $\{h(R_i)S_i : i \in A^+_{\boldsymbol{x}_j}, i = 1, \ldots, n\}$, which gives the two-step estimator of

$\mathrm{VaR}_{S^*}(\alpha|\boldsymbol{x}_j)$ as

$$\widehat{\mathrm{VaR}}_{S^*}(\alpha|\boldsymbol{x}_j) = \inf\{s : \frac{\sum_{i=1}^n I(i \in A^+_{\boldsymbol{x}_j}, h(R_i)S_i > s)}{\sum_{i=1}^n I(i \in A^+_{\boldsymbol{x}_j})} \leq \frac{1-\alpha}{\hat{p}^*_{\boldsymbol{x}_j}}\}.$$

Here $I(\cdot)$ denotes the indicator function. A simple application of the standard asymptotic theory can prove the asymptotic normality of the above VaR estimator, but the asymptotic variance is complicated and depends on the inference uncertainties of both steps. To construct a confidence interval or conduct a hypothesis test without estimating the asymptotic variance of the above two-step risk estimation, we propose the following empirical likelihood method based on the estimating equation approach in Qin and Lawless (1994).

It follows from (2.5) that

$$p_i(\boldsymbol{\beta}) = \frac{R_i}{1 + \exp\{-\boldsymbol{\beta}^T \bar{\boldsymbol{X}}_i\}}$$

and the corresponding scores are

$$\boldsymbol{Z}_i(\boldsymbol{\beta}) = \frac{I(S_i > 0)}{p_i(\boldsymbol{\beta})}\frac{\partial p_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \frac{I(S_i = 0)}{1 - p_i(\boldsymbol{\beta})}\frac{\partial p_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \qquad \text{for} \quad i = 1, \cdots, n.$$

Put $\theta = \mathrm{VaR}_{S^*}(\alpha|\boldsymbol{x}_j, S^* > 0)$. For $i \in A^+_{\boldsymbol{x}_j}$, it follows from (2.4) that $P(S_i > \theta|\boldsymbol{X}_i) = \frac{1-\alpha}{p_i/R_i}$. Hence, we define

$$Z_{i,\boldsymbol{x}_j}(\theta, \boldsymbol{\beta}) = I(S_i > \theta) - \frac{1-\alpha}{p_i(\boldsymbol{\beta})/R_i} \qquad \text{for} \quad i \in A^+_{\boldsymbol{x}_j}.$$

By noting that $\boldsymbol{Z}_i$'s are constructed from $\{I(S_i > 0)\}_{i=1}^n$, $Z_{i,\boldsymbol{x}_j}$'s are based on $\{S_i : i \in A^+_{\boldsymbol{x}_j},\ i = 1, \ldots, n\}$, and $\{I(S_i > 0)\}_{i=1}^n$ is independent of $\{S_i : i \in A^+_{\boldsymbol{x}_j},\ i = 1, \ldots, n\}$, we formulate the empirical likelihood function for $\theta = \mathrm{VaR}_{S^*}(\alpha|\boldsymbol{x}_j)$ based on the method for two

independent samples as

$$
\begin{aligned}
& L(\theta, \boldsymbol{\beta}|\boldsymbol{x}_j) \\
= \; & \sup\Big\{ \textstyle\prod_{k=1}^n (nq_k) \prod_{i\in A^+_{\boldsymbol{x}_j}} (m^+_{\boldsymbol{x}_j} q_i^*) : q_k \geq 0 \text{ for } k = 1, \cdots, n, \sum_{k=1}^n q_k = 1, \\
& q_i^* > 0 \text{ for } i \in A^+_{\boldsymbol{x}_j}, \textstyle\sum_{i\in A^+_{\boldsymbol{x}_j}} q_i^* = 1, \;\; \sum_{k=1}^n q_k \boldsymbol{Z}_k(\boldsymbol{\beta}) = 0, \sum_{i\in A^+_{\boldsymbol{x}_j}} q_i^* Z_{i,\boldsymbol{x}_j}(\theta, \boldsymbol{\beta}) = 0 \Big\},
\end{aligned}
$$

where $m^+_{\boldsymbol{x}_j} = \sum_{i=1}^n I(i \in A^+_{\boldsymbol{x}_j})$.

It follows from the Lagrange multiplier technique that

$$
\begin{aligned}
l(\theta, \boldsymbol{\beta}|\boldsymbol{x}_j) \;\; &:= -2 \log L(\theta, \boldsymbol{\beta}|\boldsymbol{x}_j) \\
&= 2 \textstyle\sum_{i=1}^n \log\{1 + \boldsymbol{\lambda}_1^T \boldsymbol{Z}_i(\boldsymbol{\beta})\} + 2 \sum_{i\in A^+_{\boldsymbol{x}_j}} \log\{1 + \lambda_2 Z_{i,\boldsymbol{x}_j}(\theta, \boldsymbol{\beta})\},
\end{aligned}
$$

where $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_1(\boldsymbol{\beta})$ and $\lambda_2 = \lambda_2(\theta, \boldsymbol{\beta})$ satisfy

$$
\sum_{i=1}^n \frac{\boldsymbol{Z}_i(\boldsymbol{\beta})}{1 + \boldsymbol{\lambda}_1^T \boldsymbol{Z}_i(\boldsymbol{\beta})} = 0 \quad \text{and} \quad \sum_{i\in A^+_{\boldsymbol{x}_j}} \frac{Z_{i,\boldsymbol{x}_j}(\theta, \boldsymbol{\beta})}{1 + \lambda_2 Z_{i,\boldsymbol{x}_j}(\theta, \boldsymbol{\beta})} = 0. \tag{2.8}
$$

Because we are interested in $\theta = \mathrm{VaR}_{S^*}(\alpha|\boldsymbol{x}_j)$, we consider the following profile log-empirical likelihood ratio

$$
l^P(\theta|\boldsymbol{x}_j) = \min_{\boldsymbol{\beta}} l(\theta, \boldsymbol{\beta}|\boldsymbol{x}_j).
$$

Theorem 2.1. *Assume $\{(S_i, \boldsymbol{X}_i^T)^T\}_{i=1}^n$ is a sequence of independent random vectors, $\{R_i\}_{i=1}^n$ are deterministic and $\boldsymbol{X}_i$ is a d-dimensional categorical vector with levels $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$. Further assume (2.2), (2.3), (2.5), $\lim_{n\to\infty} m^+_{\boldsymbol{x}_j}/n = \tau^+_{\boldsymbol{x}_j} \in (0,1)$ for $j = 1, \ldots, m$, $E\big(\boldsymbol{Z}_1(\boldsymbol{\beta}^0)\boldsymbol{Z}_1^T(\boldsymbol{\beta}^0)\big)$ is positive definite with $\boldsymbol{\beta}^0$ denoting the true value of $\boldsymbol{\beta}$, and the conditional distribution function of S_i^* given $\boldsymbol{X}_i = \boldsymbol{x}_j$ and $S_i^* > 0$ is independent of i with a positive density at $\mathrm{VaR}_{S^*}(\alpha^*_{\boldsymbol{x}_j}|\boldsymbol{x}_j)$ for $j = 1, \ldots, m$. Then, for each $j = 1, \ldots, m$, $l^P(\mathrm{VaR}_{S^*}(\alpha|\boldsymbol{x}_j)|\boldsymbol{x}_j)$ converges in distribution to a chi-squared limit with one degree of freedom as $n \to \infty$.*

Based on the above theorem, an empirical likelihood confidence interval with level a for $\mathrm{VaR}_{S^*}(\alpha|\boldsymbol{x}_j)$ is obtained as

$$I_a(\alpha|\boldsymbol{x}_j) = \{\theta : l^P(\theta|\boldsymbol{x}_j) \leq \chi^2_{1,a}\},$$

where $\chi^2_{1,a}$ denotes the a-th quantile of a chi-squared distribution function with one degree of freedom. Similarly, an empirical likelihood test for $H_0 : \mathrm{VaR}_{S^*}(\alpha|\boldsymbol{x}_j) = \theta_0$ at the level a rejects the null hypothesis whenever $l^P(\theta_0|\boldsymbol{x}_j) > \chi^2_{1,1-a}$.

Remark 2.1. *We can develop a similar empirical likelihood method for constructing a confidence region of* $(\mathrm{VaR}_{S^*}(\alpha_1|\boldsymbol{x}_1), \ldots, \mathrm{VaR}_{S^*}(\alpha_k|\boldsymbol{x}_k))^T$ *and testing*

$$H_0 : \mathrm{VaR}_{S^*}(\alpha_1|\boldsymbol{x}_1) = \theta_1, \ldots, \mathrm{VaR}_{S^*}(\alpha_k|\boldsymbol{x}_k) = \theta_k,$$

where k is a given integer. Here we allow different risk levels.

2.2 Application to an insurer database

We reexamine the dataset from an Australian automobile insurance company between the years 2004 and 2005, which is analyzed by Heras, Moreno, and Vilar-Zanón (2018) based on the two-step inference of using logistic regression at the first step and quantile regression at the second step. The total number of policies is $n = 67856$, and the categorical explanatory variable $\boldsymbol{X}_i$ involves the age of the vehicle with four levels and the driver's age with six levels. Hence, the total number of different levels of $\boldsymbol{X}_i$ is $m = 24$. We refer to De Jong and Heller (2008) for a detailed description of this dataset.

Following Heras, Moreno, and Vilar-Zanón (2018), we use $h(R_i) = 1$ without adjusting the exposure rates at the second step. Because Heras, Moreno, and Vilar-Zanón (2018) apply quantile regression to the sample in each category of $\{A^+_{\boldsymbol{x}_j}\}_{j=1}^{24}$, we suspect that their estimates may be equal to our estimates $\{\widehat{\mathrm{VaR}}_{S^*}(\alpha|\boldsymbol{x}_j)\}_{j=1}^{24}$ without modeling the conditional quantile of S^* given

$\boldsymbol{X}$ and $S^* > 0$. After computing $\{\widehat{\mathrm{VaR}}_{S^*}(0.95|\boldsymbol{x}_j)\}_{j=1}^{24}$ and reporting them in Table 2.1 below, we find that our two-step estimates are different from the two-step estimates in Heras, Moreno, and Vilar-Zanón (2018). To further investigate the effectiveness and necessity of using quantile regression, we employ the proposed empirical likelihood method to test whether the estimates in Heras, Moreno, and Vilar-Zanón (2018) are significantly different from the true values without modeling the conditional quantile of S^* given $\boldsymbol{X}$ and $S^* > 0$. The p-values reported in Table 2.1 show that the two-step estimates in Heras, Moreno, and Vilar-Zanón (2018) are not significantly different from the true values at the 5% significant level. In other words, the estimates in Heras, Moreno, and Vilar-Zanón (2018) are in the 95% confidence intervals of the true values without modeling the conditional quantile of S^* given $\boldsymbol{X}$ and $S^* > 0$, i.e., it is not significantly useful to employ quantile regression.

2.3 Simulation study

This section investigates the finite sample performance of the proposed empirical likelihood confidence interval in terms of coverage accuracy.

First, we compute the sample proportions $\tau_{\boldsymbol{x}_1}, \ldots, \tau_{\boldsymbol{x}_{24}}$ of those 24 categories in the real dataset analyzed in Section 2.2 and generate the explanatory variables $\boldsymbol{X}_i = (\mathrm{VehAge}_i, \mathrm{AgeCat}_i)^T$ to have the sample size $n_j = [n\tau_{\boldsymbol{x}_j}]$ for the j-th category, where $j = 1, \ldots, 24$. Here VehAge_i denotes the age of the vehicle with levels 1 (newest), 2, 3, and 4, and AgeCat_i denotes the driver's age with levels 1 (youngest), 2, 3, 4, 5, and 6. Note that the total sample size $\tilde{n} = n_1 + \cdots + n_{24}$ may be smaller than n due to the rounding effect.

Next, using the estimator $\hat{\boldsymbol{\beta}}$ in fitting (2.5) to the real dataset in Section 2.2, we generate independent Bernoulli random variables with $p_i = \frac{1}{1+\exp\{-\hat{\boldsymbol{\beta}}^T \bar{\boldsymbol{X}}_i\}}$ for $i = 1, \ldots, \tilde{n}$, where $\bar{\boldsymbol{X}}_i$ is the dummy variable of $\boldsymbol{X}_i$. We use $\{N_i\}_{i=1}^{\tilde{n}}$ to denote these Bernoulli variables, which indicate whether a loss is positive or zero.

For each $i = 1, \ldots, \tilde{n}$, we further generate S_i from a standard Gamma distribution with parameter being the sum of the levels of VehAge_i and AgeCat_i when $N_i = 1$, and set $S_i = 0$ when

Table 2.1: We report our two-step estimates $\{\widehat{\mathrm{VaR}}_{S^*}(0.95|\boldsymbol{x}_j)\}_{j=1}^{24}$, copy the two-step estimates $\{\widetilde{\mathrm{VaR}}_{S^*}(0.95|\boldsymbol{x}_j)\}_{j=1}^{24}$ from Heras, Moreno, and Vilar-Zanón (2018), and report the P-values of the proposed empirical likelihood test for testing whether the true Value-at-Risk equals the estimate in Heras, Moreno, and Vilar-Zanón (2018) for each group. The two numbers inside the bracket of Group represent the levels of the age of the vehicle and the driver's age, respectively.

Group	$\widetilde{\mathrm{VaR}}_{S^*}(0.95\|\boldsymbol{x}_j)$	P- value	$\widehat{\mathrm{VaR}}_{S^*}(0.95\|\boldsymbol{x}_j)$
1(2&1)	3212.78	0.6431	3277.82
2(1&1)	2534.94	0.6584	3241.63
3(3&1)	2901.37	0.4582	2636.08
4(2&2)	1726.35	0.2197	1438.54
5(4&1)	2927.59	0.4146	2442.07
6(1&2)	1407.25	0.2456	1573.67
7(2&3)	1556.17	0.7342	1615.35
8(1&3)	1327.27	0.0773	1045.51
9(2&4)	1347.33	0.9427	1346.14
10(3&2)	1691.04	0.9821	1709.33
11(1&4)	1143.21	0.6944	1203.77
12(3&3)	1487.55	0.7001	1576.59
13(4&2)	1736.64	0.5042	1859.44
14(3&4)	1283.53	0.8254	1271.97
15(4&3)	1470.00	0.9062	1478.81
16(4&4)	1311.30	0.9199	1319.02
17(2&5)	1014.82	0.7465	1073.10
18(2&6)	1146.38	0.5192	1086.51
19(1&5)	837.34	0.8487	836.90
20(1&6)	947.30	0.7327	926.72
21(3&5)	914.85	0.8660	969.59
22(3&6)	1067.53	0.1958	1193.75
23(4&5)	889.11	0.9058	882.40
24(4&6)	1062.24	0.1844	893.90

$N_i = 0$. We use $R_i = 1$ and $h(R_i) = 1$ for all $i = 1, \ldots, \tilde{n}$, i.e., all policies are observed in a full year cycle.

Based on the generated data $\{(S_i, \boldsymbol{X}_i^T)^T\}_{i=1}^{\tilde{n}}$, for each category, we first compute the true value $\mathrm{VaR}_{S^*}(0.95|\boldsymbol{x}_j)$ from the employed Gamma distribution and the corresponding $p_{\boldsymbol{x}_j} = \frac{1}{1+\exp\{-\hat{\boldsymbol{\beta}}^T\bar{\boldsymbol{x}}_j\}}$, and then calculate the profile log-empirical likelihood ratio $l^P(\mathrm{VaR}_{S^*}(0.95|\boldsymbol{x}_j)|\boldsymbol{x}_j)$. Repeat this procedure 5000 times so that the empirical coverage probabilities for the empirical likelihood confidence intervals $I_{0.90}(0.95|\boldsymbol{x}_j)$ and $I_{0.95}(0.95|\boldsymbol{x}_j)$ are obtained and reported in Table 2.2. This

table shows that the proposed empirical likelihood method produces accurate confidence intervals and the coverage accuracy improves as the sample size becomes larger.

Table 2.2: We report τ_j's, the empirical coverage probabilities for $I_{0.9}(0.95|\boldsymbol{x}_j)$ and $I_{0.95}(0.95|\boldsymbol{x}_j)$ of the proposed empirical likelihood based confidence intervals with sample sizes $n = 30,000$ and $n = 60,000$. The two numbers inside the bracket of Group represent the levels of the age of the vehicle and the driver's age, respectively.

Group	τ_j	$n = 30000$ $I_{0.9}(0.95\|\boldsymbol{x}_j)$	$n = 30000$ $I_{0.95}(0.95\|\boldsymbol{x}_j)$	$n = 60000$ $I_{0.9}(0.95\|\boldsymbol{x}_j)$	$n = 60000$ $I_{0.95}(0.95\|\boldsymbol{x}_j)$
1(2&1)	2.2165%	0.9152	0.9588	0.9086	0.9542
2(1&1)	1.8908%	0.9022	0.9522	0.9040	0.9520
3(3&1)	2.4213%	0.9064	0.9556	0.9042	0.9544
4(2&2)	4.6672%	0.9104	0.9558	0.9094	0.9572
5(4&1)	1.9335%	0.9084	0.9532	0.9110	0.9550
6(1&2)	3.1832%	0.9090	0.9524	0.9072	0.9534
7(2&3)	5.5131%	0.9126	0.9594	0.9068	0.9554
8(1&3)	3.9879%	0.9122	0.9604	0.9128	0.9550
9(2&4)	5.7755%	0.9102	0.9560	0.9144	0.9570
10(3&2)	5.8230%	0.9090	0.9544	0.9150	0.9528
11(1&4)	4.3253%	0.9190	0.9604	0.9070	0.9592
12(3&3)	7.1121%	0.9126	0.9606	0.9096	0.9556
13(4&2)	5.2936%	0.9160	0.9590	0.9142	0.9610
14(3&4)	7.0149%	0.9186	0.9580	0.9200	0.9592
15(4&3)	6.6228%	0.9120	0.9584	0.9176	0.9604
16(4&4)	6.7422%	0.9174	0.9584	0.9128	0.9618
17(2&5)	3.8832%	0.9196	0.9680	0.9076	0.9558
18(2&6)	2.3889%	0.9098	0.9602	0.9060	0.9552
19(1&5)	3.0093%	0.9112	0.9570	0.9124	0.9600
20(1&6)	1.6668%	0.9122	0.9598	0.9130	0.9576
21(3&5)	4.5508%	0.9154	0.9632	0.9124	0.9606
22(3&6)	2.6394%	0.9198	0.9642	0.9130	0.9574
23(4&5)	4.3784%	0.9218	0.9612	0.9206	0.9664
24(4&6)	2.9533%	0.9164	0.9540	0.9146	0.9580

2.4 Conclusions

To accurately forecast the Value-at-Risk of the aggregate insurance losses in insurance ratemaking, Heras, Moreno, and Vilar-Zanón (2018) propose to model the probability with nonzero claims by logistic regression at the first step and model the conditional quantile of the aggregate loss given

nonzero claims by quantile regression at the second step. When the explanatory variables are categorical, the adjusted quantile level from the first step is different for each category and quantile regression can only be applied to each category. Thus it is unnecessary to employ quantile regression at the second step. This observation motivates us to estimate the Value-at-Risk and quantify the uncertainty without using quantile regression. This study provides the explicit model assumptions and develops an empirical likelihood method to construct a confidence interval for the Value-at-Risk measure and to test whether the two-step estimates in Heras, Moreno, and Vilar-Zanón (2018) are significantly different from those without modeling the conditional quantile of the aggregate loss given the explanatory variable. Data analysis with the proposed new method does show that the second step of using quantile regression is not necessary. A simulation study confirms the good finite sample performance of the proposed empirical likelihood method.

2.5 Proof of Theorem 2.1

Before proving Theorem 2.1, we need some lemmas, where $\boldsymbol{\beta}^0$ and θ_j^0 denote the true values of $\boldsymbol{\beta}$ and $\theta_j = \mathrm{VaR}_{S^*}(\alpha|\boldsymbol{x}_j)$, respectively. We use $\xrightarrow{d}$ and $\xrightarrow{p}$ to denote the convergence in distribution and in probability, respectively. We also use the following conventional notation for partial derivatives:

- $\frac{\partial y}{\partial \boldsymbol{x}} = (\frac{\partial y}{\partial x_1}, \ldots, \frac{\partial y}{\partial x_{d_1}})^T$ when y is scalar and $\boldsymbol{x} = (x_1, \ldots, x_{d_1})^T$;
- $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_{d_1}} \\ \cdot & \cdots & \cdot \\ \frac{\partial y_{d_2}}{\partial x_1} & \cdots & \frac{\partial y_{d_2}}{\partial x_{d_1}} \end{pmatrix}$ when $\boldsymbol{x} = (x_1, \ldots, x_{d_1})^T$ and $\boldsymbol{y} = (y_1, \ldots, y_{d_2})^T$.

Lemma 2.1. *Under conditions of Theorem 2.1, for each $j = 1, \ldots, m$, we have*

$$\begin{cases} \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{Z}_i(\boldsymbol{\beta}^0) \xrightarrow{d} \boldsymbol{W}_1 \sim N(0, \Sigma), \\ \frac{1}{n} \sum_{i=1}^n \boldsymbol{Z}_i(\boldsymbol{\beta}^0) \boldsymbol{Z}_i^T(\boldsymbol{\beta}^0) \xrightarrow{p} \Sigma = E\left(\boldsymbol{Z}_1(\boldsymbol{\beta}^0) \boldsymbol{Z}_1^T(\boldsymbol{\beta}^0)\right), \\ \frac{1}{\sqrt{m_{\boldsymbol{x}_j}^+}} \sum_{i \in A_{\boldsymbol{x}_j}^+} Z_{i,\boldsymbol{x}_j}(\theta_j^0, \boldsymbol{\beta}^0) \xrightarrow{d} W_2 \sim N(0, \sigma_j^2), \\ \frac{1}{m_{\boldsymbol{x}_j}^+} \sum_{i \in A_{\boldsymbol{x}_j}^+} \{Z_{i,\boldsymbol{x}_j}(\theta_j^0, \boldsymbol{\beta}^0)\}^2 \xrightarrow{p} \sigma_j^2 = \frac{\alpha - 1 + p_{\boldsymbol{x}_j}^*}{p_{\boldsymbol{x}_j}^*} \{1 - \frac{\alpha - 1 + p_{\boldsymbol{x}_j}^*}{p_{\boldsymbol{x}_j}^*}\}, \end{cases}$$

as $n \to \infty$, where $\boldsymbol{W}_1$ and W_2 are independent, and $p^*_{\boldsymbol{x}_j}$ is defined in (2.7).

Proof. It directly follows from the central limit theorem, weak law of large numbers and the independence between $\{I(S_i > 0)\}_{i=1}^n$ and $\{S_i : i \in A^+_{\boldsymbol{x}_j}, i = 1, \ldots, n\}$. □

Lemma 2.2. *Under conditions of Theorem 2.1, as $n \to \infty$, with probability tending to one, $l(\theta_j^0, \boldsymbol{\beta}|\boldsymbol{x}_j)$ for each $j = 1, \ldots, m$ attains its minimum value at some point $\widetilde{\boldsymbol{\beta}}$ in the interior of the ball $||\boldsymbol{\beta} - \boldsymbol{\beta}^0|| \le n^{-1/3}$ and $\widetilde{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\lambda}}_1 = \boldsymbol{\lambda}_1(\widetilde{\boldsymbol{\beta}})$ and $\tilde{\lambda}_2 = \lambda_2(\widetilde{\boldsymbol{\beta}})$ satisfy that*

$$Q_{1n}(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\lambda}}_1) = 0,\ Q_{2n}(\widetilde{\boldsymbol{\beta}}, \widetilde{\lambda}_2) = 0,\ Q_{3n}(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\lambda}}_1, \widetilde{\lambda}_2) = 0, \tag{2.9}$$

where $Q_{1n}(\boldsymbol{\beta}, \boldsymbol{\lambda}_1) = \frac{1}{n}\sum_{i=1}^n \frac{\boldsymbol{Z}_i(\boldsymbol{\beta})}{1+\boldsymbol{\lambda}_1^T \boldsymbol{Z}_i(\boldsymbol{\beta})}$, $Q_{2n}(\boldsymbol{\beta}, \lambda_2) = \frac{1}{n}\sum_{i \in A^+_{\boldsymbol{x}_j}} \frac{Z_{i,\boldsymbol{x}_j}(\theta_j^0, \boldsymbol{\beta})}{1+\lambda_2 Z_{i,\boldsymbol{x}_j}(\theta_j^0, \boldsymbol{\beta})}$ and

$$Q_{3n}(\boldsymbol{\beta}, \boldsymbol{\lambda}_1, \lambda_2) = \frac{1}{n}\Big\{ \sum_{i=1}^n \frac{1}{1 + \boldsymbol{\lambda}_1^T \boldsymbol{Z}_i(\boldsymbol{\beta})} \big(\frac{\partial \boldsymbol{Z}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\big)^T \boldsymbol{\lambda}_1 + \sum_{i \in A^+_{\boldsymbol{x}_j}} \frac{1}{1 + \lambda_2 Z_{i,\boldsymbol{x}_j}(\theta_j^0, \boldsymbol{\beta})} \frac{\partial Z_{i,\boldsymbol{x}_j}(\theta_j^0, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \lambda_2 \Big\}.$$

Proof. It follows from the same arguments in the proof of Lemma 1 in Qin and Lawless (1994). □

Proof of Theorem 2.1. Note that

$$\frac{\partial Q_{1n}(\boldsymbol{\beta}, \mathbf{0})}{\partial \boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^n \frac{\partial \boldsymbol{Z}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}},\ \frac{\partial Q_{1n}(\boldsymbol{\beta}, \mathbf{0})}{\partial \boldsymbol{\lambda}_1} = -\frac{1}{n}\sum_{i=1}^n \boldsymbol{Z}_i(\boldsymbol{\beta}) \boldsymbol{Z}_i^T(\boldsymbol{\beta}),$$

$$\frac{\partial Q_{1n}(\boldsymbol{\beta}, \mathbf{0})}{\partial \lambda_2} = \mathbf{0},\ \frac{\partial Q_{2n}(\boldsymbol{\beta}, 0)}{\partial \boldsymbol{\beta}} = \frac{1}{n}\sum_{i \in A^+_{\boldsymbol{x}_j}} \frac{\partial Z_{i,\boldsymbol{x}_j}(\theta_j^0, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}},$$

$$\frac{\partial Q_{2n}(\boldsymbol{\beta}, 0)}{\partial \boldsymbol{\lambda}_1} = \mathbf{0},\ \frac{\partial Q_{2n}(\boldsymbol{\beta}, 0)}{\partial \lambda_2} = -\frac{1}{n}\sum_{i \in A^+_{\boldsymbol{x}_j}} \{Z_{i,\boldsymbol{x}_j}(\theta_j^0, \boldsymbol{\beta})\}^2,\ \frac{\partial Q_{3n}(\boldsymbol{\beta}, \mathbf{0}, 0)}{\partial \boldsymbol{\beta}} = \mathbf{0},$$

$$\frac{\partial Q_{3n}(\boldsymbol{\beta}, \mathbf{0}, 0)}{\partial \boldsymbol{\lambda}_1} = \frac{1}{n}\sum_{i=1}^n \big(\frac{\partial \boldsymbol{Z}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\big)^T,\ \frac{\partial Q_{3n}(\boldsymbol{\beta}, \mathbf{0}, 0)}{\partial \lambda_2} = \frac{1}{n}\sum_{i \in A^+_{\boldsymbol{x}_j}} \frac{\partial Z_{i,\boldsymbol{x}_j}(\theta_j^0, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}.$$

By Taylor expansion and Lemmas 2.1-2.2, we have

$$\begin{aligned}\mathbf{0} &= Q_{1n}(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\lambda}}_1) \\ &= Q_{1n}(\boldsymbol{\beta}^0, \mathbf{0}) + \tfrac{\partial Q_{1n}(\boldsymbol{\beta}^0, \mathbf{0})}{\partial \boldsymbol{\beta}}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + \tfrac{\partial Q_{1n}(\boldsymbol{\beta}^0, \mathbf{0})}{\partial \boldsymbol{\lambda}_1}\widetilde{\boldsymbol{\lambda}}_1 + \tfrac{\partial Q_{1n}(\boldsymbol{\beta}^0, \mathbf{0})}{\partial \lambda_2}\widetilde{\lambda}_2 + o_p(\delta_n),\end{aligned}$$

$$\begin{aligned}0 &= Q_{2n}(\widetilde{\boldsymbol{\beta}}, \widetilde{\lambda}_2) \\ &= Q_{2n}(\boldsymbol{\beta}^0, 0) + \big(\tfrac{\partial Q_{2n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\beta}}\big)^T(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + \big(\tfrac{\partial Q_{2n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\lambda}_1}\big)^T\widetilde{\boldsymbol{\lambda}}_1 + \tfrac{\partial Q_{2n}(\boldsymbol{\beta}^0, 0)}{\partial \lambda_2}\widetilde{\lambda}_2 + o_p(\delta_n),\end{aligned}$$

$$\begin{aligned}\mathbf{0} &= Q_{3n}(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\lambda}}_1, \widetilde{\lambda}_2) \\ &= Q_{3n}(\boldsymbol{\beta}^0, \mathbf{0}, 0) + \tfrac{\partial Q_{3n}(\boldsymbol{\beta}^0, \mathbf{0}, 0)}{\partial \boldsymbol{\beta}}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + \tfrac{\partial Q_{3n}(\boldsymbol{\beta}^0, \mathbf{0}, 0)}{\partial \boldsymbol{\lambda}_1}\widetilde{\boldsymbol{\lambda}}_1 + \tfrac{\partial Q_{3n}(\boldsymbol{\beta}^0, \mathbf{0}, 0)}{\partial \lambda_2}\widetilde{\lambda}_2 + o_p(\delta_n),\end{aligned}$$

where $\delta_n = ||\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0|| + ||\widetilde{\boldsymbol{\lambda}}_1|| + |\widetilde{\lambda}_2|$. That is,

$$\begin{pmatrix} \widetilde{\lambda}_2 \\ \widetilde{\boldsymbol{\lambda}}_1 \\ \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 \end{pmatrix} = S_n^{-1} \begin{pmatrix} -Q_{2n}(\boldsymbol{\beta}^0, 0) + o_p(\delta_n) \\ -Q_{1n}(\boldsymbol{\beta}^0, \mathbf{0}) + o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix},$$

where

$$\begin{aligned} S_n &= \begin{pmatrix} \frac{\partial Q_{2n}(\boldsymbol{\beta}^0, 0)}{\partial \lambda_2} & \big(\frac{\partial Q_{2n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\lambda}_1}\big)^T & \big(\frac{\partial Q_{2n}(\boldsymbol{\beta}^0, 0)}{\partial \boldsymbol{\beta}}\big)^T \\ \frac{\partial Q_{1n}(\boldsymbol{\beta}^0, \mathbf{0})}{\partial \lambda_2} & \frac{\partial Q_{1n}(\boldsymbol{\beta}^0, \mathbf{0})}{\partial \boldsymbol{\lambda}_1} & \frac{\partial Q_{1n}(\boldsymbol{\beta}^0, \mathbf{0})}{\partial \boldsymbol{\beta}} \\ \frac{\partial Q_{3n}(\boldsymbol{\beta}^0, \mathbf{0}, 0)}{\partial \lambda_2} & \frac{\partial Q_{3n}(\boldsymbol{\beta}^0, \mathbf{0}, 0)}{\partial \boldsymbol{\lambda}_1} & \frac{\partial Q_{3n}(\boldsymbol{\beta}^0, \mathbf{0}, 0)}{\partial \boldsymbol{\beta}} \end{pmatrix} \\ &\xrightarrow{p} S = \begin{pmatrix} -\tau^+_{\boldsymbol{x}_j} E\big(Z_{j_0, \boldsymbol{x}_j}(\theta_j^0, \boldsymbol{\beta}^0)\big)^2 & \mathbf{0}^T & \tau^+_{\boldsymbol{x}_j} E\big(\frac{\partial Z_{j_0, \boldsymbol{x}_j}(\theta_j^0, \boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}}\big)^T \\ \mathbf{0} & -E\big(\boldsymbol{Z}_1(\boldsymbol{\beta}^0)\boldsymbol{Z}_1^T(\boldsymbol{\beta}^0)\big) & E\big(\frac{\partial \boldsymbol{Z}_1(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}}\big) \\ \tau^+_{\boldsymbol{x}_j} E\big(\frac{\partial Z_{j_0, \boldsymbol{x}_j}(\theta_j^0, \boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}}\big) & E\big(\frac{\partial \boldsymbol{Z}_1(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}}\big)^T & \mathbf{0} \end{pmatrix} \end{aligned}$$

with j_0 being any element in $A^+_{\boldsymbol{x}_j}$. Put $a_1 = \tau^+_{\boldsymbol{x}_j} E\big(Z_{j_0,\boldsymbol{x}_j}(\theta^0_j, \boldsymbol{\beta}^0)\big)^2$, $A_1 = E\big(\boldsymbol{Z}_1(\boldsymbol{\beta}^0)\boldsymbol{Z}_1^T(\boldsymbol{\beta}^0)\big)$, and write

$$S_{11} = \begin{pmatrix} -a_1 & \mathbf{0}^T \\ \mathbf{0} & -A_1 \end{pmatrix},\ S_{12} = \begin{pmatrix} \tau^+_{\boldsymbol{x}_j} E\big(\frac{\partial Z_{j_0,\boldsymbol{x}_j}(\theta^0_j,\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}}\big)^T \\ E\big(\frac{\partial \boldsymbol{Z}_1(\boldsymbol{\beta}^0)}{\partial \boldsymbol{\beta}}\big) \end{pmatrix},$$

that is, we write

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^T & \mathbf{0} \end{pmatrix},$$

which gives that

$$S^{-1} = \begin{pmatrix} S_{11}^{-1} - S_{11}^{-1}S_{12}\Delta^{-1}S_{12}^T S_{11}^{-1} & S_{11}^{-1}S_{12}\Delta^{-1} \\ \Delta^{-1}S_{12}^T S_{11}^{-1} & \Delta^{-1} \end{pmatrix},$$

where $\Delta = S_{12}^T S_{11}^{-1} S_{12}$. Therefore

$$\sqrt{n}\begin{pmatrix} \widetilde{\lambda}_2 \\ \widetilde{\boldsymbol{\lambda}}_1 \end{pmatrix} = \big(S_{11}^{-1} - S_{11}^{-1}S_{12}\Delta^{-1}S_{12}^T S_{11}^{-1}\big)\sqrt{n}\begin{pmatrix} -Q_{2n}(\boldsymbol{\beta}^0, 0) \\ -Q_{1n}(\boldsymbol{\beta}^0, \mathbf{0}) \end{pmatrix} + o_p(1) \tag{2.10}$$

and

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) = \Delta^{-1}S_{12}^T S_{11}^{-1}\sqrt{n}\begin{pmatrix} -Q_{2n}(\boldsymbol{\beta}^0, 0) \\ -Q_{1n}(\boldsymbol{\beta}^0, \mathbf{0}) \end{pmatrix} + o_p(1). \tag{2.11}$$

Put $\widetilde{\boldsymbol{\lambda}} = (\widetilde{\lambda}_2, \widetilde{\boldsymbol{\lambda}}_1^T)^T$ and $\boldsymbol{W}_n = \sqrt{n}\big(Q_{2n}(\boldsymbol{\beta}^0, 0), Q_{1n}^T(\boldsymbol{\beta}^0, \mathbf{0})\big)^T$. Then, by Lemma 2.1, (2.10)–(2.11) and Taylor expansion, we have

$$\begin{aligned}
l^P(\theta_j^0|\boldsymbol{x}_j) &= 2\widetilde{\boldsymbol{\lambda}}_1^T \textstyle\sum_{i=1}^n \boldsymbol{Z}_i(\widetilde{\boldsymbol{\beta}}) - \widetilde{\boldsymbol{\lambda}}_1^T \big(\sum_{i=1}^n \boldsymbol{Z}_i(\widetilde{\boldsymbol{\beta}})\boldsymbol{Z}_i^T(\widetilde{\boldsymbol{\beta}})\big)\widetilde{\boldsymbol{\lambda}}_1 \\
&\quad + 2\widetilde{\lambda}_2 \textstyle\sum_{i\in A_{\boldsymbol{x}_j}^+} Z_{i,\boldsymbol{x}_j}(\theta_j^0, \widetilde{\boldsymbol{\beta}}) - \widetilde{\lambda}_2 \sum_{i\in A_{\boldsymbol{x}_j}^+} \big(Z_{i,\boldsymbol{x}_j}(\theta_j^0, \widetilde{\boldsymbol{\beta}})\big)^2 \widetilde{\lambda}_2 + o_p(1) \\
&= 2\sqrt{n}\widetilde{\boldsymbol{\lambda}}^T \boldsymbol{W}_n - \sqrt{n}\widetilde{\boldsymbol{\lambda}}^T \begin{pmatrix} a_1 & \mathbf{0}^T \\ \mathbf{0} & A_1 \end{pmatrix} \sqrt{n}\widetilde{\boldsymbol{\lambda}} + o_p(1) \\
&= -2\boldsymbol{W}_n^T \big(S_{11}^{-1} - S_{11}^{-1}S_{12}\Delta^{-1}S_{12}^T S_{11}^{-1}\big)\boldsymbol{W}_n \\
&\quad + \boldsymbol{W}_n^T \big(S_{11}^{-1} - S_{11}^{-1}S_{12}\Delta^{-1}S_{12}^T S_{11}^{-1}\big) S_{11} \big(S_{11}^{-1} - S_{11}^{-1}S_{12}\Delta^{-1}S_{12}^T S_{11}^{-1}\big)\boldsymbol{W}_n + o_p(1) \\
&= -\boldsymbol{W}_n^T \big(S_{11}^{-1} - S_{11}^{-1}S_{12}\Delta^{-1}S_{12}^T S_{11}^{-1}\big)\boldsymbol{W}_n + o_p(1) \\
&= \big((-S_{11})^{-1/2}\boldsymbol{W}_n\big)^T \big(I_{k\times k} - S_{11}^{-1/2}S_{12}\Delta^{-1}S_{12}^T S_{11}^{-1/2}\big)\big((-S_{11})^{-1/2}\boldsymbol{W}_n\big) + o_p(1),
\end{aligned}$$

where $I_{k\times k}$ denotes the $k \times k$ identity matrix, and k is the dimension of S_{11}. Hence the theorem follows from the facts that $(-S_{11})^{-1/2}\boldsymbol{W}_n \xrightarrow{d} N(0, I_{k\times k})$, the matrix $I_{k\times k} - S_{11}^{-1/2}S_{12}\Delta^{-1}S_{12}^T S_{11}^{-1/2}$ is idempotent and its rank is

$$k - rank(S_{12}\Delta^{-1}S_{12}^T S_{11}^{-1}) = k - rank(\Delta^{-1}\Delta) = k - (k-1) = 1.$$

□

Chapter 3

Two-Step Risk Analysis in Insurance Ratemaking

This chapter is based on my publication: Kang, S., Peng, L. and Golub, A.(2021), Two-step risk analysis in insurance ratemaking, *Scandinavian Actuarial Journal, 2021(6), 532-542*.

In property and casualty insurance ratemaking, a crucial task for actuaries is to identify common risk characteristics of policyholders (covariates) and use them to estimate equitable rates to policyholders by analyzing historical loss data. Precisely, they calculate the insurance premium based on some observed characteristics vector $\boldsymbol{X}_i$, the number of claims N_i, and the corresponding losses $L_{i,1}, \cdots, L_{i,N_i}$ for a total of n policyholders.

A standard method for calculating the insurance premium uses generalized linear models to model the expected number of claims and the expected aggregated loss (see De Jong and Heller (2008) and Anderson et al. (2004)). For more accurate premium calculations, researchers have exerted considerable efforts to model the dependence between frequency and losses. We refer to Frees, Lee, and Yang (2016) for an excellent literature review of dependence models and recent copula regression models and Fung, Badescu, and Lin (2019a), Fung, Badescu, and Lin (2019b) for mixture models.

When an underwriter forecasts the aggregated loss of a future policyholder, it is challenging to derive the conditional distribution of the aggregated loss $S_i = \sum_{j=1}^{N_i} L_{i,j}$ given $\boldsymbol{X}_i$ even if one can fit well the conditional distributions of N_i and $L_{i,j}$'s given $\boldsymbol{X}_i$ by assuming independent losses. In this case, one often computes the conditional risk measure of S_i given $\boldsymbol{X}_i$ by simulating N_i and $L_{i,j}$'s from the fitted parametric models, which is computationally intensive. Therefore, it may be computationally efficient to directly model the conditional distribution of N_i given $\boldsymbol{X}_i$ and the conditional distribution of S_i given $\boldsymbol{X}_i$ and N_i.

In this Chapter, we are interested in estimating the conditional Value-at-Risk of the aggregated loss at level $\alpha \in (0, 1)$, which is defined as

$$\mathrm{VaR}_i(\alpha|\boldsymbol{x}) = \sup\{s : P(S_i \leq s|\boldsymbol{X}_i = \boldsymbol{x}) \leq \alpha\}.$$

The Value-at-Risk measure is widely used for regulatory purpose in the Solvency II's framework for the insurance industry and internal risk management and performance measurement of an insurance company. When $\mathrm{VaR}_i(\alpha|\boldsymbol{X}_i) > 0$ for $i = 1, \cdots, n$, a direct model for the Value-at-Risk is quantile regression by assuming that

$$\mathrm{VaR}_i(\alpha|\boldsymbol{X}_i) = \exp\{\beta_0 + \boldsymbol{\beta}^\tau \boldsymbol{X}_i\}, \tag{3.1}$$

where $\boldsymbol{\beta}^\tau$ denotes the transpose of the vector $\boldsymbol{\beta}$. In this case, an underwriter can forecast the Value-at-Risk of a future policyholder with characteristics vector $\boldsymbol{x}$ by

$$\mathrm{VaR}(\alpha|\boldsymbol{x}) = \exp\{\beta_0 + \boldsymbol{\beta}^\tau \boldsymbol{x}\}. \tag{3.2}$$

Throughout, we assume

$$P(S_i = 0|N_i = 0) = 1, \tag{3.3}$$

which implies

$$P(S_i \le y) = P(N_i = 0) + P(N_i > 0)P(S_i \le y|N_i > 0) \text{ for } y > 0,$$

That is, the Value-at-Risk of S_i at the α level is that of the conditional loss of S_i given $N_i > 0$ at the $\frac{\alpha - P(N_i=0)}{P(N_i>0)}$ level. This motivates Kudryavtsev (2009) to infer model (3.1) by applying the quantile regression technique in Koenker and Bassett Jr (1978) to those S_i's and $\boldsymbol{X}_i$'s with nonzero claims at the adjusted risk level $\frac{n\alpha - \sum_{i=1}^n I(N_i=0)}{\sum_{i=1}^n I(N_i>0)}$ rather than α, where $I(A)$ denotes the indicator function, and $n^{-1}\sum_{i=1}^n I(N_i = 0)$ estimates $P(N_i = 0)$. We refer to Koenker (2005) for an overview of quantile regression.

To improve the risk forecast in Kudryavtsev (2009), Heras, Moreno, and Vilar-Zanón (2018) propose to model $P(N_i = 0)$ as a function of $\boldsymbol{X}_i$ before using quantile regression, which leads to a two-step inference procedure. This two-step inference applies to both categorical and continuous explanatory vector $\boldsymbol{X}_i$. When $\boldsymbol{X}_i$ is categorical, the right hand side of (3.1) is constant for each

category.So, the application of quantile regression to each level, not all levels simultaneously, in Heras, Moreno, and Vilar-Zanón (2018) becomes the empirical quantile estimation based on data in each category. This observation motivates Kang, Peng, and Xiao (2020) to develop an empirical likelihood method to quantify the uncertainty of the two-step risk inference with the empirical quantile estimation instead of quantile regression in the second step. When all covariates are categorical, one has to apply quantile regression to all categories simultaneously. Hence, the quantile regression estimation differs from the empirical quantile in each category as the common intercept links all categories together. Also, one may wonder how to quantify the inference uncertainty. Note that Heras, Moreno, and Vilar-Zanón (2018) do not address uncertainty quantification, and Kang, Peng, and Xiao (2020) only deal with categorical covariates by using empirical quantile for each category.

This paper has two objectives. First, we propose a random weighted bootstrap method to quantify the uncertainty of the two-step inference in Heras, Moreno, and Vilar-Zanón (2018). Although Heras, Moreno, and Vilar-Zanón (2018) do not derive the asymptotic limit of their Value-at-Risk estimation, the asymptotic variance is expected to be very complicated, and it is challenging to estimate it explicitly. To avoid estimating the complicated asymptotic variance, one may employ the parametric bootstrap method for the logistic regression and a bootstrap method for the quantile regression developed in one of Hahn (1995), Chen, Wei, and Parzen (2004), Feng, He, and Hu (2011), Davidson (2012), and Hagemann (2017). Because of its computational efficiency (see arguments in Chiang, James, and Wang (2005) and Zheng et al. (2018)) and its flexibility on the error structure such as allowing heteroscedasticity (see Zhu (2016)), we adopt the random weighted bootstrap method in Jin, Ying, and Wei (2001) to quantify the two-step inference uncertainty in insurance ratemaking.

Secondly, we propose an alternative two-step inference, where a weighted quantile regression is employed in the second stage. Unlike the quantile regression in Kudryavtsev (2009) and Heras, Moreno, and Vilar-Zanón (2018), we do not adjust the risk level. Again, the asymptotic variance of the Value-at-Risk estimation of the new two-step procedure is complex, and it is infeasible

to estimate it directly and make an analytical comparison with the two-step inference in Heras, Moreno, and Vilar-Zanón (2018). Nevertheless, the simulation study and data analysis show that both two-step inferences are comparable.

We organize this Chapter 3 as follows. Section 3.1 presents the random weighted bootstrap method for quantifying the uncertainty of the two-step inference in Heras, Moreno, and Vilar-Zanón (2018) and a new two-step inference with a random weighted bootstrap method for uncertainty quantification. Sections 3.2 and 3.3 are the empirical analysis and simulation study, respectively. Section 3.4 concludes the paper.

3.1 Methodologies

3.1.1 Uncertainty quantification: random weighted bootstrap method

The first step in Heras, Moreno, and Vilar-Zanón (2018) models the conditional probability of $p_i(\boldsymbol{X}_i) = P(N_i = 0|\boldsymbol{X}_i)$ by logistic regression (see Goldburd et al. (2016), which is

$$I(N_i > 0|\boldsymbol{X}_i) \sim Binomial(1, 1 - p_i(\boldsymbol{X}_i)) \text{ and } \log \frac{1 - p_i(\boldsymbol{X}_i)}{p_i(\boldsymbol{X}_i)} = \gamma_0 + \boldsymbol{\gamma}^\tau \boldsymbol{X}_i. \tag{3.4}$$

Using observations $\{N_i, \boldsymbol{X}_i\}_{i=1}^n$, we can estimate γ_0 and $\boldsymbol{\gamma}$ by maximizing the following log-likelihood function

$$L_1(\gamma_0, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \{I(N_i = 0) \log(\frac{1}{1 + \exp(\gamma_0 + \boldsymbol{\gamma}^\tau \boldsymbol{X}_i)}) + I(N_i > 0) \log(\frac{\exp(\gamma_0 + \boldsymbol{\gamma}^\tau \boldsymbol{X}_i)}{1 + \exp(\gamma_0 + \boldsymbol{\gamma}^\tau \boldsymbol{X}_i)})\}. \tag{3.5}$$

Denote the resulted estimators for γ_0 and $\boldsymbol{\gamma}$ by $\hat{\gamma}_0$ and $\hat{\boldsymbol{\gamma}}$.

In the second step, Heras, Moreno, and Vilar-Zanón (2018) apply quantile regression to S_i's with nonzero claims at an adjusted risk level, depending on the conditional probabilities estimated from the first step. More specifically, separate $\{N_i, S_i, \boldsymbol{X}_i\}_{i=1}^n$ into one group with zero claims and another group with nonzero claims. Denote the group with nonzero claims by $\{\tilde{N}_i, \tilde{S}_i, \tilde{\boldsymbol{X}}_i\}_{i=1}^{\tilde{n}}$, where $\tilde{n} = \sum_{i=1}^n I(N_i > 0)$. Because we assume the aggregate claim amount S_i is zero if the policy

has no claims, i.e., (3.3) holds, the Value-at-Risk of S_i at the α level is that of the conditional loss of S_i given $N_i > 0$ at the $\frac{\alpha - P(N_i=0)}{P(N_i>0)}$ level. Hence, Heras, Moreno, and Vilar-Zanón (2018) apply quantile regression to the sample $\{\tilde{N}_i, \tilde{S}_i, \tilde{\boldsymbol{X}}_i\}_{i=1}^{\tilde{n}}$ at the adjusted level $\frac{\alpha - P(N_i=0)}{P(N_i>0)}$, which minimizes

$$\min_{\beta_0, \boldsymbol{\beta}} \sum_{i=1}^{\tilde{n}} \{\tilde{S}_i - \exp(\beta_0 + \boldsymbol{\beta}^\tau \tilde{\boldsymbol{X}}_i)\} \{\frac{\alpha - \hat{\tilde{p}}_i}{1 - \hat{\tilde{p}}_i} - I(\tilde{S}_i < \exp(\beta_0 + \boldsymbol{\beta}^\tau \tilde{\boldsymbol{X}}_i))\}, \tag{3.6}$$

where

$$\hat{\tilde{p}}_i = \frac{1}{1 + \exp(\hat{\gamma}_0 + \hat{\boldsymbol{\gamma}}^\tau \tilde{\boldsymbol{X}}_i)}$$

is the estimated probability of having no claim from the first step. Denote the resulted quantile regression estimation of β_0 and $\boldsymbol{\beta}$ by $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}$. Finally, the two-step Value-at-Risk estimator for $\mathrm{VaR}(\alpha|\boldsymbol{x})$ in Heras, Moreno, and Vilar-Zanón (2018) is

$$\widehat{\mathrm{VaR}}(\alpha|\boldsymbol{x}) = \exp(\hat{\beta}_0 + \hat{\boldsymbol{\beta}}^\tau \boldsymbol{x}).$$

Note that $\widehat{\mathrm{VaR}}(\alpha|\boldsymbol{x})$ depends on $\hat{\gamma}_0$ and $\hat{\boldsymbol{\gamma}}$ because $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}$ depend on them. The asymptotic normality of $\widehat{\mathrm{VaR}}(\alpha|\boldsymbol{x})$ follows from the standard asymptotic theory, but its asymptotic variance is quite complicated. We skip the details as we do not estimate it explicitly.

To quantify the uncertainty of the above two-step Value-at-Risk estimator and construct a confidence interval for $\mathrm{VaR}(\alpha|\boldsymbol{x})$, which is not discussed in Heras, Moreno, and Vilar-Zanón (2018), one can use a bootstrap method to avoid estimating the complicated asymptotic variance. This will involve the application of the parametric bootstrap method to the logistic regression and a bootstrap method to quantile regression. Researchers have proposed several bootstrap methods for quantile regression; see Hahn (1995), Chen, Wei, and Parzen (2004), Feng, He, and Hu (2011), Davidson (2012), and Hagemann (2017).

This study proposes to use the random weighted bootstrap method in Jin, Ying, and Wei (2001) and Zhu (2016) for quantifying the two-step inference uncertainty, which allows heteroscedastic errors in quantile regression. As argued in Chiang, James, and Wang (2005) and Zheng et al. (2018),

the random weighted bootstrap method is less computationally intensive than other bootstrap methods such as the wild bootstrap method in Feng, He, and Hu (2011). Because of the computational advantage and the flexibility of error structures, we adopt the random weighted bootstrap method to the two-step Value-at-Risk inference as follows.

- *Step i).* Draw a random sample with sample size n from a distribution function with mean one and variance one, say standard exponential distribution. Denote them by $\xi_1^b, \cdots, \xi_n^b$. Independently, draw another random sample with sample size $\tilde{n}$ from a distribution function with mean one and variance one, say $\tilde{\xi}_1^b, \cdots, \tilde{\xi}_{\tilde{n}}^b$.

- *Step ii).* We estimate γ_0 and $\boldsymbol{\gamma}$ by maximizing

$$L_1^b(\gamma_0, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \xi_i^b \{ I(N_i = 0) \log(\frac{1}{1 + \exp(\gamma_0 + \boldsymbol{\gamma}^\tau \boldsymbol{X}_i)}) + I(N_i > 0) \log(\frac{\exp(\gamma_0 + \boldsymbol{\gamma}^\tau \boldsymbol{X}_i)}{1 + \exp(\gamma_0 + \boldsymbol{\gamma}^\tau \boldsymbol{X}_i)}) \},$$

  and denote the resulted estimators for γ_0 and $\boldsymbol{\gamma}$ by $\hat{\gamma}_0^b$ and $\hat{\boldsymbol{\gamma}}^b$.

- *Step iii).* We estimate β_0 and $\boldsymbol{\beta}$ by minimizing

$$L_2^b(\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{\tilde{n}} \tilde{\xi}_i^b \{ \log \tilde{S}_i - (\beta_0 + \boldsymbol{\beta}^\tau \tilde{\boldsymbol{X}}_i) \} \{ \frac{\alpha - \hat{\tilde{p}}_i^b}{1 - \hat{\tilde{p}}_i^b} - I(\log \tilde{S}_i < \beta_0 + \boldsymbol{\beta}^\tau \tilde{\boldsymbol{X}}_i) \},$$

  where

$$\hat{\tilde{p}}_i^b = \frac{1}{1 + \exp(\hat{\gamma}_0^b + \hat{\boldsymbol{\gamma}}^{b\tau} \tilde{\boldsymbol{X}}_i)}.$$

  Denote the resulted estimators for β_0 and $\boldsymbol{\beta}$ by $\hat{\beta}_0^b$ and $\hat{\boldsymbol{\beta}}^b$. Finally, we have the bootstrapped Value-at-Risk estimator

$$\widehat{\mathrm{VaR}}^b(\alpha|\boldsymbol{x}) = \exp(\hat{\beta}_0^b + \hat{\boldsymbol{\beta}}^{b\tau} \boldsymbol{x}).$$

- *Step iv)* Repeat the above steps B times and obtain $\{\widehat{\mathrm{VaR}}^b(\alpha|\boldsymbol{x})\}_{b=1}^B$.

Put $\Delta_b = \widehat{\mathrm{VaR}}^b(\alpha|\boldsymbol{x}) - \widehat{\mathrm{VaR}}(\alpha|\boldsymbol{x})$ for $b = 1, \cdots, B$. Therefore, we can estimate the asymptotic variance of $\widehat{\mathrm{VaR}}(\alpha|\boldsymbol{x})$ by $\hat{\sigma}^2 = \frac{1}{B} \sum_{b=1}^B \Delta_b^2$ and construct a confidence interval for $\mathrm{VaR}(\alpha|\boldsymbol{x})$ with

level $a \in (0,1)$ by either

$$I_1(a|\boldsymbol{x}) = (\widehat{\mathrm{VaR}}(\alpha|\boldsymbol{x}) - \Delta_{B,\lceil \frac{1+a}{2}B \rceil},\ \widehat{\mathrm{VaR}}(\alpha|\boldsymbol{x}) - \Delta_{B,\lceil \frac{1-a}{2}B \rceil})$$

or

$$I_2(a|\boldsymbol{x}) = (\widehat{\mathrm{VaR}}(\alpha|\boldsymbol{x}) - \Delta_{<\lceil B(1-a)\rceil>},\ \widehat{\mathrm{VaR}}(\alpha|\boldsymbol{x}) + \Delta_{<\lceil B(1-a)\rceil>}),$$

where $\Delta_{B,1} \le \cdots \Delta_{B,B}$ denote the order statistics of $\Delta_1, \cdots, \Delta_B$, $\Delta_{<1>} \le \cdots \le \Delta_{<B>}$ denote the order statistics of $|\Delta_1|, \cdots, |\Delta_B|$, and $\lceil x \rceil$ represents the least integer greater than or equal to x. Note that $\{\xi_i^b\}_{i=1}^n$ in Step ii) and $\{\tilde{\xi}_i^b\}_{i=1}^{\tilde{n}}$ in Step iii) are independent because $\{\tilde{S}_i\}_{i=1}^{\tilde{n}}$ are independent of $\{I(N_i = 0)\}_{i=1}^n$ given $\{\boldsymbol{X}_i\}_{i=1}^n$.

When the classical Edgeworth expansion holds, i.e.,

$$P\left(\frac{\sqrt{n}}{\sigma}(\widehat{\mathrm{VaR}}(\alpha|\boldsymbol{x}) - \mathrm{VaR}(\alpha|\boldsymbol{x})) \le x\right) = \Phi(x) + n^{-1/2}\Delta_n(x) + o(n^{-1})$$

for some $\sigma > 0$ and even functions of $\Delta_n(x)$ as $n \to \infty$, where $\Phi(x)$ is the distribution function of a standard normal random variable, we have

$$\begin{aligned} P\left(\mathrm{VaR}(\alpha|\boldsymbol{x}) \in I_1(a|\boldsymbol{x})\right) &= \Phi(\sigma\Delta_{B,\lceil \frac{1+a}{2}B \rceil}) - \Phi(\sigma\Delta_{B,\lceil \frac{1-a}{2}B \rceil}) \\ &\quad + n^{-1/2}\{\Delta_n(\sigma\Delta_{B,\lceil \frac{1+a}{2}B \rceil}) - \Delta_n(x)(\sigma\Delta_{B,\lceil \frac{1-a}{2}B \rceil})\} \end{aligned}$$

and

$$P\left(\mathrm{VaR}(\alpha|\boldsymbol{x}) \in I_2(a|\boldsymbol{x})\right) = \Phi(\sigma\Delta_{<\lceil (1-a)B\rceil>}) - \Phi(-\sigma\Delta_{<\lceil (1-a)B\rceil>}) + O(n^{-1})$$

because of symmetric $\Delta_n(x)$. Hence, one may expect that $I_2(a|\boldsymbol{x})$ is more accurate than $I_1(a|\boldsymbol{x})$. A theoretical verification requires us to derive the Edgeworth expansion of the two-step Value-at-Risk estimator, which is challenging and beyond the scope of this paper. Nevertheless, using the standard asymptotic techniques, we can show that the distribution of $\sqrt{n}(\widehat{\mathrm{VaR}}^b(\alpha|\boldsymbol{x}) - \widehat{\mathrm{VaR}}(\alpha|\boldsymbol{x}))$ approximates that of $\sqrt{n}(\widehat{\mathrm{VaR}}(\alpha|\boldsymbol{x}) - \mathrm{VaR}(\alpha|\boldsymbol{x}))$. That is, the coverage probability of the above confidence intervals is asymptotically correct.

3.1.2 Weighted Quantile Regression

Next, we propose a different two-step inference from Heras, Moreno, and Vilar-Zanón (2018) by using weighted quantile regression without adjusting the risk level. Write

$$\begin{aligned} & E\{(S_i-u)(\alpha-I(S_i-u<0))|\boldsymbol{X}_i\} \\ = \; & E\{(S_i-u)(\alpha-I(S_i<u))|N_i>0,\boldsymbol{X}_i\}\{1-p_i(\boldsymbol{X}_i)\}-u(\alpha-1)p_i(\boldsymbol{X}_i), \end{aligned} \tag{3.7}$$

where $p_i(\boldsymbol{X}_i)=P(N_i=0|\boldsymbol{X}_i)$ as before. Because the α-quantile of S_i is equivalent to minimizing the left hand side of (3.7) for u, it is equal to minimizing the right hand side of (3.7) as well, which motivates us to estimate β_0 and $\boldsymbol{\beta}$ by minimizing

$$\begin{aligned} & \textstyle\sum_{i=1}^{\tilde{n}}\Big\{\Big(\tilde{S}_i-\exp(\beta_0+\boldsymbol{\beta}^\tau\tilde{\boldsymbol{X}}_i)\Big)\Big(\alpha-I(\tilde{S}_i<\exp(\beta_0+\boldsymbol{\beta}^\tau\tilde{\boldsymbol{X}}_i))\Big)(1-\tilde{p}_i) \\ & -(\alpha-1)\tilde{p}_i\exp(\beta_0+\boldsymbol{\beta}^\tau\tilde{\boldsymbol{X}}_i)\Big\}, \end{aligned} \tag{3.8}$$

where $\tilde{p}_i=\frac{1}{1+\exp(\gamma_0+\boldsymbol{\gamma}^\tau\tilde{\boldsymbol{X}}_i)}$. Therefore, our new two-step inference estimates γ_0 and $\boldsymbol{\gamma}$ by maximizing (3.5) to get $\hat{\tilde{p}}_i=\frac{1}{1+\exp(\hat{\gamma}_0+\hat{\boldsymbol{\gamma}}^\tau\tilde{\boldsymbol{X}}_i)}$ and estimates β_0 and $\boldsymbol{\beta}$ by minimizing (3.8) with $\tilde{p}_i$ replaced by $\hat{\tilde{p}}_i$ to get $\tilde{\beta}_0$ and $\tilde{\boldsymbol{\beta}}$. Finally, the new Value-at-Risk estimator for $\mathrm{VaR}(\alpha|\boldsymbol{x})$ is

$$\widetilde{\mathrm{VaR}}(\alpha|\boldsymbol{x})=\exp(\tilde{\beta}_0+\tilde{\boldsymbol{\beta}}^\tau\boldsymbol{x}).$$

Rewrite (3.8) as

$$\begin{aligned} & \textstyle\sum_{i=1}^{\tilde{n}}\Big\{\Big((1-\tilde{p}_i)\tilde{S}_i-\exp(\log(1-\tilde{p}_i)+\beta_0+\boldsymbol{\beta}^\tau\tilde{\boldsymbol{X}}_i)\Big) \\ & \times\Big(\alpha-I((1-\tilde{p}_i)\tilde{S}_i<\exp(\log(1-\tilde{p}_i)+\beta_0+\boldsymbol{\beta}^\tau\tilde{\boldsymbol{X}}_i))\Big) \\ & -(\alpha-1)\tilde{p}_i\exp(\beta_0+\boldsymbol{\beta}^\tau\tilde{\boldsymbol{X}}_i)\Big\}. \end{aligned}$$

Therefore, the weighted quantile regression is based on the scaled positive losses $\{(1-p_i)\tilde{S}_i\}$, while the quantile regression in Heras, Moreno, and Vilar-Zanón (2018) is based on losses $\{\tilde{S}_i\}$ and the adjusted risk level $\{\frac{\alpha-p_i}{1-p_i}\}$.

To quantify the uncertainty of the above new Value-at-Risk estimator, we employ a similar random weighted bootstrap method as before, where $L_2^b(\beta_0, \boldsymbol{\beta})$ in Step iii) is replaced by

$$\sum_{i=1}^{\tilde{n}} \tilde{\xi}_i^b \Big\{ \Big(\tilde{S}_i - \exp(\beta_0 + \boldsymbol{\beta}^\tau \tilde{\boldsymbol{X}}_i)\Big) \Big(\alpha - I(\tilde{S}_i < \exp(\beta_0 + \boldsymbol{\beta}^\tau \tilde{\boldsymbol{X}}_i))\Big) (1 - \hat{\tilde{p}}_i^b) \\ -(\alpha - 1)\hat{\tilde{p}}_i^b \exp(\beta_0 + \boldsymbol{\beta}^\tau \tilde{\boldsymbol{X}}_i) \Big\}$$

with $\hat{\tilde{p}}_i^b = \frac{1}{1+\exp(\hat{\gamma}_0^b + \hat{\boldsymbol{\gamma}}^{b\tau} \tilde{\boldsymbol{X}}_i)}$.

Denote the bootstrapped risk estimators by $\{\widetilde{\mathrm{VaR}}^b(\alpha|\boldsymbol{x})\}_{b=1}^B$. Put

$$\tilde{\Delta}_b = \widetilde{\mathrm{VaR}}^b(\alpha|\boldsymbol{x}) - \widetilde{\mathrm{VaR}}(\alpha|\boldsymbol{x}) \text{ for } b = 1, \cdots, B.$$

Then, we can estimate the asymptotic variance of $\widetilde{\mathrm{VaR}}(\alpha|\boldsymbol{x})$ by $\tilde{\sigma}^2 = \frac{1}{B}\sum_{b=1}^B \tilde{\Delta}_b^2$ and construct a confidence interval for $\mathrm{VaR}(\alpha|\boldsymbol{x})$ with level $a \in (0,1)$ by either

$$\tilde{I}_1(a|\boldsymbol{x}) = (\widetilde{\mathrm{VaR}}(\alpha|\boldsymbol{x}) - \tilde{\Delta}_{\lceil B, \frac{1+a}{2} B \rceil},\ \widetilde{\mathrm{VaR}}(\alpha|\boldsymbol{x}) - \tilde{\Delta}_{\lceil B, \frac{1-a}{2} B \rceil})$$

or

$$\tilde{I}_2(a|\boldsymbol{x}) = (\widetilde{\mathrm{VaR}}(\alpha|\boldsymbol{x}) - \tilde{\Delta}_{<\lceil B(1-a) \rceil>},\ \widetilde{\mathrm{VaR}}(\alpha|\boldsymbol{x}) + \tilde{\Delta}_{<\lceil B(1-a) \rceil>}),$$

where $\tilde{\Delta}_{B,1} \leq \cdots \tilde{\Delta}_{B,B}$ denote the order statistics of $\tilde{\Delta}_1, \cdots, \tilde{\Delta}_B$ and $\tilde{\Delta}_{<1>} \leq \cdots \leq \tilde{\Delta}_{<B>}$ denote the order statistics of $|\tilde{\Delta}_1|, \cdots, |\tilde{\Delta}_B|$. Again, $\tilde{I}_2(a|\boldsymbol{x})$ may be more accurate than $\tilde{I}_1(a|\boldsymbol{x})$ as argued before.

3.2 Data Analysis

We analyze the Australian automobile insurance dataset, which has the total number of policyholders $n = 67856$; see De Jong and Heller (2008) for details. This dataset is available in the R packages "insuranceData" and "CASdatasets". Like Heras, Moreno, and Vilar-Zanón (2018), we consider these two explanatory variables: the age of the vehicle with four levels and the driver's age with six levels. That is, the categorical explanatory vector $\boldsymbol{X}_i$ has 24 different levels (or groups).

Because each of the policy has a different exposure rate, we adjust the exposure rates at the first step in fitting the model (3.4). More specifically, we replace the second equation in (3.4) by

$$\log \frac{(1-p_i(\boldsymbol{X}_i))/R_i}{1-(1-p_i(\boldsymbol{X}_i))/R_i} = \gamma_0 + \boldsymbol{\gamma}^\tau \boldsymbol{X}_i,$$

where R_i is the exposure rate of the ith policyholder. Hence,

$$p_i(\boldsymbol{X}_i) = 1 - R_i \frac{\exp(\gamma_0 + \boldsymbol{\gamma}^\tau \boldsymbol{X}_i)}{1+\exp(\gamma_0 + \boldsymbol{\gamma}^\tau \boldsymbol{X}_i)}. \tag{3.9}$$

To estimate γ_0 and $\boldsymbol{\gamma}$, we replace (3.5) by

$$\begin{aligned} L_1(\gamma_0, \boldsymbol{\gamma}) \quad &= \textstyle\sum_{i=1}^n \{ I(N_i = 0) \log(1 - R_i \frac{\exp(\gamma_0+\boldsymbol{\gamma}^\tau \boldsymbol{X}_i)}{1+\exp(\gamma_0+\boldsymbol{\gamma}^\tau \boldsymbol{X}_i)}) \\ &\quad + I(N_i > 0) \log(R_i \frac{\exp(\gamma_0+\boldsymbol{\gamma}^\tau \boldsymbol{X}_i)}{1+\exp(\gamma_0+\boldsymbol{\gamma}^\tau \boldsymbol{X}_i)}) \}. \end{aligned}$$

To estimate β_0 and $\boldsymbol{\beta}$ in the second step of quantile regression, we use the new $p_i(\boldsymbol{X}_i)$ in (3.9), adjusting the exposure rates.

For minimizing (3.6), we first use the 'optim' function in the R statistical software with the initial values being the estimates for the first group in Table 7 of Heras, Moreno, and Vilar-Zanón (2018). Using the obtained estimates as initial values, we use 'optim' to minimize (3.6) again to get the final estimates. We use the same procedure to minimize (3.8). We employ $B = 1000$ in the random weighted bootstrap method. Note that we carry out the minimization for all groups simultaneously rather than for each group in Heras, Moreno, and Vilar-Zanón (2018).

Table 3.1 reports the two-step Value-at-Risk estimates of $\widehat{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)$ and $\widetilde{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)$ in the 3rd and 5th column and their 90% confidence intervals of $I_2(0.9|\boldsymbol{x}_j)$ and $\tilde{I}_2(0.9|\boldsymbol{x}_j)$ in the 4th and 6th column for each group. We also copy the two-step Value-at-Risk estimates from Heras, Moreno, and Vilar-Zanón (2018) in the 2nd column of Table 3.1. Note that Heras, Moreno, and Vilar-Zanón (2018) apply quantile regression to each category. The two numbers inside the bracket of Group represent the levels of the age of the vehicle and the driver's age, respectively.

Table 3.1 shows that these three Value-at-Risk estimates are slightly different but not different with statistical significance, as they are within the reported two intervals. To gain some insight into the accuracy of these two two-step inferences, we conduct a simulation study below with settings close to the real dataset.

3.3 Simulation study

This section investigates the finite sample performance of the two-step inferences and the coverage accuracy of the suggested random weighted bootstrap confidence intervals.

Firstly, we compute the sample proportions $\tau_{\boldsymbol{x}_1}, \ldots, \tau_{\boldsymbol{x}_{24}}$ of those 24 categories in the real dataset analyzed above and generate the explanatory variables $\boldsymbol{X}_i = (\mathrm{VehAge}_i, \mathrm{AgeCat}_i)^\tau$ with sample size $n_j = \lceil n\tau_{\boldsymbol{x}_j} \rceil$ for the j-th category, where $j = 1, \ldots, 24$, and $n = 30000$, or 50000. Here, VehAge_i denotes the age of the vehicle with levels 1 (newest), 2, 3, and 4, and AgeCat_i denotes the driver's age with levels 1 (youngest), 2, 3, 4, 5, and 6. We remark that the total sample size $n^* = n_1 + \cdots + n_{24}$ may be smaller than n due to the rounding effect.

Secondly, using the estimates $\hat{\gamma}_0$ and $\hat{\boldsymbol{\gamma}}$ in fitting (3.4) to the real dataset in Section 3.2, we generate independent Bernoulli random variables with $p_i = \frac{1}{1+\exp\{-\hat{\gamma}_0 - \hat{\boldsymbol{\gamma}}^T \tilde{\boldsymbol{X}}_i\}}$ for $i = 1, \ldots, n^*$ and denote them by $\{N_i\}_{i=1}^{n^*}$.

Thirdly, for each $i = 1, \ldots, n^*$, we generate S_i's from the Pareto distribution $F(x) = 1 - (1 + ax)^{-b}$ when $N_i = 1$. We set $S_i = 0$ when $N_i = 0$. Hence, (3.3) holds. We take

$$a = \{(\frac{1-p_i}{1-\alpha})^{-1/b} - 1\} \exp(-\hat{\beta}_0 - \hat{\boldsymbol{\beta}}^\tau \boldsymbol{X}_i),$$

where $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}$ are the estimates from fitting (3.1) to the real dataset in Section 3.2. Therefore, the random sample $\{N_i, S_i, \boldsymbol{X}_i\}_{i=1}^{n^*}$ satisfies models (3.1) and (3.4) with true values $\hat{\gamma}_0$, $\hat{\boldsymbol{\gamma}}$, $\hat{\beta}_0$, and $\hat{\boldsymbol{\beta}}$ being estimates obtained from the real dataset.

We generate 1000 random samples with $b = 1/2$ or 4 from the above setting and use $B = 1000$ in the random weighted bootstrap method. We compute the averages of $\widehat{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)$

and $\widetilde{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)$ and calculate the coverage probabilities of intervals $I_1(0.9|\boldsymbol{x}_j)$, $I_2(0.9|\boldsymbol{x}_j)$, $\tilde{I}_1(0.9|\boldsymbol{x}_j)$, and $\tilde{I}_2(0.9|\boldsymbol{x}_j)$. We report these results in Tables 3.2–3.5.

From Tables 3.2–3.5, we conclude that both two-step inferences are comparable, $I_2(0.9|\boldsymbol{x}_j)$ ($\tilde{I}_2(0.9|\boldsymbol{x}_j)$) is more accurate than $I_1(0.9|\boldsymbol{x}_j)$ ($\tilde{I}_1(0.9|\boldsymbol{x}_j)$), the performance of both methods improves as n becomes larger, and intervals are not accurate in some cases. Hence, it is useful to develop more efficient methods for quantifying the uncertainty of the two-step Value-at-Risk inferences. On the other hand, because bootstrap methods are efficient in general, it may be useful to model the positive aggregated losses parametrically instead of using quantile regression. This will be our future research for improving risk forecasts in insurance ratemaking.

3.4 Conclusions

This study proposes a random weighted bootstrap method to quantify the estimation uncertainty of the two-step inference in Heras, Moreno, and Vilar-Zanón (2018). Further, this study presents an alternative two-step inference using weighted quantile regression in the second step and a random weighted bootstrap method for uncertainty quantification. Unlike the two-step inference in Heras, Moreno, and Vilar-Zanón (2018), the new two-step inference does not need to adjust the risk level. A simulation study shows that both methods are comparable, and interval estimation is not accurate in some situations. To forecast risk at a higher risk level and improve the accuracy of these two-step inferences, we will investigate the possibility of replacing quantile regression in the second step by using generalized Pareto distributions to model the conditional distribution function dynamically like Hall and Tajvidi (2000), Chavez-Demoulin and Embrechts (2004), and Kelly and Jiang (2014).

Table 3.1: We report the two-step estimates of $\{\widehat{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)\}_{j=1}^{24}$ and $\{\widetilde{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)\}_{j=1}^{24}$ in the 3rd and 5th column and their 90% confidence intervals of $I_2(0.9|\boldsymbol{x}_j)$ and $\tilde{I}_2(0.9|\boldsymbol{x}_j)$ in the 4th and 6th columns for each group. We copy the two-step estimates from Heras, Moreno, and Vilar-Zanón (2018) in the 2nd column. The two numbers inside the bracket of Group represent the levels of the age of the vehicle and the driver's age, respectively.

Group	$\mathrm{VaR}(0.95\|\boldsymbol{x}_j)$	$\widehat{\mathrm{VaR}}(0.95\|\boldsymbol{x}_j)$	$I_2(0.9\|\boldsymbol{x}_j)$	$\widetilde{\mathrm{VaR}}(0.95\|\boldsymbol{x}_j)$	$\tilde{I}_2(0.9\|\boldsymbol{x}_j)$
1(2&1)	3212.78	2986.92	(2363.61, 3610.24)	3095.41	(2474.62, 3716.21)
2(1&1)	2534.94	2468.17	(1785.45, 3150.89)	2649.37	(2010.07, 3288.06)
3(3&1)	2901.37	2934.04	(2235.68, 3632.39)	2909.68	(2312.79, 3506.57)
4(2&2)	1726.35	1760.81	(1469.21, 2052.40)	1760.10	(1465.26, 2054.95)
5(4&1)	2927.59	2809.39	(1981.11, 3637.68)	2893.27	(2035.49, 3751.04)
6(1&2)	1407.25	1455.00	(1237.05, 1672.95)	1506.47	(1282.28, 1730.67)
7(2&3)	1556.17	1535.49	(1334.96, 1736.02)	1538.24	(1321.90, 1754.57)
8(1&3)	1327.27	1268.82	(1092.28, 1445.35)	1316.58	(1124.76, 1508.39)
9(2&4)	1347.33	1359.98	(1200.96, 1518.99)	1365.51	(1204.10, 1526.92)
10(3&2)	1691.04	1729.63	(1425.96, 2033.30)	1654.49	(1330.73, 1978.25)
11(1&4)	1143.21	1123.78	(972.85, 1274.72)	1168.74	(991.14, 1346.34)
12(3&3)	1487.55	1508.30	(1252.26, 1764.35)	1445.94	(1167.14, 1724.73)
13(4&2)	1736.64	1565.15	(1377.09, 1935.22)	1645.16	(1358.63, 1931.68)
14(3&4)	1283.53	1335.90	(1139.35, 1532.45)	1283.57	(1082.32, 1484.83)
15(4&3)	1470.00	1444.23	(1215.72, 1672.74)	1437.78	(1178.80,1696.76)
16(4&4)	1311.30	1279.15	(1092.78, 1465.52)	1276.33	(1073.48, 1479.19)
17(2&5)	1014.82	979.38	(837.47, 1121.30)	981.70	(848.92, 1114.49)
18(2&6)	1146.38	1123.59	(924.41, 1322.77)	1130.98	(928.85, 1333.11)
19(1&5)	837.34	809.29	(689.30, 929.28)	840.24	(697.78, 982.70)
20(1&6)	947.30	928.45	(760.88, 1096.02)	968.01	(795.35, 1140.67)
21(3&5)	914.85	962.04	(827.09, 1097.00)	922.80	(783.78, 1061.81)
22(3&6)	1067.53	1103.69	(931.10, 1276.29)	1063.12	(876.76, 1249.47)
23(4&5)	889.11	921.17	(782.09, 1060.26)	917.59	(761.29, 1073.89)
24(4&6)	1062.24	1056.81	(858.59, 1255.03)	1057.12	(846.61, 1267.63)

Table 3.2: $n = 30000$ *and* $b = 1/2$. We report the averages of $\widehat{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)$ and $\widetilde{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)$ and coverage probabilities of $I_1(0.9|\boldsymbol{x}_j)$, $I_2(0.9|\boldsymbol{x}_j)$, $\tilde{I}_1(0.9|\boldsymbol{x}_j)$, and $\tilde{I}_2(0.9|\boldsymbol{x}_j)$. The two numbers inside the bracket of Group represent the levels of the age of the vehicle and the driver's age, respectively.

Group	$\mathrm{VaR}(0.95\|\boldsymbol{x}_j)$	$\widehat{\mathrm{VaR}}(0.95\|\boldsymbol{x}_j)$	$I_1(0.9\|\boldsymbol{x}_j)$	$I_2(0.9\|\boldsymbol{x}_j)$	$\widetilde{\mathrm{VaR}}(0.95\|\boldsymbol{x}_j)$	$\tilde{I}_1(0.9\|\boldsymbol{x}_j)$	$\tilde{I}_2(0.9\|\boldsymbol{x}_j)$
1(2&1)	2986.92	2996.24	0.849	0.917	2997.16	0.858	0.913
2(1&1)	2468.17	2542.60	0.859	0.904	2558.50	0.861	0.914
3(3&1)	2934.04	3029.72	0.866	0.913	3002.17	0.860	0.913
4(2&2)	1760.81	1765.09	0.884	0.933	1765.52	0.886	0.930
5(4&1)	2809.39	2894.18	0.870	0.923	2890.28	0.871	0.921
6(1&2)	1455.00	1497.73	0.888	0.929	1505.08	0.891	0.934
7(2&3)	1535.49	1527.90	0.868	0.913	1533.03	0.879	0.919
8(1&3)	1268.82	1297.20	0.878	0.915	1307.79	0.875	0.911
9(2&4)	1359.98	1343.90	0.868	0.915	1351.40	0.879	0.923
10(3&2)	1729.63	1785.67	0.862	0.899	1769.12	0.878	0.909
11(1&4)	1123.78	1141.69	0.881	0.925	1153.28	0.873	0.929
12(3&3)	1508.30	1544.43	0.867	0.914	1535.47	0.862	0.931
13(4&2)	1565.15	1704.89	0.884	0.926	1701.07	0.879	0.926
14(3&4)	1335.90	1359.44	0.874	0.915	1353.97	0.881	0.933
15(4&3)	1444.23	1475.67	0.876	0.914	1476.64	0.884	0.918
16(4&4)	1279.15	1299.22	0.877	0.916	1302.76	0.883	0.935
17(2&5)	979.38	940.05	0.842	0.906	910.45	0.816	0.873
18(2&6)	1123.59	1131.54	0.872	0.917	1160.68	0.866	0.934
19(1&5)	809.29	798.60	0.882	0.936	780.01	0.869	0.929
20(1&6)	928.45	961.91	0.878	0.913	990.78	0.864	0.935
21(3&5)	962.04	952.40	0.865	0.927	915.70	0.871	0.916
22(3&6)	1103.69	1145.27	0.864	0.909	1165.26	0.852	0.907
23(4&5)	921.17	909.13	0.882	0.934	880.81	0.873	0.933
24(4&6)	1056.81	1093.40	0.869	0.911	1119.17	0.854	0.926

Table 3.3: $n = 30000$ *and* $b = 4$. We report the averages of $\widehat{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)$ and $\widetilde{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)$ and coverage probabilities of $I_1(0.9|\boldsymbol{x}_j)$, $I_2(0.9|\boldsymbol{x}_j)$, $\tilde{I}_1(0.9|\boldsymbol{x}_j)$, and $\tilde{I}_2(0.9|\boldsymbol{x}_j)$. The two numbers inside the bracket of Group represent the levels of the age of the vehicle and the driver's age, respectively.

Group	$\mathrm{VaR}(0.95\|\boldsymbol{x}_j)$	$\widehat{\mathrm{VaR}}(0.95\|\boldsymbol{x}_j)$	$I_1(0.9\|\boldsymbol{x}_j)$	$I_2(0.9\|\boldsymbol{x}_j)$	$\widetilde{\mathrm{VaR}}(0.95\|\boldsymbol{x}_j)$	$\tilde{I}_1(0.9\|\boldsymbol{x}_j)$	$\tilde{I}_2(0.9\|\boldsymbol{x}_j)$
1(2&1)	2986.92	2975.40	0.851	0.918	2967.50	0.864	0.911
2(1&1)	2468.17	2482.52	0.860	0.911	2475.31	0.856	0.904
3(3&1)	2934.04	2954.47	0.856	0.909	2927.52	0.875	0.916
4(2&2)	1760.81	1757.67	0.879	0.925	1756.45	0.880	0.910
5(4&1)	2809.39	2824.89	0.876	0.918	2809.41	0.869	0.916
6(1&2)	1455.00	1466.43	0.903	0.937	1464.96	0.893	0.925
7(2&3)	1535.49	1529.40	0.880	0.919	1531.74	0.863	0.906
8(1&3)	1268.82	1276.12	0.872	0.909	1277.61	0.887	0.913
9(2&4)	1359.98	1349.13	0.871	0.914	1352.68	0.888	0.928
10(3&2)	1729.63	1745.26	0.872	0.904	1733.03	0.878	0.905
11(1&4)	1123.78	1125.86	0.871	0.912	1128.55	0.879	0.926
12(3&3)	1508.30	1518.42	0.873	0.917	1511.20	0.873	0.916
13(4&2)	1565.15	1668.99	0.897	0.929	1662.62	0.891	0.924
14(3&4)	1335.90	1339.61	0.863	0.909	1334.89	0.869	0.911
15(4&3)	1444.23	1452.22	0.885	0.919	1449.99	0.872	0.912
16(4&4)	1279.15	1281.19	0.877	0.920	1280.70	0.878	0.929
17(2&5)	979.38	957.30	0.827	0.896	953.85	0.824	0.925
18(2&6)	1123.59	1124.32	0.842	0.909	1130.40	0.841	0.912
19(1&5)	809.29	798.81	0.866	0.935	796.10	0.866	0.950
20(1&6)	928.45	938.17	0.855	0.917	942.96	0.867	0.915
21(3&5)	962.04	950.84	0.838	0.904	942.18	0.838	0.929
22(3&6)	1103.69	1116.46	0.863	0.903	1115.76	0.839	0.901
23(4&5)	921.17	909.06	0.881	0.940	903.64	0.841	0.943
24(4&6)	1056.81	1067.03	0.860	0.912	1070.15	0.847	0.921

Table 3.4: $n = 50000$ *and* $b = 1/2$. We report the averages of $\widehat{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)$ and $\widetilde{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)$ and coverage probabilities of $I_1(0.9|\boldsymbol{x}_j)$, $I_2(0.9|\boldsymbol{x}_j)$, $\tilde{I}_1(0.9|\boldsymbol{x}_j)$, and $\tilde{I}_2(0.9|\boldsymbol{x}_j)$. The two numbers inside the bracket of Group represent the levels of the age of the vehicle and the driver's age, respectively.

Group	$\mathrm{VaR}(0.95\|\boldsymbol{x}_j)$	$\widehat{\mathrm{VaR}}(0.95\|\boldsymbol{x}_j)$	$I_1(0.9\|\boldsymbol{x}_j)$	$I_2(0.9\|\boldsymbol{x}_j)$	$\widetilde{\mathrm{VaR}}(0.95\|\boldsymbol{x}_j)$	$\tilde{I}_1(0.9\|\boldsymbol{x}_j)$	$\tilde{I}_2(0.9\|\boldsymbol{x}_j)$
1(2&1)	2986.92	3001.04	0.879	0.918	2977.82	0.843	0.903
2(1&1)	2468.17	2523.58	0.877	0.914	2534.26	0.842	0.901
3(3&1)	2934.04	2995.55	0.869	0.906	2974.75	0.848	0.911
4(2&2)	1760.81	1764.54	0.871	0.915	1748.42	0.845	0.901
5(4&1)	2809.39	2868.19	0.883	0.917	2862.67	0.855	0.905
6(1&2)	1455.00	1489.19	0.871	0.905	1487.76	0.855	0.915
7(2&3)	1535.49	1533.26	0.868	0.915	1516.82	0.869	0.925
8(1&3)	1268.82	1294.24	0.859	0.904	1291.63	0.850	0.905
9(2&4)	1359.98	1353.49	0.876	0.908	1347.25	0.855	0.915
10(3&2)	1729.63	1761.19	0.866	0.899	1746.43	0.842	0.901
11(1&4)	1123.78	1141.61	0.862	0.908	1146.16	0.853	0.913
12(3&3)	1508.30	1530.57	0.867	0.904	1515.69	0.847	0.914
13(4&2)	1565.15	1687.25	0.834	0.901	1680.65	0.845	0.913
14(3&4)	1335.90	1350.72	0.875	0.910	1345.78	0.862	0.926
15(4&3)	1444.23	1466.63	0.864	0.894	1459.37	0.837	0.903
16(4&4)	1279.15	1293.48	0.881	0.913	1295.91	0.828	0.906
17(2&5)	979.38	945.20	0.846	0.917	914.75	0.769	0.803
18(2&6)	1123.59	1143.67	0.857	0.901	1160.69	0.871	0.936
19(1&5)	809.29	797.77	0.879	0.924	779.72	0.868	0.921
20(1&6)	928.45	964.62	0.873	0.891	987.04	0.873	0.920
21(3&5)	962.04	944.19	0.865	0.917	915.30	0.827	0.872
22(3&6)	1103.69	1141.45	0.874	0.910	1159.83	0.859	0.933
23(4&5)	921.17	904.26	0.856	0.912	881.06	0.840	0.897
24(4&6)	1056.81	1093.91	0.854	0.897	1116.56	0.844	0.920

Table 3.5: $n = 50000$ *and* $b = 4$. We report the averages of $\widehat{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)$ and $\widetilde{\mathrm{VaR}}(0.95|\boldsymbol{x}_j)$ and coverage probabilities of $I_1(0.9|\boldsymbol{x}_j)$, $I_2(0.9|\boldsymbol{x}_j)$, $\tilde{I}_1(0.9|\boldsymbol{x}_j)$, and $\tilde{I}_2(0.9|\boldsymbol{x}_j)$. The two numbers inside the bracket of Group represent the levels of the age of the vehicle and the driver's age, respectively.

Group	$\mathrm{VaR}(0.95\|\boldsymbol{x}_j)$	$\widehat{\mathrm{VaR}}(0.95\|\boldsymbol{x}_j)$	$I_1(0.9\|\boldsymbol{x}_j)$	$I_2(0.9\|\boldsymbol{x}_j)$	$\widetilde{\mathrm{VaR}}(0.95\|\boldsymbol{x}_j)$	$\tilde{I}_1(0.9\|\boldsymbol{x}_j)$	$\tilde{I}_2(0.9\|\boldsymbol{x}_j)$
1(2&1)	2986.92	2980.41	0.863	0.910	2975.16	0.871	0.907
2(1&1)	2468.17	2483.89	0.855	0.884	2475.33	0.851	0.887
3(3&1)	2934.04	2947.93	0.864	0.900	2931.98	0.870	0.909
4(2&2)	1760.81	1755.30	0.856	0.908	1756.22	0.862	0.898
5(4&1)	2809.39	2820.27	0.884	0.911	2815.07	0.874	0.908
6(1&2)	1455.00	1462.79	0.869	0.907	1461.04	0.875	0.923
7(2&3)	1535.49	1528.98	0.876	0.916	1530.14	0.898	0.927
8(1&3)	1268.82	1274.43	0.858	0.894	1273.19	0.863	0.899
9(2&4)	1359.98	1357.09	0.889	0.922	1355.68	0.882	0.910
10(3&2)	1729.63	1736.04	0.867	0.907	1730.67	0.881	0.908
11(1&4)	1123.78	1130.96	0.876	0.911	1127.76	0.883	0.915
12(3&3)	1508.30	1512.98	0.873	0.911	1507.97	0.877	0.918
13(4&2)	1565.15	1660.94	0.873	0.913	1661.76	0.869	0.905
14(3&4)	1335.90	1342.25	0.896	0.930	1335.99	0.888	0.921
15(4&3)	1444.23	1446.93	0.875	0.912	1448.05	0.872	0.907
16(4&4)	1279.15	1284.26	0.891	0.921	1282.86	0.875	0.911
17(2&5)	979.38	961.11	0.820	0.894	961.03	0.815	0.919
18(2&6)	1123.59	1128.32	0.866	0.907	1131.51	0.864	0.919
19(1&5)	809.29	801.00	0.850	0.920	799.76	0.869	0.938
20(1&6)	928.45	940.07	0.880	0.913	941.28	0.854	0.910
21(3&5)	962.04	950.75	0.849	0.918	947.45	0.840	0.934
22(3&6)	1103.69	1115.89	0.869	0.908	1115.02	0.873	0.927
23(4&5)	921.17	909.62	0.847	0.898	909.74	0.847	0.941
24(4&6)	1056.81	1067.78	0.858	0.905	1070.70	0.858	0.912

Chapter 4

Three-Step Risk Inference In Insurance Ratemaking

This chapter is based on my publication: Hou, Y., Kang, S.K., Lo, C.C. and Peng, L.(2022), Three-step risk inference in insurance ratemaking, *Insurance: Mathematics and Economics, 105, 1-13*.

Consider an insurance dataset $\{\mathbf{X}_i, N_i, \{L_{i,j}\}_{j=1}^{N_i}\}_{i=1}^{n}$ in insurance ratemaking in a given year, where $\mathbf{X}_i$ is an explanatory vector representing some characteristics of the ith policyholder (e.g., age of the policyholder and age of car), N_i is the number of claims, and $\{L_{i,j}\}_{j=1}^{N_i}$ are the corresponding observed losses. Then $S_i = \sum_{j=1}^{N_i} L_{i,j}$ is the aggregate loss of the ith policyholder. A practical question in risk management is to forecast the risk of the aggregate loss S of a future policyholder with characteristic vector $\mathbf{x}$.

Two widely employed conditional risk measures of S given $\mathbf{x}$ in the financial industry and insurance business are the conditional Value-at-Risk (VaR) and conditional Expected Shortfall (ES), defined as

$$\mathrm{VaR_S}(\alpha|\mathbf{x}) = \inf\{\mathrm{s} : \mathrm{P}(\mathrm{S} \leq \mathrm{s}|\mathbf{x}) \geq \alpha\} \text{ and } \mathrm{ES_S}(\alpha|\mathbf{x}) = \mathrm{E}(\mathrm{S}|\mathrm{S} > \mathrm{VaR_S}(\alpha|\mathbf{x}), \mathbf{x}), \text{ respectively.}$$

Practically, regulators often require the risk level α to be high such as 0.99 for VaR and 0.975 for ES, making nonparametric inference inefficient. On the other hand, a parametric inference may lead to an unstable risk forecast due to the higher risk level and the fact that the parametric inference mainly employs the information around the distribution center.

To better appreciate the proposed study, we describe the recent two-step inference procedure in Heras, Moreno, and Vilar-Zanón (2018) for predicting the Value-at-Risk of the aggregate loss. The first step uses the logistic regression model to estimate the probability of having no claim, i.e., $p_i = P(N_i = 0|X_i)$. When the semicontinuous property of S_i admits the following decomposition

$$P(S_i \leq s|\mathbf{X}_i) = P(N_i = 0|\mathbf{X}_i) + P(N_i > 0|\mathbf{X}_i)P(S_i \leq s|\mathbf{X}_i, N_i > 0) \text{ for } s > s_0, \qquad (4.1)$$

the Value-at-Risk of S_i at level α is equal to that of the conditional loss S_i given $N_i > 0$ at the adjusted level $\tilde{\alpha}_i = \frac{\alpha - p_i}{1 - p_i}$ as long as the VaR is above s_0. Therefore, the second step in Heras, Moreno, and Vilar-Zanón (2018) applies quantile regression to those $(S_i, \mathbf{X}_i)$ with positive N_i at the adjusted risk level $\tilde{\alpha}_i$, which depends on the estimator for p_i in the first step. To quantify the uncertainty of this two-step risk forecast, Kang, Peng, and Golub (2021) develops a random weighted bootstrap method when the second step uses empirical quantile estimation rather than quantile regression. Kang, Peng, and Xiao (2020) develops another two-step procedure using weighted quantile regression. A drawback of quantile regression is the infeasible application to other risk measures such as Expected Shortfall or Expectile.

For applications to more general risk measures at a high risk level, this paper proposes to model the conditional excess function of S_i over a dynamic threshold $u_i = u(\mathbf{X}_i)$ given $\mathbf{X}_i$ and $N_i > 0$ parametrically. That is, we need a model for the dynamic threshold and a parametric model for the excess function of S_i (i.e., a semiparametric model for the distribution of S_i). When $F_{i,\mathbf{x}}(s) = P(S_i \leq s | \mathbf{X}_i = \mathbf{x}, N_i > 0)$ is in the domain of attraction of extreme value distribution with index $\xi_{i,\mathbf{x}}$ and right endpoint z_0 (see Resnick (2008) and Embrechts, Klüppelberg, and Mikosch (1999) for an overview about Extreme Value Theory), there exists a function $\sigma_{i,\mathbf{x}}(u) > 0$ such that

$$\lim_{u \to z_0} \sup_{0 \leq z < z_0 - u} |F_{i,\mathbf{x},u}(z) - G_{\xi_{i,\mathbf{x}}, \sigma_{i,\mathbf{x}}(u)}(z)| = 0, \tag{4.2}$$

where

$$F_{i,\mathbf{x},u}(z) = 1 - \frac{1 - F_{i,\mathbf{x}}(u+z)}{1 - F_{i,\mathbf{x}}(u)}$$

is the excess function, and

$$G_{\xi,\sigma}(z) = 1 - (1 + \xi z/\sigma)^{-1/\xi} \tag{4.3}$$

for $1 + \xi z/\sigma > 0$ with $\sigma > 0$ and $\xi \in \mathrm{R}$ is called the generalized Pareto distribution (GPD); see Balkema and De Haan (1974). Therefore, we propose to model the conditional excess function of S_i given $\mathbf{X}_i$ by a GPD with parameters ξ and σ in (4.3) depending on the covariate vector $\mathbf{X}_i$.

Because we do not model losses below the threshold, the resulted risk forecast is robust against high risk levels.

It is not new to model the parameters in the GPD as parametric functions of some covariates; see Chavez-Demoulin, Davison, and McNeil (2005), Kelly and Jiang (2014), Chavez-Demoulin, Embrechts, and Sardy (2014), and Massacci (2017) for financial returns and Hall and Tajvidi (2000) for climate data. However, the threshold in these papers is independent of the covariates. Because of using a dynamic GPD, we employ a quantile regression model to estimate the dynamic threshold and study the following three-step procedure for forecasting risk at a high risk level: i) using logistic regression to model the probability $p_i = P(N_i = 0|\mathbf{X}_i)$ at the first stage, ii) using quantile regression to model the dynamic threshold u_i at the 90% or 95% level at the second stage as a rule of thumb in fitting a generalized Pareto distribution (see Hull (2012)), iii) and fitting a dynamic generalized Pareto distribution to exceedances over the selected dynamic threshold u_i based on those S_i's and $\mathbf{X}_i$'s with $N_i > 0$. Combining these three steps leads to a robust forecast for a given risk measure at a high risk level because we model the distribution of S_i semiparametrically, i.e., we model losses over the threshold parametrically and below the threshold nonparametrically. In contrast, the two-step inference in Heras, Moreno, and Vilar-Zanón (2018) only works for Value-at-Risk. To quantify the uncertainty of the derived risk forecast, we develop a random weighted bootstrap method.

We organize this Chapter 4 as follows. Section 4.1 presents the three-step inference for estimating risk measures at a high level and a random weighted bootstrap method for uncertainty quantification. Section 4.2 analyzes an automobile dataset. We conclude the Chapter 4 in Section 4.3. All theoretical derivations appear in Section 4.4.

4.1 Methodologies and Main Results

4.1.1 Three-step inference for risk measures at a high level

We observe the actuarial dataset $\{\mathbf{X}_i, N_i, \{L_{i,j}\}_{j=1}^{N_i}\}_{i=1}^{n}$ in a given year, where $\mathbf{X}_i$ is the characteristic vector representing the ith policyholder, $N_i \geq 0$ is the number of claims, and $\{L_{i,j} \geq 0\}_{j=1}^{N_i}$ are the

corresponding losses. Define the aggregate loss $S_i = \sum_{j=1}^{N_i} L_{i,j}$. For forecasting a conditional risk measure of S_i given $\mathbf{X}_i$ at a high risk level α such as 0.99, we employ equation (4.1) by modeling $P(N_i = 0|\mathbf{X}_i)$ parametrically in the first step and $P(S_i \leq s|\mathbf{X}_i, N_i > 0)$ semiparametrically in the second and third steps.

Throughout, we write $\bar{\mathbf{X}}_i = (1, \mathbf{X}_i^\tau)^\tau$ and $\bar{\mathbf{x}} = (1, \mathbf{x}^\tau)^\tau$, where A^τ denotes the transpose of matrix or vector A. Like Heras, Moreno, and Vilar-Zanón (2018), the first step models the conditional probability of $p(\mathbf{X}_i) = P(N_i = 0|\mathbf{X}_i)$ by logistic regression:

$$I(N_i = 0)|\mathbf{X}_i \sim \text{Bin}(1, p(\mathbf{X}_i)) \quad \text{and} \quad p(\mathbf{X}_i) = p(\mathbf{X}_i; \boldsymbol{\theta}_1) = \frac{1}{1 + \exp(\boldsymbol{\theta}_1^\tau \bar{\mathbf{X}}_i)}, \tag{4.4}$$

where $I(A)$ is the indicator function of A. The above chosen parametric form ensures $p(\mathbf{X}_i) \in (0, 1)$. The maximum likelihood estimation of $\boldsymbol{\theta}_1$ is

$$\hat{\boldsymbol{\theta}}_1 = \arg\max_{\boldsymbol{\theta}} \prod_{i=1}^{n} \left\{ \frac{1}{1 + \exp(\boldsymbol{\theta}^\tau \bar{\mathbf{X}}_i)} \right\}^{I(N_i=0)} \left\{ 1 - \frac{1}{1 + \exp(\boldsymbol{\theta}^\tau \bar{\mathbf{X}}_i)} \right\}^{I(N_i>0)}.$$

Using these estimators, we estimate $p(\mathbf{x})$ for $\mathbf{X}_i = \mathbf{x}$ by

$$\hat{p}(\mathbf{x}) = p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1) = \frac{1}{1 + \exp(\hat{\boldsymbol{\theta}}_1^\tau \bar{\mathbf{x}})}. \tag{4.5}$$

Further, we estimate the adjusted level $\alpha^*(\mathbf{x}) = \frac{\alpha - p(\mathbf{x})}{1 - p(\mathbf{x})}$ for $\mathbf{X}_i = \mathbf{x}$ by

$$\hat{\alpha}^*(\mathbf{x}) = \frac{\alpha - \hat{p}(\mathbf{x})}{1 - \hat{p}(\mathbf{x})}.$$

To model $P(S_i \leq s|\mathbf{X}_i, N_i > 0)$ semiparametrically, the second and third steps select the threshold by quantile regression and fit the exceedances by a GPD, respectively. For ease of presentation, we write the observations in $\{\mathbf{X}_i, N_i, S_i\}_{i=1}^n$ with nonzero claims as $\{\mathbf{X}_i, N_i, \tilde{S}_i\}_{i=1}^{\tilde{n}}$, i.e., the first $\tilde{n}$ of N_i's are nonzero. Thus, $\tilde{S}_i$ is the conditional loss of S_i given $N_i > 0$. Using the data set $\{\mathbf{X}_i, N_i, \tilde{S}_i\}_{i=1}^{\tilde{n}}$, the second step models the dynamic threshold by the conditional quantile at a

chosen risk level α_0 as

$$u(\mathbf{X}_i) = u(\mathbf{X}_i; \boldsymbol{\theta}_2) = \mathrm{VaR}_{\tilde{S}_i}(\alpha_0|\mathbf{X}_i) \tag{4.6}$$

for $i = 1, \cdots, \tilde{n}$. Note that $\boldsymbol{\theta}_2$ is related to the quantile level α_0 above, but α_0 is a chosen level, independent of the predictor $\mathbf{X}_i$'s and less than the targeted risk level α. As a rule of thumb in Hull (2012), we employ $\alpha_0 = 90\%$ or 95% in practice.

Using quantile regression inference, we estimate $\boldsymbol{\theta}_2$ by

$$\hat{\boldsymbol{\theta}}_2 = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{\tilde{n}} \rho_{\alpha_0}(\tilde{S}_i - u(\mathbf{X}_i; \boldsymbol{\theta})) \tag{4.7}$$

with $\rho_{\alpha_0}(s) = s(\alpha_0 - I(s < 0))$. Further, we estimate the dynamic threshold by

$$\hat{u}(\mathbf{X}_i) = u(\mathbf{X}_i; \hat{\boldsymbol{\theta}}_2). \tag{4.8}$$

The third step models the conditional excess function of $\tilde{S}_i$ over the threshold $u(\mathbf{X}_i)$ in (4.6) given $\mathbf{X}_i$ by the generalized Pareto distribution

$$P(\tilde{S}_i > u(\mathbf{X}_i) + z|\mathbf{X}_i) = (1 - \alpha_0)\{1 + \xi(\mathbf{X}_i)z/\sigma(\mathbf{X}_i)\}^{-1/\xi(\mathbf{X}_i)}, \tag{4.9}$$

where $z > 0$ and $i = 1, \cdots, \tilde{n}$. Note that all $u(\mathbf{X}_i), \xi(\mathbf{X}_i)$ and $\sigma(\mathbf{X}_i)$ depend on the predictors, indicating the dynamic structures in the proposed three-step method. Like (4.6), we assume that

$$\xi(\mathbf{X}_i) = \xi(\mathbf{X}_i; \boldsymbol{\theta}_3) \quad \text{and} \quad \sigma(\mathbf{X}_i) = \sigma(\mathbf{X}_i; \boldsymbol{\theta}_4), \tag{4.10}$$

for $i = 1, \cdots, \tilde{n}$. We will specify the parametric forms later. To infer the GPD, we denote $\boldsymbol{\eta}_3 = (\boldsymbol{\theta}_3^\tau, \boldsymbol{\theta}_4^\tau)^\tau$ and define

$$l_i(\boldsymbol{\eta}_3|z) \quad = -\{1 + \tfrac{1}{\xi(\mathbf{X}_i;\boldsymbol{\theta}_3)}\} \log\left(1 + \tfrac{\xi(\mathbf{X}_i;\boldsymbol{\theta}_3)}{\sigma(\mathbf{X}_i;\boldsymbol{\theta}_4)} z\right) - \log \sigma(\mathbf{X}_i; \boldsymbol{\theta}_4)$$

for $i = 1, \cdots, \tilde{n}$. Then, we estimate $\boldsymbol{\eta}_3$, $\xi(\mathbf{x})$, and $\sigma(\mathbf{x})$ by

$$\hat{\boldsymbol{\eta}}_3 = (\hat{\boldsymbol{\theta}}_3^\tau, \hat{\boldsymbol{\theta}}_4^\tau)^\tau = \arg\max_{\boldsymbol{\eta}} \sum_{i=1}^{\tilde{n}} I(\tilde{S}_i > \hat{u}(\mathbf{X}_i)) l_i(\boldsymbol{\eta} | \tilde{S}_i - \hat{u}(\mathbf{X}_i)),$$

$$\hat{\xi}(\mathbf{x}) = \xi(\mathbf{x}; \hat{\boldsymbol{\theta}}_3), \quad \text{and} \quad \hat{\sigma}(\mathbf{x}) = \sigma(\mathbf{x}; \hat{\boldsymbol{\theta}}_4).$$

Finally, we predict the conditional Value-at-Risk and conditional Expected Shortfall of S given $\mathbf{x}$ by combining the above three steps. Because for high risk level α, we consider the case of $\alpha^*(\mathbf{x}) \geq \alpha_0$ and assume $N_i > 0$ when $S_i > \mathrm{VaR}_S(\alpha|\mathbf{x})$. Therefore,

$$\mathrm{VaR}_S(\alpha|\mathbf{x}) = \mathrm{VaR}_{\tilde{S}}(\alpha^*(\mathbf{x})|\mathbf{x}) = u(\mathbf{x}) + \frac{\sigma(\mathbf{x})}{\xi(\mathbf{x})} \left\{ \left(\frac{1-\alpha^*(\mathbf{x})}{1-\alpha_0} \right)^{-\xi(\mathbf{x})} - 1 \right\}$$

and

$$\begin{aligned} \mathrm{ES}_S(\alpha|\mathbf{x}) &= \tfrac{1-\alpha^*(\mathbf{x})}{1-\alpha} \mathrm{VaR}_S(\alpha|\mathbf{x}) + \tfrac{1-\alpha_0}{1-\alpha} \tfrac{\sigma(\mathbf{x})}{1-\xi(\mathbf{x})} \left\{ 1 + \xi(\mathbf{x}) \tfrac{\mathrm{VaR}_S(\alpha|\mathbf{x}) - u(\mathbf{x})}{\sigma(\mathbf{x})} \right\}^{-1/\xi(\mathbf{x})+1} \\ &= \tfrac{1-\alpha^*(\mathbf{x})}{1-\alpha} \left\{ \tfrac{\mathrm{VaR}_S(\alpha|\mathbf{x})}{1-\xi(\mathbf{x})} + \tfrac{\sigma(\mathbf{x}) - \xi(\mathbf{x}) u(\mathbf{x})}{1-\xi(\mathbf{x})} \right\} \end{aligned}$$

when $0 < \xi(\mathbf{x}) < 1$. Because of our focus on insurance losses, we assume $\xi(\mathbf{x}) > 0$. The existence of expected shortfall requires $\xi(\mathbf{x}) < 1$. Plugging estimates in the three steps into the above risk measures leads to our risk forecasts

$$\widehat{\mathrm{VaR}}_S(\alpha|\mathbf{x}) = \hat{u}(\mathbf{x}) + \frac{\hat{\sigma}(\mathbf{x})}{\hat{\xi}(\mathbf{x})} \left\{ \left(\frac{1-\hat{\alpha}^*(\mathbf{x})}{1-\alpha_0} \right)^{-\hat{\xi}(\mathbf{x})} - 1 \right\} \tag{4.11}$$

and

$$\begin{aligned} \widehat{\mathrm{ES}}_S(\alpha|\mathbf{x}) &= \tfrac{1-\hat{\alpha}^*(\mathbf{x})}{1-\alpha} \widehat{\mathrm{VaR}}_S(\alpha|\mathbf{x}) + \tfrac{1-\alpha_0}{1-\alpha} \tfrac{\hat{\sigma}(\mathbf{x})}{1-\hat{\xi}(\mathbf{x})} \left\{ 1 + \hat{\xi}(\mathbf{x}) \tfrac{\widehat{\mathrm{VaR}}_S(\alpha|\mathbf{x}) - \hat{u}(\mathbf{x})}{\hat{\sigma}(\mathbf{x})} \right\}^{-1/\hat{\xi}(\mathbf{x})+1} \\ &= \tfrac{1-\hat{\alpha}^*(\mathbf{x})}{1-\alpha} \left\{ \tfrac{\widehat{\mathrm{VaR}}_S(\alpha|\mathbf{x})}{1-\hat{\xi}(\mathbf{x})} + \tfrac{\hat{\sigma}(\mathbf{x}) - \hat{\xi}(\mathbf{x}) \hat{u}(\mathbf{x})}{1-\hat{\xi}(\mathbf{x})} \right\}. \end{aligned} \tag{4.12}$$

When $\alpha^*(\mathbf{x}) < \alpha_0$, we estimate $P(\tilde{S} \leq s|\mathbf{X} = \mathbf{x})$ for $s \leq u(\mathbf{x})$ nonparametrically and $P(\tilde{S} > s|\mathbf{X} = \mathbf{x})$ for $s > u(\mathbf{x})$ parametrically by the fitted GPD; see Remark 4.1 below.

Remark 4.1. *If $\hat{\alpha}^*(\mathbf{x}) < \alpha_0$ and $\mathbf{X}_i$ is categorical, we estimate $\mathrm{VaR}_S(\alpha|\mathbf{x})$ and $\mathrm{ES}_S(\alpha|\mathbf{x})$ by*

$$\widehat{\mathrm{VaR}}_S(\alpha|\mathbf{x}) = \tilde{G}_{\mathbf{x}}^{-1}(\hat{\alpha}^*(\mathbf{x}))$$

and

$$\begin{aligned}\widehat{\mathrm{ES}}_S(\alpha|\mathbf{x}) &= \frac{\sum_{i=1}^{\tilde{n}} \tilde{S}_i I(\tilde{G}_{\mathbf{x}}^{-1}(\hat{\alpha}^*(\mathbf{x})) < \tilde{S}_i < \tilde{G}_{\mathbf{x}}^{-1}(\alpha_0)) I(\mathbf{X}_i = \mathbf{x})}{(1-\alpha)\sum_{i=1}^{\tilde{n}} I(\mathbf{X}_i = \mathbf{x})} \\ &\quad + \frac{1-\alpha_0}{1-\alpha}\widehat{\mathrm{VaR}}_S(\alpha|\mathbf{x}) + \frac{1-\alpha_0}{1-\alpha}\frac{\hat{\sigma}(\mathbf{x})}{1-\hat{\xi}(\mathbf{x})},\end{aligned}$$

respectively, where

$$\tilde{G}_{\mathbf{x}}(s) = \frac{\sum_{i=1}^{\tilde{n}} I(\tilde{S}_i \le x) I(\mathbf{X}_i = \mathbf{x})}{\sum_{i=1}^{\tilde{n}} I(\mathbf{X}_i = \mathbf{x})}$$

and $\tilde{G}_{\mathbf{x}}^{-1}$ denotes the generalized inverse of $\tilde{G}_{\mathbf{x}}$. When $\mathbf{X}_i$ is continuous, we replace $I(\mathbf{X}_i = \mathbf{x})$ with kernel smoothing estimation; see Fan and Gijbels (1996) for kernel smoothing techniques.

Remark 4.2. *We model $u(\mathbf{x};\boldsymbol{\theta}_2), \xi(\mathbf{x};\boldsymbol{\theta}_3)$, and $\sigma(\mathbf{x};\boldsymbol{\theta}_4)$ parametrically. Because $u(\mathbf{x};\boldsymbol{\theta}_2), \xi(\mathbf{x};\boldsymbol{\theta}_3)$, and $\sigma(\mathbf{x};\boldsymbol{\theta}_4)$ are positive, and the linear function is the simplest approximation, we assume*

$$u(\mathbf{x};\boldsymbol{\theta}_2) = \exp(\bar{\mathbf{x}}^\tau\boldsymbol{\theta}_2), \quad \xi(\mathbf{x};\boldsymbol{\theta}_3) = \exp(\bar{\mathbf{x}}^\tau\boldsymbol{\theta}_3), \quad \textit{and} \quad \sigma(\mathbf{x};\boldsymbol{\theta}_4) = \exp(\bar{\mathbf{x}}^\tau\boldsymbol{\theta}_4). \tag{4.13}$$

Our theorems below use the parameters above, but they are valid for general parametric forms under some regularity conditions. For developing a goodness-of-fit test for the above parametric forms, it is necessary to study the nonparametric smoothing inference of the proposed three-step procedure, which is beyond the scope of this paper.

We state the following regularity conditions before deriving the asymptotic limit of the proposed risk forecasts.

Assumption 4.1. Assume $\{\mathbf{X}_i\}$ is a sequence of independent and identically distributed random vectors with bounded support. Given $\{\mathbf{X}_i\}$, $\{N_i\}$ is a sequence of independent random variables satisfying model (4.4). Given $\{\mathbf{X}_i\}$, $\{S_i\}$ is a sequence of independent random variables satisfying model (4.1). Let $\tilde{S}_i$ denote the conditional variable of S_i given $N_i > 0$. We use $F_i(s|\mathbf{X}_i)$ and

$f_i(s|\mathbf{X}_i)$ to denote the conditional distribution function and conditional density function of $\tilde{S}_i$ given $\mathbf{X}_i$, respectively, satisfying models (4.6), (4.9), and (4.10) with (4.13).

Assumption 4.2. Assume $Ep(\mathbf{X}_i) = p_0 \in (0,1)$, and $\Sigma_1 = E\{p(\mathbf{X}_i)(1-p(\mathbf{X}_i))\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau\}$,

$$\Gamma_2 = E\{f_i(F_i^{-1}(\alpha_0|\mathbf{X}_i)|\mathbf{X}_i)\exp(2\boldsymbol{\theta}_2^\tau\bar{\mathbf{X}}_i)\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau\},\ \Sigma_2 = E\{\exp(2\boldsymbol{\theta}_2^\tau\bar{\mathbf{X}}_i)\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau\},$$

and

$$\Gamma_4 = (1-\alpha_0)E\begin{pmatrix} \frac{2\xi(\mathbf{X}_i)^2}{(1+\xi(\mathbf{X}_i))(1+2\xi(\mathbf{X}_i))}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau & \frac{2+3\xi(\mathbf{X}_i)}{(1+\xi(\mathbf{X}_i))(1+2\xi(\mathbf{X}_i))}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau \\ \frac{2+3\xi(\mathbf{X}_i)}{(1+\xi(\mathbf{X}_i))(1+2\xi(\mathbf{X}_i))}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau & \frac{1}{1+2\xi(\mathbf{X}_i)}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau \end{pmatrix}$$

are positive definite.

Theorem 4.1. *Suppose Assumptions 1 and 2 hold. Then, as $n \to \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) \xrightarrow{d} \Sigma_1^{-1}\mathbf{W}_1,\ \sqrt{n}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) \xrightarrow{d} \frac{1}{\sqrt{1-p_0}}\Gamma_2^{-1}\mathbf{W}_2,$$

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_3 - \boldsymbol{\eta}_3) \xrightarrow{d} \frac{1}{\sqrt{1-p_0}}\Gamma_4^{-1}\mathbf{W}_5,$$

where $\mathbf{W}_5 = (\mathbf{W}_{5,1}^\tau, \mathbf{W}_{5,2}^\tau)^\tau$, $\mathbf{W}_{5,1} = \mathbf{W}_3 + \Gamma_{3,1}\Gamma_2^{-1}\mathbf{W}_2$, $\mathbf{W}_{5,2} = \mathbf{W}_4 + \Gamma_{3,2}\Gamma_2^{-1}\mathbf{W}_2$,

$$\Gamma_{3,1} = (1-\alpha_0)E\left\{\frac{\xi^2(\mathbf{X}_i)}{\sigma(\mathbf{X}_i)(1+\xi(\mathbf{X}_i))(1+2\xi(\mathbf{X}_i))}\exp(\boldsymbol{\theta}_2^\tau\bar{\mathbf{X}}_i)\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau\right\},$$

$$\Gamma_{3,2} = -(1-\alpha_0)E\left\{\frac{1+\xi(\mathbf{X}_i)}{\sigma(\mathbf{X}_i)(1+2\xi(\mathbf{X}_i))}\exp(\boldsymbol{\theta}_2^\tau\bar{\mathbf{X}}_i)\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau\right\},$$

and $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{W}_4$ *are defined in Lemma 4.1 below.*

Define $G(\alpha^*, u, \xi, \sigma) = u + \frac{\sigma}{\xi}\left(\left(\frac{1-\alpha^*}{1-\alpha_0}\right)^{-1/\xi} - 1\right)$ and let $\nabla G(\alpha^*, u, \xi, \sigma)$ denote the gradient of G at $(\alpha^*, u, \xi, \sigma)$. Write $\nabla G_{\mathbf{x}} := \nabla G(\alpha^*(\mathbf{x}), u(\mathbf{x}), \xi(\mathbf{x}), \sigma(\mathbf{x}))$, which has nonzero coordinates. Similarly, define $\tilde{G}(\alpha^*, u, \xi, \sigma) = \frac{1-\alpha^*}{1-\alpha}G(\alpha^*, u, \xi, \sigma) + \frac{1-\alpha_0}{1-\alpha}\frac{\sigma}{1-\xi}\left(1 + \xi\frac{G(\alpha^*, u, \xi, \sigma) - u}{\sigma}\right)^{-1/\xi+1}$, denote $\nabla\tilde{G}(\alpha^*, u, \xi, \sigma)$ as the gradient of $\tilde{G}$ at $(\alpha^*, u, \xi, \sigma)$, and write $\nabla\tilde{G}_{\mathbf{x}} = \nabla\tilde{G}(\alpha^*(\mathbf{x}), u(\mathbf{x}), \xi(\mathbf{x}), \sigma(\mathbf{x}))$.

Thus,

$$\text{VaR}_S(\alpha|\mathbf{x}) = G(\alpha^*(\mathbf{x}), u(\mathbf{x}), \xi(\mathbf{x}), \sigma(\mathbf{x})) \text{ and } \text{ES}_S(\alpha|\mathbf{x}) = \tilde{G}(\alpha^*(\mathbf{x}), u(\mathbf{x}), \xi(\mathbf{x}), \sigma(\mathbf{x})).$$

An application of the delta method to Theorem 4.1 yields the following asymptotic limits of our risk forecasts.

Theorem 4.2. *Suppose conditions in Theorem 4.1 hold. Further assume* $\alpha^*(\mathbf{x}) > \alpha_0$*,* $N_i > 0$ *whenever* $S_i > \text{VaR}_S(\alpha|\mathbf{x})$*, and* $\xi(\mathbf{x}) < 1$ *for estimating* $\text{ES}_S(\alpha|\mathbf{x})$*. Then, as* $n \to \infty$*,*

$$\sqrt{n}\{\widehat{\text{VaR}}_S(\alpha|\mathbf{x}) - \text{VaR}_S(\alpha|\mathbf{x})\} \xrightarrow{d} N(0, \nabla G_{\mathbf{x}}^{\tau} \Sigma_{\mathbf{x}} \nabla G_{\mathbf{x}})$$

and

$$\sqrt{n}\{\widehat{\text{ES}}_S(\alpha|\mathbf{x}) - \text{ES}_S(\alpha|\mathbf{x})\} \xrightarrow{d} N(0, \nabla G'^{\tau}_{\mathbf{x}} \Sigma_{\mathbf{x}} \nabla G'_{\mathbf{x}}),$$

where $\Sigma_{\mathbf{x}}$ *is defined in Lemma 4.5 below.*

4.1.2 Uncertainty quantification

Because the asymptotic variances in Theorem 4.2 above are very complicated, we do not estimate them directly for quantifying our forecast uncertainty. Instead, we adopt the random weighted bootstrap method in Jin, Ying, and Wei (2001), Chiang, James, and Wang (2005), and Zhu (2016) as follows. The idea is to repeat the three-step inference many times using random weighted likelihood and distance.

- *Step 1)* Independently draw a random sample with size n from a distribution function with mean one and variance one, such as standard exponential distribution. Denote them by $\{\delta_i^b\}_{i=1}^n$. Write $\{\tilde{\delta}_i^b, \mathbf{X}_i, N_i, \tilde{S}_i\}_{i=1}^{\tilde{n}}$ as those of $(\delta_i^b, \mathbf{X}_i, N_i, S_i)$'s with nonzero claim.

- *Step 2)* Maximize

$$\sum_{i=1}^{n} \delta_i^b \left\{ I(N_i = 0) \log \left(\frac{1}{1 + \exp(\boldsymbol{\theta}_1^\tau \bar{\mathbf{X}}_i)} \right) + I(N_i > 0) \log \left(\frac{\exp(\boldsymbol{\theta}_1^\tau \bar{\mathbf{X}}_i)}{1 + \exp(\boldsymbol{\theta}_1^\tau \bar{\mathbf{X}}_i)} \right) \right\}$$

and denote the resulted estimator by $\hat{\boldsymbol{\theta}}_1^b$. Compute

$$\hat{p}^b(\mathbf{x}) = p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1^b) = \frac{1}{1 + \exp((\hat{\boldsymbol{\theta}}_1^b)^\tau \bar{\mathbf{x}})} \quad \text{and} \quad \hat{\alpha}^{*b}(\mathbf{x}) = \frac{\alpha - \hat{p}^b(\mathbf{x})}{1 - \hat{p}^b(\mathbf{x})}.$$

- *Step 3)* Minimize

$$\sum_{i=1}^{\tilde{n}} \delta_i^b \rho_{\alpha_0}(\tilde{S}_i - u(\mathbf{X}_i; \boldsymbol{\theta}^b)).$$

Denote the resulted estimator by $\hat{\boldsymbol{\theta}}_2^b$ and estimate the threshold by

$$\hat{u}^b(\mathbf{x}) = u(\mathbf{x}; \hat{\boldsymbol{\theta}}_2^b).$$

- *Step 4)* Maximize

$$\sum_{i=1}^{\tilde{n}} \delta_i^b I(\tilde{S}_i > \hat{u}^b(\mathbf{X}_i)) l_i(\boldsymbol{\eta} | \tilde{S}_i - \hat{u}^b(\mathbf{X}_i))$$

and denote the resulted estimator by $\hat{\boldsymbol{\eta}}_3^b$.

- *Step 5)* Compute

$$\widehat{\mathrm{VaR}}_S^b(\alpha|\mathbf{x}) = \hat{u}^b(\mathbf{x}) + \frac{\hat{\sigma}^b(\mathbf{x})}{\hat{\xi}^b(\mathbf{x})} \left\{ \left(\frac{1 - \hat{\alpha}^{*b}(\mathbf{x})}{1 - \alpha_0} \right)^{-\hat{\xi}^b(\mathbf{x})} - 1 \right\}$$

and

$$\begin{aligned} \widehat{\mathrm{ES}}_S^b(\alpha|\mathbf{x}) &= \tfrac{1-\hat{\alpha}^{*b}(\mathbf{x})}{1-\alpha} \widehat{\mathrm{VaR}}_S^b(\alpha|\mathbf{x}) + \tfrac{1-\alpha_0}{1-\alpha} \tfrac{\hat{\sigma}^b(\mathbf{x})}{1-\hat{\xi}^b(\mathbf{x})} \left\{ 1 + \hat{\xi}^b(\mathbf{x}) \tfrac{\widehat{\mathrm{VaR}}_S^b(\alpha|\mathbf{x}) - \hat{u}^b(\mathbf{x})}{\hat{\sigma}^b(\mathbf{x})} \right\}^{-1/\hat{\xi}^b(\mathbf{x})+1} \\ &= \tfrac{1-\hat{\alpha}^{*b}(\mathbf{x})}{1-\alpha} \left\{ \tfrac{\widehat{\mathrm{VaR}}_S^b(\alpha|\mathbf{x})}{1-\hat{\xi}^b(\mathbf{x})} + \tfrac{\hat{\sigma}^b(\mathbf{x}) - \hat{\xi}^b(\mathbf{x}) \hat{u}^b(\mathbf{x})}{1-\hat{\xi}^b(\mathbf{x})} \right\}, \end{aligned}$$

where

$$\hat{\xi}^b(\mathbf{x}) = \xi(\mathbf{x}; \hat{\boldsymbol{\theta}}_3^b) = \exp\{\bar{\mathbf{x}}^\tau \hat{\boldsymbol{\theta}}_3^b\} \text{ and } \hat{\sigma}^b(\mathbf{x}) = \sigma(\mathbf{x}; \hat{\boldsymbol{\theta}}_4^b) = \exp\{\bar{\mathbf{x}}^\tau \hat{\boldsymbol{\theta}}_4^b\}.$$

- *Step 6)* Repeat the above steps B times and obtain $\{\widehat{\mathrm{VaR}}_S^b(\alpha|\mathbf{x})\}_{b=1}^B$ and $\{\widehat{\mathrm{ES}}_S^b(\alpha|\mathbf{x})\}_{b=1}^B$.

- *Step 7)* Let $\Delta_b = \widehat{VaR}_S^b(\alpha|\mathbf{x}) - \widehat{VaR}_S(\alpha|\mathbf{x})$ for $b = 1, \cdots, B$ and estimate the asymptotic variance of $\widehat{VaR}_S(\alpha|\mathbf{x})$ by $\hat{\sigma}^2 = \frac{1}{B}\sum_{b=1}^B \Delta_b^2$. Hence, the confidence intervals with level a for $VaR_S(\alpha|\mathbf{x})$ are either

$$I_1(a|\mathbf{x}) = (\widehat{VaR}_S(\alpha|\mathbf{x}) - \Delta_{B,\lceil \frac{1+a}{2}B \rceil},\ \widehat{VaR}_S(\alpha|\mathbf{x}) - \Delta_{B,\lceil \frac{1-a}{2}B \rceil})$$

or

$$I_2(a|\mathbf{x}) = (\widehat{VaR}_S(\alpha|\mathbf{x}) - \Delta_{<\lceil B(1-a)\rceil>},\ \widehat{VaR}_S(\alpha|\mathbf{x}) + \Delta_{<\lceil B(1-a)\rceil>}),$$

where $\Delta_{B,1} \leq \cdots \leq \Delta_{B,B}$ denote the order statistics of $\Delta_1, \cdots, \Delta_B$, $\Delta_{<1>} \leq \cdots \leq \Delta_{<B>}$ denote the order statistics of $|\Delta_1|, \cdots, |\Delta_B|$, and $\lceil x \rceil$ represents the least integer greater than or equal to x. Similarly, we can construct confidence intervals for the conditional Expected Shorfall.

The theorem below shows that the coverage probabilities of the above proposed intervals are asymptotically correct.

Theorem 4.3. *Under the conditions of Theorem 4.2, as $B \to \infty$ and $n \to \infty$,*

$$\frac{\widehat{\mathrm{VaR}}_S(\alpha|\mathbf{x}) - \mathrm{VaR}_S(\alpha|\mathbf{x})}{\sqrt{B^{-1}\sum_{b=1}^B (\widehat{\mathrm{VaR}}_S^b(\alpha|\mathbf{x}) - \widehat{\mathrm{VaR}}_S(\alpha|\mathbf{x}))^2}} \xrightarrow{d} N(0,1)$$

and

$$\frac{\widehat{\mathrm{ES}}_S(\alpha|\mathbf{x}) - \mathrm{ES}_S(\alpha|\mathbf{x})}{\sqrt{B^{-1}\sum_{b=1}^B (\widehat{\mathrm{ES}}_S^b(\alpha|\mathbf{x}) - \widehat{\mathrm{ES}}_S(\alpha|\mathbf{x}))^2}} \xrightarrow{d} N(0,1).$$

4.2 Data Analysis

In this section, we analyze the Australian automobile insurance data by using the proposed three-step inference method. The data set includes 67,856 one-year vehicle insurance policies in Australia between 2004 and 2005, which is available in the R package 'InsuranceData' (see Wolny-Dominiak and Trzesiok (2014)). We refer to De Jong and Heller (2008) for a detailed description.

Our goal is to predict the conditional VaR at level 99% and the conditional ES at level 97.5% of the aggregate loss given two influential dependent variables, the age of the vehicle and the driver's age, following the variable selection in Heras, Moreno, and Vilar-Zanón (2018). That is, the dimension of $\mathbf{X}_i$ is two. These two categorical variables have four and six levels, respectively, the combination of which results in a total of 24 distinct levels of explanatory vector $\mathbf{X}_i$. We select a dynamic threshold using $\alpha_0 = 90\%$ and employ $B = 5000$ in the proposed random weighted bootstrap method.

Table 4.1 reports estimates in fitting logistic regression in the first step and quantile regression in the second step. The upper panel displays $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ in fitting logistic regression and quantile regressions, respectively. The lower panel shows $\hat{p}(\mathbf{X}_i)$ and $\hat{u}(\mathbf{X}_i)$ for each category in fitting logistic regression and quantile regression, respectively. Table 4.2 reports estimates in the third step, where we choose $\xi(\mathbf{x}; \boldsymbol{\theta}_3)$ as constant following Hall and Tajvidi (2000), i.e., ξ is independent of $\mathbf{X}_i$. Using these fittings above, Tables 4.3 and 4.4 report the predictions for the conditional VaR at level 99% and conditional ES at level 97.5% within each category, respectively. The two numbers inside the bracket of the first column represent the combination of the levels of the two explanatory variables. In Table 4.3, the 2nd column represents the number of observations of each category, the 3rd column is the naive estimates of VaR at level 99% (i.e., nonparametric estimation without using the second and third steps), the 4th column is the GPD estimates with a static threshold chosen as the 90% quantile of all positive losses, the 5th column is the three-step estimates $\{\widehat{\mathrm{VaR}}(0.99|\mathbf{x}_j)\}_{j=1}^{24}$, and the 6th and 7th columns are the proposed 90% confidence intervals $I_1(0.9|\mathbf{x}_j)$ and $I_2(0.9|\mathbf{x}_j)$,

respectively. Likewise, Table 4.4 reports estimates and intervals for the conditional ES at 97.5% level.

Our observations from Tables 4.3 and 4.4 are as follows.The naive estimates of VaR and ES are smaller than those computed from the GPD estimates with a static and dynamic thresholds, except for the first five categories in Table 4.4, which may mean that naive estimators tend to underestimate high risks. Because the naive estimators of VaR and ES are outside the intervals except for category 10 for ES, the three-step forecast is significantly different from the naive forecast. Also, the GPD estimates for VaR with a static threshold are outside the intervals for some categories, implying that the GPD estimates with a static and dynamic threshold forecast VaR differently. In contrast, there is no significant difference between the GPD estimates with a static and dynamic threshold for forecasting ES. Further, intervals $I_2(0.9|\mathbf{x}_j)$ are slightly more skewed to the right than $I_1(0.9|\mathbf{x}_j)$. To check the GPD fit visually, we use the PP-plots computed from the GPD estimates with a fixed (90% quantile) and dynamic threshold (quantile regression). It is seen from Figure 4.1 that the GPD estimate with a dynamic threshold fits better than that with a static threshold.

In summary, our data analysis shows that the developed new three-step inference procedure with the random weighted bootstrap uncertainty quantification can provide different insights. The PP-plot indicates that using GPD semiparametrically offers a good fit. Developing an efficient goodness-of-fit test for the new method is necessary to address the concern of model misspecification, which requires corresponding nonparametric inferences for all three steps and is beyond the scope of this paper.

4.3 Conclusions

This study develops an effective three-step inference procedure for forecasting a risk measure at a high level in insurance ratemaking. The first step uses logistic regression to estimate the probability of having no claim accurately. Conditional on nonzero claims, the second step employs quantile regression to model and estimate the dynamic threshold for a robust fit to loss distribution. The third step uses extreme value theory to fit a generalized Pareto distribution to exceedances over the

selected threshold in the second step. Furthermore, this study adopts a random weighted bootstrap method to quantify the risk forecast derived from the above three steps. Finally, we reexamine the Australian automobile data for forecasting Value-at-Risk at level 99% and Expected Shortfall at 97.5% and find that the three-step method provides significantly different forecasts from the naive approach without modeling the losses. One future research is to develop a goodness-of-fit test for the proposed models.

4.4 Proofs

Put

$$\begin{cases} \mathbf{Z}_{i,1} = \bar{\mathbf{X}}_i(I(N_i = 0) - p(\mathbf{X}_i)), \\ \mathbf{Z}_{i,2} = \bar{\mathbf{X}}_i \exp(\boldsymbol{\theta}_2^\tau \mathbf{X}_i)(\alpha_0 - I(\tilde{S}_i \le \exp(\boldsymbol{\theta}_2^\tau \bar{\mathbf{X}}_i))), \\ S_i(u_i) = 1 + \frac{\xi(\mathbf{X}_i)}{\sigma(\mathbf{X}_i)}(\tilde{S}_i - u(\mathbf{X}_i)), \\ \mathbf{Z}_{i,3}(u_i) = \bar{\mathbf{X}}_i\big\{\frac{1}{\xi(\mathbf{X}_i)} \log S_i(u_i) - (1 + \frac{1}{\xi(\mathbf{X}_i)})(1 - \frac{1}{S_i(u_i)})\big\} I(\tilde{S}_i > u(\mathbf{X}_i)), \\ \mathbf{Z}_{i,4}(u_i) = \bar{\mathbf{X}}_i\big\{(1 + \frac{1}{\xi(\mathbf{X}_i)})(1 - \frac{1}{S_i(u_i)}) - 1\big\} I(\tilde{S}_i > u(\mathbf{X}_i)). \end{cases}$$

Lemma 4.1. *Under conditions of Theorem 4.1,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{Z}_{i,1} = \mathbf{W}_1 + o_p(1), \ \frac{1}{\sqrt{\tilde{n}}} \sum_{i=1}^{\tilde{n}} \mathbf{Z}_{i,2} = \mathbf{W}_2 + o_p(1),$$

$$\frac{1}{\sqrt{\tilde{n}}} \sum_{i=1}^{\tilde{n}} \mathbf{Z}_{i,3}(u_i) = \mathbf{W}_3 + o_p(1), \ \frac{1}{\sqrt{\tilde{n}}} \sum_{i=1}^{\tilde{n}} \mathbf{Z}_{i,4}(u_i) = \mathbf{W}_4 + o_p(1),$$

where the joint distribution of $\mathbf{W}_1, \cdots, \mathbf{W}_4$ *is a multivariate normal distribution with*

$$E(\mathbf{W}_1 \mathbf{W}_1^\tau) = E\{p(\mathbf{X}_i)(1 - p(\mathbf{X}_i))\bar{\mathbf{X}}_i \bar{\mathbf{X}}_i^\tau\},$$

$$E(\mathbf{W}_2 \mathbf{W}_2^\tau) = \alpha_0(1 - \alpha_0) E\{\exp(2\boldsymbol{\theta}_2^\tau \bar{\mathbf{X}}_i)\bar{\mathbf{X}}_i \bar{\mathbf{X}}_i^\tau\},$$

$$E(\mathbf{W}_3 \mathbf{W}_3^\tau) = (1 - \alpha_0) E\left\{\frac{2\xi(\mathbf{X}_i)^2}{(1 + 2\xi(\mathbf{X}_i))(1 + \xi(\mathbf{X}_i))} \bar{\mathbf{X}}_i \bar{\mathbf{X}}_i^\tau\right\},$$

$$E(\mathbf{W}_4\mathbf{W}_4^\tau) = (1-\alpha_0)E\left\{\frac{1}{1+2\xi(\mathbf{X}_i)}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau\right\},$$

$$E(\mathbf{W}_1\mathbf{W}_j^\tau) = \mathbf{0} \textit{ for } j = 2,3,4,$$

$$E(\mathbf{W}_2\mathbf{W}_3^\tau) = \mathbf{0},\ E(\mathbf{W}_2\mathbf{W}_4^\tau) = \mathbf{0},$$

$$E(\mathbf{W}_3\mathbf{W}_4^\tau) = (1-\alpha_0)E\left\{\frac{\xi(\mathbf{X}_i)}{(1+\xi(\mathbf{X}_i))(1+2\xi(\mathbf{X}_i))}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau\right\}.$$

Proof. When $\tilde{S}_i > u(\mathbf{X}_i)$, we have

$$P(S_i(u_i) \geq z|\mathbf{X}_i) = (1-\alpha_0)z^{-1/\xi(\mathbf{X}_i)} \text{ for } z \geq 1,$$

implying that

$$E(\log S_i(u_i)|\mathbf{X}_i) = (1-\alpha_0)\xi(\mathbf{X}_i),\ E((\log S_i(u_i))^2|\mathbf{X}_i) = 2(1-\alpha_0)\xi^2(\mathbf{X}_i),$$

$$E(1-\frac{1}{S_i(u_i)}|\mathbf{X}_i) = (1-\alpha_0)\frac{\xi(\mathbf{X}_i)}{1+\xi(\mathbf{X}_i)},$$

$$E((1-\frac{1}{S_i(u_i)})^2|\mathbf{X}_i) = (1-\alpha_0)\frac{2\xi^2(\mathbf{X}_i)}{(1+\xi(\mathbf{X}_i))(1+2\xi(\mathbf{X}_i))},$$

$$E(\log(S_i(u_i))(1-\frac{1}{S_i(u_i)})|\mathbf{X}_i) = (1-\alpha_0)\frac{\xi(\mathbf{X}_i)^2(\xi(\mathbf{X}_i)+2)}{(1+\xi(\mathbf{X}_i))^2}.$$

Using these equations, straightforward calculations give that

$$E(\mathbf{Z}_{i,3}(u_i)|\mathbf{X}_i) = \mathbf{0},\ E(\mathbf{Z}_{i,4}(u_i)|\mathbf{X}_i) = \mathbf{0},$$

$$E(\mathbf{Z}_{i,3}(u_i)\mathbf{Z}_{i,3}^\tau(u_i)|\mathbf{X}_i) = (1-\alpha_0)\frac{2\xi(\mathbf{X}_i)^2}{(1+\xi(\mathbf{X}_i))(1+2\xi(\mathbf{X}_i))}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau,$$

$$E(\mathbf{Z}_{i,3}(u_i)\mathbf{Z}_{i,4}^\tau(u_i)|\mathbf{X}_i) = (1-\alpha_0)\frac{\xi(\mathbf{X}_i)}{(1+\xi(\mathbf{X}_i))(1+2\xi(\mathbf{X}_i))}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau,$$

$$E(\mathbf{Z}_{i,4}(u_i)\mathbf{Z}_{i,4}^\tau(u_i)|\mathbf{X}_i) = (1-\alpha_0)\frac{1}{1+2\xi(\mathbf{X}_i)}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau.$$

Further, we have

$$E(\mathbf{Z}_{i,1}|\mathbf{X}_i) = \mathbf{0},\ E(\mathbf{Z}_{i,1}\mathbf{Z}_{i,1}^\tau|\mathbf{X}_i) = p(\mathbf{X}_i)(1-p(\mathbf{X}_i))\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau,$$

$$E(\mathbf{Z}_{i,2}|\mathbf{X}_i) = \mathbf{0},\ E(\mathbf{Z}_{i,2}\mathbf{Z}_{i,2}^\tau|\mathbf{X}_i) = \alpha_0(1-\alpha_0)\exp(2\boldsymbol{\theta}_2^\tau\bar{\mathbf{X}}_i)\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau,$$

$$E(\mathbf{Z}_{i,2}\mathbf{Z}_{i,3}^\tau(u_i)|\mathbf{X}_i) = E(\mathbf{Z}_{i,2}E(\mathbf{Z}_{i,3}^\tau(u_i)|\tilde{S}_i > u(\mathbf{X}_i), \mathbf{X}_i)|\mathbf{X}_i) = \mathbf{0},$$

$$E(\mathbf{Z}_{i,2}\mathbf{Z}_{i,4}^\tau(u_i)|\mathbf{X}_i) = E(\mathbf{Z}_{i,2}E(\mathbf{Z}_{i,4}^\tau(u_i)|\tilde{S}_i > u(\mathbf{X}_i), \mathbf{X}_i)|\mathbf{X}_i) = \mathbf{0}.$$

Because $\tilde{S}_i$ is the conditional variable of one of S_j given $N_j > 0$, the conditional variable of $\sum_{i=1}^n \mathbf{Z}_{i,1}$ given $\mathbf{X}_i$'s is mutually independent of the conditional variables of

$$\sum_{i=1}^{\tilde{n}} \mathbf{Z}_{i,2},\ \sum_{i=1}^{\tilde{n}} \mathbf{Z}_{i,3}(u_i),\ \text{and}\ \sum_{i=1}^{\tilde{n}} \mathbf{Z}_{i,4}(u_i)\ \text{given}\ \mathbf{X}_i's.$$

Therefore, Lemma 4.1 follows from the central limit theorem.

□

Lemma 4.2. *Under conditions of Theorem 4.1, as $n \to \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) = \Sigma_1^{-1}\mathbf{W}_1 + o_p(1).$$

Proof. Define

$$L_1(\boldsymbol{\theta}_1) = \sum_{i=1}^n \{I(N_i = 0)\log p(\mathbf{X}_i) + I(N_i > 0)\log(1 - p(\mathbf{X}_i))\}.$$

Then,

$$\frac{\partial p(\mathbf{X}_i)}{\partial \boldsymbol{\theta}_1} = p(\mathbf{X}_i)(1-p(\mathbf{X}_i))\bar{\mathbf{X}}_i,$$

$$\frac{1}{\sqrt{n}}\frac{\partial L_1(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{Z}_{i,1} \xrightarrow{d} \mathbf{W}_1,$$

$$\frac{1}{n}\frac{\partial^2 L_1(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1^\tau} = -\frac{1}{n}\sum_{i=1}^{n} p(\mathbf{X}_i)(1-p(\mathbf{X}_i))\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau \xrightarrow{p} -\Sigma_1.$$

Hence, it follows from Theorem 5.39 of Van der Vaart (2000) that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) = -\left\{\frac{\partial^2 L_1(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_1^\tau}\right\}^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_1(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\theta}_1} + o_p(1) = \Sigma_1^{-1}\mathbf{W}_1 + o_p(1).$$

□

Lemma 4.3. *Under conditions of Theorem 1, as $n \to \infty$,*

$$\tilde{n}/n \xrightarrow{p} 1 - p_0 \text{ and } \sqrt{\tilde{n}}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) = \Gamma_2^{-1}\mathbf{W}_2 + o_p(1).$$

Proof. The first equation follows from the weak law of large numbers by noting that

$$\tilde{n} = \sum_{i=1}^{n} I(N_i > 0) \text{ and } EI(N_i > 0) = E(1 - p(\mathbf{X}_i)) = 1 - p_0.$$

Define

$$Q_{\tilde{n}}(\mathbf{z}) = \sum_{t=1}^{\tilde{n}}\{\rho_{\alpha_0}(\tilde{S}_i - \exp(\boldsymbol{\theta}_2^\tau\bar{\mathbf{X}}_i) - \exp(\boldsymbol{\theta}_2^\tau\bar{\mathbf{X}}_i)(\exp(\mathbf{z}^\tau\bar{\mathbf{X}}_i/\sqrt{\tilde{n}}) - 1)) - \rho_{\alpha_0}(\tilde{S}_i - \exp(\boldsymbol{\theta}_2^\tau\bar{\mathbf{X}}_i))\},$$

$$Q_{\tilde{n},1}(\mathbf{z}) = -\sum_{i=1}^{\tilde{n}}\exp(\boldsymbol{\theta}_2^\tau\mathbf{X}_i)(\exp(\mathbf{z}^\tau\mathbf{X}_i/\sqrt{\tilde{n}}) - 1)(\alpha_0 - I(\tilde{S}_i - \exp(\boldsymbol{\theta}_2^\tau\mathbf{X}_i) < 0)),$$

$$Q_{\tilde{n},2}(\mathbf{z}) = \sum_{i=1}^{\tilde{n}}\int_0^{\exp(\boldsymbol{\theta}_2^\tau\mathbf{X}_i)(\exp(\mathbf{z}^\tau\mathbf{X}_i/\sqrt{\tilde{n}})-1)}\{I(\tilde{S}_i - \exp(\boldsymbol{\theta}_2^\tau\mathbf{X}_i) \le s) - I(\tilde{S}_i - \exp(\boldsymbol{\theta}_2^\tau\mathbf{X}_i) \le 0)\}\,ds.$$

It follows from Knight's equality that

$$Q_{\tilde{n}}(\mathbf{z}) = Q_{\tilde{n},1}(\mathbf{z}) + Q_{\tilde{n},2}(\mathbf{z}).$$

Note that

$$Q_{\tilde{n},1}(\mathbf{z}) = -\mathbf{z}^\tau \frac{1}{\sqrt{n}} \sum_{i=1}^{\tilde{n}} \mathbf{Z}_{i,2}(u_i) + o_p(1)$$

and

$$E\{Q_{\tilde{n},2}(\mathbf{z})\} = E\{E(Q_{\tilde{n},2}(\mathbf{z})|\{\mathbf{X}_i\})\} = \frac{1}{2}\mathbf{z}^\tau \Gamma_1 \mathbf{z} + o_p(1).$$

Using the above expansions and the standard techniques in Section 4.4 of Koenker (2005) for nonlinear quantile regression, we can show that

$$\sqrt{\tilde{n}}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) = \Gamma_2^{-1} \frac{1}{\sqrt{\tilde{n}}} \sum_{i=1}^{\tilde{n}} \mathbf{Z}_{i,2} + o_p(1) = \Gamma_2^{-1}\mathbf{W}_2 + o_p(1).$$

□

Lemma 4.4. *Under conditions of Theorem 4.1, as $n \to \infty$,*

$$\sqrt{\tilde{n}}(\hat{\boldsymbol{\eta}}_3 - \boldsymbol{\eta}_3) = \Gamma_4^{-1}\mathbf{W}_5 + o_p(1).$$

Proof. Define

$$\hat{L}_3(\boldsymbol{\eta}_3) = \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} I(\tilde{S}_i > \hat{u}(\mathbf{X}_i)) l_i(\boldsymbol{\eta}_3 | \tilde{S}_i - \hat{u}(\mathbf{X}_i)).$$

Then,

$$\sqrt{\tilde{n}}\frac{\partial \hat{L}_3(\boldsymbol{\eta}_3)}{\partial \boldsymbol{\theta}_3} = \frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{\tilde{n}} \mathbf{Z}_{i,3}(\hat{u}_i) = \frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{\tilde{n}} \mathbf{Z}_{i,3}(u_i) + \frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{\tilde{n}} \{\mathbf{Z}_{i,3}(\hat{u}_i) - \mathbf{Z}_{i,3}(u_i)\},$$

$$\begin{aligned}
& \frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{\tilde{n}} \{\mathbf{Z}_{i,3}(\hat{u}_i) - \mathbf{Z}_{i,3}(u_i)\} \\
= \; & \frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{\tilde{n}} \bar{\mathbf{X}}_i \left\{ -\frac{1}{\sigma(\mathbf{X}_i)S_i(u_i)} + \frac{1+\xi(\mathbf{X}_i)}{\sigma(\mathbf{X}_i)S_i^2(u_i)} \right\} (\hat{u}_i - u(\mathbf{X}_i)) I(\tilde{S}_i > u(\mathbf{X}_i)) + o_p(1) \\
= \; & \frac{1}{\tilde{n}}\sum_{i=1}^{\tilde{n}} \left\{ -\frac{1}{\sigma(\mathbf{X}_i)S_i(u_i)} + \frac{1+\xi(\mathbf{X}_i)}{\sigma(\mathbf{X}_i)S_i^2(u_i)} \right\} \exp(\boldsymbol{\theta}_2^\tau \bar{\mathbf{X}}_i) \bar{\mathbf{X}}_i \bar{\mathbf{X}}_i^\tau I(\tilde{S}_i > u(\mathbf{X}_i)) \sqrt{\tilde{n}}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + o_p(1) \\
= \; & (1-\alpha_0) E\left\{ \frac{\xi^2(\mathbf{X}_i)}{\sigma(\mathbf{X}_i)(1+\xi(\mathbf{X}_i))(1+2\xi(\mathbf{X}_i))} \exp(\boldsymbol{\theta}_2^\tau \mathbf{X}_i) \bar{\mathbf{X}}_i \bar{\mathbf{X}}_i^\tau \right\} \sqrt{\tilde{n}}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + o_p(1),
\end{aligned}$$

implying that

$$\sqrt{\tilde{n}}\frac{\partial \hat{L}_3(\boldsymbol{\eta}_3)}{\partial \boldsymbol{\theta}_3} = \mathbf{W}_3 + \Gamma_{3,1}\Gamma_2^{-1}\mathbf{W}_2 + o_p(1). \tag{4.14}$$

Similarly,

$$\sqrt{\tilde{n}}\frac{\partial \hat{L}_3(\boldsymbol{\eta}_3)}{\partial \boldsymbol{\theta}_4} = \frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{\tilde{n}} \mathbf{Z}_{i,4}(\hat{u}_i) = \frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{\tilde{n}} \mathbf{Z}_{i,4}(u_i) + \frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{\tilde{n}} \{\mathbf{Z}_{i,4}(\hat{u}_i) - \mathbf{Z}_{i,4}(u_i)\},$$

$$\begin{aligned}
& \frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{\tilde{n}} \{\mathbf{Z}_{i,4}(\hat{u}_i) - \mathbf{Z}_{i,4}(u_i)\} \\
= & \frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{\tilde{n}} \bar{\mathbf{X}}_i \frac{-(1+\xi(\mathbf{X}_i))}{\sigma(\mathbf{X}_i)S_i^2(u_i)}(\hat{u}_i - u(\mathbf{X}_i))I(\tilde{S}_i > u(\mathbf{X}_i)) + o_p(1) \\
= & -\frac{1}{\tilde{n}}\sum_{i=1}^{\tilde{n}} \frac{1+\xi(\mathbf{X}_i)}{\sigma(\mathbf{X}_i)S_i^2(u_i)} \exp(\boldsymbol{\theta}_2^\tau \bar{\mathbf{X}}_i)\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau I(\tilde{S}_i > u(\mathbf{X}_i))\sqrt{\tilde{n}}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + o_p(1) \\
= & -(1-\alpha_0)E\left\{\frac{1+\xi(\mathbf{X}_i)}{\sigma(\mathbf{X}_i)(1+2\xi(\mathbf{X}_i))} \exp(\boldsymbol{\theta}_2^\tau \bar{\mathbf{X}}_i)\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau\right\}\sqrt{\tilde{n}}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + o_p(1),
\end{aligned}$$

implying that

$$\frac{1}{\sqrt{\tilde{n}}}\frac{\partial L_3(\boldsymbol{\eta}_3)}{\partial \boldsymbol{\theta}_4} = \mathbf{W}_4 + \Gamma_{3,2}\Gamma_2^{-1}\mathbf{W}_2 + o_p(1). \tag{4.15}$$

Because

$$\begin{aligned}
\frac{\partial^2 L_3(\boldsymbol{\eta}_3)}{\partial \boldsymbol{\theta}_3 \partial \boldsymbol{\theta}_3^\tau} &= \frac{1}{\tilde{n}}\sum_{i=1}^{\tilde{n}} \bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau \left\{-\frac{1}{\xi(\mathbf{X}_i)}\log S_i(u_i) + \frac{2}{\xi(\mathbf{X}_i)}(1 - \frac{1}{S_i(u_i)})\right. \\
&\quad \left. -\frac{1+\xi(\mathbf{X}_i)}{\xi(\mathbf{X}_i)}(\frac{1}{S_i(u_i)} - \frac{1}{S_i^2(u_i)})\right\} I(\tilde{S}_i > u(\mathbf{X}_i)) + o_p(1) \\
&= -(1-\alpha_0)E\left\{\frac{2\xi(\mathbf{X}_i)^2}{(1+\xi(\mathbf{X}_i))(1+2\xi(\mathbf{X}_i))}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau\right\} + o_p(1),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_3(\boldsymbol{\eta}_3)}{\partial \boldsymbol{\theta}_3 \partial \boldsymbol{\theta}_4^\tau} &= \frac{1}{\tilde{n}}\sum_{i=1}^{\tilde{n}} \bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau \left\{\frac{1}{\xi(\mathbf{X}_i)}\frac{1-S_i(u_i)}{S_i(u_i)} + \frac{1+\xi(\mathbf{X}_i)}{\xi(\mathbf{X}_i)}\frac{1-S_i(u_i)}{S_i^2(u_i)}\right\} I(\tilde{S}_i > u(\mathbf{X}_i)) + o_p(1) \\
&= -(1-\alpha_0)E\left\{\frac{2+3\xi(\mathbf{X}_i)}{(1+\xi(\mathbf{X}_i))(1+2\xi(\mathbf{X}_i))}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau\right\} + o_p(1),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_3(\boldsymbol{\eta}_3)}{\partial \boldsymbol{\theta}_4 \partial \boldsymbol{\theta}_4^\tau} &= \frac{1}{\tilde{n}}\sum_{i=1}^{\tilde{n}} \bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau \frac{1+\xi(\mathbf{X}_i)}{\xi(\mathbf{X}_i)}\frac{1-S_i(u_i)}{S_i^2(u_i)} I(\tilde{S}_i > u(\mathbf{X}_i)) + o_p(1) \\
&= -(1-\alpha_0)E\left\{\frac{1}{1+2\xi(\mathbf{X}_i)}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i^\tau\right\} + o_p(1),
\end{aligned}$$

we have

$$\frac{1}{\tilde{n}}\sum_{i=1}^{\tilde{n}} \frac{\partial^2 L_3(\boldsymbol{\eta}_3)}{\partial \boldsymbol{\eta}_3 \partial \boldsymbol{\eta}_3^\tau} = -\Gamma_4 + o_p(1). \tag{4.16}$$

Hence, the lemma follows from (4.14)-(4.16).

□

Proof of Theorem 4.1. The theorem follows from Lemmas 4.1–4.4. □

Lemma 4.5. *Under conditions of Theorem 4.1, given any* $\mathbf{x}$*, as* $n \to \infty$*,*

$$\sqrt{n}(\hat{\alpha}^*(\mathbf{x}) - \alpha^*(\mathbf{x}), \hat{u}(\mathbf{x}) - u(\mathbf{x}), \hat{\xi}(\mathbf{x}) - \xi(\mathbf{x}), \hat{\sigma}(\mathbf{x}) - \sigma(\mathbf{x})) \xrightarrow{d} N(\mathbf{0}, \Sigma_{\mathbf{x}}),$$

where $\Sigma_{\mathbf{x}} = \mathbf{D}_{\mathbf{x}} \Sigma_{\mathbf{W}} \mathbf{D}_{\mathbf{x}}^{\tau}$*,* $\Sigma_{\mathbf{W}}$ *is the covariance matrix of* $(\mathbf{W}_1^{\tau}, \mathbf{W}_2^{\tau}, \mathbf{W}_5^{\tau})^{\tau}$*, and*

$$\mathbf{D}_{\mathbf{x}} = \mathit{diag}\left(\frac{1-\alpha}{p(\mathbf{x})} \bar{\mathbf{x}}^{\tau} \Sigma_1^{-1}, \frac{u(\mathbf{x})}{\sqrt{1-p_0}} \bar{\mathbf{x}}^{\tau} \Gamma_2^{-1}, \mathit{diag}\left(\frac{\xi(\mathbf{x})}{\sqrt{1-p_0}} \bar{\mathbf{x}}^{\tau}, \frac{\sigma(\mathbf{x})}{\sqrt{1-p_0}} \bar{\mathbf{x}}^{\tau} \right) \Gamma_4^{-1} \right).$$

Proof. It follows from Lemmas 4.1 - 4.4 and the delta method that

$$\begin{aligned}
\sqrt{n}(\hat{\alpha}^*(\mathbf{x}) - \alpha^*(\mathbf{x})) =& \frac{1-\alpha}{p(\mathbf{x})} \bar{\mathbf{x}}^{\tau} \sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) + o_{\mathbb{P}}(1) = \frac{1-\alpha}{p(\mathbf{x})} \bar{\mathbf{x}}^{\tau} \Sigma_1^{-1} \mathbf{W}_1 + o_p(1), \\
\sqrt{n}(\hat{u}(\mathbf{x}) - u(\mathbf{x})) =& u(\mathbf{x}) \bar{\mathbf{x}}^{\tau} \sqrt{n}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2) + o_{\mathbb{P}}(1) = \frac{u(\mathbf{x})}{\sqrt{1-p_0}} \bar{\mathbf{x}}^{\tau} \Gamma_2^{-1} \mathbf{W}_2 + o_p(1), \\
\sqrt{n}(\hat{\xi}(\mathbf{x}) - \xi(\mathbf{x}), \hat{\sigma}(\mathbf{x}) - \sigma(\mathbf{x}))^{\tau} =& \begin{pmatrix} \frac{\xi(\mathbf{x})}{\sqrt{1-p_0}} \bar{\mathbf{x}}^{\tau} & 0 \\ 0 & \frac{\sigma(\mathbf{x})}{\sqrt{1-p_0}} \bar{\mathbf{x}}^{\tau} \end{pmatrix} \sqrt{n}(\hat{\boldsymbol{\eta}}_3 - \boldsymbol{\eta}_3) + o_p(1) \\
=& \begin{pmatrix} \frac{\xi(\mathbf{x})}{\sqrt{1-p_0}} \bar{\mathbf{x}}^{\tau} & 0 \\ 0 & \frac{\sigma(\mathbf{x})}{\sqrt{1-p_0}} \bar{\mathbf{x}}^{\tau} \end{pmatrix} \Gamma_4^{-1} \mathbf{W}_5 + o_p(1),
\end{aligned}$$

implying that

$$\sqrt{n} \begin{pmatrix} \hat{\alpha}^*(\mathbf{x}) - \alpha^*(\mathbf{x}) \\ \hat{u}(\mathbf{x}) - u(\mathbf{x}) \\ \hat{\xi}(\mathbf{x}) - \xi(\mathbf{x}) \\ \hat{\sigma}(\mathbf{x}) - \sigma(\mathbf{x}) \end{pmatrix} = \mathbf{D}_{\mathbf{x}} \begin{pmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \mathbf{W}_5 \end{pmatrix} + o_p(1).$$

□

Proof of Theorem 4.2. The theorem follows immediately from Lemma 4.5 and the delta method.

□

Proof of Theorem 4.3. Like the proof of Lemma 4.1, we can show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i^b\mathbf{Z}_{i,1}=\mathbf{W}_1^b+o_p(1),\ \frac{1}{\sqrt{\tilde{n}}}\sum_{i=1}^{\tilde{n}}\delta_i^b\mathbf{Z}_{i,j}=\mathbf{W}_j^b+o_p(1)\text{ for }j=2,3,4,$$

and $\{\mathbf{W}_j^b-\mathbf{W}_j\}_{j=1}^4$ and $\{\mathbf{W}_j\}_{j=1}^4$ are independent with the same distribution.

Following the proof of Lemma 4.2, we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_1^b-\boldsymbol{\theta}_1)=\Sigma_1^{-1}\mathbf{W}_1^b+o_p(1),$$

implying that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_1^b-\hat{\boldsymbol{\theta}}_1)=\Sigma_1^{-1}(\mathbf{W}_1^b-\mathbf{W}_1)+o_p(1).$$

Similarly, we have

$$\sqrt{\tilde{n}}(\hat{\boldsymbol{\theta}}_2^b-\hat{\boldsymbol{\theta}}_2)=\Gamma_2^{-1}(\mathbf{W}_2^b-\mathbf{W}_2)+o_p(1)$$

and

$$\sqrt{\tilde{n}}(\hat{\boldsymbol{\eta}}_3^b-\hat{\boldsymbol{\eta}}_3)=\Gamma_4^{-1}(\mathbf{W}_5^b-\mathbf{W}_5)+o_p(1),$$

where

$$\mathbf{W}_5^b=((\mathbf{W}_3^b+\Gamma_{3,1}\Gamma_2^{-1}\mathbf{W}_2^b)^\tau,(\mathbf{W}_4^b+\Gamma_{3,2}\Gamma_2^{-1}\mathbf{W}_2^b)^\tau)^\tau.$$

Therefore, the joint limit of $\sqrt{n}(\hat{\boldsymbol{\theta}}_j^b-\hat{\boldsymbol{\theta}}_j)$ for $j=1,\cdots,4$ is the same as that of $\sqrt{n}(\hat{\boldsymbol{\theta}}_j-\boldsymbol{\theta}_j)$ for $j=1,\cdots,4$, implying that $\sqrt{n}\{\widehat{\mathrm{VaR}}_S^b(\alpha|\mathbf{x})-\widehat{\mathrm{VaR}}_S(\alpha|\mathbf{x})\}$ and $\sqrt{n}\{\widehat{\mathrm{ES}}_S^b(\alpha|\mathbf{x})-\widehat{\mathrm{ES}}_S(\alpha|\mathbf{x})\}$ have the same limit as $\sqrt{n}\{\widehat{\mathrm{VaR}}_S(\alpha|\mathbf{x})-\mathrm{VaR}_S(\alpha|\mathbf{x})\}$ and $\sqrt{n}\{\widehat{\mathrm{ES}}_S(\alpha|\mathbf{x})-\mathrm{ES}_S(\alpha|\mathbf{x})\}$, respectively. Furthermore, we can show that

$$\frac{n}{B}\sum_{b=1}^{B}\{\widehat{\mathrm{VaR}}_S^b(\alpha|\mathbf{x})-\widehat{\mathrm{VaR}}_S(\alpha|\mathbf{x})\}^2\text{ and }\frac{n}{B}\sum_{b=1}^{B}\{\widehat{\mathrm{ES}}_S^b(\alpha|\mathbf{x})-\widehat{\mathrm{ES}}_S(\alpha|\mathbf{x})\}^2$$

converge in probability to the asymptotic variances of

$$\sqrt{n}\{\widehat{\mathrm{VaR}}_S(\alpha|\mathbf{x}) - \mathrm{VaR}_S(\alpha|\mathbf{x})\} \text{ and } \sqrt{n}\{\widehat{\mathrm{ES}}_S(\alpha|\mathbf{x}) - \mathrm{ES}_S(\alpha|\mathbf{x})\}, \text{ respectively.}$$

That is, the theorem holds.

□

Table 4.1: This table reports the estimates of logistic regression in the first step and quantile regression in the second step. The upper panel displays $\hat{\theta}_1$ and $\hat{\theta}_2$ in fitting logistic regression and quantile regressions, respectively. The lower panel shows $\hat{p}(\mathbf{X}_i)$ and $\hat{u}(\mathbf{X}_i)$ for each categories in fitting logistic regression and quantile regression, respectively. The probability level in selecting threshold is $\alpha_0 = 0.90$.

Parameter estimates		
	Logistic Regression ($\hat{\boldsymbol{\theta}}_1$)	Quantile Regression ($\hat{\boldsymbol{\theta}}_2$)
(Intercept)	-1.907	8.240
Veh Age: 1	-0.031	-0.181
Veh Age: 3	-0.127	0.110
Veh Age: 4	-0.221	0.257
Agecat: 1	0.533	0.587
Agecat: 2	0.334	0.189
Agecat: 3	0.272	0.123
Agecat: 4	0.230	0.127
Agecat: 6	-0.003	-0.057

Probability and threshold estimates			
Category	Size	Logistic Regression ($\hat{p}(\mathbf{X}_i)$)	Quantile Regression ($\hat{u}(\mathbf{X}_i)$)
1(2&1)	1504	0.798	6817.390
2(1&1)	1283	0.803	5688.863
3(3&1)	1643	0.818	7609.343
4(2&2)	3167	0.828	4577.220
5(4&1)	1312	0.831	8816.629
6(1&2)	2160	0.833	3819.523
7(2&3)	3741	0.837	4284.405
8(1&3)	2706	0.841	3575.180
9(2&4)	3919	0.843	4301.956
10(3&2)	3956	0.846	5108.940
11(1&4)	2935	0.847	3589.826
12(3&3)	4826	0.853	4782.110
13(4&2)	3592	0.857	5919.516
14(3&4)	4760	0.859	4801.700
15(4&3)	4494	0.865	5540.832
16(4&4)	4575	0.870	5563.530
17(2&5)	2635	0.871	3789.770
18(2&6)	1621	0.871	3580.935
19(1&5)	2042	0.874	3162.425
20(1&6)	1131	0.875	2988.160
21(3&5)	3088	0.884	4230.015
22(3&6)	1791	0.885	3996.920
23(4&5)	2971	0.894	4901.142
24(4&6)	2004	0.894	4631.065

Table 4.2: This table reports estimates in fitting the Generalized Pareto Distribution in the third step, where $\xi(\mathbf{x}, \boldsymbol{\theta}_3)$ is constant, i.e., independent of $\mathbf{x}$, and $\sigma(\mathbf{x}, \boldsymbol{\theta}_4) = \exp\{\bar{\mathbf{x}}^\tau \boldsymbol{\theta}_4\}$.

Estimates in fitting GPD		
$\hat{\xi}$		-1.817
$\hat{\boldsymbol{\theta}}_4$	(Intercept)	8.370
	Veh Age: 1	0.114
	Veh Age: 3	-0.161
	Veh Age: 4	-0.241
	Agecat: 1	0.025
	Agecat: 2	0.238
	Agecat: 3	0.028
	Agecat: 4	0.088
	Agecat: 6	0.174

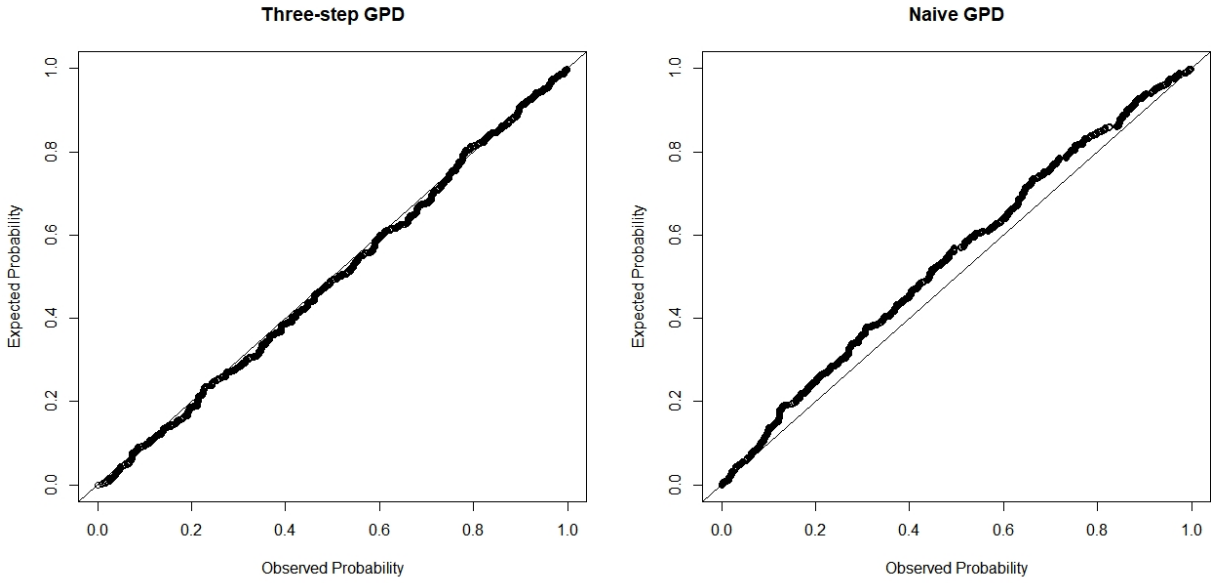
Table 4.3: This table reports sample size, nonparametric estimate, Naive GPD estimate, and three-step estimate of the conditional VaR at 99% level, and the two 90% confidence intervals using the three-step inference and random weighted bootstrap method with $B = 5000$ for each category.

				Three-step Estimate		
Category	Size	Naive $\mathrm{VaR}(0.99\|\boldsymbol{x}_j)$	GPD $\mathrm{VaR}(0.99\|\boldsymbol{x}_j)$	$\widehat{\mathrm{VaR}}(0.99\|\boldsymbol{x}_j)$	$I_1(0.9\|\boldsymbol{x}_j)$	$I_2(0.9\|\boldsymbol{x}_j)$
1(2&1)	1504	7341.31	8616.69	10109.95	(8223.92, 11673.32)	(8389.59, 11830.31)
2(1&1)	1283	5697.99	9038.44	9238.36	(6855.95, 10847.83)	(7269.99, 11206.72)
3(3&1)	1643	5594.83	7447.37	9982.25	(8290.21, 11385.22)	(8456.35, 11508.15)
4(2&2)	3167	4089.01	8171.66	7670.78	(6238.09, 8991.58)	(6312.78, 9028.78)
5(4&1)	1312	6712.36	6831.47	10712.61	(8834.67, 12255.24)	(9017.61, 12407.6)
6(1&2)	2160	2701.22	8495.82	7111.36	(5233.15, 8531.48)	(5505.27, 8717.46)
7(2&3)	3741	3567.86	7450.26	6546.33	(5483.03, 7492.56)	(5547.66, 7545.01)
8(1&3)	2706	2396.47	7685.67	5967.99	(4569.29, 7034.11)	(4761.99, 7123.98)
9(2&4)	3919	3146.32	7458.32	6520.53	(5441.17, 7507.4)	(5494.93, 7546.13)
10(3&2)	3956	3855.48	7014.56	7206.24	(6017.29, 8223.99)	(6118.41, 8294.07)
11(1&4)	2935	2613.19	7681.10	5925.3	(4501.7, 6951.36)	(4706.04, 7144.55)
12(3&3)	4826	3923.08	6507.43	6272.36	(5425.78, 7043.76)	(5470.94, 7073.79)
13(4&2)	3592	4032.21	6407.01	7488.88	(6320.13, 8564.61)	(6371.31, 8606.45)
14(3&4)	4760	3779.84	6474.93	6228.21	(5250.68, 6983.21)	(5374.47, 7081.96)
15(4&3)	4494	3661.65	6012.94	6618.93	(5679.58, 7459.69)	(5731.9, 7505.96)
16(4&4)	4575	3997.81	5989.61	6564.52	(5684.77, 7434.84)	(5687.25, 7441.79)
17(2&5)	2635	2813.81	6164.60	4922.43	(3891.81, 5729.72)	(4010.59, 5834.27)
18(2&6)	1621	1994.16	6729.40	4914.38	(3530.94, 5814.8)	(3787.45, 6041.3)
19(1&5)	2042	1768.52	6205.05	4293.94	(3188.51, 5029.5)	(3372.71, 5215.18)
20(1&6)	1131	2188.47	6786.32	4318.53	(2824.69, 5168.59)	(3149.93, 5487.13)
21(3&5)	3088	2277.86	5500.48	4771.73	(3617.16, 5637.14)	(3753.07, 5790.38)
22(3&6)	1791	2860.75	5749.03	4629.36	(3285.56, 5255.69)	(3674.89, 5583.83)
23(4&5)	2971	3249.02	5154.53	5112.70	(4059.73, 6014.14)	(4123.92, 6101.48)
24(4&6)	2004	2489.78	5238.38	4871.62	(3496.9, 5645.11)	(3795.25, 5947.98)

Table 4.4: This table reports the sample size, nonparametric estimate, GPD estimate with a static threshold, and three-step estimate of the conditional ES at 97.5% level, and the two 90% confidence intervals using the three-step inference and random weighted bootstrap method with $B = 5000$ for each category.

Three-step Estimate						
Category	Size	Naive $\mathrm{ES}(0.975\|\boldsymbol{x}_j)$	GPD $\mathrm{ES}(0.975\|\boldsymbol{x}_j)$	$\widehat{\mathrm{ES}}(0.975\|\boldsymbol{x}_j)$	$I_1(0.9\|\boldsymbol{x}_j)$	$I_2(0.9\|\boldsymbol{x}_j)$
1(2&1)	1504	340068	213577	211909	(201877, 219747)	(203154, 220663)
2(1&1)	1283	297957	288678	285943	(272539, 294759)	(275098, 296788)
3(3&1)	1643	245718	209312	209001	(199509, 215671)	(201110, 216892)
4(2&2)	3167	244731	208108	207959	(198627, 216187)	(199315, 216603)
5(4&1)	1312	254996	218490	219198	(209343, 226876)	(210520, 227877)
6(1&2)	2160	183670	285244	284079	(271570, 294098)	(273312, 294847)
7(2&3)	3741	179336	203247	201845	(194831, 207806)	(195428, 208263)
8(1&3)	2706	143020	279994	277566	(267811, 285216)	(269098, 286033)
9(2&4)	3919	169734	205230	203212	(195449, 209448)	(196323, 210101)
10(3&2)	3956	201986	202112	203263	(195999, 210006)	(196311, 210216)
11(1&4)	2935	170819	282267	279076	(268986, 286825)	(270222, 287931)
12(3&3)	4826	189645	197844	197751	(192449, 202328)	(192856, 202646)
13(4&2)	3592	187990	209337	211548	(203613, 218502)	(204214, 218882)
14(3&4)	4760	176240	199457	198932	(192902, 204176)	(193321, 204544)
15(4&3)	4494	184675	205249	206149	(200382, 211158)	(200777, 211521)
16(4&4)	4575	188146	206674	207259	(200802, 212638)	(201369, 213148)
17(2&5)	2635	150802	199861	199277	(192023, 205908)	(192435, 206119)
18(2&6)	1621	114702	209234	202354	(189099, 212079)	(191088, 213620)
19(1&5)	2042	91678	276712	275253	(265056, 283765)	(265910, 284597)
20(1&6)	1131	187416	287839	278940	(263703, 290461)	(265781, 292098)
21(3&5)	3088	136175	194506	195040	(188937, 200369)	(189296, 200784)
22(3&6)	1791	127544	201763	197437	(186030, 205772)	(187788, 207087)
23(4&5)	2971	153992	201709	203127	(196707, 208532)	(197246, 209007)
24(4&6)	2004	155001	207747	205121	(194499, 213067)	(195975, 214267)

Figure 4.1: PP-plots for the GPD estimates with a dynamic threshold (quantile regression) in the left panel and static threshold (90% quantile of all positive losses) in the right panel.



Chapter 5

Conclusion

In this dissertation, I focus on developing efficient statistical methods for making inferences and forecasting risk for insurance data. Especially, I am interested in forecasting quantile risk measures, such as Value-at-Risk or Expected shortfall, quantifying uncertainty regarding such inferences, and applying them to insurance premium calculation and ratemaking. However, many previous pieces of research, which also study quantile risk measures for insurance data have some limitations. My dissertation research has focused on resolving such limitations and contributing to this research discipline theoretically.

My dissertation research is strongly motivated by the paper Heras, Moreno, and Vilar-Zanón (2018). To overcome the usual simple quantile estimation methods and make a better forecast, Heras, Heras, Moreno, and Vilar-Zanón (2018) propose a two-step inference to predict the Value-at-Risk of aggregated losses in non-life insurance, which uses logistic regression and quantile regression in their first and second steps, respectively. However, their application of the quantile regression in the second step is unclear. They apply quantile regression in their data analysis to each category, which has a different quantile level, at the second step, which leads to the failure of pooling information from all levels of the categorical explanatory variables. Hence it is suspected that the quantile regression estimation in Heras, Moreno, and Vilar-Zanón (2018) is the same as the sample quantile based on the positive losses in each category.

To confirm it, in Chapter 2, I propose an alternative two-step inference for the VaR via logistic regression and the sample quantile rather than quantile regression and see how the estimator from the proposed method is different from the one in Heras, Moreno, and Vilar-Zanón (2018). Since the computation of the sample quantile is much cheaper than the quantile regression, it is not arguable that my methodology is superior to the method in Heras, Moreno, and Vilar-Zanón (2018) in terms of the computation efficiency. To quantify the uncertainty of the proposed inference, I employ an empirical likelihood method and test whether the two-step estimates in Heras, Moreno, and

Vilar-Zanón (2018) are significantly different from my alternative inference of VaR. Data analysis and simulation study confirm that the second step of using quantile regression is not necessary.

In Chapter 3, I expand the previous research by proposing the new uncertainty quantification method of the two-step inference of the Value-at-Risk of aggregated losses of insurance claims for both categorical and continuous explanatory variables and developing alternative two-step inference that can be applied to generalized model setting. One objective of this research is to quantify the uncertainty of the two-step inference more in a general way. I propose to use the random weighted bootstrap method in Jin, Ying, and Wei (2001) and Zhu (2016) for uncertainty quantification, which allows heteroscedastic errors in quantile regression and is less computationally intensive than other bootstrap methods. Another objective of the research is to develop a newly designed two-step inference method for Value-at-Risk, which can be employed in more universal model settings. This proposed method applies a weighted quantile regression in the second stage. Since I use the original risk level for the quantile, not the adjusted risk one, the new two-step inference is not equal to sample quantile estimation when the explanatory vector is categorical, which is unlike the two-step inference in Heras, Moreno, and Vilar-Zanón (2018). By performing the simulation study, I show that the result from Heras, Moreno, and Vilar-Zanón (2018) and the proposed method are comparable, and the performance of both methods improves as the number of simulated data becomes larger. However, the new two-step inference is more versatile since it does not need to adjust the risk level, and it is not the empirical quantile estimation when the explanatory vector is categorical and the minimization is done for each category.

Lastly, I develop a novel way of forecasting risk at a higher level existing in the insurance portfolio in Chapter 4. In many cases, insurance companies require to hold capital by regulators for their financial security. Two widely employed conditional risk measures of aggregate loss in the financial industry and insurance business are the conditional Value-at-Risk (VaR) and conditional Expected Shortfall (ES). Practically, regulators often require the risk level α to be high, such as 0.99 for VaR and 0.975 for ES, making nonparametric inference inefficient. On the other hand, a parametric inference may lead to an unstable risk forecast due to the higher risk level and the

fact that the parametric inference mainly employs the information around the distribution center. For dealing with a higher risk level and applications to more general risk measures, I propose to model the conditional excess function of aggregate loss over a threshold by a generalized Pareto distribution (GPD). It is not new to model the parameters in the GPD as parametric functions of some covariates. However, the threshold in the previous research is independent of the covariates. Because of using a dynamic GPD, it has no reason to employ a static threshold, which motivates me to study the following three-step procedure for forecasting risk at a high-risk level: i) using logistic regression to model the probability of having positive loss at the first stage, ii) using quantile regression to model the dynamic threshold u_i at the 90% or 95% level at the second stage as a rule of thumb in fitting a generalized Pareto distribution, iii) and fitting the dynamic generalized Pareto distribution for modeling excess losses over the dynamic threshold u_i. To quantify the risk forecast uncertainty, I further propose a random weighted bootstrap method. I show that the nonparametric forecast for VaR and ES are statistically different from the three-step forecasts, which implies that employing the nonparametric estimation for risk measures fails to perform well with a higher risk level.

My dissertation contributes to several aspects of related pieces of literature. First, I continue to develop a novel way of quantifying uncertainty regarding the inference methodology for quantile risk measures that many of the previous works of literature have been unable to address. Even though a novel statistical methodology is suggested for an efficient forecast, no consideration of making quantification of uncertainty regarding the methodology can make it harder to prove their efficiency and sound implication. I aim to overcome such a limitation of the existing literature and make the statistical data analysis lucrative. Second, I make a contribution to enlarging the inference methods in a more general setting. Not only my proposed inference methodologies are able to apply in more general cases, such as data with continuous explanatory variables, but they are also more efficient in the objective that requires extra caution, such as the forecast of risk at an extreme probability level.

References

Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E., and Thandi, N. (2004). A Practitioner's Guide To Generalized Linear Models. In: *CAS Discussion Paper Program*, pp. 1–116.

Baione, F. and Biancalana, D. (2019). An individual risk model for premium calculation based on quantile: a comparison between generalized linear models and quantile regression. In: *North American Actuarial Journal* 23 (4), pp. 573–590.

Baione, F. and Biancalana, D. (2021). An application of parametric quantile regression to extend the two-stage quantile regression for ratemaking. In: *Scandinavian Actuarial Journal* 2021 (2), pp. 156–170.

Balkema, A.A. and De Haan, L. (1974). Residual lifetime at great age. In: *The Annals of Probability* 2 (5), pp. 792–804.

Barbe, P. and Bertail, P. (1995). The Weighted Bootstrap. Monograph, Lecture Notes in Statistics, Springer-Verlag, New York.

Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. In: *The annals of statistics* 9 (6), pp. 1196–1217.

Chatterjee, S. and Bose, A (2005). Generalized bootstrap for estimating equations. In: *The Annals of Statistics* 33 (1), pp. 414–436.

Chavez-Demoulin, V., Davison, A.C., and McNeil, A.J. (2005). A point process approach to value-at-risk estimation. In: *Quantitative Finance* 5 (2), pp. 227–234.

Chavez-Demoulin, V. and Embrechts, P. (2004). Smooth extremal models in finance and insurance. In: *Journal of Risk and Insurance* 71 (2), pp. 183–199.

Chavez-Demoulin, V., Embrechts, P., and Sardy, S. (2014). Extreme-quantile tracking for financial time series. In: *Journal of Econometrics* 181 (1), pp. 44–52.

Chen, K., Ying, Z., Zhang, H., and Zhao, L. (2008). Analysis of least absolute deviation. In: *Biometrika* 95 (1), pp. 107–122.

Chen, L., Wei, L.J., and Parzen, M.I. (2004). Quantile regression for correlated observations. In Proceedings of the Second Seattle Symposium in Biostatistics, pp. 51-69. Springer, New York.

Chiang, C.T., James, L.F., and Wang, M.C. (2005). Random weighted bootstrap method for recurrent events with informative censoring. In: *Lifetime Data Analysis* 11 (4), pp. 489–509.

Davidson, R. (2012). Statistical inference in the presence of heavy tails. In: *The Econometrics Journal* 15 (1), pp. C31–C53.

De Jong, P. and Heller, G.Z. (2008). Generalized linear models for insurance data. Cambridge: Cambridge University Press.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1999). Modelling extremal events. In: *British Actuarial Journal* 5 (2), pp. 465–465.

EU Commission (2009). Directive 2009/138/EC of the European Parliament and of the Council, Official Journal of the European Union.

Fan, J. and Gijbels, I. (1996). Local Polynomial Modelling and its Applications. Chapman and Hall, London.

Feng, X., He, X., and Hu, J. (2011). Wild bootstrap for quantile regression. In: *Biometrika* 98 (4), pp. 995–999.

Frees, E.W. (2010). Regression modeling with actuarial and financial applications. New York: Cambridge University Press.

Frees, E.W., Lee, G., and Yang, L. (2016). Multivariate frequency-severity regression models in insurance. In: *Risks* 4 (1), p. 4.

Frumento, P. and Bottai, M. (2016). Parametric modeling of quantile regression coefficient functions. In: *Biometrics* 72 (1), pp. 74–84.

Fung, T.C., Badescu, A.L., and Lin, X.S. (2019a). A class of mixture of experts models for general insurance: Application to correlated claim frequencies. In: *ASTIN Bulletin: The Journal of the IAA* 49 (3), pp. 647–688.

Fung, T.C., Badescu, A.L., and Lin, X.S. (2019b). A class of mixture of experts models for general insurance: Theoretical developments. In: *Insurance: Mathematics and Economics* 89, pp. 111–127.

Goldburd, M., Khare, A., Tevet, D., and Guller, D. (2016). Generalized Linear Models For Insurance Rating. In: *CAS Monograph, Series Number 5*.

Hagemann, A. (2017). Cluster-robust bootstrap inference in quantile regression models. In: *Journal of the American Statistical Association* 112 (517), pp. 446–456.

Hahn, J. (1995). Bootstrapping quantile regression estimators. In: *Econometric Theory* 11 (1), pp. 105–121.

Hall, G.P. and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. In: *Statistical Science*, pp. 153–167.

Hao, L. and Naiman, D.Q. (2007). Quantile Regression(No. 149). SAGE Publication.

He, X. and Shao, Q.M. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. In: *The Annals of Statistics* 24 (6), pp. 2608–2630.

Henckaerts, R., Antonio, K., and Clijsters, M. (2018). A data driven binning strategy for the construction of insurance tariff classes. In: *Scandinavian Actuarial Journal* 2018 (8), pp. 681–705.

Heras, A., Moreno, I., and Vilar-Zanón, J.L. (2018). An application of two-stage quantile regression to insurance ratemaking. In: *Scandinavian Actuarial Journal* 2018 (9), 753–769.

Hou, Y., Kang, S., Lo, C.C., and Peng, L. (2022). Three-step risk inference in insurance ratemaking. In: *Insurance: Mathematics and Economics* 105, pp. 1–13.

Hull, J. (2012). Risk management and financial institutions. John Wiley Sons.

IAA (2009). Measurement of Liabilities for Insurance Contracts: Current Estimates and Risk Margins. An International Actuarial Research Paper prepared by the ad hoc Risk Margin Working Group, International Actuarial Association. April 15, 2009.

IASB (2007). International Accounting Standards Board (2007). Discussion Paper, Preliminary Views on Insurance Contracts.

IASB (2017). IFRS 17 Insurance Contracts. International Accounting Standards Board. International Actuarial Association. 2018. Risk Adjustments for Insurance Contracts Under IFRS 17. Educational Monograph. ISBN: 978-0-981396.

Jin, Z., Ying, Z., and Wei, L.J. (2001). A simple resampling method by perturbing the minimand. In: *Biometrika* 88 (2), pp. 381–390.

Kang, S., Peng, L., and Golub, A. (2021). Two-step risk analysis in insurance ratemaking. In: *Scandinavian Actuarial Journal* 2021 (6), pp. 532–542.

Kang, S., Peng, L., and Xiao, H. (2020). Risk analysis with categorical explanatory variables. In: *Insurance: Mathematics and Economics* 91, pp. 238–243.

Kelly, B. and Jiang, H. (2014). Tail risk and asset prices. In: *The Review of Financial Studies* 27 (10), pp. 2841–2871.

Koenker, R. (2005). Quantile Regression(Econometric Society Monographs). Cambridge: Cambridge University Press.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. In: *Econometrica* 46 (1), pp. 33–50.

Koenker, R. and Hallock, K.F. (2001). Quantile regression. In: *Journal of Economic Perspectives* 15 (4), pp. 143–156.

Koenker, R., Chernozhukov, V., He, X., and Peng, L. (2017). Handbook of quantile regression.

Kudryavtsev, A.A. (2009). Using quantile regression for rate-making. In: *Insurance: Mathematics and Economics* 45 (2), pp. 296–304.

Liu, R.Y. (1988). Bootstrap procedures under some non-iid models. In: *The Annals of Statistics* 16 (4), pp. 1696–1708.

Massacci, D. (2017). Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. In: *Management Science* 63 (9), pp. 3072–3089.

McNeil, A.J., Frey, R., and Embrechts, P. (2015). Quantitative risk management: concepts, techniques and tools-revised edition. Princeton university press.

Owen, A. B. (2001). Empirical likelihood. Chapman and Hall/CRC.

Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. In: *Biometrika* 75, pp. 237–249.

Owen, A.B. (1990). Empirical likelihood ratio confidence regions. In: *The annals of statistics* 18 (1), pp. 90–120.

Pelkiewicz, A.J., Ahmed, S.W., Fulcher, P., Johnson, K.L., Reynolds, S.M., Schneider, R.J., and Scott, A.J. (2020). A review of the risk margin–Solvency II and beyond. In: *British Actuarial Journal* 25.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. In: *the Annals of Statistics* 22 (1), pp. 300–325.

Rao, C.R. and Zhao, L.C. (1992). Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap. In: *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 323–331.

Resnick, S.I. (2008). Extreme values, regular variation, and point processes (Vol. 4). Springer Science Business Media.

Rockafellar, R.T. (2007). Coherent approaches to risk in optimization under uncertainty. In: *OR Tools and Applications: Glimpses of Future Technologies*, pp. 38–61.

Rockafellar, R.T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. In: *Journal of risk* 2, pp. 21–42.

Rolski, T., Schmidli, H., Schmidt, V., and Teugels, J.L. (2009). Stochastic processes for insurance and finance. John Wiley & Sons.

Smyth, G.K. and Jørgensen, B. (2002). Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. In: *ASTIN Bulletin: The Journal of the IAA* 32 (1), pp. 143–157.

Van der Vaart, A.W. (2000). Asymptotic Statistics. Cambridge University Press.

Wolny-Dominiak, A. and Trzesiok, M. (2014). InsuranceData: A Collection of Insurance Datasets Useful in Risk Classification in Non–life Insurance. R package version, 1.

Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. In: *The Annals of Statistics* 14 (4), pp. 1261–1295.

Zheng, Y., Zhu, Q., Li, G., and Xiao, Z. (2018). Hybrid quantile regression estimation for time series models with conditional heteroscedasticity. In: *Journal of the Royal Statistical Society Series B* 80 (5), pp. 975–993.

Zhu, K. (2016). Bootstrapping the portmanteau tests in weak auto-regressive moving average models. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78 (2), pp. 463–485.