Three-Step Risk Inference In Insurance Ratemaking

Yanxi Hou[1], Seul Ki Kang[2], Chia Chun Lo[2] and Liang Peng[2]

Abstract. As catastrophic events happen more and more frequently, accurately forecasting risk at a high level is vital for the financial stability of the insurance industry. This paper proposes an efficient three-step procedure to deal with the semicontinuous property of insurance claim data and forecast extreme risk. The first step uses a logistic regression model to estimate the nonzero claim probability. The second step employs a quantile regression model to select a dynamic threshold for fitting the loss distribution semiparametrically. The third step fits a generalized Pareto distribution to exceedances over the selected dynamic threshold. Combining these three steps leads to an efficient risk forecast. Furthermore, a random weighted bootstrap method is employed to quantify the uncertainty of the derived risk forecast. Finally, we apply the proposed method to an automobile insurance data set.

Keywords: Generalized Pareto distribution, Insurance loss, Logistic regression, Quantile regression, Random weighted bootstrap

1 Introduction

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

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

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

[1]School of Data Science, FuDan University, China

[2]Department of Risk Management and Insurance, Georgia State University, Atlanta, GA 30303, USA

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 [9] 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

$$P(S_i \leq s|\mathbf{X}_i) = P(N_i = 0|\mathbf{X}_i) + P(N_i > 0|\mathbf{X}_i)P(S_i \leq s|\mathbf{X}_i, N_i > 0) \text{ for } s > s_0, \tag{1}$$

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 [9] 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, [13] develops a random weighted bootstrap method when the second step uses empirical quantile estimation rather than quantile regression. [12] 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 [18] and [6] for an overview about Extreme Value Theory), there exists a function $\sigma_{i,\mathbf{x}}(u) > 0$ such that

$$\lim_{u \to z_0} \sup_{0 \leq z < z_0 - u} |F_{i,\mathbf{x},u}(z) - G_{\xi_{i,\mathbf{x}}, \sigma_{i,\mathbf{x}}(u)}(z)| = 0, \tag{2}$$

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

$$G_{\xi,\sigma}(z) = 1 - (1 + \xi z/\sigma)^{-1/\xi} \tag{3}$$

for $1 + \xi z/\sigma > 0$ with $\sigma > 0$ and $\xi \in \mathrm{R}$ is called the generalized Pareto distribution (GPD); see [1]. 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 (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 [2], [14], [3], and [17] for financial returns and [8] 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 [10]), 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 [9] only works for Value-at-Risk. To quantify the uncertainty of the derived risk forecast, we develop a random weighted bootstrap method.

We organize the paper as follows. Section 2 presents the three-step inference for estimating risk measures at a high level and a random weighted bootstrap method for uncertainty quantification. Section 3 analyzes an automobile dataset. We conclude the paper in Section 4. All theoretical derivations appear in Section 5.

2 Methodologies and Main Results

2.1 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 (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 [9], the first step models the conditional probability of $p(\mathbf{X}_i) = P(N_i =$

$0|\mathbf{X}_i)$ by logistic regression:

$$I(N_i = 0)|\mathbf{X}_i \sim \text{Bin}(1, p(\mathbf{X}_i)) \quad \text{and} \quad p(\mathbf{X}_i) = p(\mathbf{X}_i; \boldsymbol{\theta}_1) = \frac{1}{1 + \exp(\boldsymbol{\theta}_1^\tau \bar{\mathbf{X}}_i)}, \tag{4}$$

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

$$\hat{p}(\mathbf{x}) = p(\mathbf{x}; \hat{\boldsymbol{\theta}}_1) = \frac{1}{1 + \exp(\hat{\boldsymbol{\theta}}_1^\tau \bar{\mathbf{x}})}. \tag{5}$$

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

$$u(\mathbf{X}_i) = u(\mathbf{X}_i; \boldsymbol{\theta}_2) = \text{VaR}_{\tilde{S}_i}(\alpha_0|\mathbf{X}_i) \tag{6}$$

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 [10], we employ $\alpha_0 = 90\%$ or 95% in practice.

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

$$\hat{\boldsymbol{\theta}}_2 = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{\tilde{n}} \rho_{\alpha_0}(\tilde{S}_i - u(\mathbf{X}_i; \boldsymbol{\theta})) \tag{7}$$

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

$$\hat{u}(\mathbf{X}_i) = u(\mathbf{X}_i; \hat{\boldsymbol{\theta}}_2). \tag{8}$$

The third step models the conditional excess function of $\tilde{S}_i$ over the threshold $u(\mathbf{X}_i)$ in (6) given $\mathbf{X}_i$ by the generalized Pareto distribution

$$P(\tilde{S}_i > u(\mathbf{X}_i) + z|\mathbf{X}_i) = (1 - \alpha_0)\{1 + \xi(\mathbf{X}_i)z/\sigma(\mathbf{X}_i)\}^{-1/\xi(\mathbf{X}_i)}, \tag{9}$$

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 (6), we assume that

$$\xi(\mathbf{X}_i) = \xi(\mathbf{X}_i; \boldsymbol{\theta}_3) \quad \text{and} \quad \sigma(\mathbf{X}_i) = \sigma(\mathbf{X}_i; \boldsymbol{\theta}_4), \tag{10}$$

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 of high risk level α, we consider the case of $\alpha^*(\mathbf{x}) \geq \alpha_0$ and assume $N_i > 0$ when $S_i > \text{VaR}_S(\alpha|\mathbf{x})$. Therefore,

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

and

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

when $0 < \xi(\mathbf{x}) < 1$. Because we 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

$$\widehat{\text{VaR}_S}(\alpha|\mathbf{x}) = \hat{u}(\mathbf{x}) + \frac{\hat{\sigma}(\mathbf{x})}{\hat{\xi}(\mathbf{x})} \left\{ \left(\frac{1 - \hat{\alpha}^*(\mathbf{x})}{1 - \alpha_0} \right)^{-\hat{\xi}(\mathbf{x})} - 1 \right\} \tag{11}$$

and

$$\begin{aligned} \widehat{\text{ES}}_S(\alpha|\mathbf{x}) \quad &= \tfrac{1-\hat{\alpha}^*(\mathbf{x})}{1-\alpha} \widehat{\text{VaR}}_S(\alpha|\mathbf{x}) + \tfrac{1-\alpha_0}{1-\alpha} \tfrac{\hat{\sigma}(\mathbf{x})}{1-\hat{\xi}(\mathbf{x})} \left\{ 1 + \hat{\xi}(\mathbf{x}) \tfrac{\widehat{\text{VaR}}_S(\alpha|\mathbf{x}) - \hat{u}(\mathbf{x})}{\hat{\sigma}(\mathbf{x})} \right\}^{-1/\hat{\xi}(\mathbf{x})+1} \\ &= \tfrac{1-\hat{\alpha}^*(\mathbf{x})}{1-\alpha} \left\{ \tfrac{\widehat{\text{VaR}}_S(\alpha|\mathbf{x})}{1-\hat{\xi}(\mathbf{x})} + \tfrac{\hat{\sigma}(\mathbf{x}) - \hat{\xi}(\mathbf{x}) \hat{u}(\mathbf{x})}{1-\hat{\xi}(\mathbf{x})} \right\}. \end{aligned} \tag{12}$$

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 1 below.

Remark 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 [7] for kernel smoothing techniques.

Remark 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*

$$u(\mathbf{x};\boldsymbol{\theta}_2) = \exp(\bar{\mathbf{x}}^\tau \boldsymbol{\theta}_2), \quad \xi(\mathbf{x};\boldsymbol{\theta}_3) = \exp(\bar{\mathbf{x}}^\tau \boldsymbol{\theta}_3), \quad \textit{and} \quad \sigma(\mathbf{x};\boldsymbol{\theta}_4) = \exp(\bar{\mathbf{x}}^\tau \boldsymbol{\theta}_4). \tag{13}$$

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 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). Given $\{\mathbf{X}_i\}$, $\{S_i\}$ is a sequence of independent random variables satisfying model (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 (6), (9), and (10) with (13).

Assumption 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 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 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,

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

An application of the delta method to Theorem 1 yields the following asymptotic limits of our risk forecasts.

Theorem 2. *Suppose conditions in Theorem 1 hold. Further assume $\alpha^*(\mathbf{x}) > \alpha_0$, $N_i > 0$ whenever $S_i > \mathrm{VaR}_S(\alpha|\mathbf{x})$, and $\xi(\mathbf{x}) < 1$ for estimating $\mathrm{ES}_S(\alpha|\mathbf{x})$. Then, as $n \to \infty$,*

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

and

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

where $\Sigma_{\mathbf{x}}$ is defined in Lemma 5 below.

2.2 Uncertainty quantification

Because the asymptotic variances in Theorem 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 [11], [4], and [21] 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 $\mathrm{V}aR_S(\alpha|\mathbf{x})$ 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 $\Delta_{B,1} \le \cdots \le \Delta_{B,B}$ 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. 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 3. *Under the conditions of Theorem 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

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

3 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 [20]). We refer to [5] 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 [9]. 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 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 2 reports estimates in the third step, where we choose $\xi(\mathbf{x};\boldsymbol{\theta}_3)$ as constant following [8], i.e., ξ is independent of $\mathbf{X}_i$. Using these fittings above, Tables 3 and 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 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 reports estimates and intervals for the conditional ES at 97.5% level.

Our observations from Tables 3 and 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 threshold, except for the first five categories in Table 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 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 Conclusions

This paper 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 paper 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.

5 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 1. *Under conditions of Theorem 1,*

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

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

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 1 follows from the central limit theorem.

□

Lemma 2. *Under conditions of Theorem 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 [19] 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 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 [15] 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. *Under conditions of Theorem 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,

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

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

implying that

$$\sqrt{\tilde{n}}\frac{\partial \hat{L}_3(\boldsymbol{\eta}_3)}{\partial \boldsymbol{\theta}_3} = \mathbf{W}_3 + \Gamma_{3,1}\Gamma_2^{-1}\mathbf{W}_2 + o_p(1). \tag{14}$$

Similarly,

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

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

implying that

$$\frac{1}{\sqrt{\tilde{n}}}\frac{\partial L_3(\boldsymbol{\eta}_3)}{\partial \boldsymbol{\theta}_4} = \mathbf{W}_4 + \Gamma_{3,2}\Gamma_2^{-1}\mathbf{W}_2 + o_p(1). \tag{15}$$

Because

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

we have

$$\frac{1}{\tilde{n}}\sum_{i=1}^{\tilde{n}}\frac{\partial^2 L_3(\boldsymbol{\eta}_3)}{\partial\boldsymbol{\eta}_3\partial\boldsymbol{\eta}_3^\tau} = -\Gamma_4 + o_p(1). \tag{16}$$

Hence, the lemma follows from (14)-(16).

□

Proof of Theorem 1. The theorem follows from Lemmas 1–4. □

Lemma 5. *Under conditions of Theorem 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*

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

Proof. It follows from Lemmas 1 - 4 and the delta method that

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

$$\begin{aligned}
\sqrt{n}(\hat{\xi}(\mathbf{x}) - \xi(\mathbf{x}), \hat{\sigma}(\mathbf{x}) - \sigma(\mathbf{x}))^\tau =& \begin{pmatrix} \frac{\xi(\mathbf{x})}{\sqrt{1-p_0}}\bar{\mathbf{x}}^\tau & 0 \\ 0 & \frac{\sigma(\mathbf{x})}{\sqrt{1-p_0}}\bar{\mathbf{x}}^\tau \end{pmatrix} \sqrt{n}(\hat{\boldsymbol{\eta}}_3 - \boldsymbol{\eta}_3) + o_p(1) \\
=& \begin{pmatrix} \frac{\xi(\mathbf{x})}{\sqrt{1-p_0}}\bar{\mathbf{x}}^\tau & 0 \\ 0 & \frac{\sigma(\mathbf{x})}{\sqrt{1-p_0}}\bar{\mathbf{x}}^\tau \end{pmatrix} \Gamma_4^{-1}\mathbf{W}_5 + o_p(1),
\end{aligned}$$

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 2. The theorem follows immediately from Lemma 5 and the delta method. □

Proof of Theorem 3. Like the proof of Lemma 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 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.

$\square$

Acknowledgments

We thank two reviewers for their helpful comments. Yanxi Hou's research was partially supported by the National Natural Science Foundation of China grants 72171055, 71803026, and 71991471, and by the Natural Science Foundation of Shanghai grant 20ZR1403900. Peng's research was partly supported by the Simons Foundation and the NSF grant of DMS-2012448.

References

[1] Balkema, A. A., and De Haan, L. (1974). Residual lifetime at great age. *Annals of Probability, 2(5), 792-804.*

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

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

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

[5] De Jong, P., and Heller, G. Z. (2008). Generalized Linear Models for Insurance Data. *Cambridge University Press.*

[6] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1999). Modelling extremal events. *British Actuarial Journal, 5(2), 465-465.*

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

[8] Hall, P., and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science, 15(2), 153-167.*

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

[10] Hull, J. C. (2018), Risk Management and Financial Institutions. *Hoboken: Wiley.*

[11] Jin, Z., Ying, Z., and Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika, 88(2), 381-390.*

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

[13] Kang, S., Peng, L., and Xiao, H. (2020). Risk analysis with categorical explanatory variables. *Insurance: Mathematics and Economics, 91, 238-243.*

[14] Kelly, B., and Jiang, H. (2014). Tail risk and asset prices. *The Review of Financial Studies, 27(10), 2841-2871.*

[15] Koenker, R. (2005). Quantile Regression (Econometric Society Monographs). *Cambridge University Press.*

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

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

[18] Resnick, S. I. (2013). Extreme Values, Regular Variation and Point Processes. *Springer.*

[19] Van der Vaart, A. W. (2000). Asymptotic Statistics (Vol. 3). *Cambridge University Press.*

[20] 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.*

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

Table 1: This table reports the estimates of logistic regression in the first step and quantile regression in the second step. The upper panel displays $\hat{\boldsymbol{\theta}}_1$ 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 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 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 3: This table reports sample size, nonparametric estimate, GPD estimate with a static threshold, 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 VaR$(0.99\|\boldsymbol{x}_j)$	GPD VaR$(0.99\|\boldsymbol{x}_j)$	$\widehat{\text{VaR}}(0.99\|\boldsymbol{x}_j)$	$I_1(0.9\|\boldsymbol{x}_j)$	$I_2(0.9\|\boldsymbol{x}_j)$
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: 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 ES$(0.975\|\boldsymbol{x}_j)$	GPD ES$(0.975\|\boldsymbol{x}_j)$	$\widehat{\text{ES}}(0.975\|\boldsymbol{x}_j)$	$I_1(0.9\|\boldsymbol{x}_j)$	$I_2(0.9\|\boldsymbol{x}_j)$
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 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.