# Regression Tree Credibility Model

Liqun Diao and Chengguo Weng

Department of Statistics and Actuarial Science, University of Waterloo

Advances in Predictive Analytics Conference, Waterloo, Ontario
Dec 1, 2017

# Overview

$$\underbrace{\text{Statistical Method}}_{\text{Regression Trees}} + \underbrace{\text{Actuarial Model}}_{\text{Credibility Model}}$$

# Outline

# Bühlmann-Straub Credibility Model

- CREDIBILITY THEORY has become the paradigm for insurance experience rating and widely used by actuaries.

  Bühlmann model (1967, 1969), and the Bühlmann-Straub model (1970).
- Consider a portfolio of $I$ risks, where each individual risk $i$ has $n_i$ years of claim experiences, $i = 1, 2, \ldots, I$.
- Let $Y_{i,j}$ denote the claim ratio of individual risk $i$ in year $j$, and $m_{i,j}$ be the associated volume measure, also known as weight variable.
- Collect all the claims experience of individual risk $i$ into a vector $\mathbf{Y}_i$, i.e., $\mathbf{Y}_i = (Y_{i,1}, \ldots, Y_{i,n_i})^T$.
- The profile of individual risk $i$ is characterized by $\theta_i$, which is the realization of a random element $\Theta_i$ (usually either a random variable or a random vector).

- Assume that the following conditions are satisfied:

  H01. Conditionally given $\Theta_i = \theta_i$, $\{Y_{i,j} : j = 1, 2, \ldots, n_i\}$ are independent

$$\mathbb{E}[Y_{i,j}|\Theta_i = \theta_i] = \mu(\theta_i) \text{ and } \mathrm{Var}[Y_{i,j}|\Theta_i = \theta_i] = \frac{\sigma^2(\theta_i)}{m_{i,j}}$$

  for some unknown but deterministic functions $\mu(\cdot)$ and $\sigma^2(\cdot)$;

  H02. The pairs $(\Theta_1, \mathbf{Y}_1), \ldots, (\Theta_I, \mathbf{Y}_I)$ are independent, and $\{\Theta_1, \ldots, \Theta_I\}$ are independent and identically distributed.

- Define structural parameters $\sigma^2 = \mathbb{E}[\sigma^2(\Theta_i)]$ and $\tau^2 = \mathrm{Var}[\mu(\Theta_i)]$ for risks within the collective $\mathcal{I} := \{1, 2, \ldots, I\}$, and denote

$$m_i = \sum_{j=1}^{n_i} m_{i,j}, \quad m = \sum_{i=1}^{I} m_i, \quad \overline{Y}_i = \sum_{j=1}^{n_i} \frac{m_{i,j}}{m_i} Y_{i,j}, \quad \overline{Y} = \sum_{i=1}^{I} \frac{m_i}{m} \overline{Y}_i,$$

- Let $\mu = \mathbb{E}[Y_{i,j}]$ denote the collective net premium.
- We are interested an estimator $\hat{\mu}(\Theta_i)$ of $\mu(\Theta_i)$, is called the correct individual premium of the individual risk (the fair risk premium), such that it makes the following quadratic loss as small as possible

$$E\left[(\hat{\mu}(\Theta_i) - \mu(\Theta_i))^2\right].$$

# Inhomogeneous Credibility Premium

- The (inhomogeneous) credibility premium for an individual risk is defined as the best premium predictor $P_i$ among the class

$$\left\{ Q : \ Q = a_0 + \sum_{i=1}^{I} \sum_{j=1}^{n_i} a_{i,j} Y_{i,j}, \ \ a_0, a_{i,j} \in \mathbb{R} \right\}$$

to minimize the quadratic loss

$$\mathbb{E}\left[ \left( Q_i - \mu(\Theta_i) \right)^2 \right]$$

and its formula is given by

$$P_i = \alpha_i \overline{Y}_i + (1 - \alpha_i)\mu, \tag{1}$$

where

$$\alpha_i = \frac{m_i}{m_i + \sigma^2/\tau^2} \tag{2}$$

is called *credibility factor* for an individual risk.

- The corresponding minimum quadratic loss for $\mathbb{E}\left[\left(Q_i - \mu(\Theta_i)\right)^2\right]$ is given by

$$L_i = \frac{\sigma^2}{m_i + \sigma^2/\tau^2}. \tag{3}$$

- The minimum quadratic loss for $\mathbb{E}\left[\left(Q_i - Y_{i,n+1}\right)^2\right]$ is given by

$$L_i = \frac{\sigma^2}{m_i + \sigma^2/\tau^2} + \sigma^2. \tag{4}$$

# Homogeneous Credibility Premium

- The homogeneous credibility premium for an individual risk $i$ from the collective $\mathcal{I}$ is defined as the best premium predictor $P_i$ among the class

$$\left\{ Q_i : \ Q_i = \sum_{i=1}^{I} \sum_{j=1}^{n_i} a_{i,j} Y_{i,j}, \ \mathbb{E}[Q_i] = \mathbb{E}[\mu(\Theta_i)], \ a_{i,j} \in \mathbb{R} \right\}$$

to minimize the quadratic loss

$$\mathbb{E}\left[ (Q_i - \mu(\Theta_i))^2 \right]$$

and its formula is given by

$$P_i = \alpha_i \overline{Y_i} + (1 - \alpha_i) \overline{Y}, \tag{5}$$

where

$$\alpha_i = \frac{m_i}{m_i + \sigma^2/\tau^2} \tag{6}$$

is called *credibility factor* for individual risk $i$.

- The corresponding minimum quadratic loss for $\mathbb{E}\left[\left(Q_i - \mu(\Theta_i)\right)^2\right]$ is given by

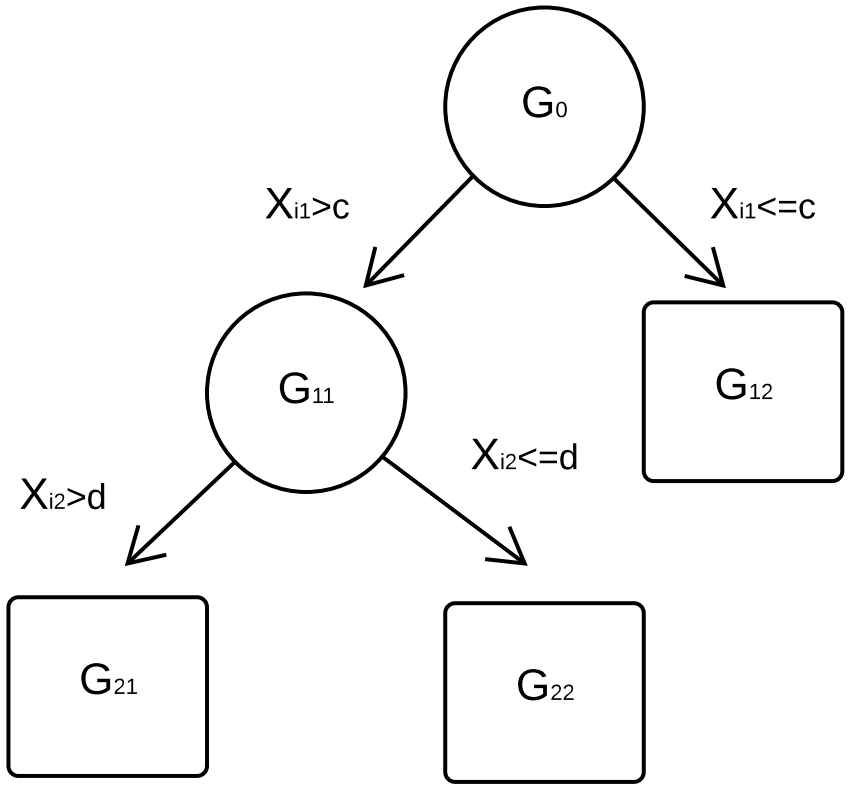
$$L_i = \tau^2(1-\alpha_i)\left(1 + \frac{1-\alpha_i}{\alpha_\bullet}\right), \tag{7}$$

  where $\alpha_\bullet = \sum_{i=1}^{I} \alpha_i$.

- The minimum quadratic loss for $\mathbb{E}\left[\left(Q_i - Y_{i,n+1}\right)^2\right]$ is given by

$$L_i = \tau^2(1-\alpha_i)\left(1 + \frac{1-\alpha_i}{\alpha_\bullet}\right) + \sigma^2, \tag{8}$$

  where $\alpha_\bullet = \sum_{i=1}^{I} \alpha_i$.




Better if we partition the collective?

The answer is artificially YES but genuinely not necessary

- For $n_i = n$ and $m_{i,j} = 1$ for and $i = 1, \ldots, I$ and $j = 1, \ldots, n$. Consider the loss function of homogenous premium and $\mu(\Theta_i)$.
  - Arbitrarily partition a given collective of individuals into two sub-collectives, and then apply credibility formula separately for each of the two resulting sub-collectives.
  - Let $L_1$ and $L_2$ be the total credibility losses we have for the two sub-collectives respectively, while $L$ denotes the credibility loss with premium prediction apply with the whole collective without any partitioning.
  - We formally proved $L_1 + L_2 \leq L$.

- The above result relies on an artificial assumption:

  We know the credibility factor $\alpha$ for each sub-collective".

- The premium prediction is given by

$$\widehat{P}_i = \widehat{\alpha}_i \overline{Y}_i + (1 - \widehat{\alpha}_i)\overline{Y},$$

where $\widehat{\alpha}_i$ is an estimator for $\alpha_i$, $i = 1, \ldots, I$.

# Overview of Regression Trees

(Breiman, Friedman, Stone and Olshen, 1984)

- CLASSIFICATION AND REGRESSION TREES (CART; Breiman, Friedman, Stone and Olshen, 1984) and RANDOM FORESTS (RF; Breiman, 2001) are the most popular single-tree and ensemble recursive partitioning methods respectively.
- Any tree building process can be broadly described in three steps:
  1. Choosing a criterion for making splitting decisions;
  2. Generating a corresponding sequence of candidate trees;
  3. Selecting best candidate tree.
- It's common to allow each step to depend on a given loss function
- Prevailing software implementation of CART: `rpart` by `R`

# Covariate-Dependent Model Setup

- Consider a portfolio of $I$ risks numbered with $1, \ldots, I$.
- Let $\mathbf{Y}_i = (Y_{i,1}, \ldots, Y_{i,n_i})^T$ be the vector of claim ratios, $\mathbf{m}_i = (m_{i,1}, \ldots, m_{i,n_i})$ be the corresponding weight vector, and $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,p})^T$ be the covariate vector associated with individual risk $i$, $i = 1, \ldots, I$.
- The risk profile of each individual risk $i$ is characterized by a scalar $\theta_i$, which is a realization of a random element $\Theta_i$.
- The following two conditions are further assumed:

H11. The triplets $(\Theta_1, \mathbf{Y}_1, \mathbf{X}_1), \ldots, (\Theta_I, \mathbf{Y}_I, \mathbf{X}_I)$ are independent;

H12. Conditionally given $\Theta_i = \theta_i$ and $\mathbf{X}_i = \mathbf{x}_i$, the entries $Y_{i,j}$, $j = 1, \ldots, n$, are independent with

$$\mathbb{E}[Y_{i,j} | \mathbf{X}_i = \mathbf{x}_i, \Theta_i = \theta_i] = \mu(\mathbf{x}_i, \theta_i) \text{ and}$$

$$\text{Var}[Y_{i,j} | \mathbf{X}_i = \mathbf{x}_i, \Theta_i = \theta_i] = \frac{\sigma^2(\mathbf{x}_i, \theta_i)}{m_{i,j}}$$

for some unknown but deterministic functions $\mu(\cdot, \cdot)$ and $\sigma^2(\cdot, \cdot)$.

- We approximate $\mu(\mathbf{X}_i, \Theta_i)$ and $\sigma^2(\mathbf{X}_i, \Theta_i)$:

$$\sum_{k=1}^{K} \mathrm{I}\{\mathbf{X}_i \in A_k\}\mu_{(k)}(\Theta_i), \text{ and } \sum_{k=1}^{K} \mathrm{I}\{\mathbf{X}_i \in A_k\}\frac{\sigma^2_{(k)}(\Theta_i)}{m_{i,j}},$$

  where

  - $\{A_1, A_2, \ldots, A_K\}$ is a partition of the covariate space
  - $\mu_{(k)}(\Theta_i)$ and $\sigma^2_{(k)}(\Theta_i)$ respectively represent the net premium and the variance of an individual risk $i$ from the $k$th sub-collective with a risk profile $\Theta_i$, i.e.,

$$\mu_{(k)}(\theta_i) = \mathbb{E}(Y_{i,j}|\mathbf{X}_i \in A_k, \Theta_i = \theta_i) \text{ and}$$

$$\sigma^2_{(k)}(\theta_i) = \mathrm{Var}(Y_{i,j}|\mathbf{X}_i \in A_k, \Theta_i = \theta_i).$$

  - The condition of $\mathbf{X}_i \in A_k$ means that the individual risk $i$ is classified into the $k$th sub-collective based on its covariate information.

# Regression Tree Credibility Premium

- We target to find a “Good" partition $\{A_1, A_2, \ldots, A_K\}$ and apply credibility formula for each sub-collective $A_i$ separately.
- By “Good", we should minimize the true prediction error $\mathbb{E}\left[\left(\mu(\Theta_i) - \widehat{P}_i\right)^2\right]$ as much as possible
- **Credibility Regression Trees**
  - Adopt one of the four credibility loss function with plugged-in estimates of structure parameters;
  - Invent a heuristic longitudinal cross-validation.
- Credibility formula is applied for premium prediction for each terminal node separately, which is Regression Tree Credibility Premium.

# Simulation Studies

- Consider a collective of $I = 300$ individual risks
- For each $i = 1, \ldots, I$, independently simulate covariate vector

$$\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,p}) \overset{\text{i.i.d.}}{\sim} U\{1, \ldots, 100\}^p$$

  when $p = 10$ and $p = 50$.
- Balanced Claims Model: individual risks with $n = 5$, 10 and 20 years of claims experience, respectively. Unbalanced claims model is also explored.
- $1,000$ independent samples

# Simulation Scheme 1

Interaction effect

For each $i = 1, \ldots, I$, independently simulate $\varepsilon_{i,j}$ from a given distribution function $F(\cdot)$, for $j = 1, \ldots, n$, and define the $n$ claims of individual risk $i$ as

$$Y_{i,j} = \mathrm{e}^{f(\mathbf{X}_i)} + \varepsilon_{i,j}, \quad j = 1, \ldots, n, \tag{9}$$

where

$$f(\mathbf{X}_i) = 0.01 \left( X_{i,1} + 2X_{i,2} - X_{i,3} + 2\sqrt{X_{i,1}X_{i,3}} - \sqrt{X_{i,2}X_{i,4}} \right). \tag{10}$$

In our simulation, $F$ takes one of the following three distributions:

- EXP(1.6487)
- LN (0, 1)
- PAR(3, 3.2974)

# Simulation Scheme 2

Interactive effect+heterogeneous variance

For each $i = 1, \ldots, I$, independently simulate $\varepsilon_{i,j}$ from a distribution function $F(\cdot; \mathbf{X}_i)$ which depends on the covariate $\mathbf{X}_i$ of the individual risk $i$, for $j = 1, \ldots, n$, and define the $n$ claims of individual risk $i$ as

$$Y_{i,j} = \mathrm{e}^{f(\mathbf{X}_i)} + \varepsilon_{i,j}, \quad j = 1, \ldots, n, \tag{11}$$

where $f(\mathbf{X}_i)$ is given by (10). We respectively consider three distinct distributions for $F(\cdot; \mathbf{X}_i)$:

$$\text{(1) EXP}\left(\mathrm{e}^{\gamma(\mathbf{X}_i)/2}\right), \qquad \text{(2) LN}(0, \gamma(\mathbf{X}_i)), \qquad \text{(3) PAR}(3, 2\mathrm{e}^{\gamma(\mathbf{X}_i)/2}), \tag{12}$$

where $\gamma(\mathbf{X}_i) = \frac{1}{102}\left|2X_{i,1} - X_{i,2} + \sqrt{X_{i,1}X_{i,2}}\right|$.

# Simulation Scheme 3

Interactive effect + heterogeneous variance + multiplication random effect

For each $i = 1, \ldots, I$,

(1) independently simulate random effect variable $\Theta_i$ from the uniform distribution $U(0.9,\ 1.1)$

(2) independently simulate $\varepsilon_{i,j}$ from a distribution function $F(\cdot; \mathbf{X}_i)$ which depends on the covariates $\mathbf{X}_i$ associated with risk $i$, for $j = 1, \ldots, n$, and

(3) define the $n$ claims of individual risk $i$ as

$$Y_{i,j} = \Theta_i \left[ e^{f(\mathbf{X}_i)} + \varepsilon_{i,j} \right], \quad j = 1, \ldots, n, \tag{13}$$

where $f(\mathbf{X}_i)$ is given by (10).

We consider each of the three distributions described in Scheme 2 for the distribution $F(\cdot; \mathbf{X}_i)$.

# Simulation Scheme 4

Interactive effect + heterogeneous variance + complex random effect structure

- For each $i = 1, \ldots, I$, $\Theta_i = (\xi_{i,1}, \xi_{i,2})^T$,
  (1) independently simulate random effect variables $\xi_{i,1}$ and $\xi_{i,2}$ from the uniform distribution $U(0.9,\ 1.1)$.
  (2) independently simulate $\varepsilon_{i,j}$ from a distribution function $F(\cdot; \mathbf{X}_i, \xi_{i,2})$ for $j = 1, \ldots, n$, and
  (3) define the $n$ claims of individual risk $i$ as

$$Y_{i,j} = \mathrm{e}^{\xi_{i,1} \cdot f(\mathbf{X}_i)} + \varepsilon_{i,j}, \quad j = 1, \ldots, n, \tag{14}$$

  where $f(\mathbf{X}_i)$ is defined in (10).

- We respectively consider the three distributions as defined in equation (12) in Scheme 2 with $\gamma(\mathbf{X}_i)$ replaced by $\xi_{i,2} \cdot \gamma(\mathbf{X}_i)$ so that the distribution of $\varepsilon_{i,j}$ depends on the random effect variable $\xi_{i,2}$ in addition to covariate variable $\mathbf{X}_i$.

- Data-driven covariate-dependent partitioning: For each simulated sample, we grow and prune regression trees built using four credibility loss functions (R1-4) and $L_2$ (R$L_2$)loss function and the best tree is selected using longitudinal cross-validation.
- In addition, we consider ad hoc covariate-dependent partitioning, which is defined via the following notation:

$$\mathcal{R}(X_j) = \Big\{\{i \in \mathcal{I} : X_{i,j} \leq 50\},\ \{i \in \mathcal{I} : X_{i,j} > 50\}\Big\},\ j = 1, \cdots, 5.$$

and

$$\mathcal{R}(X_{j_1}, X_{j_2}) = \Big\{\ \{i \in \mathcal{I} : X_{i,j_1} \leq 50, X_{i,j_2} \leq 50\},\ \{i \in \mathcal{I} : X_{i,j_1} \leq 50, X_{i,j_2} > 50\},$$
$$\{i \in \mathcal{I} : X_{i,j_1} > 50, X_{i,j_2} \leq 50\},\ \{i \in \mathcal{I} : X_{i,j_1} > 50, X_{i,j_2} > 50\}\Big\},$$

for $j_1, j_2 = 1, \ldots, 5$ with $j_1 \neq j_2$.

- We consider
  - $\mathcal{R}(X_2), \mathcal{R}(X_4)$,
  - $\mathcal{R}(X_1, X_2, X_3), \mathcal{R}(X_1, X_2, X_4), \mathcal{R}(X_2, X_3, X_4), \mathcal{R}(X_1, X_3, X_4)$, and
  - $\mathcal{R}(X_1, X_2, X_3, X_4)$.

- Prediction error for a given partitioning $\{\mathcal{I}_1, \ldots, \mathcal{I}_K\}$:

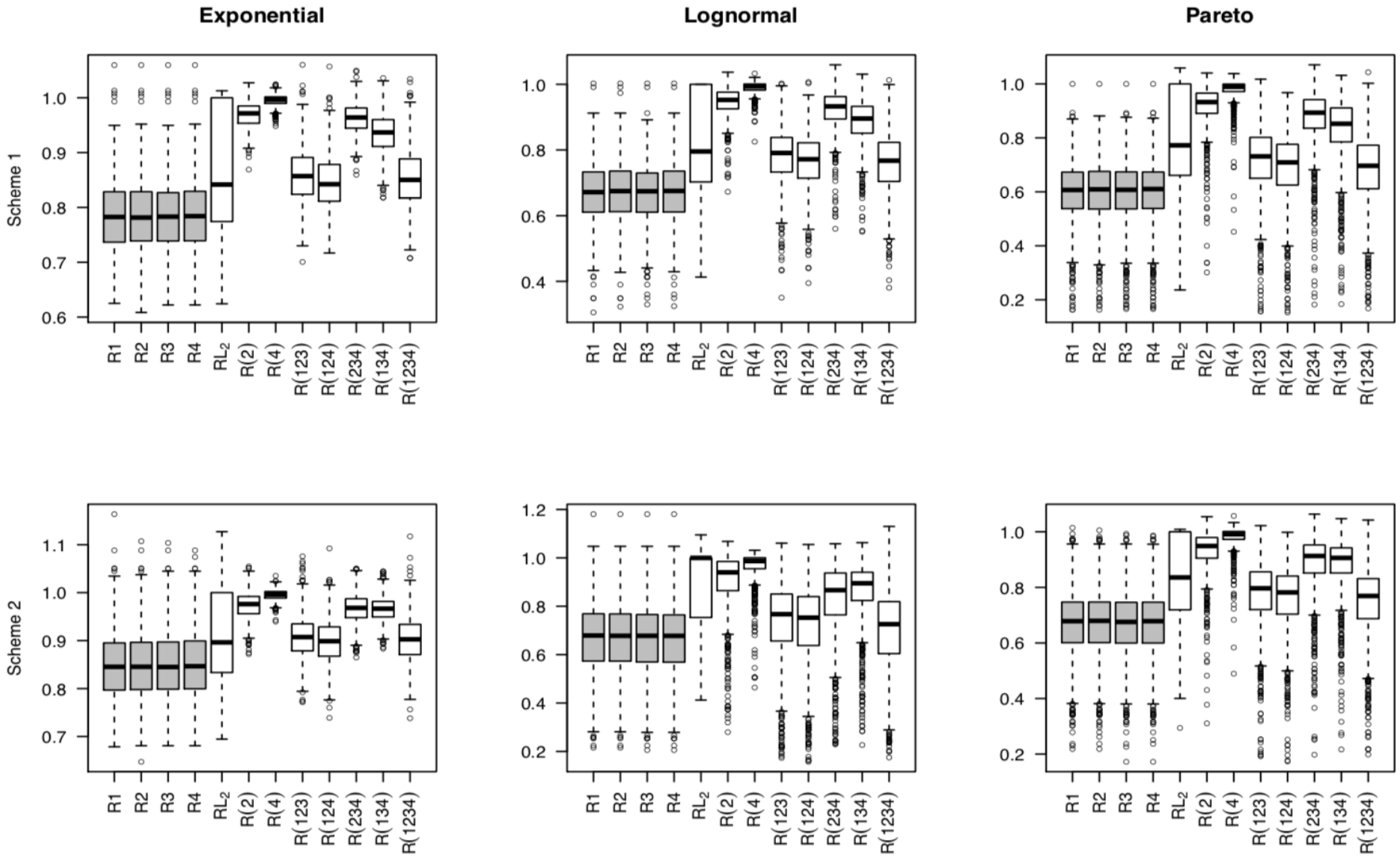
$$\mathrm{PE} = \frac{1}{I} \sum_{i=1}^{I} \sum_{k=1}^{K} \mathrm{I}\{\mathbf{X}_i \in \mathcal{I}_k\} \left( \pi_i^{(\mathrm{H})(k)} - \mu(\mathbf{X}_i, \Theta_i) \right)^2, \tag{15}$$

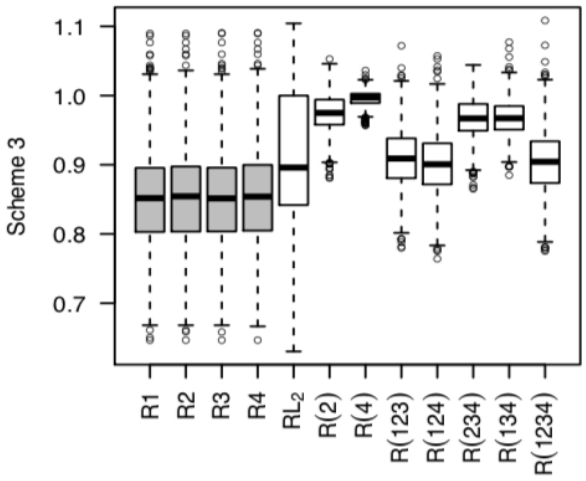
  where
  - $\pi_i^{(\mathrm{H})(k)}$ is the resulting premium prediction
  - $\mu(\mathbf{X}_i, \Theta_i)$ is the true net premium
- The collective prediction error:

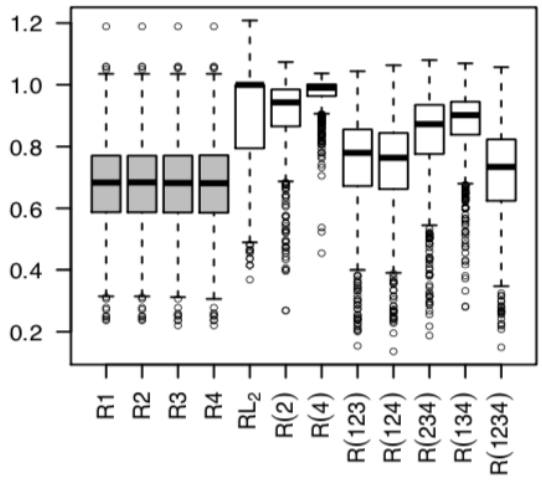
$$\mathrm{PE}_0 = \frac{1}{I} \sum_{i=1}^{I} \left( P_i^{(\mathrm{H})} - \mu(\mathbf{X}_i, \Theta_i) \right)^2,$$

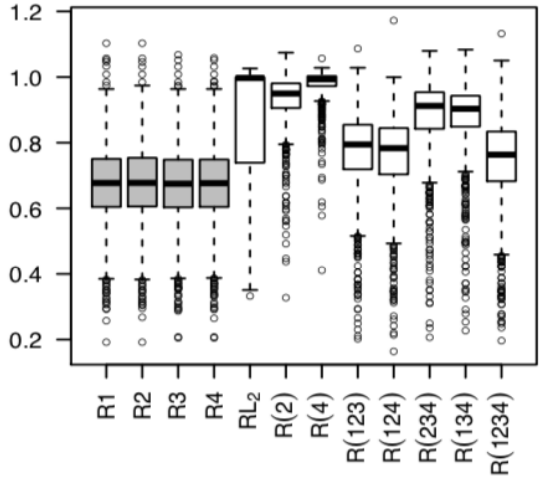
  which does not use any covariate information and is anticipated to underperform compared to various kinds of covariate-dependent partitioning.
- The relative prediction error (RPE):

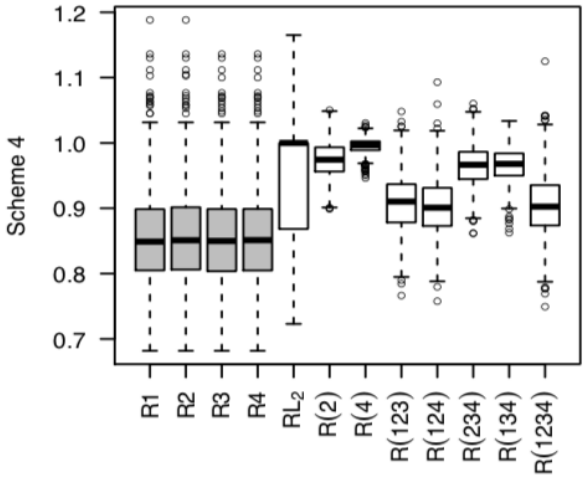
$$\mathrm{R} = \mathrm{PE}/\mathrm{PE}_0.$$

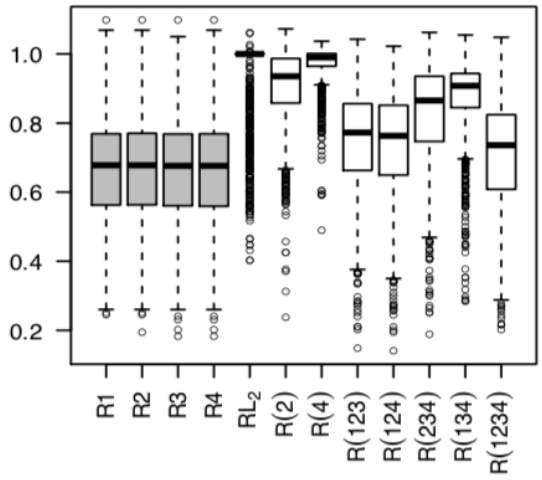

Exponential
Lognormal
Pareto
Scheme 1
Scheme 2
R1
R2
R3
R4
RL2
R(2)
R(4)
R(123)
R(124)
R(234)
R(134)
R(1234)

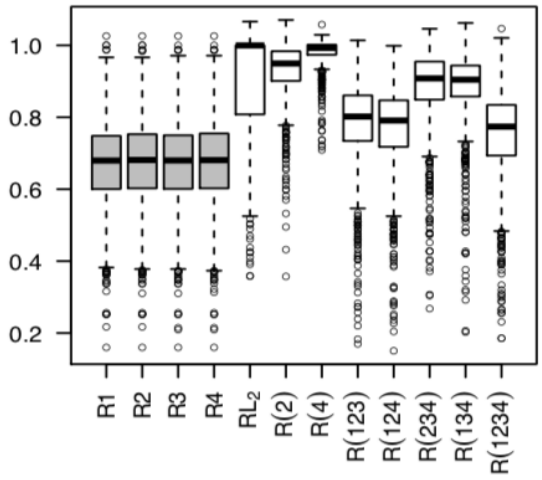

Scheme 3
1.1
1.0
0.9
0.8
0.7
R1
R2
R3
R4
RL2
R(2)
R(4)
R(123)
R(124)
R(234)
R(134)
R(1234)


1.2
1.0
0.8
0.6
0.4
0.2
R1
R2
R3
R4
RL2
R(2)
R(4)
R(123)
R(124)
R(234)
R(134)
R(1234)


1.2
1.0
0.8
0.6
0.4
0.2
R1
R2
R3
R4
RL2
R(2)
R(4)
R(123)
R(124)
R(234)
R(134)
R(1234)


Scheme 4
1.2
1.1
1.0
0.9
0.8
0.7
R1
R2
R3
R4
RL2
R(2)
R(4)
R(123)
R(124)
R(234)
R(134)
R(1234)


1.0
0.8
0.6
0.4
0.2
R1
R2
R3
R4
RL2
R(2)
R(4)
R(123)
R(124)
R(234)
R(134)
R(1234)


1.0
0.8
0.6
0.4
0.2
R1
R2
R3
R4
RL2
R(2)
R(4)
R(123)
R(124)
R(234)
R(134)
R(1234)

# Concluding Remarks

- We propose novel regression tree credibility (RTC) model, and bring machine learning techniques into the framework of credibility theory to enhance the prediction accuracy of credibility premium.
- In our proposed model, no ex ante analysis on the relationship between individual net premium and covariate variables is necessary, and the designed regression tree algorithm automatically selects influential covariate variables and informative cutting points to form a partition of data space, upon which a well-performed premium prediction rule can be consequently established.
- Our simulation studies and data analysis show that the proposed RTC model performs very well compared to no partitioning, ad-hoc partitioning and the $L_2$ loss based binary partition procedure.

- Although only the Classification and Regression Trees is introduced in this paper, it will be fruitful to pursue further research by considering other recursive partitioning methods, e.g.,
  - partDSA (partitioning deletion/substitution/addition algorithm) and
  - MARS (multivariate adaptive regression splines)
- It will be even more promising to consider the applications of ensemble algorithms, such as bagging, boosting, and random forests.
- It will be useful to develop an algorithm which can adopt time-dependent covaraites.
- It will be even more fruitful to consider their applications in various other insurance problems in addition to the premium rating, since many practical insurance problems amount to quantifying the relationship between insureds' claims and their demographic information.
- It is the authors' hope that the present paper will stimulate more actuarial applications of these machine learning techniques and eventually contribute to the development of insurance predictive analytics in general.