Article

Credibility Distribution Estimation with Weighted or Grouped Observations

Georgios Pitselis [1,2]

[1] Department of Statistics & Insurance Science, University of Piraeus, 80 Karaoli & Dimitriou Str. T. K., 18534 Piraeus, Greece; pitselis@unipi.gr; Tel.: +30-2104142026; Fax: +30-2104142340

[2] Department of Mathematics & Statistics, Concordia University, 1455 De Maisonneuve Blvd. W., Montreal, QC H3G 1M8, Canada

Abstract: In non-life insurance practice, actuaries are often faced with the challenge of predicting the number of claims and claim amounts to be incurred at any given time, which serve to implement fair pricing and reserves given the nature of the risk. This paper extends Jewell's credible distribution in terms of forecasting the distribution of individual risk in cases where the observations are weighted or are grouped in intervals. More specifically, we show how empirical distribution functions can be embedded within Bühlmann's and Straub's credibility model. The optimal projection theorem is applied for credibility estimation and more insight into the derivation of the credibility distribution estimators is also provided. In addition, distribution credibility estimators are established and numerical illustrations are presented herein. Two examples of distribution credibility estimation are given, one with insurance loss data and the other with industry financial data.

Keywords: credibility distribution estimation; empirical Bayes

1. Introduction

In actuarial science, one of the fundamental problems is that of predicting future claims of individual risk given one's past experience of a collective of heterogeneous risks. Credibility is a ratemaking technique that serves to forecast future premiums for a group of insurance contracts for which we have experience, whilst we have a lot more experience for a collection of contracts that are similar but not exactly the same.

In the insurance industry, some legislated rules indicate that some changes over time occurred across the claim distribution. Therefore, it is essential to examine these changes at different points of the distribution. An empirical distribution function provides a way to model and sample cumulative probabilities for a data sample that does not fit a standard probability distribution. Its value at a given point is equal to the proportion of observations from the sample that are less than or equal to that point.

In non-life insurance practice, actuaries are often faced with the challenge of predicting the number of claims and the claim amounts to be incurred at any given time, which serve to implement fair pricing and reserves given the nature of the risk. Actuaries usually deal with events that are uncertain and their economic consequences. The aim of this paper is to carry out the credibility estimation of empirical distribution functions in measuring and managing these uncertainties.

In the first part of this paper, we extend the work of Jewell (1974b) in terms of forecasting the distribution of individual risk in cases where the observations are weighted for the non-homogeneous and homogeneous models. Here, the weights (sizes) w_{ij}, $i = 1, \ldots, n_j$, $j = 1, \ldots, K$ are now changing in time. The contract j might result from a grouping and averaging of w_{ij} observations in a contract with several independent and identically distributed observations S_{lij}, $l = 1, \cdots w_{ij}$, during the year i, i.e., $X_{ij} = \frac{1}{w_{ij}} \sum_{l=1}^{w_{ij}} S_{lij}$, and then taking the conditional mean of the identity function $E[I(X_{ij} \leq x)|\Theta_j]$. Alternatively, in the case of

raw data, the contract j might result from the grouping and averaging of identity functions within the year i, $\bar{I}(S_{ij} \leq x) = \frac{1}{w_{ij}} \sum_{l=1}^{w_{ij}} I(S_{lij} \leq x)$ and then taking the $E[\bar{I}(S_{ij} \leq x)|\Theta_j]$.

Here, we proceed with the former considering the credibility distribution estimation as a point estimate approach of $F_{X_{ij}|\Theta_j}(x|\Theta_j) = E[I(X_{ij} \leq x)|\Theta_j]$. Optimal linearized estimators of $F_{X_{ij}\Theta_j}(x|\Theta_j)$ are obtained by the classical least squares approach as well as by the optimal projection theorem of random variables on planes as presented by De Vylder (1976, 1996).

In the second part of this paper, we consider credibility distribution estimation based on grouped data formed by aggregating the individual observations of a variable into groups. The construction of the empirical distribution based on grouped data can be performed by obtaining the point values of the empirical distribution function whenever is possible. Then, we approximate the distribution functions by connecting those points with straight lines and applying premium estimation in a credibility framework. An alternative model of credibility estimation is also obtained similarly as in Bühlmann and Straub (1970) model.

Related Works

Bühlmann (1967) and Bühlmann and Straub (1970) established the theoretical foundation of modern credibility theory, presented as a distribution-free credibility estimation. The method was extended in the regression model by Hachemeister (1975), where the credibility premium depends linearly on a number of risk characteristics.

Jewell (1974a) has shown that credibility is exactly Bayesian for a certain exponential family of distributions with natural conjugate priors. Furthermore, Landsman and Makov (1998, 1999) extended the results on the exponential family to the exponential dispersion family. The following key references are related to new developments in credibility estimation: Makov et al. (1996), Christiansen and Schinzinger (2016), Tsai and Lin (2017), Gong et al. (2018), Xacur and Garrido (2018), Tsai and Wu (2020), Tsai and Zhang (2019), Bozikas and Pitselis (2020, 2021), Youn et al. (2021), Wang et al. (2021), Yan and Song (2022), and Kim et al. (2022).

Credibility distribution estimation is closely connected to the area of quantile credibility estimation. The quantile function is the inverse of the distribution function. It specifies the value of the random variable such that the probability of the variable that is less than or equal to that value is equal to the given probability. Kim and Jeon (2013) proposed a credibility theory by truncating the loss data based on quantiles. Some other references related to quantile estimation are: Pitt (2006), Pitselis (2009, 2013, 2017), Kudryavtsev (2009), Gebizlioglu and Yagci (2008), Denuit (2008) and Landsman (1996).

Jewell (1974b) extended the classical Bühlmann (1967) model to the problem of forecasting the distribution of individual risk based upon collective statistics and individual experience data and solved the problem by finding a Bayesian conditional distribution. Jewell (1974b) also obtain an additional insight into the nature of credibility estimation assuming that the true value of Θ_j is known and obtained credible distributions and credible densities by carrying out simulations for some conjugate prior families of distributions (e.g., Poisson–gamma, etc.). He also considered the problem of founding a credibility approximation to the true distribution of the next observation.

Korwar and Hollander (1973) defined a sequence of empirical Bayes estimators for estimating a distribution function. Zehnwirth (1981) established the asymptotic optimality of the empirical Bayes distribution function created from the Bayes rule relative to the Dirichlet process prior with unknown parameter. Cai et al. (2015) combined Bühlmann's credibility theory and Ferguson's (1973) nonparametric Bayes analysis to develop a completely nonparametric estimation for loss distributions and established a unified distribution-free approach to experience rating for arbitrary premium principles.

This paper is organized as follows. In Section 2, both the linearized non-homogeneous and homogeneous estimators in the weighted credibility distribution model are obtained, and the credibility parameters are estimated. Optimal credibility distribution estimators are also obtained using the optimal projection theorem. In Section 3, the credibility distribution

estimation for grouped data is presented. In Section 4, an alternative model of credibility distribution estimation is obtained when the observations are grouped in intervals. Applications to real data are presented in Section 5, one with insurance loss data and the other with industry financial data. Some concluding remarks are presented in Section 6.

2. Weighted Credibility Distribution Estimation

In the following, we consider the credibility model with several contracts and weighted observations. For an insurance portfolio, X_{ij} are the average losses of w_{ij} observations for contract $j = 1, \cdots, K$ and period $i = 1, \cdots n_j$. For industry portfolios, X_{ij} denotes the average returns (losses/gains) of w_{ij} firms in $j = 1, \cdots K$ portfolios for period $i = 1, \cdots, n_j$.

2.1. Assumptions

We have the following assumptions:

(i)   The contracts are independent and the variables $\Theta_1, \cdots, \Theta_K$ are identically distributed.

(ii) $F_{X_{ij}|\Theta_j}(x|\Theta_j) = P(X_{ij} \le x|\Theta_j) = E[I(X_{ij} \le x)|\Theta_j]$, where $I(X_{ij} \le x)$ is an indicator function that is equal to 1 if $X_{ij} \le x$ and 0 otherwise.

(iii) $Cov[I(X_{ij} \le x), I(X_{rj} \le x)|\Theta_j] = \delta_{ir}\frac{1}{w_{ij}}\sigma^2_{F_x}(\Theta_j)$, where $\delta_{ir} = 1$ if $i = r$ and 0 otherwise.

2.2. Structural Parameters

The structural parameters $F_{X_{ij}}(x)$, $s^2_{F_x}$, and a_{F_x} are as follows:

(SP1) $F_{X_{ij}}(x) = E[F_{X_{ij}|\Theta_j}(x|\Theta_j)] = E[P(X_{ij} \le x|\Theta_j)]$;

(SP2) $s^2_{F_x} = E[\sigma^2_{F_x}(\Theta_j)]$;

(SP3) $a_{F_x} = Var\{E[I(X_{ij} \le x)|\Theta_j]\} = Var[F_{X_{ij}|\Theta_j}(x|\Theta_j)]$.

2.3. Notation

Here, we present the weighted empirical distribution function as well as some notations that are useful for the derivation of the credibility distribution estimation.

$$F_{n_{wj}}(x) = \sum_{i=1}^{n_j} \frac{w_{ij}}{w_{\cdot j}} I(X_{ij} \le x), \ F_{n_{ww}}(x) = \sum_{j=1}^{K} \frac{w_{\cdot j}}{w_{\cdot\cdot}} F_{n_{wj}}(x), \ w_{\cdot j} = \sum_{i=1}^{n_j} w_{ij},$$

$$w_{\cdot\cdot} = \sum_{i=1}^{K} w_{\cdot j}, \ F_{n_{wz}}(x) = \sum_{j=1}^{K} \frac{Z_j^{F_x}}{z_{\cdot}^{F_x}} F_{n_{wj}}(x), \ z_{\cdot}^{F_x} = \sum_{j=1}^{K} Z_j^{F_x},$$

$$Z_j^{F_x} = \frac{a_{F_x} w_{\cdot j}}{a_{F_x} w_{\cdot j} + s^2_{F_x}}. \tag{1}$$

Lemma 1 (**Expectation and Covariance Relations**). *Based on the above assumptions, we can obtain expressions for the conditional expectations and covariances as follows,*

$$E[I(X_{ij} \le x)] = E[F_{n_{wj}}(x)] = E[F_{n_{ww}}(x)] = E[F_{n_{wz}}(x)] = F_{X_{ij}}(x), \tag{2}$$

$$Cov[I(X_{ij} \le x), I(X_{rj} \le x)] = \delta_{ir}\frac{1}{w_{ij}}s^2_{F_x} + a_{F_x}, \tag{3}$$

$$Cov[I(X_{ij} \le x), F_{n_{wj}}(x)] = Cov[F_{n_{wj}}(x), F_{n_{wj}}(x)] = \frac{1}{w_{\cdot j}}s^2_{F_x} + a_{F_x} = \frac{a_{F_x}}{Z_j^{F_x}}, \tag{4}$$

$$Cov[I(X_{ij} \leq x), F_{n_{wz}}(x)] = Cov[F_{n_{wj}}(x), F_{n_{wz}}(x)] = Cov[F_{n_{wz}}(x), F_{n_{wz}}(x)] = \frac{a_{F_x}}{z_{\cdot}^{F_x}}, \quad (5)$$

$$Cov[F_{X_{ij}|\Theta_j}(x|\Theta_j), I(X_{ij'} \leq x)] = \delta_{jj'} a_{F_x}. \quad (6)$$

Proof. Relation (2) is straightforward. Relation (3) results from

$$Cov[I(X_{ij} \leq x), I(X_{rj} \leq x)]$$

$$= E\Big(Cov[I(X_{ij} \leq x), I(X_{rj} \leq x)|\Theta_j]\Big) + Cov\Big(E[I(X_{ij} \leq x)|\Theta_j], E[I(X_{rj} \leq x)|\Theta_j]\Big)$$

$$= \delta_{ir} \frac{1}{w_{ij}} s_{F_x}^2 + a_{F_x}.$$

The first part of (4) results from

$$Cov[I(X_{ij} \leq x), F_{n_{w_j}}(x)] = Cov\Big(I(X_{ij} \leq x), \sum_{r=1}^{n_j} \frac{w_{rj}}{w_{\cdot j}} I(X_{rj} \leq x)\Big)$$

$$= \sum_{r=1}^{n_j} \frac{w_{rj}}{w_{ij}} Cov[I(X_{ij} \leq x), I(X_{rj} \leq x)] = \sum_{r=1}^{n_j} \frac{w_{rj}}{w_{\cdot j}}\Big(\delta_{ir} \frac{1}{w_{ij}} s_{F_x}^2 + a_{F_x}\Big)$$

$$= \frac{1}{w_{\cdot j}} s_{F_x}^2 + \sum_{r \neq i}^{n_j} \delta_{ir} \frac{1}{w_{ij}} s_{F_x}^2 + a_{F_x}$$

$$= \frac{1}{w_{\cdot j}} s_{F_x}^2 + a_{F_x}.$$

Similarly, we can prove the second and third parts of (4). For the proof of the first part of relation (5), we have

$$Cov[I(X_{ij} \leq x), F_{n_{wz}}(x)] = Cov\Big(I(X_{ij} \leq x), \sum_{j'=1}^{K} \frac{Z_{j'}^{F_x}}{z_{\cdot}^{F_x}} F_{n_{wj'}}(x)\Big)$$

$$= \sum_{j'=1}^{K} \sum_{r=1}^{n_{j'}} \frac{Z_{j'}^{F_x}}{z_{\cdot}^{F_x}} Cov[I(X_{ij} \leq x), I(X_{rj'} \leq x)]$$

$$= \sum_{r=1}^{n_j} \frac{Z_j^{F_x}}{z_{\cdot}^{F_x}} Cov[I(X_{ij} \leq x), I(X_{rj} \leq x)] + \sum_{j' \neq j}^{K} \sum_{r=1}^{n_{j'}} \frac{Z_{j'}^{F_x}}{z_{\cdot}^{F_x}} Cov[I(X_{ij} \leq x), I(X_{rj'} \leq x)]$$

$$= \frac{Z_j^{F_x}}{z_{\cdot}^{F_x}} Cov[I(X_{ij} \leq x), I(X_{ij} \leq x)] + \sum_{r \neq i}^{n_j} \frac{Z_j^{F_x}}{z_{\cdot}^{F_x}} Cov[(I(X_{ij} \leq x), I(X_{rj} \leq x)]$$

$$= \frac{a_{F_x}}{z_{\cdot}^{F_x}}.$$

In the same way, we can prove the second and third parts of (5). Finally, (6) can be proved as

$$Cov[F_{X_{ij}|\Theta_j}(x|\Theta_j), I(X_{ij'} \leq x)]$$

$$= E\{Cov[F_{X_{ij}}(x|\Theta_j), I(X_{ij'} \leq x)|\Theta_j]\}$$

$$+Cov\{E[F_{X_{ij}|\Theta_j}(x|\Theta_j)|\Theta_j], E[I(X_{ij'} \leq x)|\Theta_j]\}$$

$$= \delta_{jj'}a_{F_x}.$$

□

Similarly, as in Bühlmann and Straub (1970), by the following theorems, we will provide the optimal linearized non-homogeneous (as well as the homogeneous) credibility estimators and provide some useful estimators for the structure parameters.

Theorem 1 (**Linearized non-homogeneous credibility distribution estimator**)**.** *Under the assumptions (i)–(iii), the optimal linearized non-homogeneous estimator of $F_{X_{ij}}(x|\Theta_j)$ is obtained by*

$$F_{X_{ij}}^{Cred}(x|\Theta_j) = Z_j^{F_x} F_{n_{wj}}(x) + (1 - Z_j^{F_x})F_{X_{ij}}(x), \tag{7}$$

with $F_{n_{wj}}(x)$ and $Z_j^{F_x}$ as in (1).

Proof. We have to find $c_0^j, c_{11}^j, ..., c_{n_jK}^j$ in

$$g_j[I(X_{11} \leq x), \cdots, I(X_{n_jK} \leq x)] = c_0^j + \sum_{l=1}^{K} \sum_{i=1}^{n_j} c_{il}^j I(X_{ij} \leq x), \tag{8}$$

such that

$$Q = E\left(P(X_{ij} \leq x|\Theta_j) - c_0^j - \sum_{l=1}^{K} \sum_{i=1}^{n_j} c_{il}^j I(X_{ij} \leq x)\right)^2 \tag{9}$$

is minimum. Differentiating (9) with respect to c_0^j, we have

$$\frac{\partial Q}{\partial c_0^j} = E\left(F_{X_{ij}|\Theta_j}(x|\Theta_j) - c_0^j - \sum_{l=1}^{K} \sum_{i=1}^{n_j} c_{il}^j I(X_{ij} \leq x)\right) = 0$$

$$\Rightarrow c_0^j = E[F_{X_{ij}|\Theta_j}(x|\Theta_j)] - \sum_{l=1}^{K} \sum_{i=1}^{n_j} c_{il}^j E[I(X_{ij} \leq x)]$$

$$= F_{X_{ij}}(x) - \sum_{l=1}^{K} \sum_{i=1}^{n_j} c_{il}^j F_{X_{ij}}(x). \tag{10}$$

Substituting the value of c_0^j in (9) and differentiating with respect to $c_{rl'}$, we obtain

$$\frac{\partial Q}{\partial c_{rl'}} = \frac{\partial}{\partial c_{rl'}} E\left[F_{X_{ij}\Theta_j}(x|\Theta_j) - F_{X_{ij}}(x) + \sum_{l=1}^{K}\sum_{i=1}^{n_j} c_{il}^j F_{X_{ij}}(x) - \sum_{l=1}^{K}\sum_{i=1}^{n_j} c_{il}^j I(X_{ij} \le x)\right]^2$$

$$= E\left(F_{X_{ij}|\Theta_j}(x|\Theta_j) - F_{X_{ij}}(x) - \sum_{l=1}^{K}\sum_{i=1}^{n_j} c_{il}^j [I(X_{il} \le x) - F_{X_{ij}}(x)][I(X_{rl'} \le x) - F_{X_{ij}}(x)]\right)$$

$$= 0$$

$$\Rightarrow Cov[F_{X_{ij}|\Theta_j}(x|\Theta_j), I(X_{rl'} \le x)] = \sum_{l=1}^{K}\sum_{i=1}^{n_j} c_{il}^j Cov[I(X_{il} \le x), I(X_{rl'} \le x)]. \tag{11}$$

The right-hand side of (11) becomes

$$\sum_{l=1}^{K}\sum_{i=1}^{n_j} c_{il}^j Cov[I(X_{il} \le x), I(X_{rl'} \le x)]$$

$$= \sum_{i=1}^{n_j} c_{ij}^j Cov[I(X_{ij} \le x), I(X_{rj} \le x)] + \sum_{l \ne j}^{K}\sum_{i=1}^{n_j} Cov[I(X_{il} \le x), I(X_{rj} \le x)]$$

$$= \sum_{i=1}^{n_j} c_{ij}^j\left(\delta_{ir}\frac{1}{w_{rj}}s_{F_x}^2 + a_{F_x}\right).$$

Then, (11) implies that

$$a_{F_x} = \sum_{i=1}^{n_j} c_{ij}^j a_{F_x} + c_{ij}^j \frac{1}{w_{ij}}s_{F_x}^2. \tag{12}$$

Multiplying (12) by w_{ij} and summing with respect to $i = 1, \ldots, n_j$, (i.e., $\sum_{i=1}^{n_j} c_{ij}^j = c_{.j}^j$), we obtain

$$c_{.j}^j = \frac{a_{F_x}w_{.j}}{a_{F_x}w_{.j} + s_{F_x}^2} = Z_j^{F_x} \Rightarrow c_{ij}^j = \frac{w_{ij}}{w_{.j}}Z_j^{F_x}.$$

Since the probability distribution of $I(X_{11} \le x), \cdots, I(X_{n_jK} \le x)$ is invariant under permutations of $I(X_{ij} \le x)$ and $F_{X_{ij}|\Theta_j}(x|\Theta_j)$ is uniquely defined, it must hold that $c_{1j}^j = c_{2j}^j = \cdots = c_{n_jj}^j$. Then, (8) becomes

$$F_{X_{ij}}^{Cred}(x|\Theta_j) = F_{X_{ij}}(x) - \sum_{l=1}^{K}\sum_{i=1}^{n_j} c_{il}^j F_{X_{ij}}(x) + \sum_{l=1}^{K}\sum_{i=1}^{n_j} c_{il}^j I(X_{il} \le x)$$

$$= F_{X_{ij}}(x) - \sum_{i=1}^{n_j} c_{ij}^j F_{X_{ij}}(x) - \sum_{l \ne j}^{K}\sum_{i=1}^{n_j} c_{il}^j F_{X_{ij}}(x)$$

$$+ \sum_{i=1}^{n_j} c_{il}^j I(X_{ij} \le x) + \sum_{l \ne l}^{K}\sum_{i=1}^{n_j} c_{il}^j I(X_{il} \le x)$$

$$= Z_j^{F_x} \sum_{i=1}^{n_j} \frac{w_{ij}}{w_{.j}} I(X_{ij} \le x) + (1 - Z_j^{F_x})F_{X_{ij}}(x),$$

which provides (7). $\quad\square$

Theorem 2 (Linearized homogeneous credibility distribution estimator). *Under the assumptions (i)–(iii), the optimal linearized homogeneous estimator of $F_{X_{ij}}(x|\Theta_j)$ is obtained by*

$$F^{Cred}_{X_{ij}|\Theta_j}(x|\Theta_j) = Z_j^{F_x} F_{n_{wj}}(x) + (1 - Z_j^{F_x}) F_{n_{wz}}(x), \tag{13}$$

with $F_{n_{wj}}(x)$, $F_{n_{wz}}(x)$ and $Z_j^{F_x}$ as defined in (1).

Proof. Letting

$$g_j[I(X_{11} \leq x), \cdots, I(X_{n_lK} \leq x)] = \sum_{l=1}^K \sum_{i=1}^{n_l} c_{il}^j I(X_{il} \leq x), \tag{14}$$

we have to minimize

$$Q = E\left(F_{X_{ij}|\Theta_j}(x|\Theta_j) - \sum_{l=1}^K \sum_{i=1}^{n_l} c_{il}^j I(X_{il} \leq x) \right)^2, \tag{15}$$

such that

$$E[F_{X_{il}}(x|\Theta_j)] = \sum_{l=1}^K \sum_{i=1}^{n_l} c_{il}^j E[I(X_{il} \leq x)] \tag{16}$$

holds under the restrictions $\sum_{l=1}^K \sum_{i=1}^{n_l} c_{il}^j = 1$, with the Lagrange multiplier 2λ. The following quantity leads to

$$Q = E\left(F_{X_{ij}|\Theta_j}(x|\Theta_j) - F_{X_{ij}}(x) - \sum_{l=1}^K \sum_{i=1}^{n_l} c_{il}^j [I(X_{ij} \leq x) - F_{X_{ij}}(x)] \right)^2$$
$$- 2\lambda\left(\sum_{l=1}^K \sum_{i=1}^{n_l} c_{il}^j F_{X_{ij}}(x) - F_{X_{ij}}(x) \right). \tag{17}$$

From (16), we obtain

$$F_{X_{ij}}(x) = E[F_{X_{ij}|\Theta_j}(x|\Theta_j)] = \sum_{l=1}^K \sum_{i=1}^{n_l} c_{il}^j E[I(X_{il} \leq x)] = F_{X_{ij}}(x) \sum_{l=1}^K \sum_{i=1}^{n_l} c_{il}^j$$
$$\Rightarrow \sum_{l=1}^K \sum_{i=1}^{n_l} c_{il}^j = 1. \tag{18}$$

Differentiating (17) with respect to $c_{i'l'}$, we obtain

$$Cov[F_{X_{ij}|\Theta_j}(x|\Theta_j), I(X_{i'l'} \leq x)] - \sum_{l=1}^K \sum_{i=1}^{n_l} c_{il}^j Cov[I(X_{il} \leq x), I(X_{i'l'} \leq x)] - \lambda F_{X_{ij}}(x) = 0$$

$$\Rightarrow \delta_{jl'} a_{F_x} + \lambda F_{X_{ij}}(x) = \sum_{i=1}^{n_l} c_{il'}^j Cov[I(X_{il'} \leq x), I(X_{i'l'} \leq x)]$$

$$= \sum_{i=1}^{n_l} c_{il'}^j \left(a_{F_x} + \delta_{ii'} \frac{1}{w_{i'l'}} s_{F_x}^2 \right)$$

$$= c_{\cdot l'}^j a_{F_x} + c_{i'l'}^j \frac{1}{w_{i'l'}} s_{F_x}^2. \tag{19}$$

Multiplying both sides by $w_{i'l'}$ and taking the sum over i', we obtain for each l':

$$c^j_{.l'} = \frac{[\delta_{jl'} a_{F_x} + \lambda F_{X_{ij}}(x)]w_{.l'}}{\left(a_{F_x} w_{.l'} + s^2_{F_x}\right]} = [\delta_{jl'} + \frac{\lambda}{a_{F_x}} F_{X_{ij}}(x)]Z^{F_x}_{l'}. \tag{20}$$

Substituting (20) into (19), we obtain

$$c^j_{i'l'} = \frac{[\delta_{jl'} a_{F_x} + \lambda F_{X_{ij}}(x)](1 - Z^{F_x}_{l'})}{s^2_{F_x}} w_{i'l'} \tag{21}$$

$$\Rightarrow 1 = \sum_{l=1}^{K} \sum_{i=1}^{n_j} c^j_{il} = \sum_{l=1}^{K} c^j_{.l} = \sum_{l=1}^{K} \left(\delta_{jl} + \frac{\lambda}{a_{F_x}} F_{X_{ij}}(x)\right) Z^{F_x}_l$$

$$= \frac{\lambda}{a_{F_x}} F_{X_{ij}}(x) \sum_{l=1}^{K} Z^{F_x}_l + \sum_{l \neq j}^{K} Z^{F_x}_l \delta_{jl} + \delta_{jj} Z^{F_x}_j$$

$$\Rightarrow \lambda F_{X_{ij}}(x) = a_{F_x} \frac{(1 - Z^{F_x}_j)}{z^{F_x}_{.}}. \tag{22}$$

Then, the optimal linearized homogeneous estimator of $F_{X_{ij}}(x|\Theta_j)$ becomes

$$F^{Cred}_{X_{ij}}(x|\Theta_j) = \sum_{l=1}^{K} \sum_{i=1}^{n_j} c^j_{il} I(X_{il} \leq x)$$

$$= \sum_{l=1}^{K} \sum_{i=1}^{n_j} \frac{[\delta_{jl'} a_{F_x} + \lambda F_{X_{ij}}(x)](1 - Z^{F_x}_l)}{s^2_{F_x}} w_{il} I(X_{il} \leq x)$$

$$= \sum_{l=1}^{K} \frac{Z^{F_x}_j}{z^{F_x}_{.}} \sum_{i=1}^{n_j} \frac{w_{il}}{w_{.l}} (1 - Z^{F_x}_j) I(X_{il} \leq x) + \sum_{l=1}^{K} \sum_{i=1}^{n} Z^{F_x}_l \frac{w_{il}}{w_{.l}} \delta_{jl} I(X_{il} \leq x)$$

$$= (1 - Z^{F_x}_j) \sum_{l=1}^{K} \frac{Z^{F_x}_j}{z^{F_x}_{.}} F_{n_{wj}}(x) + \sum_{l \neq j}^{K} \sum_{i=1}^{n_j} Z^{F_x}_l \frac{w_{il}}{w_{.l}} \delta_{jl} I(X_{il} \leq x)$$

$$+ Z^{F_x}_j \sum_{i=1}^{n_j} \frac{w_{ij}}{w_{.j}} I(X_{ij} \leq x)$$

resulting in (13). □

The following theorem will prove that $F_{n_{wz}}(x)$ has a smaller variance than $F_{n_{wj}}(x)$, i.e., based on the heterogeneity and the fluctuation of the risk, $F_{n_{wz}}(x)$ has a minimal mean square error.

Theorem 3. *The* $Var\left(\sum_{j=1}^{K} \sum_{i=1}^{n_j} c_{ij} I(X_{ij} \leq x)\right)$ *is the minimum for all c_{ij}, such that $\sum_{j=1}^{K} \sum_{i=1}^{n_j} c_{ij} = 1$, for $c_{ij} = \frac{w_{ij}}{w_{.j}} \frac{Z^{F_x}_j}{z^{F_x}_{.}}$.*

Proof. We have to minimize the following quantity

$$
Q = E\left(\sum_{j=1}^{K}\sum_{i=1}^{n_j} c_{ij} I(X_{ij} \leq x) - E[\sum_{j=1}^{K}\sum_{i=1}^{n_j} c_{ij}^j I(X_{ij} \leq x)]\right)^2
$$

$$
-2\lambda F_{X_{ij}}(x)\left(\sum_{j=1}^{K}\sum_{i=1}^{n_j} c_{ij}^j - 1\right). \tag{23}
$$

Taking the derivative of (23) with respect $c_{i'l'}$ for $i' = 1, \cdots, n$ and $j' = 1, \cdots, K$, we obtain

$$
\frac{\partial Q}{\partial c_{i'j'}} = 2E\left(\sum_{j=1}^{K}\sum_{i=1}^{n_j} c_{ij} I(X_{ij} \leq x) - E[\sum_{j=1}^{K}\sum_{i=1}^{n_j} c_{ij} I(X_{ij} \leq x)]\right)\left(I(X_{i'j'} \leq x) - F_{X_{ij}}(x)\right)
$$

$$
-2\lambda F_{X_{ij}}(x) = 0
$$

$$
\Rightarrow E\left(\sum_{j=1}^{K}\sum_{i=1}^{n_j} c_{ij}[I(X_{ij} \leq x) - F_{X_{ij}}(x)][I(X_{i'j'} \leq x) - F_{X_{ij}}(x)]\right) = \lambda F_{X_{ij}}(x)
$$

$$
\Rightarrow \sum_{i=1}^{n_j} c_{ij} Cov[I(X_{ij} \leq x), I(X_{i'j'} \leq x)] = \lambda F_{X_{ij}}(x). \tag{24}
$$

This is the same as

$$
\sum_{\substack{j \neq j' \\ }}^{K}\sum_{i=1}^{n_j} c_{ij} Cov[I(X_{ij} \leq x), I(X_{i'j'} \leq x)] +
$$

$$
+ \sum_{i=1}^{n_j} c_{ij} Cov[I(X_{ij} \leq x), I(X_{i'j'} \leq x)] = \lambda F_{X_{ij}}(x)
$$

$$
\Rightarrow \sum_{i=1}^{n_j} c_{ij}\left(a_{F_x} + \delta_{ii'}\frac{s_{F_x}^2}{w_{ij}}\right) = \lambda F_{X_{ij}}(x). \tag{25}
$$

This gives

$$
\Rightarrow a_{F_x}\sum_{i=1}^{n_j} c_{ij} + \sum_{i \neq i'}^{n} c_{ij}\delta_{ii'}\frac{s_{F_x}^2}{w_{ij}} + c_{ij}\delta_{ii}\frac{s_{F_x}^2}{w_{ij}} = \lambda F_{X_{ij}}(x)
$$

$$
\Rightarrow a_{F_x}\sum_{i=1}^{n_j} c_{ij} + c_{ij}\frac{s_{F_x}^2}{w_{ij}} = \lambda F_{X_{ij}}(x)
$$

$$
\Rightarrow a_{F_x} w_{ij} c_{.j} + c_{ij} s_{F_x}^2 = \lambda F_{X_{ij}}(x) w_{ij}
$$

$$
\Rightarrow c_{ij} = \frac{w_{ij}[\lambda F_{X_{ij}}(x) - a_{F_x} c_{.j}]}{s_{F_x}^2} \tag{26}
$$

$$\Rightarrow \sum_{i=1}^{n_j} a_{F_x}\, w_{ij} c_{\cdot j} + \sum_{i=1}^{n_j} c_{ij} s_{F_x}^2 = \sum_{i=1}^{n_j} \lambda F_{X_{ij}}(x) w_{ij}$$

$$\Rightarrow a_{F_x} w_{\cdot j} c_{\cdot j} + c_{\cdot j} s_{F_x}^2] = \lambda F_{X_{ij}}(x) w_{\cdot j}$$

$$\Rightarrow c_{\cdot j} = \frac{w_{\cdot j} \lambda F_{X_{ij}}(x)}{a_{F_x}\, w_{\cdot j} + s_{F_x}^2} = \lambda F_{X_{ij}}(x) \frac{Z_j^{F_x}}{a_{F_x}}. \tag{27}$$

We know that

$$\sum_{j=1}^{K} \sum_{i=1}^{n_j} c_{ij} = 1 \;\; \Rightarrow \sum_{j=1}^{K} c_{\cdot j} = 1$$

$$\Rightarrow \sum_{j=1}^{K} \lambda F_{X_{ij}}(x) \frac{Z_j^{F_x}}{a_{F_x}} = \lambda F_{X_{ij}}(x) \frac{z_{\cdot}^{F_x}}{a_{F_x}} = 1 \;\; \Rightarrow \lambda = \frac{a_{F_x}}{F_{X_{ij}}(x)\, z_{\cdot}^{F_x}}.$$

We therefore obtain

$$c_{\cdot j} = \frac{a_{F_x} F_{X_{ij}}(x)\, Z_j^{F_x}}{F_{X_{ij}}(x)\, z_{\cdot}^{F_x}} = \frac{Z_j^{F_x}}{z_{\cdot}^{F_x}} \tag{28}$$

and from (26), we have

$$c_{ij} = \frac{w_{ij}(\lambda F_{X_{ij}}(x) - a_{F_x} c_{\cdot j})}{s_{F_x}^2} = \frac{w_{ij}\left(\frac{a_{F_x}}{z_{\cdot}^{F_x}} - \frac{a_{F_x} Z_j^{F_x}}{z_{\cdot}^{F_x}}\right)}{s_{F_x}^2} = \frac{w_{ij}}{w_{\cdot j}} \frac{Z_j^{F_x}}{z_{\cdot}^{F_x}}.$$

$\square$

Theorem 4. *Under assumptions (i)–(iii), the quadratic loss for the credibility distribution estimator is given by*

$$E[F_{X_{ij}}^{Cred}(x|\Theta_j) - F_{X_{ij}}(x|\Theta_j)]^2 = a_{F_x}(1 - Z_j^{F_x}). \tag{29}$$

Proof. We have

$$E[F_{X_{ij}}^{Cred}(x|\Theta_j) - F_{X_{ij}}(x|\Theta_j)]^2$$

$$= E\left(Z_j^{F_x}[F_{wj}(x) - F_{X_{ij}}(x)] - [F_{X_{ij}}(x|\Theta_j) - F_{X_{ij}}(x)] \right)^2$$

$$= (Z_j^{F_x})^2 \left(Var[F_{wj}(x)] + Var[F_{X_{ij}}(x|\Theta_j)] - 2 Z_j^{F_x} Cov[F_{wj}(x), F_{X_{ij}}(x|\Theta_j)] \right)$$

$$= Z_j^{F_x} \left(\frac{a_{F_x} w_{\cdot j}}{a_{F_x} w_{\cdot j} + s_{F_x}^2} \right) \left(\frac{s_{F_x}^2}{w_{\cdot j}} + a_{F_x} \right) + a_{F_x} - 2 Z_j^{F_x} a_{F_x}$$

that provides (29). $\square$

2.4. Optimal Projection Theorem

In the following, De Vylder's (1976, 1996) optimal projection theorem of random variables in the plane is applied in order to derive the optimal estimator of $F_{X_{ij}}(x)$ and $F_{X_{ij}}(x|\Theta_j)$. Practically, $F_{X_{ij}}(x)$ is replaced by $F_{n_{wz}}(x)$ in (7).

Theorem 5. *The optimal estimator of $F_{X_{ij}}(x)$ in the plane $H(I(X_{ij} \leq x), i = 1, \cdots, n_j, j = 1, \cdots, K)$ is*

$$
\begin{aligned}
F_{X_{ij}}(x)_{Proj} &= Proj[F_{X_{ij}}(x)|H_F(I(X_{ij} \leq x), i = 1, \cdots, n_j, j = 1, \cdots, K)] \\
&= F_{n_{wz}}(x).
\end{aligned}
$$

Proof. Directly from (2) and (5). $\square$

Theorem 6. *The optimal credibility estimator of $F_{X_{ij}}(x|\Theta_j)$ based on $I(X_{11} \leq x), \cdots, I(X_{n_jK} \leq x)$ is*

$$
\begin{aligned}
F^{Cred}_{X_{ij}|\Theta_j}(x|\Theta_j) &= Proj[F_{X_{ij}|\Theta_j}(x|\Theta_j)|H_F(I(X_{ij} \leq x), i = 1, \cdots, n_j, j = 1, \cdots, K)] \\
&= Z_j^{F_x} F_{n_{wj}}(x) + (1 - Z_j^{F_x}) F_{n_{wz}}(x).
\end{aligned} \tag{30}
$$

Proof. In order to prove (30), it is sufficient to prove the unbiasedness and covariance conditions of the optimal projection theorem of random variables on planes not through the origin (see De Vylder (1976, 1996)), that is

$$
E[F^{Cred}_{X_{ij}|\Theta_j}(x|\Theta_j)] = E[F_{X_{ij}|\Theta_j}(x|\Theta_j)] = F_{X_{ij}}(x)
$$

and

$$
Cov[F^{Cred}_{X_{ij}|\Theta_j}(x|\Theta_j) - F_{X_{ij}|\Theta_j}(x|\Theta_j), I(X_{ij'} \leq x)] = const.
$$

The unbiasedness condition results from (2) and

$$
E[F^{Cred}_{X_{ij}|\Theta_j}(x|\Theta_j)] = E\left(Z_j^{F_x} F_{n_{wj}}(x) + (1 - Z_j^{F_x}) F_{n_{wz}}(x) \right) = F_{X_{ij}}(x).
$$

The covariance condition results from the independence of the contracts and the covariance relations of Lemma 1, which gives

$$
\begin{aligned}
&Cov[F^{Cred}_{X_{ij}|\Theta_j}(x|\Theta_j) - F_{X_{ij}|\Theta_j}(x|\Theta_j), I(X_{ij'} \leq x)] \\
&= Cov\left(Z_j^{F_x} F_{n_{wj}}(x) + (1 - Z_j^{F_x}) F_{n_{wz}}(x) - F_{X_{ij}|\Theta_j}(x|\Theta_j), I(X_{ij'} \leq x) \right) \\
&= Z_j^{F_x} \delta_{jj'} Cov[F_{n_{wj}}(x), I(X_{ij'} \leq x)] + (1 - Z_j^{F_x}) Cov[F_{n_{wz}}(x), I(X_{rj'} \leq x)] \\
&\quad - \delta_{jj'} Cov[F_{X_{ij}|\Theta_j}(x|\Theta_j), I(X_{ij'} \leq x)] \\
&= Z_j^{F_x} \delta_{jj'} \frac{a_{F_x}}{Z_j^{F_x}} + (1 - Z_j^{F_x}) \frac{a_{F_x}}{z^{F_x}} - \delta_{jj'} a_{F_x} = (1 - Z_j^{F_x}) \frac{a_{F_x}}{z^{F_x}}.
\end{aligned} \tag{31}
$$

$\square$

2.5. Unbiased Estimators

Below, we provide unbiased estimators analogous to the Bühlmann and Straub (1970) model.

Lemma 2. *The following estimators of the structural parameters $F_{X_{ij}}(x)$, $s^2_{F_x}$ and a_{F_x}, presented in Section 2.2, are unbiased.*

$$\widehat{F}_{X_{ij}}(x) = F_{n_{ww}(x)} \ \ or \ \ \widehat{F}_{X_{ij}}(x) = F_{n_{wz}(x)}, \tag{32}$$

$$\widehat{s}^2_{F_x} = \frac{\sum_{j=1}^{K}\sum_{i=1}^{n_j} w_{ij}[I(X_{ij}\le x) - F_{n_{wj}}(x)]^2}{\sum_{j=1}^{K}(n_j - 1)}, \tag{33}$$

$$\widehat{a}_{F_x} = \frac{w_{..}}{w_{..}^2 - \sum_{j=1}^{K} w_{.j}^2}\left[\sum_{j=1}^{K} w_{.j}[F_{n_{wj}}(x) - F_{n_{ww}}(x)]^2 - (K-1)\widehat{s}^2_{F_x}\right]. \tag{34}$$

Based on De Vylder *(1978), an unbiased estimator of a_{F_x} can take the form*

$$\widehat{a}_{F_x} = \frac{\sum_{j=1}^{K} Z_j^{F_x}[F_{n_{wj}}(x) - F_{n_{wz}}(x)]^2}{(K-1)}, \ \ (pseudo\text{-}estimator). \tag{35}$$

Proof. The unbiasedness of $\widehat{F}_{X_{ij}}(x)$ is straightforward and is omitted. The unbiasedness of $\widehat{s}^2_F$ follows from

$$\sum_{j=1}^{K}(n_j - 1)E(\widehat{s}^2_{F_x}) = E\left[\sum_{j=1}^{K}\sum_{i=1}^{n_j} w_{ij}[I(X_{ij}\le x) - F_{n_{wj}}(x)]^2\right]$$

$$= \sum_{j=1}^{K}\sum_{i=1}^{n_j}\left[w_{ij}Var[I(X_{ij}\le x)] + Var[F_{n_{wj}}(x)] - 2Cov[I(X_{ij}\le x), F_{n_{wj}}(x)]\right]$$

$$= \sum_{j=1}^{K}\sum_{i=1}^{n_j} w_{ij}\left(\frac{s^2_{F_x}}{w_{ij}} + a_{F_x} + \frac{a_{F_x}}{Z_j^{F_x}} - 2\frac{a_{F_x}}{Z_j^{F_x}}\right) = \sum_{j=1}^{K}(n_j - 1)s^2_{F_x}, \tag{36}$$

resulting in (33). For the proof of the unbiasedness of (34), we refer to Bühlmann and Straub (1970). Finally, the unbiasedness of $\widehat{a}_F$ in (35) results from

$$(K-1)E(\widehat{a}_{F_x}) = E\left[\sum_{j=1}^{K} Z_j^{F_x}[F_{n_{wj}}(x) - F_{n_{wz}}(x)]^2\right]$$

$$= \sum_{j=1}^{K} Z_j^{F_x}\left(Var[F_{n_{wj}}(x)] + Var[F_{n_{wz}}(x)] - 2Cov[F_{n_{wj}}(x), F_{n_{wz}}(x)]\right)$$

$$= \sum_{j=1}^{K} Z_j^{F_x}\left[\left(\frac{a_{F_x}}{Z_j^{F_x}}\right) + \left(\frac{a_{F_x}}{z_.^{F_x}}\right) - 2\left(\frac{a_{F_x}}{z_.^{F_x}}\right)\right] = (K-1)a_{F_x}$$

which implies (35). □

3. Credible Distribution for Grouped Data

Grouped data are formed by aggregating the individual observations of a variable into groups. For example, a histogram is a density approximation for grouped data. The construction of the empirical distribution based on grouped data can be achieved by obtaining the point values of the empirical distribution function whenever possible. Then, we can approximate the distribution function by connecting those point values with straight lines.

Empirical distribution for grouped data is evaluated at a point estimate x. We consider the case where the point estimate x is at a boundary and the case where the value of x is between the boundaries.

3.1. Empirical Distribution for Grouped Data at Boundary

For contract j, let the group boundaries be $c_{0j} < c_{1j} < \cdots < c_{nj}$, where $c_{0j} = 0$ and $c_{n+1,j} = \infty$. Let m_{ij} be the number of observations in the interval $(c_{i-1,j}, c_{ij})$, $i = 1, 2, \ldots, n_j$, $j = 1, 2, \ldots, K$ and $m_{.j} = \sum_{i=1}^{n_j} m_{ij}$ be the total number of observations for the j contract. For grouped data, the empirical distribution function at each group boundary c_{ij} is defined as

$$F_{mj}(c_{rj}) = \frac{1}{m_{.j}} \sum_{i=1}^{r} m_{ij}. \tag{37}$$

For grouped data, there is no problem if the distribution function has to be estimated at a boundary. When all of the information is available, working with the empirical estimate of the distribution function is straightforward (see Klugman et al. (2012)). We have the following assumptions:

3.1.1. Assumptions

(i*) The contracts are independent and the variables $\Theta_1, \cdots, \Theta_K$ are identically distributed. The observations X_{ij} have finite variance,

(ii*) $E[I(X_{ij} \leq x)|\Theta_j)] = E[F_{mj}(x|\Theta_j)] = F_{X_{ij}|\Theta_j}(x|\Theta_j)$,

(iii*) $Var[I(X_{ij} \leq x)|\Theta_j] = \frac{1}{m_{ij}}\sigma_x^2(\Theta_j)$ and $Var[F_{mj}(x|\Theta_j)] = \frac{1}{m_{.j}}\sigma_x^2(\Theta_j)$.

3.1.2. Structural Parameters

$$\mu_x = E[F_{mj}(x)] = F_{X_{ij}}(x), \ s_x^2 = E[\sigma_x^2(\Theta_j)], \ a_x = Var\{E[I(X_{ij} \leq x)|\Theta_j]\}. \tag{38}$$

3.1.3. Notation

Here, we adopt the following notation:

$$F_{mj}(x) = \sum_{j=1}^{n_j} \frac{m_{ij}}{m_{.j}} I(X_{ij} \leq x), \ F_{mm}(x) = \sum_{j=1}^{K} \frac{m_{.j}}{m_{..}} F_{mj}(x), \ F_{mz}(x) = \sum_{j=1}^{K} \frac{Z_j^x}{z_.^x} F_{mj}(x),$$

$$m_{.j} = \sum_{i=1}^{n_j} m_{ij}, m_{..} = \sum_{i=1}^{K} m_{.j}, \ z_.^x = \sum_{j=1}^{K} Z_j^x, \ Z_j^x = \frac{m_{.j} a_x}{m_{.j} a_x + s_x^2}. \tag{39}$$

Based on the above assumptions, a credibility distribution estimator for $F_{X_{ij}}(x|\Theta_j)$ is obtained as

$$F_{X_{ij}}^{Cred}(x|\Theta_j) = Z_j^x F_{mj}(x) + (1 - Z_j^x) F_{X_{ij}}(x). \tag{40}$$

With the following theorem, we can obtain the credibility distribution estimator of $F_{X_{ij}}(x|\Theta_j)$.

Theorem 7. *Under the assumptions $(i^*)-(iii^*)$, the credibility factor in (40) is given by*

$$Z_j^x = \frac{m_{.j} a_x}{m_{.j} a_x + s_x^2},$$

with a_x as in (38) and $m_{.j}$ as in (39).

Proof. The proof of the theorem can be obtained by minimizing the expression

$$Q = E\left(F_{X_{ij}}^{Cred}(x|\Theta_j) - Z_j^x F_{mj}(x) - (1 - Z_j^x) F_{X_{ij}}(x)\right)^2,$$

with respect to Z_j^x. $\quad \square$

3.1.4. Credibility Estimators

Lemma 3. *The credibility point estimators of $F_{X_{ij}}(x)$, s_x^2 and a_x are given as follows:*

$$\widehat{F}_{X_{ij}}(x) \;=\; F_{mm}(x), \text{or} \;\; \widehat{F}_{X_{ij}}(x) = F_{mz}(x)$$

$$\widehat{s}_x^2 \;=\; \frac{\sum_{j=1}^{K} \sum_{i=1}^{n_j} m_{ij}[I(X_{ij} \leq x) - F_{mj}(x)]^2}{\sum_{j=1}^{K}(n_j - 1)}$$

$$\widehat{a}_x \;=\; \left(\frac{m_{..}}{m_{..}^2 - \sum_{j=1}^{K} m_{.j}^2}\right)\left[\sum_{j=1}^{K} m_{.j}[F_{mj}(x) - F_{mm}(x)]^2 - (K-1)\widehat{s}_x^2\right]$$

or

$$\widehat{a}_x \;=\; \frac{\sum_{j=1}^{K} Z_j^x[F_{mj}(x) - F_{mz}(x)]^2}{(K-1)}.$$

Proof. Similarly to the proof of Lemma 2. $\quad \square$

3.2. Empirical Distribution for Grouped Data at Value x between Boundaries

Now, suppose that the value of x is between the boundaries $c_{i-1,j}$ and c_{ij}. Then, for contract j, the empirical distribution function is given by

$$F_{mj}(x) = \begin{cases} 0, & x \leq c_0, \\ \frac{(c_{ij}-x)F_{mj}(c_{i-1,j}) + (x-c_{i-1,j})F_{mj}(c_{ij})}{c_{ij}-c_{i-1,j}}, & c_{i-1,j} \leq x \leq c_{ij}, \\ 1, & x > c_n. \end{cases} \tag{41}$$

This function is differentiable at all values except for the group boundaries. Based on (41), we can obtain the following

$$E[F_{mj}(x|\Theta_j)] = \frac{(c_{ij}-x)F_{X_{ij}}(c_{i-1,j}|\Theta_j) + (x-c_{i-1,j})F_{X_{ij}}(c_{ij}|\Theta_j)}{c_{ij}-c_{i-1,j}}$$

and

$$F_{X_{ij}}(x) = E[F_{mj}(x)] = \frac{(c_{ij}-x)F_{X_{ij}}(c_{i-1,j}) + (x-c_{i-1,j})F_{X_{ij}}(c_{ij})}{c_{ij}-c_{i-1,j}}.$$

Note that the above estimator is biased although it is an unbiased estimator of the true interpolated value (see Klugman et al. (2012)).

The conditional variance of the empirical distribution is

$$Var[F_{mj}(x|\Theta_j)] \;=\; \frac{(c_{ij}-x)^2 \, Var[F_{m_j}(c_{i-1,j}|\Theta_j)] + (x-c_{i-1,j})^2 \, Var[F_{mj}(c_{ij}|\Theta_j)]}{(c_{ij}-c_{i-1,j})^2}$$

$$+\frac{2Cov[F_{mj}(c_{i-1,j}|\Theta_j), F_{mj}(c_{ij}|\Theta_j)]}{(c_{ij}-c_{i-1,j})^2},$$

where

$$Var[F_{mj}(c_{i-1,j}|\Theta_j)] = \frac{1}{m_{.j}}F_{X_{ij}}(c_{i-1,j}|\Theta_j)[1 - F_{X_{ij}}(c_{i-1,j}|\Theta_j)],$$

$$Var[F_{mj}(c_{ij}|\Theta_j)] = \frac{1}{m_{.j}} F_{X_{ij}}(c_{ij}|\Theta_j)[1 - F_{X_{ij}}(c_{ij}|\Theta_j)]$$

and

$$Cov[F_{mj}(c_{i-1,j}|\Theta_j), F_{mj}(c_{ij}|\Theta_j)] = \frac{1}{m_{.j}} \left(F_{X_{ij}}(\min\{c_{i-1,j}, c_{ij}\}|\Theta_j) \right.$$
$$\left. - F_{X_{ij}}(c_{i-1,j}|\Theta_j) F_{X_{ij}}(c_{ij}|\Theta_j) \right).$$

Then, we can proceed as in Section 3.1 for obtaining the credibility distribution estimator of $F_{X_{ij}}(x|\Theta_j)$, when the value of x is between boundaries.

4. Alternative Credibility Distribution Approach for Grouped Data

For grouped data, the previous approaches yield credibility point estimates. If we want to find the credibility estimation in the framework of Bühlmann and Straub (1970), we may apply the concept of uniform distribution within each interval $(c_{i-1,j}, c_{ij})$ and the first two moments can be estimated from

$$\widehat{\mu}_j^{(k)} = \sum_{i=1}^{r} \left(\frac{m_{ij}}{m_{.j}} \right) \left(\frac{c_{ij}^{k+1} - c_{i-1,j}^{k+1}}{(k+1)(c_{ij} - c_{i-1,j})} \right),$$

for $k = 1, 2$. Thus, for contract j, the empirical estimate of the mean ($k = 1$) is the weighted average of the interval midpoints where the weight m_{ij} for an interval is the proportion of the observations that are in the interval (histogram), i.e.,

$$\widehat{\mu}_j = \sum_{i=1}^{r} \frac{m_{ij}}{m_{.j}} \left(\frac{c_{ij} + c_{i-1,j}}{2} \right).$$

Letting $C_{ij} = \frac{c_{ij} + c_{i-1,j}}{2}$ and assuming that $E(C_{ij}|\Theta_j) = \mu(\Theta_j)$ and $\text{Cov}(C_{rj}, C_{ij}|\Theta_j) = \delta_{ri} \frac{1}{m_{ij}} \sigma^2(\Theta_j)$, the credibility estimation based on grouped data can be obtained similarly as in the Bühlmann and Straub (1970) model

$$\mu^{Cred}(\Theta_j) = Z_j \mu_j + (1 - Z_j)\mu, \tag{42}$$

with parameters

$$\mu = E[\mu(\Theta_j)], \ s^2 = E[\sigma^2(\Theta_j)], \ a = Var[\mu(\Theta_j)], \ Z_j = \frac{am_{.j}}{am_{.j} + s^2}. \tag{43}$$

Theorem 8. *The following are unbiased estimators for μ, s^2, and a:*

$$\widehat{\mu} = \bar{C}_{..} = \sum_{j=1}^{K} \frac{m_{.j}}{m_{..}} \bar{C}_{.j}, \ \text{with} \ \widehat{\mu}_j = \bar{C}_{.j} = \sum_{j=1}^{K} \frac{m_{ij}}{m_{.j}} C_{ij},$$

$$\widehat{s}^2 = \frac{1}{K} \sum_{j=1}^{K} \widehat{s}_j^2, \ \text{with} \ \widehat{s}_j^2 = \widehat{\mu}_j^{(2)} - (\widehat{\mu}_j)^2$$

and

$$\widehat{a} = \left(\frac{m_{..}}{m_{..}^2 - \sum_{j=1}^{K} m_{.j}^2} \right) \left[\sum_{j=1}^{K} m_{.j}(\widehat{\mu}_j - \widehat{\mu})^2 - (K-1)\widehat{s}^2 \right],$$

or

$$\widehat{a} = \frac{1}{K-1} \sum_{j=1}^{K} Z_j (\bar{C}_{.j} - \bar{C}_z)^2, \; where \; \bar{C}_z = \sum_{j=1}^{K} \frac{Z_j}{z_.} \bar{C}_{.j}.$$

Proof. See Bühlmann and Straub (1970) and De Vylder (1978). □

5. Numerical Illustrations

In this section, we use two datasets, one with insurance motor claims data and a second with monthly returns financial data.

5.1. Numerical Example with Insurance Data

The dataset is provided by Insurance Europe (2022) and includes a database with figures on the European insurance industry during the period 2004–2020 for 32 EU countries. Our numerical illustration is based on a complete dataset of 10 selected countries for the years 2004–2018. Our dataset also contains the motor claims paid and the number of motor claims for each country and each year. The selected countries are the following: Austria (AT); Germany (DE); Finland (FI); Greece (GR); Hungary (HR); Italy (IT); Norway (NO); Poland (PL); Portugal (PT); and Sweden (SE). Table 1 shows the summary statistics of the motor claim amounts and the claim numbers for countries $j = 1, \cdots, 10$ and years $i = 1, \cdots, 15$.

Table 1. Summary statistics for 10 selected European countries.

Motor Claims and Number of Motor Claims for the Years 2004–2018										
X_{ij}: Motor claims amount in millions for the years $i = 1, \cdots, 15$ and countries $j = 1, \cdots, 10$										
Country	AT	DE	FI	GR	HR	IT	NO	PL	PT	SE
Min.	1923	18,789	692	438	207	12,791	1062	1582	987	311
1st Qu.	1989	19,322	803	530	226	12,968	1205	1935	1187	912
Median	2032	20,222	914	978	248	15,239	1396	2219	1261	987
Mean	2104	20,692	909	840	256	15,112	1359	2322	1218	1098
3rd Qu.	2201	21,828	1024	1083	288	16,492	1514	2620	1282	1504
Max.	2430	23,897	1141	1224	330	18,210	1637	3235	1362	1670
w_{ij}: Weights–number of motor claims										
Min.	1,177,269	8,673,000	368,898	403,604	170,205	3,389,677	562,981	1,307,003	643,713	890,304
1st Qu.	1,238,392	9,002,000	420,122	427,265	199,179	3,467,180	659,063	1,465,112	712,819	1,043,110
Median	1,279,586	9,247,000	497,201	474,875	204,421	4,541,671	758,814	1,749,483	837,694	1,098,411
Mean	1,282,093	9,220,067	493,392	475,977	208,084	4,317,710	728,372	2,177,715	800,425	1,111,064
3rd Qu.	1,323,949	9,425,500	573,942	515,984	218,964	5,026,480	783,394	1,891,870	876,323	1,168,650
Max.	1,396,250	9,750,000	641,513	580,985	238,904	5,249,558	908,663	4,515,087	959,781	1,328,331

Note: X_{ij} are the average claims per year and w_{ij} represents the number of motor claims that correspond to each X_{ij}.

Table 2 illustrates the results of a credibility distribution function for motor claims amount data during the years 2004–2018. More analytically, the upper part of the table shows the individual empirical distribution $\widehat{F}_{n_{wj}}(x)$ of claim amounts $X_{ij} \leq x, X_{ij} \leq x$, ($x = 320, 800, 1000, 2000, 3000, 23,800, 23,896, 23,897$) and the corresponding credibility distribution estimators $\widehat{F}_{X_{ij}}^{Cred}(x|\Theta_j)$ are shown in the middle part of the table. The estimated credibility factors $\widehat{Z}_j^{F_x}$, as well the estimated parameters $\widehat{F}_{n_{ww}}(x), \widehat{s}_F^2, \widehat{a}_F$, are presented in the lower part of Table 2. Note that $\widehat{F}_{n_{wj}}(x) = 0$ means that the value of all claims $X_{ij} > x$ and $\widehat{F}_{n_{wj}}(x) = 1$ if claims $X_{ij} \leq x$.

In Table 2, we observe a lack of monotonicity of the estimated credibility distributions for all contracts. In order to obtain monotonicity, we similarly proceed as in Cai et al. (2015) by restricting the credibility factor Z_j^F to be a constant free of x. The results are shown in Table 3. Although monotonicity has been restored from a risk management perspective (which serves to fair pricing and reserves given the nature of the risk), more investigation is required, especially in the points where monotonicity breaks down.

Remark 1. *Another way of obtaining monotonicity of the credibility estimated by distribution functions is by sorting the resulting credibility by estimated distribution functions. In the relevant literature, there are methods for extracting a monotone function from non-monotonic data. Such a method is the monotonic regression that achieves the monotonicity and smoothness of the regression by introducing a regularization term, and solving an optimization problem with constraints. Some key references are: Friedman and Tibshirani (1984), Mukerjee (1988), Shively et al. (2009) and Zhang (2004). Similarly, the above approaches could be applied to our model.*

By letting the values of motor claims be larger than $x = 23{,}800$ and less than or equal to $x = 23{,}897$ ($x = 23{,}897$ is the maximum threshold of contract DE, which is the contract with the largest values of motor claims, as shown in Table 1), whilst the values of the estimated credibility distribution $\widehat{F}_{X_{ij}}^{Cred}(x|\Theta_j)$ remain the same up to the fifth decimal place. By letting $x > 23{,}897$, the estimated credibility distribution goes to 1 (see Table 2).

Remark 2. *Similarly to in Bühlmann and Straub (1970) model, $\widehat{a}_{F_x}$ can possibly be negative. This means that there is no detectable difference between the risks. In this case we put $\widehat{a}_{F_x} = 0$, as in our cases for $x = 23{,}800, 23{,}896, 23{,}897$.*

Figure 1 displays the individual empirical distribution in each contract. Note that the red bullets indicate the corresponding credibility estimate at specific points presented in Table 2.

Credibility Coefficients for Motor Claims Data

In the following, we provide an intuitive interpretation for the form of the credibility distribution estimator given in Theorem 1 for motor claims by presenting the following coefficients in Table 4, which were derived based on the lower part of Table 2. These are: the coefficient of variation $BRV = \frac{\sqrt{\widehat{a}_{F_x}}}{\widehat{F}_{n_{ww}}(x)}$, which is a good measure for the heterogeneity of the portfolio, (i.e., a good measure for the between-risk variability) and the average within the risk coefficient of variation $WRV = \frac{\widehat{s}_{F_x}}{\widehat{F}_{n_{ww}}(x)}$, which is a good measure of the within risk variability. The smaller the credibility coefficient $CC = \frac{\widehat{s}_{F_x}^2}{\widehat{a}_{F_x}}$, the greater the $\widehat{Z}_j^{F_x}$.
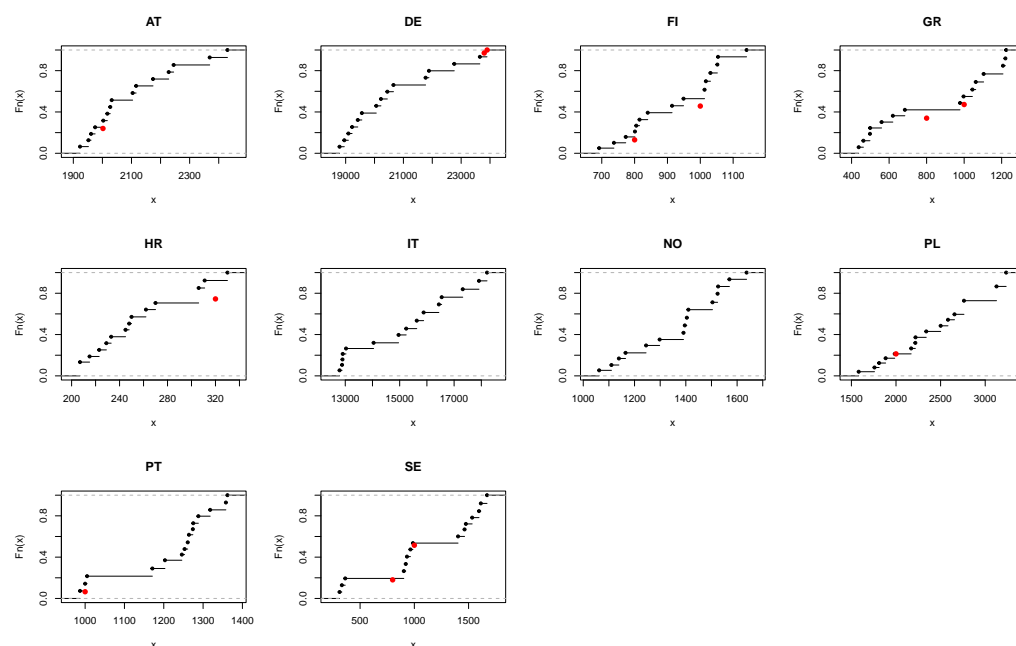

Figure 1. Individual empirical distribution and credibility distribution point estimates (in red) for motor claims per contract.

Table 2. Credibility distribution estimation for motor claims.

Motor Claim Amounts and Number of Motor Claims from 10 Selected European Countries during the Period 2004–2018									
Individual empirical distribution with claim amount $X_{ij} \leq x$, ($x = 320, 800, 1000, 2000, 3000, 23{,}800, 23{,}896, 23{,}897$)									
Country	AT	DE	FI	GR	HR	IT	NO	PL	PT	SE
$\widehat{F}_{n_{wj}}(320)$	0.0000	0.0000	0.0000	0.0000	0.9235	0.0000	0.0000	0.0000	0.0000	0.0624
$\widehat{F}_{n_{wj}}(800)$	0.0000	0.0000	0.1584	0.4206	1.0000	0.0000	0.0000	0.0000	0.0000	0.1940
$\widehat{F}_{n_{wj}}(1000)$	0.0000	0.0000	0.5286	0.5501	1.0000	0.0000	0.0000	0.0000	0.1428	0.5359
$\widehat{F}_{n_{wj}}(2000)$	0.2518	0.0000	1.0000	1.0000	1.0000	0.0000	1.0000	0.2128	1.0000	1.0000
$\widehat{F}_{n_{wj}}(3000)$	1.0000	0.0000	1.0000	1.0000	1.0000	0.0000	1.0000	0.7278	1.0000	1.0000
$\widehat{F}_{n_{wj}}(23{,}800)$	1.0000	0.9339	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\widehat{F}_{n_{wj}}(23{,}896)$	1.0000	0.9339	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\widehat{F}_{n_{wj}}(23{,}897)$	1	1	1	1	1	1	1	1	1	1
Credibility distribution estimation with claim amount $X_{ij} \leq x$, ($x = 320, 800, 1000, 2000, 3000, 23{,}800, 23{,}896, 23{,}897$)									
Country	AT	DE	FI	GR	HR	IT	NO	PL	PT	SE
$\widehat{F}^{Cred}_{X_{ij}}(320\Theta_j)$	0.01256	0.01256	0.01256	0.01256	0.01256	0.01256	0.01256	0.01256	0.01256	0.01256
$\widehat{F}^{Cred}_{X_{ij}}(800\Theta_j)$	0.02790	0.01348	0.04299	0.06155	0.06539	0.01980	0.03015	0.02490	0.02983	0.05827
$\widehat{F}^{Cred}_{X_{ij}}(1000\Theta_j)$	0.02754	0.00583	0.23820	0.24210	0.25270	0.01136	0.03722	0.01939	0.10500	0.33390
$\widehat{F}^{Cred}_{X_{ij}}(2000\Theta_j)$	0.25080	0.00105	0.93630	0.93420	0.86420	0.00223	0.95570	0.21300	0.95950	0.97040
$\widehat{F}^{Cred}_{X_{ij}}(3000\Theta_j)$	0.95520	0.00313	0.89460	0.89140	0.79400	0.00660	0.92490	0.71150	0.93090	0.94880
$\widehat{F}^{Cred}_{X_{ij}}(23{,}800\Theta_j)$	0.9707	0.9707	0.9707	0.9707	0.9707	0.9707	0.9707	0.9707	0.9707	0.9707
$\widehat{F}^{Cred}_{X_{ij}}(23{,}896\Theta_j)$	0.9707	0.9707	0.9707	0.9707	0.9707	0.9707	0.9707	0.9707	0.9707	0.9707
$\widehat{F}^{Cred}_{X_{ij}}(23{,}897\Theta_j)$	1	1	1	1	1	1	1	1	1	1
Credibility factor $X_{ij} \leq x$, ($x = 320, 800, 1000, 2000, 3000, 23{,}800, 23{,}896, 23{,}897$)									
Parameter	$\widehat{Z}^{F_x}_1$	$\widehat{Z}^{F_x}_2$	$\widehat{Z}^{F_x}_3$	$\widehat{Z}^{F_x}_4$	$\widehat{Z}^{F_x}_5$	$\widehat{Z}^{F_x}_6$	$\widehat{Z}^{F_x}_7$	$\widehat{Z}^{F_x}_8$	$\widehat{Z}^{F_x}_9$	$\widehat{Z}^{F_x}_{10}$
$x = 320$	0	0	0	0	0	0	0	0	0	0
$x = 800$	0.1727	0.6002	0.0744	0.0719	0.0328	0.4128	0.1060	0.2618	0.1153	0.1532
$x = 1000$	0.6020	0.9158	0.3679	0.3596	0.1971	0.8359	0.4622	0.7198	0.4857	0.5672
$x = 2000$	0.9669	0.9953	0.9182	0.9155	0.8256	0.9899	0.9431	0.9802	0.9479	0.9619
$x = 3000$	0.9340	0.9903	0.8448	0.8400	0.6965	0.9794	0.8893	0.9600	0.8983	0.9246
$x = 23{,}800$	0	0	0	0	0	0	0	0	0	0
$x = 23{,}896$	0	0	0	0	0	0	0	0	0	0
$x = 23{,}897$	0	0	0	0	0	0	0	0	0	0
Parameter estimation $X_{ij} \leq x$, ($x = 320, 800, 1000, 2000, 3000, 23{,}800, 23{,}896, 23{,}897$)									
$x = 320$	$\widehat{F}_{n_{ww}}(x) = 0.01256$			$\widehat{a}_{F_x} = 0.00000$			$\widehat{s}^2_{F_x} = 456{,}695$			
$x = 800$	$\widehat{F}_{n_{ww}}(x) = 0.03372$			$\widehat{a}_{F_x} = 0.00412$			$\widehat{s}^2_{F_x} = 379{,}951$			
$x = 1000$	$\widehat{F}_{n_{ww}}(x) = 0.06920$			$\widehat{a}_{F_x} = 0.02147$			$\widehat{s}^2_{F_x} = 273{,}007$			
$x = 2000$	$\widehat{F}_{n_{ww}}(x) = 0.22117$			$\widehat{a}_{F_x} = 0.09854$			$\widehat{s}^2_{F_x} = 64{,}969$			
$x = 3000$	$\widehat{F}_{n_{ww}}(x) = 0.32113$			$\widehat{a}_{F_x} = 0.14983$			$\widehat{s}^2_{F_x} = 203{,}743$			
$x = 23{,}800$	$\widehat{F}_{n_{ww}}(x) = 0.97070$			$\widehat{a}_{F_x} = 0.00000$			$\widehat{s}^2_{F_x} = 1{,}610{,}559$			
$x = 23{,}896$	$\widehat{F}_{n_{ww}}(x) = 0.97070$			$\widehat{a}_{F_x} = 0.00000$			$\widehat{s}^2_{F_x} = 1{,}610{,}559$			
$x = 23{,}897$	$\widehat{F}_{n_{ww}}(x) = 1.00000$			$\widehat{a}_{F_x} = 0.00000$			$\widehat{s}^2_{F_x} = 1{,}736{,}923$			

Table 3. Credibility distribution estimation for motor claims.

Motor Claim Amounts and Number of Motor Claims from 10 Selected European Countries during the Period 2004–2018, Z_F Free of x									
Individual empirical distribution with claim amount $X_{ij} \leq x$, (x = 320, 800, 1000, 2000, 3000, 23,800, 23,896, 23,897)									
Country	AT	DE	FI	GR	HR	IT	NO	PL	PT	SE
$\widehat{F}_{n_{wj}}(320)$	0.0000	0.0000	0.0000	0.0000	0.9235	0.0000	0.0000	0.0000	0.0000	0.0624
$\widehat{F}_{n_{wj}}(800)$	0.0000	0.0000	0.1584	0.4206	1.0000	0.0000	0.0000	0.0000	0.0000	0.1940
$\widehat{F}_{n_{wj}}(1000)$	0.0000	0.0000	0.5286	0.5501	1.0000	0.0000	0.0000	0.0000	0.1428	0.5359
$\widehat{F}_{n_{wj}}(2000)$	0.2518	0.0000	1.0000	1.0000	1.0000	0.0000	1.0000	0.2128	1.0000	1.0000
$\widehat{F}_{n_{wj}}(3000)$	1.0000	0.0000	1.0000	1.0000	1.0000	0.0000	1.0000	0.7278	1.0000	1.0000
$\widehat{F}_{n_{wj}}(23,800)$	1.0000	0.9339	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\widehat{F}_{n_{wj}}(23,896)$	1.0000	0.9339	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\widehat{F}_{n_{wj}}(23,897)$	1	1	1	1	1	1	1	1	1	1
Credibility distribution estimation with claim amount $X_{ij} \leq x$, (x = 320, 800, 1000, 2000, 3000, 23,800, 23,896, 23,897)									
Country	AT	DE	FI	GR	HR	IT	NO	PL	PT	SE
$\widehat{F}^{Cred}_{X_{ij}}(320\Theta_j)$	0.00706	0.00190	0.00966	0.00974	0.11476	0.00347	0.00871	0.00541	0.00845	0.03262
$\widehat{F}^{Cred}_{X_{ij}}(800\Theta_j)$	0.01896	0.00511	0.06246	0.12047	0.14212	0.00931	0.02338	0.01452	0.02269	0.09829
$\widehat{F}^{Cred}_{X_{ij}}(1000\Theta_j)$	0.03891	0.01049	0.17511	0.17703	0.17362	0.01911	0.04798	0.02979	0.09327	0.25723
$\widehat{F}^{Cred}_{X_{ij}}(2000\Theta_j)$	0.23458	0.03352	0.40072	0.39581	0.30854	0.06106	0.46002	0.21640	0.47592	0.53495
$\widehat{F}^{Cred}_{X_{ij}}(3000\Theta_j)$	0.61831	0.04866	0.477641	0.47336	0.39729	0.08866	0.52932	0.55269	0.54317	0.59463
$\widehat{F}^{Cred}_{X_{ij}}(23,800\Theta_j)$	0.98352	0.93947	0.97745	0.97727	0.97398	0.99191	0.97968	0.98738	0.98028	0.98250
$\widehat{F}^{Cred}_{X_{ij}}(23,896\Theta_j)$	0.98352	0.93947	0.97745	0.97727	0.97398	0.99191	0.97968	0.98738	0.98028	0.98250
$\widehat{F}^{Cred}_{X_{ij}}(23,897\Theta_j)$	1	1	1	1	1	1	1	1	1	1
$\widehat{F}_{n_{ww}}(320) = 0.01256$, $\widehat{F}_{n_{ww}}(800) = 0.03372$, $\widehat{F}_{n_{ww}}(1000) = 0.06920$, $\widehat{F}_{n_{ww}}(2000) = 0.22117$, $\widehat{F}_{n_{ww}}(3000) = 0.32113$, $\widehat{F}_{n_{ww}}(23,800) = 0.97070$, $\widehat{F}_{n_{ww}}(23,896) = 0.97070$, $\widehat{F}_{n_{ww}}(23,897) = 1.000$									
Parameter estimation free of x									
$\widehat{a}_F = 0.008457228$ $\qquad$ $\widehat{s}^2_F = 208898.6$									
Credibility factor Z^F_j free of x	$\widehat{Z}^F_1$ 0.43775	$\widehat{Z}^F_2$ 0.84846	$\widehat{Z}^F_3$ 0.23054	$\widehat{Z}^F_4$ 0.22423	$\widehat{Z}^F_5$ 0.11218	$\widehat{Z}^F_6$ 0.72391	$\widehat{Z}^F_7$ 0.30667	$\widehat{Z}^F_8$ 0.56942	$\widehat{Z}^F_9$ 0.32708	$\widehat{Z}^F_{10}$ 0.40288

Table 4. Credibility coefficients.

Credibility Coefficients						
x	320	800	1000	2000	23,800	23,900
WRV	23,262.8543	5787.8587	3718.7216	1152.5669	254.4523	1317.9237
BRV	8.448499	4.140172	3.170754	1.976505	0	0
CC	7,581,705	1,954,336	1,375,506	340,045	∞	∞

Remark 3. *The results of Table 2 and Remark 2, for x = 23,800, $\widehat{a}_{F_x} = 0$ and $\widehat{s}^2_{F_x} = 61,020$, imply that BRV = 0 and CC = ∞. Similarly, setting x = 23,900, $\widehat{a}_{F_x} = 0$ and $\widehat{s}^2_{F_x} = 1,736,923$ implies that BRV = 0 and CC = ∞.*

5.2. Example of Credibility Distribution Estimation with Financial Data

The dataset was created (see Fama and French (2022)) as follows: each NYSE, AMEX, and NASDAQ stock was assigned to an industry portfolio at the end of June of year t based on its four-digit SIC code at that time. Compustat SIC codes have been used for the fiscal year ending in the calendar year $t − 1$. Whenever Compustat SIC codes are not available, CRSP SIC codes for June of year t were used. Then, returns from July of year t to June of year $t + 1$ are computed. The weights are the number of firms in portfolios.

In particular, the portfolios are constructed with monthly returns from July 1926 to July 2022 and it contains value returns for 10 industry portfolios. The credibility distribution for each of these portfolios needs to be estimated. As a profit (P), we consider a random variable X, with positive returns values and as a loss (L) with negative return values. The 10 industry portfolios are as follows:

(1)   NoDur: consumer non-durables—food, tobacco, textiles, apparel, leather, and toys.

(2)   Durbl: Consumer durables—cars, TVs, furniture, household appliances.

(3)   Manuf: Manufacturing—machinery, trucks, planes, chemicals, off-furn, and paper.

(4)   Enrgy: Oil, gas, and coal extraction and products.

(5)   HiTec: Business equipment—computers, software, and electronic equipment.

(6)   Telcm: Telephone and television transmission.

(7)   Shops: Wholesale, retail, and some services (laundries, repair shops).

(8)   Hlth: healthcare, medical equipment, and drugs.

(9)   Utils: Utilities.

(10)   Other: Other—mines, construction, building material, transportation, hotels, bus service, entertainment, and finance.

Table 5 provides some descriptive statistics of the (P/L) monthly returns of the 10 industry portfolios. The number of observations in each portfolio is $n = 1155$.

Table 5. Summary statistics for 10 industry portfolios.

Monthly Returns and Number of Firms in Portfolios from July 1926–July 2022										
X_{ij}: Value returns for $i = 1, \cdots, 1155, j = 1, \cdots, 10$										
Portfolio	NoDur	Durbl	Manuf	Enrgy	HiTec	Telcm	Shops	Hlth	Utils	Other
Min.	−24.6900	−34.800	−29.820	−34.490	−33.870	−21.5600	−30.240	−34.080	−33.0500	−30.0200
1st Qu.	−1.3900	−2.855	−2.000	−2.470	−2.600	−1.5200	−2.030	−1.920	−1.6850	−2.0750
Median	1.0800	0.980	1.350	0.890	1.320	0.9000	1.090	1.100	1.0500	1.2800
Mean	0.9524	1.158	1.002	1.027	1.116	0.8198	1.014	1.072	0.8738	0.9013
3rd Qu.	3.6450	4.845	4.235	4.590	5.030	3.2400	4.090	4.060	3.6200	4.1850
Max.	34.1700	79.790	57.200	38.990	53.490	28.1700	42.450	37.130	43.4600	58.6700
w_{ij}: Weight–number of firms in portfolios for $i = 1, \cdots, 1155, j = 1, \cdots, 10$										
Min.	87.0	37.0	125.0	45.0	18.0	4.00	41.0	4.0	21.0	110.0
1st Qu.	136.0	56.0	313.0	55.0	44.0	8.00	84.0	18.0	72.0	156.0
Median	173.0	92.0	449.0	116.0	358.0	41.00	276.0	122.0	102.0	1002.0
Mean	230.2	101.2	507.7	131.1	428.1	53.76	298.4	237.2	106.9	887.5
3rd Qu.	334.0	148.0	772.5	173.0	797.5	99.00	472.5	509.0	179.5	1619.0
Max.	547.0	213.0	967.0	404.0	1465.0	189.00	823.0	868.0	204.0	2249.0

Note: X_{ij} denotes the values of monthly returns per year and w_{ij} represents the number of firms in portfolios that correspond to each X_{ij}.

Table 6 illustrates the results of credibility distribution function for monthly returns for 10 industry portfolios from July 1926 to July 2022. More analytically, the upper part of the table shows the individual empirical distribution $\widehat{F}_{n_{wj}}(x)$ of the returns $X_{ijh} \leq x$, ($x = -15, -10, -5, 0, 10, 15, 34.17, 59, 60, 79.79$) and the corresponding credibility distribution estimators $\widehat{F}^{Cred}_{X_{ij}|\Theta_j}(x|\Theta_j)$ are shown in the middle part of the table. The estimated credibility factors $\widehat{Z}^{F_x}_j$, as well the estimated parameters $\widehat{F}_{n_{ww}}(x)$, $\widehat{s}^2_{F_x}$, $\widehat{a}_{F_x}$, are presented in the lower part of Table 6. The monotonicity of the estimated distribution function is shown in Table 6. By letting the values of returns be larger than $x = 59$ and less than or equal to $x = 79.79$ ($x = 79.79$ is the maximum threshold of portfolio Durbl, which is the portfolio with the largest return values, as shown in Table 5), the values of the estimated credibility distribution $\widehat{F}^{Cred}_{X_{ij}}(x = 0|\Theta_j)$ remain the same up to the fifth decimal place. By letting $x > 79.79$, $\widehat{F}^{Cred}_{X_{ij}}(x > 70.79|\Theta_j)$ goes to 1 (see Table 6).

Figure 2 displays the individual empirical distribution in each contract. Again, note that the red bullets indicate the corresponding credibility estimate at specific points presented in Table 6.

Table 6. Credibility distribution estimation for industry portfolios.

| Monthly returns for 10 industry portfolios from July 1926–July 2022										
Individual empirical distribution with returns $X_{ij} \leq x$, ($x = -15, -10, 5, 0, 10, 15, 34.17, 59, 60, >79.79$)										
Portfolios	NoDur	Durbl	Manuf	Enrgy	HiTec	Telcm	Shops	Hlth	Utils	Other
$\widehat{F}_{n_{wj}}(-15)$	0.00309	0.01670	0.00745	0.01000	0.01840	0.00498	0.00630	0.00213	0.00177	0.00952
$\widehat{F}_{n_{wj}}(-10)$	0.02113	0.04486	0.02680	0.04053	0.05738	0.02831	0.02171	0.01005	0.01472	0.03030
$\widehat{F}_{n_{wj}}(-5)$	0.07267	0.13872	0.10760	0.12174	0.16253	0.10611	0.10199	0.10553	0.07147	0.11525
$\widehat{F}_{n_{wj}}(0)$	0.38587	0.42946	0.40825	0.41781	0.43866	0.39040	0.39827	0.37940	0.39013	0.40579
$\widehat{F}_{n_{wj}}(10)$	0.97618	0.93303	0.96468	0.95040	0.90779	0.97479	0.95921	0.97628	0.98455	0.96935
$\widehat{F}_{n_{wj}}(15)$	0.99500	0.97700	0.99300	0.98200	0.96200	0.99800	0.99500	0.99700	0.99600	0.99200
$\widehat{F}_{n_{wj}}(34.17)$	0.99959	0.99676	0.99896	0.99929	0.99985	1.00000	0.99980	0.99996	0.99982	0.99974
$\widehat{F}_{n_{wj}}(59)$	1.000000	0.9996	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
$\widehat{F}_{n_{wj}}(60)$	1.000000	0.9996	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
$\widehat{F}_{n_{wj}}(x > 79.79)$	1.000000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
Credibility distribution estimation with returns $X_{ij} \leq x$, ($x = -15, -10, 5, 0, 10, 15, 34.17, 59, 60, >79.79$)										
Portfolios	NoDur	Durbl	Manuf	Enrgy	HiTec	Telcm	Shops	Hlth	Utils	Other
$\widehat{F}_{X_{ij}}^{Cred}(-15\Theta_j)$	0.00496	0.01270	0.00771	0.00952	0.01650	0.00762	0.00700	0.00426	0.00537	0.00946
$\widehat{F}_{X_{ij}}^{Cred}(-10\Theta_j)$	0.02260	0.04073	0.02711	0.03813	0.05501	0.02939	0.02283	0.01323	0.01926	0.03032
$\widehat{F}_{X_{ij}}^{Cred}(-5\Theta_j)$	0.07919	0.13160	0.10810	0.12000	0.15820	0.10980	0.10360	0.10480	0.08371	0.11520
$\widehat{F}_{X_{ij}}^{Cred}(0\Theta_j)$	0.40100	0.40800	0.40500	0.40700	0.41400	0.40400	0.40300	0.40000	0.40400	0.40600
$\widehat{F}_{X_{ij}}^{Cred}(10\Theta_j)$	0.97500	0.93600	0.96500	0.94900	0.91000	0.97100	0.95900	0.97500	0.98100	0.96900
$\widehat{F}_{X_{ij}}^{Cred}(15\Theta_j)$	0.99440	0.97890	0.99280	0.98280	0.96320	0.99530	0.99460	0.99630	0.99470	0.99190
$\widehat{F}_{X_{ij}}^{Cred}(34.17\Theta_j)$	0.99952	0.99951	0.99951	0.99952	0.99953	0.99952	0.99952	0.99952	0.99952	0.99953
$\widehat{F}_{X_{ij}}^{Cred}(59\Theta_j)$	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999
$\widehat{F}_{X_{ij}}^{Cred}(60\Theta_j)$	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999	0.99999
$\widehat{F}_{X_{ij}}^{Cred}(x > 79.79\Theta_j)$	1.000000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
Credibility factor with returns $X_{ij} \leq x$, ($x = -15, -10, 5, 0, 10, 15, 34.17, 59, 60, >79.79$)										
Parameter	$\widehat{Z}_1^{F_x}$	$\widehat{Z}_2^{F_x}$	$\widehat{Z}_3^{F_x}$	$\widehat{Z}_4^{F_x}$	$\widehat{Z}_5^{F_x}$	$\widehat{Z}_6^{F_x}$	$\widehat{Z}_7^{F_x}$	$\widehat{Z}_8^{F_x}$	$\widehat{Z}_9^{F_x}$	$\widehat{Z}_{10}^{F_x}$
$x = -15$	0.68105	0.48412	0.82488	0.54873	0.79887	0.33276	0.73462	0.68754	0.49796	0.89169
$x = -10$	0.84475	0.70514	0.92310	0.75602	0.91008	0.55964	0.87584	0.84866	0.71652	0.95451
$x = -5$	0.84567	0.70659	0.92359	0.75731	0.91066	0.56137	0.87660	0.84955	0.71794	0.95481
$x = 0$	0.20416	0.10132	0.36138	0.12747	0.32303	0.05653	0.24956	0.20908	0.10648	0.49727
$x = 10$	0.92860	0.85111	0.96632	0.88105	0.96031	0.75234	0.94401	0.93057	0.85799	0.98045
$x = 15$	0.91814	0.83134	0.96115	0.86463	0.95426	0.72373	0.93565	0.92037	0.83897	0.97740
$x = 34.17$	0.09671	0.04494	0.19106	0.05747	0.16608	0.02440	0.12188	0.09937	0.04738	0.29220
$x = 59$	0	0	0	0	0	0	0	0	0	0
$x = 60$	0	0	0	0	0	0	0	0	0	0
$x > 79.79$	0	0	0	0	0	0	0	0	0	0
Parameter estimation $X_{ij} \leq x$, ($x = -15, -10, 5, 0, 10, 15, 34.17, 59, 60, >79.79$)										
$x = -15$	$\widehat{F}_{n_{ww}}(x) = 0.00894$ $\widehat{a}_{F_x} = 2.118 \times 10^{-5}$ $\widehat{s}_{F_x}^2 = 2.637$									
$x = -10$	$\widehat{F}_{n_{ww}}(x) = 0.03077$ $\widehat{a}_{F_x} = 0.00018110$ $\widehat{s}_{F_x}^2 = 8.848$									
$x = -5$	$\widehat{F}_{n_{ww}}(x) = 0.11460$ $\widehat{a}_{F_x} = 0.00062020$ $\widehat{s}_{F_x}^2 = 30.09$									
$x = 0$	$\widehat{F}_{n_{ww}}(x) = 0.40530$ $\widehat{a}_{F_x} = 6.935 \times 10^{-5}$ $\widehat{s}_{F_x}^2 = 71.87$									
$x = 10$	$\widehat{F}_{n_{ww}}(x) = 0.95820$ $\widehat{a}_{F_x} = 0.00057730$ $\widehat{s}_{F_x}^2 = 11.80$									
$x = 15$	$\widehat{F}_{n_{ww}}(x) = 0.98810$ $\widehat{a}_{F_x} = 0.00014610$ $\widehat{s}_{F_x}^2 = 3.463$									
$x = 34.17$	$\widehat{F}_{n_{ww}}(x) = 0.99952$ $\widehat{a}_{F_x} = 5.731 \times 10^{-8}$ $\widehat{s}_{F_x}^2 = 0.1423$									
$x = 59$	$\widehat{F}_{n_{ww}}(x) = 0.99999$ $\widehat{a}_{F_x} = 0.00000000$ $\widehat{s}_{F_x}^2 = 0.00424$									
$x = 60$	$\widehat{F}_{n_{ww}}(x) = 0.99999$ $\widehat{a}_{F_x} = 0.00000000$ $\widehat{s}_{F_x}^2 = 0.00424$									
$x > 79.79$	$\widehat{F}_{n_{ww}}(x) = 1.00000$ $\widehat{a}_{F_x} = 0.00000000$ $\widehat{s}_{F_x}^2 = 0.00000$									
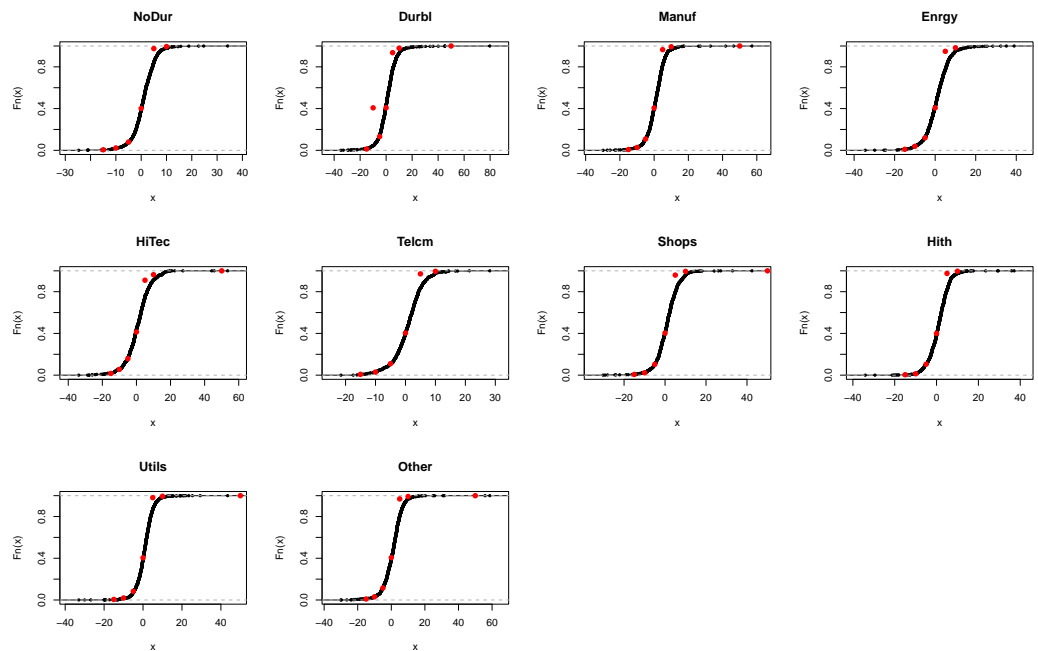
Figure 2. Individual empirical distribution and credibility distribution point estimates (in red) for industry portfolios per contract.

Credibility Coefficients for Industry Portfolios Data

Here, we provide an intuitive interpretation for the form of the credibility distribution estimator for the monthly returns for the 10 industry portfolios, by presenting the following credibility coefficients. Table 7 illustrates the coefficient of variation $BRV = \frac{\sqrt{\widehat{a}_{F_x}}}{\widehat{F}_{n_{ww}}(x)}$, the average within-risk coefficient of variation $WRV = \frac{\widehat{s}_{F_x}}{\widehat{F}_{n_{ww}}(x)}$, and the credibility coefficient $CC = \frac{\widehat{s}^2_{F_x}}{\widehat{a}_{F_x}}$ for the industry portfolio data.

Table 7. Credibility coefficients.

			Credibility Coefficients for Industry Portfolios				
x	-15	-10	-5	0	10	15	50
WRV	181.642532	96.670744	47.865927	20.916895	3.584964	1.883325	21.656408
BRV	0.51478450	0.43735262	0.21731078	0.02054692	0.02507521	0.01223275	0
CC	124504.25	48856.99	48516.61	1036337.42	20439.98	23702.94	∞

Remark 4. *The results of Table 6 and Remark 2, for $x = 50$, $\widehat{a}_{F_x} = 0$ and $\widehat{s}^2_{F_x} = 0.0469$, imply that $BRV = 0$ and $CC = \infty$.*

5.3. Example of Credibility Distribution Estimation with Financial Grouped Data

The empirical distribution function for the grouped data was depicted by the step function of Fama and French (2022) data. The grouping (see Table 8) is a subjective element in this fit and other persons would have different ones. The total number of observations in each portfolio is the same ($m_j = 1155$).

Table 9 illustrates the results of the credibility distribution function for monthly returns for 10 industry portfolios from July 1926 to July 2022. Analytically, the upper part of the table shows the individual empirical distribution $\widehat{F}_{mj}(x)$ of returns $X_{ij} \leq x$, ($x = -15, -10, -5, 0, 10, 15$) and the corresponding credibility distribution estimators $\widehat{F}^{Cred}_{X_{ij}|\Theta_j}(x|\Theta_j)$ are shown in the middle part of the table. The estimated credibility fac-

tors Z_j^x, as well the estimated parameters $\widehat{F}_{mm}(x)$, $\widehat{s}_x^2$, $\widehat{a}_x$, are presented in the lower part of Table 9. The monotonicity of the estimated distribution function is shown in Table 9, but the convergence to one of the estimated credibility distribution for grouped data should be further investigated.

Table 8. Credibility distribution estimation for grouped data.

10 Industry Portfolios									
Grouped Monthly Returns in 10 Intervals from July 1926 to July 2022									
m_{ij}: Number of data points in the interval for each portfolio									
Interval of return	NoDur	Durbl	Manuf	Enrgy	HiTec	Telcm	Shops	Hlth	Utils	Other
$-35 \leq X_{ij} \leq -20$	4	12	6	4	10	1	6	4	3	8
$-20 \leq X_{ij} \leq -13$	2	20	19	16	21	9	13	8	13	18
$-13 \leq X_{ij} \leq -6$	60	105	84	92	108	61	73	70	64	94
$-6 \leq X_{ij} \leq -4$	58	86	67	80	77	64	78	77	68	65
$-4 \leq X_{ij} \leq -1$	199	188	195	210	181	205	197	204	200	195
$-1 \leq X_{ij} \leq 2$	383	245	273	274	234	394	300	304	351	272
$2 \leq X_{ij} \leq 8$	406	357	434	364	380	362	388	408	391	415
$8 \leq X_{ij} \leq 10$	20	55	30	52	58	29	52	42	28	48
$10 \leq X_{ij} \leq 22$	20	72	41	56	81	29	43	33	33	34
$22 \leq X_{ij} \leq 80$	3	15	6	7	5	1	5	5	4	6
Total # of observations	1155	1155	1155	1155	1155	1155	1155	1155	1155	1155

Table 9. Grouped credibility distribution estimation.

Grouped monthly returns in 10 intervals for 10 industry portfolios									
from July 1926–July 2022									
Individual empirical distribution with returns $X_{ij} \leq x$, ($x = -15, -10, -5, 0, 10, 15$)									
Portfolios	NoDur	Durbl	Manuf	Enrgy	HiTec	Telcm	Shops	Hlth	Utils	Other
$\widehat{F}_{mj}(-15)$	0.00470	0.02276	0.01694	0.01336	0.02165	0.00643	0.01323	0.00841	0.01064	0.01800
$\widehat{F}_{mj}(-10)$	0.02746	0.06667	0.05281	0.05145	0.06691	0.03129	0.04354	0.03636	0.03760	0.05739
$\widehat{F}_{mj}(-5)$	0.08225	0.15580	0.12340	0.13160	0.15370	0.08918	0.11340	0.10430	0.09870	0.13200
$\widehat{F}_{mj}(0)$	0.39020	0.42660	0.40000	0.42710	0.41130	0.40810	0.40430	0.40200	0.40260	0.40750
$\widehat{F}_{mj}(10)$	0.98009	0.92468	0.95931	0.94545	0.92554	0.97403	0.95844	0.96710	0.96797	0.96537
$\widehat{F}_{mj}(15)$	0.98730	0.95065	0.97410	0.96566	0.95476	0.98449	0.97395	0.97900	0.97987	0.97763
Credibility distribution estimation with returns $X_{ij} \leq x$, ($x = -15 - 10, -5, 0, 10, 15$)									
Portfolios	NoDur	Durbl	Manuf	Enrgy	HiTec	Telcm	Shops	Hlth	Utils	Other
$\widehat{F}_{X_{ij}}^{Cred}(-15\Theta_j)$	0.00472	0.02270	0.01690	0.01340	0.02160	0.00645	0.01320	0.00842	0.01060	0.01810
$\widehat{F}_{X_{ij}}^{Cred}(-10\Theta_j)$	0.02764	0.06649	0.05276	0.05141	0.06673	0.03143	0.04357	0.03646	0.03769	0.0573
$\widehat{F}_{X_{ij}}^{Cred}(-5\Theta_j)$	0.08268	0.15540	0.12330	0.13140	0.15330	0.08953	0.11350	0.10450	0.09893	0.13180
$\widehat{F}_{X_{ij}}^{Cred}(0\Theta_j)$	0.39200	0.42470	0.40080	0.42520	0.41100	0.40810	0.40470	0.40260	0.40310	0.40705
$\widehat{F}_{X_{ij}}^{Cred}(10\Theta_j)$	0.97986	0.92499	0.95929	0.94556	0.92584	0.97386	0.95842	0.96700	0.96786	0.96529
$\widehat{F}_{X_{ij}}^{Cred}(15\Theta_j)$	0.98717	0.95084	0.97409	0.96572	0.95492	0.98439	0.97394	0.97895	0.97981	0.97759
Parameter estimation $X_{ij} \leq x$, ($x = -15, -10, -5, 0, 10, 15$)									
$x = -15$	$\widehat{F}_{mm}(x) = 0.013617$		$\widehat{a}_x = 3.8391 \times 10^{-5}$		$\widehat{s}_x^2 = 7.9089 \times 10^{-7}$		$\widehat{Z}_j^x = 0.997944$			
$x = -10$	$\widehat{F}_{mm}(x) = 0.047149$		$\widehat{a}_x = 0.000196743$		$\widehat{s}_x^2 = 1.7862 \times 10^{-5}$		$\widehat{Z}_j^x = 0.991002$			
$x = -5$	$\widehat{F}_{mm}(x) = 0.011843$		$\widehat{a}_x = 0.0006369784$		$\widehat{s}_x^2 = 7.6405 \times 10^{-5}$		$\widehat{Z}_j^x = 0.950530$			
$x = 0$	$\widehat{F}_{mm}(x) = 0.407970$		$\widehat{a}_x = 0.000118016$		$\widehat{s}_x^2 = 0.1327 \times 10^{-3}$		$\widehat{Z}_j^x = 0.9881473$			
$x = 10$	$\widehat{F}_{mm}(x) = 0.956798$		$\widehat{a}_x = 0.000362038$		$\widehat{s}_x^2 = 3.5504 \times 10^{-5}$		$\widehat{Z}_j^x = 0.990288$			
$x = 15$	$\widehat{F}_{mm}(x) = 0.972741$		$\widehat{a}_x = 0.000146344$		$\widehat{s}_x^2 = 1.2874 \times 10^{-5}$		$\widehat{Z}_j^x = 0.9912793$			

Figure 3 displays the smoothed individual empirical distribution for grouped data in each contract. Again, the red bullets indicate the corresponding credibility estimate at specific points presented in Table 9.
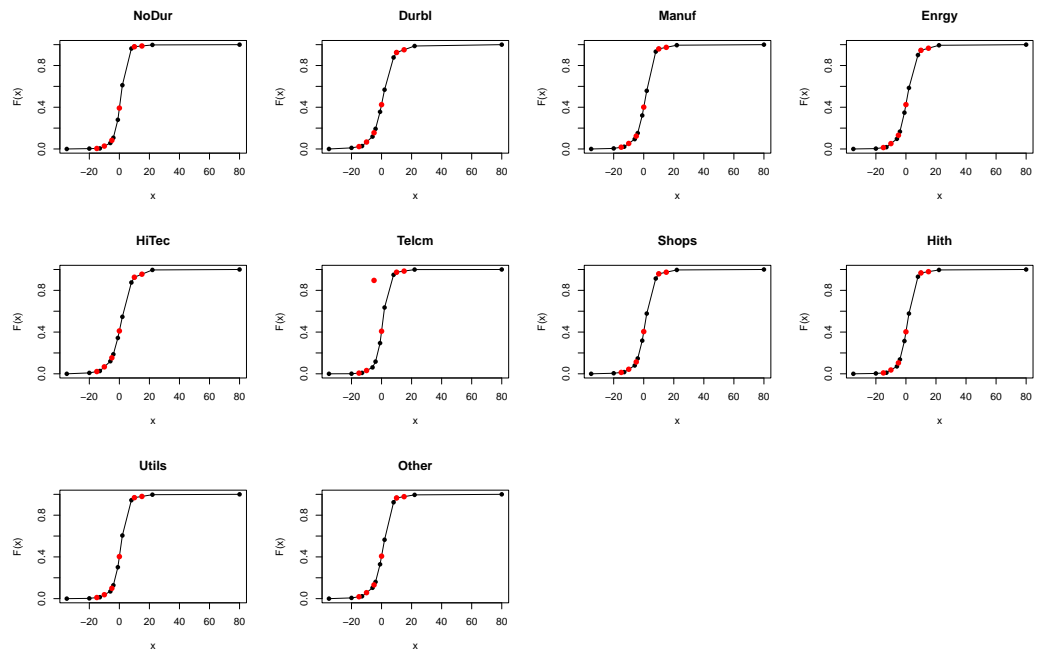

Figure 3. Individual empirical distribution and credibility distribution point estimates (in red) for grouped data per contract.

Credibility Coefficients for Financial Grouped Data

Table 10 illustrates the coefficient of variation *BRV*, the average within-risk coefficient of variation *WRV*, and the credibility coefficient *CC* for the industry portfolios of grouped data.

Table 10. Credibility coefficients.

	Credibility Coefficients for Industry Portfolios for Grouped Data					
x	−15	−10	−5	0	10	15
WRV	0.06530954	0.08963808	0.11209926	0.05720951	0.00622752	0.00368866
BRV	0.45502292	0.29749328	0.15540177	0.09426977	0.01988643	0.01243626
CC	0.02060092	0.09078849	0.52034766	0.36829137	0.09806570	0.08797441

5.4. Example of the Classical Credibility Estimation with Financial Grouped Data

For grouped data, the previous approach gives a credibility point estimate. If we want to derive the classical credibility estimation, we can apply the concept of uniform distribution within each interval of returns and take the interval midpoints as the value of return. The weights are the number of observations in each interval. Table 11 shows the individual average return for the 10 industry portfolios $\widehat{\mu}_j$, the credibility estimation of returns for these portfolios $\mu(\Theta_j)^{Cred}$, along with the credibility factor Z_j and the estimated parameters $\widehat{\mu}$, $\widehat{a}$ and $\widehat{s}^2$.

Table 11. Classical credibility model for grouped data.

Individual Average Return for the 10 Industry Portfolios									
Portfolios	NoDur	Durbl	Manuf	Enrgy	HiTec	Telcm	Shops	Hlth	Utils	Other
$\widehat{\mu}_j$	1.15022	1.52554	1.20217	1.18052	1.34199	1.01862	1.27099	1.25368	1.12597	1.06277
Credibility estimation of returns for the 10 industry portfolios									
$\mu(\Theta_j)^{Cred}$	1.18166	1.36974	1.20769	1.19685	1.27776	1.11571	1.24219	1.23351	1.16951	1.13784
Credibility parameter estimation									
$\widehat{\mu} = 1.213247 \quad \widehat{a} = 4.807552 \quad \widehat{s}^2 = 5528.034 \quad Z_j = 0.501114$									

6. Concluding Remarks

The objective of this paper was to present the appropriate credibility distribution model that adequately describes the insurance losses, a model that can be used for risk management purposes.

The main contribution of the paper is that it embedded the empirical distribution into credibility modeling in the form of the Bühlmann and Straub (1970) model. In the first part of the paper, we present the model of the weighted credibility distribution, and in the second part, a model that applies to a grouped data in intervals.

With our models, we examine two datasets, one with motor claim amounts and the number of motor claims from 10 selected European countries during the period 2004–2020, and a second with monthly returns from July 1926 to July 2022 for 10 industry portfolios. For applying our credibility distribution model with grouped data, we grouped the second dataset (Fama/French financial data) into intervals of claim amounts. Under this setting, the grouping is subjective and the weights are the number of points within each interval and the total weights in each interval are the same.

The monotonicity (or non-monotonicity) and the convergence to one of the estimated distribution functions are shown numerically in Tables 2, 3, 6 and 9. From a theoretical point of view, the monotonicity, as well as the convergence of the estimated distribution functions need further investigation. Furthermore, the sufficient conditions for the asymptotic optimality of the empirical credibility distribution estimators can be also investigated, providing some good ideas for a new project.

References

Bozikas, Apostolos, and Georgios Pitselis. 2020. Incorporating crossed classification credibility into the Lee–Carter model for multi-population mortality data. *Insurance: Mathematics and Economics* 93: 353–68. [CrossRef]

Bozikas, Apostolos, and Georgios Pitselis. 2021. Multi-population mortality modelling and forecasting: A hierarchical credibility regression approach. *European Actuarial Journal* 11: 231–67. [CrossRef]

Bühlmann, Hans. 1967. Experience rating and credibility. *ASTIN Bulletin* 4: 199–207.

Bühlmann, Hans, and Erwin Straub. 1970. Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* 70: 111–33.

Cai, Xiaoqiang, Limin Wen, Xianyi Wu, and Xian Zhou. 2015. Credibility Estimation of distribution functions with applications to experience rating in general insurance. *North American Actuarial Journal* 19: 311–35. [CrossRef]

Christiansen, Marcus, and Edo Schinzinger. 2016. A Credibility Approach for Combining Likelihoods Generalized Linear Models. *Astin Bulletin* 46: 531–69. [CrossRef]

Denuit, Michel. 2008. Comonotonic approximations to quantiles of life annuity conditional expected present value. *Insurance: Mathematics and Economics* 42: 831–38. [CrossRef]

De Vylder, Etienne F. 1996. *Advanced Risk Theory-A Self-Contained Introduction*. Brussels: Editions de L'Universite de Bruxelles.

De Vylder, Etienne F. 1978. Parameter Estimation in Credibility Theory. *ASTIN Bulletin* 10: 99–112. [CrossRef]

De Vylder, Etienne F. 1976. Geometrical Credibility. Scandinavian *Actuarial Journal* 3: 121–49. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 2022. CRSP Data. Available online: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (accessed on 20 September 2022).

Ferguson, Thomas 1973. A Bayesian analysis of some non-parametric problems. *Annals of Statistics* 1: 209–30.

Friedman, Jerome, and Robert Tibshirani. 1984. The monotone smoothing of scatterplots. *Technometrics* 26: 243–50. [CrossRef]

Gebizlioglu, Omer L., and Banu Yagci. 2008. Tolerance intervals for quantiles of bivariate risks and risk measurement. *Insurance: Mathematics and Economics* 42: 1022–27. [CrossRef]

Gong, Yikai (Maxwell), Zhuangdi Li, Maria Milazzo, Kristen Moore, and Matthew Provencher. 2018. Credibility methods for individual life insurance. *Risks* 6: 144. [CrossRef]

Hachemeister, Charles. A. 1975. Credibility for regression models with application to trend. In *Credibility, Theory and Applications*. Edited by P. Kahn. New York: Academic Press, Inc., pp. 307–48.

Insurance Europe. 2022. Available online: https://www.insuranceeurope.eu/statistics (accessed on 10 September 2022).

Jewell, William S. 1974a. Credible means are exact Bayesian for exponential families. *ASTIN Bulletin* 8: 77–90. [CrossRef]

Jewell, William S. 1974b. The Credible distribution. *ASTIN Bulletin* 7: 237–69. [CrossRef]

Kim, Joseph H.T., and Yongho Jeon. 2013. Credibility theory based on trimming. *Insurance: Mathematics and Economics* 53: 46–57. [CrossRef]

Kim, Minwoo, Himchan Jeong, and Dipak Dey. 2022. Approximation of Zero-Inflated Poisson Credibility Premium via Variational Bayes Approach. *Risks* 10: 54. [CrossRef]

Klugman, Stuart A., Harry Panjer, and Gordon E. Willmot. 2012. *Loss Models: From Data to Decisions*. New York: Wiley.

Korwar, Ramesh, M., and Myles Hollander. 1973. Contributions to the theory of Dirichlet processes. *Annals of Statistics* 1: 705–11. [CrossRef]

Kudryavtsev, Andrey 2009. Using quantile regression for rate-making. *Insurance: Mathematics and Economics* 45: 296–304.

Landsman, Zinoviy 1996. Sample quantiles and additive statistics: Information, sufficiency, estimation. *Journal of Statistical Planning and Inference* 52: 93–108.

Landsman, Zinoviy, M., and Udi E. Makov. 1998. Exponential dispersion models and credibility. *Scandinavian Actuarial Journal* 1: 89–96. [CrossRef]

Landsman, Zinoviy, M., and Udi E. Makov. 1999. Credibility evaluations for exponential dispersion families. *Insurance: Mathematics and Economics* 24: 33–9. [CrossRef]

Makov, Udi E., Adrian F. M. Smith, and Yu H. Liu. 1996. Bayesian methods in actuarial science. *Journal of the Royal Statistical Society Series D* 45: 503–15. [CrossRef]

Mukerjee, Hari. 1988. Monotone nonparametric regression. *Annals of Statistics* 16: 741–50. [CrossRef]

Pitselis, Georgios. 2009. Solvency Supervision based on a total balance sheet approach. *Journal of Computational and Applied Mathematics* 233: 83–96. [CrossRef]

Pitselis, Georgios. 2013. Quantile credibility models. *Insurance: Mathematics and Economics* 52: 477–89. [CrossRef]

Pitselis, Georgios. 2017. Risk measures in a quantile regression credibility framework with Fama/French data applications. *Insurance: Mathematics and Economics* 74: 122–34. [CrossRef]

Pitt, David, G. W. 2006. Regression quantile analysis of claim termination rates for income protection insurance. *Annals of Actuarial Science* 1: 345–57. [CrossRef]

Shively, Thomas S., Thomas W. Sager, and Stephen G. Walker. 2009. A Bayesian Approach to Non-Parametric Monotone Function Estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 71: 159–75. [CrossRef]

Tsai, Cary Chi-Liang, and Adelaide Di Wu. 2020. Bühlmann credibility-based approaches to modelling mortality rates for multiple populations. *North American Actuarial Journal* 24: 290–315. [CrossRef]

Tsai, Cary Chi-Liang, and Tzuling Lin. 2017. Incorporating the Bühlmann credibility into mortality models to improve forecasting performances. *Scandinavian Actuarial Journal* 5: 419–40. [CrossRef]

Tsai, Cary Chi-Liang, and Ying Zhang. 2019. A multi-dimensional Bühlmann credibility approach to modelling multi-population mortality rates. *Scandinavian Actuarial Journal* 5: 406–31. [CrossRef]

Wang, Wei, Limin Wen, Zhixin Yang, and Quan Yuan. 2021. Quantile Credibility Models with Common Effects. *Risks* 8: 100. [CrossRef]

Xacur, Oscar Alberto Quijano, and José Garrido. 2018. Bayesian credibility for GLMs. *Insurance: Mathematics and Economics* 83: 180–89. [CrossRef]

Yan, Yujie, and Kai-Sheng Song. 2022. A general optimal approach to Bühlmann credibility theory. *Insurance: Mathematics and Economics* 104: 262–82. [CrossRef]

Youn, Ahn Jae, Himchan Jeong, and Yang Lu. 2021. On the ordering of credibility factors. *Insurance: Mathematics and Economics* 101: 626–38. [CrossRef]

Zehnwirth, Benjamin. 1981. A Note on the Asymptotic Optimality of the Empirical Bayes Distribution Function. *Annals of Statistics* 9: 221–24. [CrossRef]

Zhang, Jin-Ting. 2004. A simple and efficient monotone smoother using smoothing splines. *Journal of Nonparametric Statistics* 16: 779–96. [CrossRef]