Comparison of Severity Estimators' Efficiency Based on Different Data Aggregation Levels

Pavel Zimmermann[1] | *University of Economics, Prague, Czech Republic*

Abstract

Estimates of the ultimate claim value occur in many actuarial models. Detailed data about each claim are available for estimation: each claim is at first booked at an initial value and processed over a random number of years, during which it is adjusted until closure. The ultimate value can be estimated based on observations of the ultimate value directly, which in this context, means using aggregated data. A more detailed, distribution-free estimator based on estimates of the initial claim value, the closure probability, and development factors is constructed in this article. It is proved that this estimator is asymptotically unbiased and an approximate analytical formula is derived for its variance. The efficiency of this estimator is compared to the efficiency of the simple arithmetic average of the ultimate claim value. Results are illustrated on an example and complemented with a simulation. The example results in significantly lower variability of the detailed estimator.

Keywords

Severity estimators, efficiency, data aggregation

JEL code

C13, G22

INTRODUCTION

In many actuarial tasks such as reserving or pricing, an estimate of the claim value is necessary. Usually, the focus is on the ultimate claim value that is the value at which the claim is closed. Prior to claim closure, the claim passes through the settlement process. Non-life insurers often collect detailed data about a variety of variables from the settlement process. In (Arjas, 1989), a mathematical description and a list of important variables is presented. Insurers, however, prefer traditional approaches and quite often aggregate their data prior to modelling. Three basic levels of aggregation can be distinguished: 1) Models based on aggregates from multiple claims. For example triangle schemes. 2) Models based on data from individual claims at its 'ultimate' state. 3) Models based on data collected throughout the settlement process, i.e. data containing whole claim 'trajectories' from its registration until its closure. Such detailed data are nowadays commonly available, however, rarely used in full detail. On the one hand, aggregation is usually connected with loss of information that can be used for efficient estimates. On the other hand,

[1] Faculty of Informatics and Statistics, W. Churchill Square 4, 130 67 Prague 3, Czech Republic. E-mail: zimmerp@vse.cz.

if models are based on more granular data, more parameters are usually involved and, hence, higher estimation error may appear. In this article we derive and compare properties of two ultimate claim value (claim severity) estimators based on level 2 and level 3 aggregation of the above mentioned typology. The term 'ultimate claim value' is preferred here to the term 'claim severity' to distinguish the value at claim closure from its value during the settlement process.

In general insurance, models based on triangles, i.e. level 1 aggregation, are presently most popular and have been studied by many authors. See, for example, (England and Verral, 2002) for an extensive list. Estimates based on less aggregated data (level 2 or even 3) are studied by far fewer authors. The research is often focused on stochastic processes underlying the claim occurrence and its development. The theoretical background of individual claim level models was originally set in (Norberg, 1993) and extended in (Norberg, 1999). The author considered a full time-continuous model of the settlement process using a non-homogeneous marked Poisson process. Another model based on the marked processes using simulation techniques was published in (Larsen, 2007). A potential bootstrap algorithm to asses the sampling error is also outlined. A simulation model based on individual claims was also developed in (Antonio and Plat, 2014). In (Herbst, 1999), the author applies survival analysis to derive an analytical formula for the estimate of incurred but not yet reported claims. Estimates based on fitting the multivariate skewed normal distribution were developed in (Pigeon, Antonio, Denuit, 2013) and (Pigeon, Antonio, Denuit, 2014). The topic of individual claim modeling was, from a practical point of view, also analyzed in several consultancy articles such as (Taylor, McGuire, Sullivan, 2008) or (Murphy and McLennan, 2006) in the context of large claims. A similar model was also assumed in (Drieskens et al., 2012). Simulation studies such as (Pigeon, Antonio, Denuit, 2014) or (Antonio and Plat, 2014) proved, on real examples, that higher efficiency of prediction of liabilities can be achieved using an individual claims approach.

Estimators of the ultimate claim value based on level 2 aggregated data appear in many actuarial models. They appear in a variety of simple frequency severity models, in collective risk models, and in more complex schemes such as (Herbst, 1999) or (Huang et al., 2015). Many of the individual claim models mentioned above, such as (Pigeon, Antonio, Denuit, 2013) or (Pigeon, Antonio, Denuit, 2014), are based on level 3 detailed data. To the author's knowledge, a comparison of efficiency of severity estimators based on these two levels of data aggregation has not been tackled previously. We assume that claims follow similar process specification as in (Murphy and McLennan, 2006) and (Drieskens et al., 2012). Each claim consists of a random initial registered value which is further adjusted by random number of random development factors that are independent but not identically distributed. See Formula (2). The first estimator considered is the simple arithmetic average of the ultimate claim value of all observed claims. This means it is calculated based only on data aggregated at level 2 of the above mentioned typology. This estimator does not consider the knowledge of the data from the whole settlement process, just the ultimate values. The second estimator (referred to as detailed) is based on the more granular level 3 data. It is constructed as the estimate of the initial value of the claim at reporting multiplied by a weighted average of estimates of development factors, from initial to a particular development year, where the weights are estimated probabilities of the claim being settled in a particular development year. See Formula (41). The estimate is constructed as an empirical counterpart of the variable defined in Formula (2). There are no specific requirements on the distribution so the estimator can be considered distribution-free.

The detailed estimator requires much more variables to be estimated (probabilities of claim settlement in each development year, development factors for each development year and the initial claim value). On the other hand, it uses more data than the simple average. The main task is to quantify to which extent such granularity contributes to the efficiency of the estimate of the ultimate claim value. The questions answered in this article are: Is it worth to construct more detailed estimate? What is the gain in efficiency?

An approximate formula of the variance of the detailed estimator is derived and compared to the variance of the simple average. Although it was not proved that the variance of the second estimate is always lower than in the case of the simple average, the presented realistic application suggests that the simple average is much less efficient under practical conditions. If the true process follows our assumptions, the detailed estimator shows, in the example case, approximately 55% lower variance.

The article is structured as follows: In the next section, the components of the ultimate claim value are introduced and the moments of the variables are derived. In Section 3, estimators of these components are constructed and their properties are derived. The main results are in Section 4 where the estimator of the ultimate claim value is constructed, its expected value and approximate formula for its variance are derived. It is further compared to the variance of simple arithmetic average. The formulas derived are applied on a realistic example in Section 5.

1 ULTIMATE CLAIM VALUE

We first define the ultimate claim value and some associated variables. Some relations and properties of these variables are stated. At the end, the first two moments of the ultimate claim value are derived.

1.1 Basic Notation and Assumptions

The following notation is used:

1. Maximum development year is denoted ω.
2. The initial value is denoted X_0. Its expected value and variance are denoted $\mathrm{E}(X_0)=\mu_0$ and $\mathrm{Var}(X_0) = \sigma_0^2$.
3. Vector $\mathbf{I} = (I_1,I_2,\ldots I_\omega)'$ is a vector indicating in which development year was the claim closed. For a claim closed in k-th development year $I_j = 1$ for $j = k$ and $I_j = 0$ otherwise. For simplicity no re-openings are assumed and therefore $\sum_{j=1}^{\omega} I_j = 1$. Expected value of I_j is denoted $\mathrm{E}(I_j)=p_j$.
4. In every development year, the claim value is updated by a random development factor. Vector of these incremental development factors is denoted $\mathbf{D} = (D_1,D_2, \ldots D_\omega)'$.
5. Vectors of cumulative development factors is denoted $\mathbf{F} = (F_1,F_2, \ldots ,F_\omega)'$ where F_j is defined as $F_j = \prod_{u=1}^{j} D_u$. The adjective 'cumulative' will often be omitted. The expected value andvariance of F_j are denoted $\mathrm{E}(F_j) = \mu_j$ and $\mathrm{Var}(F_j) = \sigma_j^2$.
6. Development factor from a period j to a period k is denoted:

$$ {}_jF_k = \prod_{u=j+1}^{k} D_u \,. \tag{1} $$

7. The ultimate claim value (the severity of the claim) is denoted X. It is defined as:

$$ X = X_0 \sum_{j=1}^{\omega} I_j\, F_j \,. \tag{2} $$

Variables X_0, $\mathbf{F}$ and $\mathbf{I}$ will be referred to as the components of the ultimate claim value. Further notation for corresponding estimators is presented in Section 3.

The following is assumed:

A 1. Maximum development year ω is known and deterministic.
A 2. Development factors D_j are mutually independent.
A 3. Vector of development factors $\mathbf{D}$ is independent on the vector of indicators $\mathbf{I}$.
A 4. Initial value X_0 is independent on $\mathbf{I}$ and $\mathbf{D}$.
A 5. The moments μ_0,σ_0^2,μ_j, and σ_j^2 are all finite.

All these assumptions are simplification of reality. Assumption 1 means that 'reasonably' high maximum number of development years have to be chosen in order to cover almost all reasonably observable cases on one hand and to have reasonable number of observations for the latest development years, on the other hand. Assumption 2 is also a simplifying assumption. Similar assumption is often assumed in aggregate models. This assumption allows derivation of analytic formulas for the variance of the estimators. Independence for given portfolio has to be tested prior to application of the estimators.

1.2 Properties of *F* and *I*

Assumption A3 means that technically we assume that claims develop even after the closure. Such development however does not influence the ultimate claim value. Due to Assumption A4 X_0 is also independent on **F**. Given the independence of D_j stated in Assumption A2, mean of j-th development factor is:

$$ {}_jF_k = \prod_{u=j+1}^{k} D_u \, . \tag{3}$$

The expected value of ${}_jF_k$ is then:

$$ \mathrm{E}\big({}_jF_k\big) = \prod_{u=j+1}^{k} \mathrm{E}\,(D_u) = \frac{\mu_k}{\mu_j}. \tag{4}$$

Independence of the incremental development factors D_j means that factors F_j and ${}_jF_k$ are also independent. For any $j < k$ the relation $F_k = F_j \cdot {}_jF_k$ holds. We can write for the covariance of F_j and F_k, $j < k$

$$ \begin{aligned} \mathrm{Cov}(F_j, F_k) &= \mathrm{Cov}\big(F_j, F_j \cdot {}_jF_k\big) = \mathrm{E}\big(F_jF_j\big)\mathrm{E}\big({}_jF_k\big) - \mathrm{E}\big(F_j\big)\mathrm{E}\big(F_j\big)\mathrm{E}\big({}_jF_k\big) \\ &= \mathrm{E}\big({}_jF_k\big)\mathrm{Var}(F_j) = \frac{\mu_k}{\mu_j}\sigma_j^2 . \end{aligned} \tag{5}$$

Vector I always contains only one element equal to one and $\omega - 1$ elements equal to zero. Vector I has multinomial distribution with parameters $v = 1$ and $p = (p_1, p_2, \dots, p_\omega)$ and therefore $\mathrm{E}(I_j) = p_j$,

$$ \mathrm{Var}(I_j) = p_j(1 - p_j). \tag{6}$$

and

$$ \mathrm{Cov}(I_j, I_k) = -p_j p_k. \tag{7}$$

A possible modification of the assumed model considering growth curves was published in (Pešta and Okhrin, 2014). The development factors are replaced by some parametric growth curves with number of parameters usually lower than ω. Lowering thenumber of parameters would lead to a more precise prediction, but it also requires backtesting of the growth curve's fit.

1.3 Properties of the Ultimate Claim Value

In this subsection, moments of the ultimate claim value are derived based on moments of the components. Firstly, it is necessary to derive the variance of the sum of products $I_j F_j$ which is further denoted ψ.

Lemma 1. Under Assumptions A1, A2, A3, and A5 variance of the sum of products $I_j F_j$ equals:

$$\begin{aligned}\psi &= \mathrm{Var}\left(\sum_{j=1}^{\omega} I_j\, F_j\right)\\ &= \sum_{j=1}^{\omega} p_j\, \sigma_j^2 + p_j(1-p_j)\mu_j^2 - 2\sum_{j<k} \mu_k\, p_j p_k \mu_j.\end{aligned} \tag{8}$$

Proof. The proof of this lemma is in the Appendix in subsection A2.

Theorem 1. Under Assumptions A1–A5 the expected value of the ultimate claim value is:

$$\mathrm{E}(X) = \mathrm{E}(X_0)\sum_{j=1}^{\omega} \mathrm{E}\,(I_j F_j) = \mu_0 \sum_{j=1}^{\omega} p_j\, \mu_j, \tag{9}$$

and the variance is:

$$\mathrm{Var}(X) = \sigma_0^2\psi + \sigma_0^2\left(\sum_{j=1}^{\omega} p_j\, \mu_j\right)^2 + \mu_0^2\psi. \tag{10}$$

Proof. The variance of the variable X may be written using Formula (A6) from the Appendix as:

$$\begin{aligned}\mathrm{Var}(X) = {} & \mathrm{Var}(X_0)\mathrm{Var}\left(\sum_{j=1}^{\omega} I_j\, F_j\right) + \mathrm{Var}(X_0)\mathrm{E}^2\left(\sum_{j=1}^{\omega} I_j\, F_j\right)\\ & + \mathrm{Var}\left(\sum_{j=1}^{\omega} I_j\, F_j\right)\mathrm{E}^2(X_0).\end{aligned} \tag{11}$$

After some algebraic operations this formula can be simplified to Formula (10).

2 ESTIMATORS OF THE COMPONENTS

As a first step to derive properties of the detailed estimator of the ultimate claim value, moments and covariances of estimators of the components and its multiples are derived.

2.1 Random Sample and its Notation

All estimators are denoted with a ‘hat’ sign. Observations are denoted by adding additional index u to the variable. Random sample of a fixed size of n claims is assumed. Random vector of numbers of claims closed in each development year $j = 1,2, \dots ,\omega$ is denoted $\mathbf{N} = (N_1,N_2, \dots , N_\omega)'$. Sum of the elements $\sum N_j = n$. It is automatically assumed that conditioning by a random event (e.g., in case of conditional expectation) means conditioning by an indicator of this random event.

Random vector $\boldsymbol{N}$ may be thought of as the sum of the n independent observations of the vector I and therefore has also multinomial distribution, this time with parameters $\nu = n$ and again $\mathbf{p} = (p_1,p_2, \dots ,p_\omega)'$. The following relations hold:

$$\mathrm{E}(N_j) = np_j, \tag{12}$$

$$\mathrm{Var}(N_j) = np_j(1 - p_j), \tag{13}$$

and

$$\mathrm{Cov}(N_j, N_k) = -np_j p_k = n\mathrm{Cov}(I_j, I_k). \tag{14}$$

Development factors F_j are only observed for claims which are closed in j or later. This means we have a random number of observations denoted $\bar{N}_j$ defined as:

$$\bar{N}_j = N_j + N_{j+1} + \cdots + N_\omega. \tag{15}$$

By definition the following implications hold for any $k > j$:

$$\bar{N}_j = 0 \Rightarrow \bar{N}_k = 0, \tag{16}$$

$$\bar{N}_k > 0 \Rightarrow \bar{N}_j > 0, \tag{17}$$

$$\bar{N}_j = 0 \Rightarrow N_j = 0. \tag{18}$$

In theory we have to consider a case for which all n claims are closed prior to the development year j and hence no observation of F_j is available for j, i.e. $\bar{N}_j = 0$. The probability of such event is denoted $\pi_j^{(0)}$ and equals:

$$\pi_j^{(0)} = \left(\sum_{k=1}^{j-1} p_k\right)^n. \tag{19}$$

In practical cases $\pi_j^{(0)}$ will be very close to 0 and also for all j if the sum in (19) is less then 1,

$$\lim_{n\to\infty} \pi_j^{(0)} = 0. \tag{20}$$

As $\bar{N}_j$ might be thought of as an outcome of n independent trials with the probability of being closed in development year j or later, equal to $\bar{p}_j = \sum_{k=j}^{\omega} p_k$, we may state that $\bar{N}_j$ has binomial distribution. First negative moment of $\bar{N}_j$ truncated at $\bar{N}_j = 0$, denoted as $\bar{n}_j^{-1}$ is defined as:

$$\bar{n}_j^{-1} = \mathrm{E}\left(\bar{N}_j^{-1} | \bar{N}_j > 0\right) = \frac{1}{1 - \pi_j^{(0)}} \sum_{u=1}^{n} \frac{1}{u}\binom{n}{u} \bar{p}_j^u \left(1 - \bar{p}_j\right)^{n-u}. \tag{21}$$

2.2 Estimator of Probability of Claim Closure

For multinomial distribution, the maximum likelihood estimate of p_j is the average of the observed indicators, i.e.

$$\hat{p}_j = \frac{\sum_{u=1}^{n} I_{j,u}}{n} = \frac{N_j}{n}. \tag{22}$$

Being the simple average, this estimator is unbiased:

$$\mathrm{E}(\hat{p}_j) = \mathrm{E}(I_j) = p_j, \tag{23}$$

and its variance is:

$$\mathrm{Var}(\hat{p}_j) = \frac{1}{n}\mathrm{Var}(I_j) = \frac{1}{n}p_j(1-p_j). \tag{24}$$

Estimates of the elements of vector $\boldsymbol{p}$ are not independent as, if in some development year more claims are closed, in other development years the number of closed claims will tend to decrease. The covariance of the estimates is, using (14),

$$\mathrm{Cov}\big(\hat{p}_j, \hat{p}_k\big) = \frac{1}{n^2}\mathrm{Cov}(N_j, N_k) = -\frac{1}{n}p_j p_k = \frac{1}{n}\mathrm{Cov}(I_j, I_k). \tag{25}$$

Lemma 2. The expected value of $\hat{p}_j$ conditional on $\bar{N}_j > 0$ equals:

$$\mathrm{E}\big(\hat{p}_j \mid \bar{N}_j > 0\big) = \frac{1}{\Big(1-\pi_j^{(0)}\Big)}p_j, \tag{26}$$

and covariance conditional on $\bar{N}_j > 0$ and $\bar{N}_k > 0$ equals:

$$\mathrm{Cov}\big(\hat{p}_j, \hat{p}_k \mid \bar{N}_j > 0, \bar{N}_k > 0\big) = \frac{\Big(1-n\pi_j^{(0)}\Big)}{n\Big(1-\pi_k^{(0)}\Big)}\mathrm{Cov}\big(\hat{p}_j, \hat{p}_k\big). \tag{27}$$

Proof. Proof of this lemma is presented in the Appendix in subsection 5.2.

2.3 Estimator of the Initial Value and Development Factors

The initial value X_0 can be observed for every loss in the sample. Simple average over all individual losses observed, denoted as $\hat{X}_0$, is considered as a predictor of X_0. Analogous approach may, however, be used for more advanced predictors, if necessary. The moments of this predictor are:

$$\mathrm{E}(\hat{X}_0) = \mathrm{E}(X_0) = \mu_0, \tag{28}$$

i.e. the predictor is unbiased, and

$$\mathrm{Var}(\hat{X}_0) = \frac{Var(X_0)}{n} = \frac{\sigma_0^2}{n}. \tag{29}$$

There are two sources of randomness in the estimate of the development ratio F_j:

1. The number of observations $\bar{N}_j$ defined in (15) available for the estimate of F_j which is the number of losses that were closed in j-th development year and later.
2. The actual observations of development factors $F_{j,u},\ u = 1, \dots, \bar{N}_j$.

The estimator assumed if $\bar{N}_j = 0$ is the average observed ratio, i.e.

$$\hat{\mu}_j = \frac{\sum_{u=1}^{\bar{N}_j} F_{j,u}}{\bar{N}_j}. \tag{30}$$

The number of observations $\bar{N}_j$ can also be 0 which slightly complicates the inference. We assume for the (theoretical) situation when $\bar{N}_j = 0$ that there is an estimate $\hat{\mu}_j$ available from some external source, for which:

$$\mathrm{E}(\hat{\mu}_j|\bar{N}_j = 0) = \alpha_j < \infty, \tag{31}$$

$$\mathrm{Var}(\hat{\mu}_j|\bar{N}_j = 0) = \beta_j < \infty. \tag{32}$$

As mentioned above $\pi_j^{(0)}$ will in practical tasks be very close to 0 and hence consideration of these external estimates is more formal than practical issue.

Lemma 3. Under Assumptions A3 and A5 the expected value of the estimator $\hat{\mu}_j$ equals:

$$\mathrm{E}(\hat{\mu}_j) = \pi_j^{(0)}\alpha_j + \left(1 - \pi_j^{(0)}\right)\mu_j, \tag{33}$$

and the variance equals:

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_j) = \pi_j^{(0)}\left(1 - \pi_j^{(0)}\right)\left(\mu_j - \alpha_j\right)^2 + \pi_j^{(0)}\beta_j \\ + \left(1 - \pi_j^{(0)}\right)\bar{n}_j^{-1}\sigma_j^2.\end{aligned} \tag{34}$$

Proof. Proof of this lemma is presented in the Appendix in subsection A2.

In the special case, where the external estimate is unbiased, i.e. $\alpha_j = \mu_j$, the estimate $\hat{\mu}_j$ is also unbiased. As the limit of $\pi_j^{(0)}$ is 0, the estimate is asymptotically unbiased (even if $\alpha_j \neq \mu_j$). Further more, in the special case where the external estimate is unbiased, the first term of (34) equals to 0 and the formula reduces to somewhat intuitive form where the variance of the estimator is weighted average of the variance in the case of the external estimate and the variance of the simple average. Due to limit (20), the estimator is consistent as the influence of the variance of the external estimator vanishes as n is increasing.

Lemma 4. Conditional covariance of the estimators $\hat{\mu}_j, \hat{\mu}_k$, $j < k$, conditioning on $\bar{N}_j > 0$, $\bar{N}_k > 0$ equals under Assumptions A2, A3, and A5:

$$\mathrm{Cov}(\hat{\mu}_j, \hat{\mu}_k|\bar{N}_j, \bar{N}_k > 0) = \bar{n}_j^{-1}\mathrm{Cov}(F_j, F_k). \tag{35}$$

Proof. Proof of this lemma is presented in the Appendix in subsection A2.

Lemma 5. The estimators $\hat{p}_j$ and $\hat{\mu}_k$ $j, k = 1, \dots, \omega$ are under Assumptions A3 and A5 conditioning on $\bar{N}_j > 0, \bar{N}_k > 0$ uncorrelated, i.e.

$$\mathrm{Cov}(\hat{p}_j, \hat{\mu}_k|\bar{N}_j, \bar{N}_k > 0) = 0. \tag{36}$$

Proof. Proof of this lemma is presented in the Appendix in subsection A2.

2.4 Properties of Products of the Estimators

In this section properties of the product of the estimators $\hat{p}_j$ and $\hat{\mu}_j$ are stated. First order approximation of the variance of the the product $\hat{p}_j\hat{\mu}_j$ is further denoted ϕ_j. First order approximation of the conditional covariance $\mathrm{Cov}\left(\hat{p}_j\hat{\mu}_j, \hat{p}_k\hat{\mu}_k|\bar{N}_j, \bar{N}_k > 0\right)$ is denoted as $\xi_{j,k}$.

Lemma 6. Under Assumptions A3 and A5 the expected value of the product $\hat{p}_j \hat{\mu}_j$ equals:

$$\mathrm{E}\left(\hat{p}_j \hat{\mu}_j\right) = p_j \mu_j, \tag{37}$$

and the approximate formula for the variance equals:

$$\phi_j = \mathrm{Var}\left(\hat{p}_j \hat{\mu}_j\right) \approx \frac{1}{n}\mathrm{Var}(I_j)\left(\mathrm{Var}(\hat{\mu}_j) + \mathrm{E}^2(\hat{\mu}_j)\right) + \mathrm{Var}(\hat{\mu}_j)\mathrm{E}^2(I_j), \tag{38}$$

where $\mathrm{E}(\hat{\mu}_j)$ is derived in (33) and $\mathrm{Var}(\hat{\mu}_j)$ is derived in (34). The product $\hat{p}_j \hat{\mu}_j$ is consistent.
Proof. Proof of this lemma is in the Appendix in Subsection 5.3.

For covariance of the product, $\mathrm{Cov}(\hat{p}_j \hat{\mu}_j, \hat{p}_k \hat{\mu}_k)$, $j < k$ is assumed. In order to derive the formula, it is necessary to cover different possible constellations of $\bar{N}_j$ and $\overline{N}_k$ being zero or grater than zero. It is necessary to cover situations $\left(\overline{N}_j > 0, \overline{N}_k > 0\right)$, $\left(\overline{N}_j > 0, \overline{N}_k = 0\right)$ and $\left(\overline{N}_j = 0, \overline{N}_k = 0\right)$. The combination $\left(\overline{N}_j = 0, \overline{N}_k > 0\right)$ cannot appear for $j < k$ due to implication (16). Using the Law of total covariance and approximate formula for the covariance of the product of random variables, the following lemma is proved:

Lemma 7. Under Assumptions A2, A3, and A5 the first order approximation of the conditional covariance $\mathrm{Cov}\left(\hat{p}_j \hat{\mu}_j, \hat{p}_k \hat{\mu}_k | \overline{N}_j, \overline{N}_k > 0\right)$ equals:

$$\begin{aligned}&\mathrm{Cov}\left(\hat{p}_j \hat{\mu}_j, \hat{p}_k \hat{\mu}_k | \overline{N}_j, \overline{N}_k > 0\right) \approx \\ &\quad \frac{p_j p_k n_j^{-1} \mathrm{Cov}\left(F_j F_k\right)}{\left(1 - \pi_j^{(0)}\right)\left(1 - \pi_k^{(0)}\right)} + \frac{\mu_j \mu_k \mathrm{Cov}\left(I_j I_k\right)\left(1 - n\pi_j^{(0)}\right)}{n\left(1 - \pi_k^{(0)}\right)},\end{aligned} \tag{39}$$

and the unconditional covariance equals approximately:

$$\begin{aligned}&\xi_{j,k} = \mathrm{Cov}\left(\hat{p}_j \hat{\mu}_j, \hat{p}_k \hat{\mu}_k\right) \approx \\ &\quad p_j p_k \mu_j \mu_k \left(\frac{n_j^{-1}\sigma_j^2}{\left(1 - \pi_j^{(0)}\right)\mu_j^2} - \frac{\left(1 - n\pi_j^{(0)}\right)}{n} + \left(1 - \pi_k^{(0)}\right)\pi_j^{(0)}\right).\end{aligned} \tag{40}$$

Proof. Proof of this lemma is in the Appendix in subsection A4.

On one hand, more precise approximations could be achieved by using stochastic expansions shown in (Hudecová and Pešta, 2013). On the other hand, more restrictive assumptions would be required.

3 ESTIMATOR OF ULTIMATE CLAIM VALUE

In this section we define the ultimate claim value estimator based on data collected over the whole claim settlement trajectory and derive its properties. The detailed estimator proposed is constructed as an empirical counterpart of Formula (2):

$$\hat{X} = \hat{X}_0 \sum_{j=1}^{\omega} \hat{p}_j \hat{\mu}_j. \tag{41}$$

3.1 Properties of the estimator

Mean and first order approximation of the variance are derived based on the properties of the estimators derived in Section 1.

Theorem 2. Estimator $\hat{X}$ is under Assumptions A1–A5 unbiased:

$$\mathrm{E}(\hat{X}) = \mu_0 \sum_{j=1}^{\omega} p_j\, \mu_j, \tag{42}$$

and the approximate variance is:

$$\mathrm{Var}(\hat{X}) \approx \frac{\sigma_0^2}{n}\phi + \frac{\sigma_0^2}{n}\left(\sum_{j=1}^{\omega} p_j\, \mu_j\right)^2 + \mu_0^2\phi, \tag{43}$$

where ϕ denotes approximate variance $\mathrm{Var}\left(\sum_{j=1}^{\omega} \hat{p}_j\, \hat{\mu}_j\right)$ and equals:

$$\mathrm{Var}\left(\sum_{j=1}^{\omega} \hat{p}_j\, \hat{\mu}_j\right) \approx \phi = \sum_{j=1}^{\omega} \phi_j + 2\sum_{j<k} \xi_{j,k}. \tag{44}$$

Proof. Using the assumption of independence of *Ij*, F_j and X_0 we can write for the mean:

$$\mathrm{E}(\hat{X}) = \mathrm{E}(\widehat{X_0}) \sum_{j=1}^{\omega} \mathrm{E}\left(\hat{p}_j \hat{\mu}_j\right) \tag{45}$$

Using (37) and (28) we get Formula (42).

The approximate variance (44) of the sum of $\hat{p}_j\, \hat{\mu}_j$ is calculated using the first order approximations ϕ_j derived in (38) and $\xi_{j,k}$ derived in (40) for the variances or covariances respectively. $\mathrm{Var}(\hat{X})$ is then calculated using Formula (A6) from the Appendix and inserting (28) and (29).

3.2 Asymptotic Relative Efficiency

Now, we compare the asymptotic efficiency of the detailed estimator of the ultimate claim value $\hat{X}$ defined in (41) with the simple average of the ultimate values of the claims observed. The simple average estimator is denoted $\tilde{X}$. The variance of $\tilde{X}$ is simply the variance of the ultimate claim value derived in (10) divided with the number of observed claims n. The asymptotic relative efficiency is then:

$$\lim_{n\to\infty} \frac{Var(\hat{X})}{Var(\tilde{X})} = \frac{\sigma_0^2\left(\sum_{j=1}^{\omega} p_j\, \mu_j\right)^2 + \mu_0^2 K}{Var(X)}, \tag{46}$$

where:

$$K = \lim_{n\to\infty} n\,\phi =$$

$$\sum_{j=1}^{\omega} \mu_j^2\, p_j(1-p_j) + \frac{\sigma_j^2}{\bar{p}_j} p_j^2 + 2\sum_{j<k} p_j\, p_k \mu_j \mu_k \left(\frac{\sigma_j^2}{\bar{p}_j \mu_j^2} - 1\right) \tag{47}$$

and $\mathrm{Var}(X)$ is defined in (10). The first term of the numerator of (46) is identical to the middle term of $\mathrm{Var}(X)$ hence $\hat{X}$ is more efficient than simple average in case the following relation holds:

$$\frac{\sigma_0^2}{\mu_0^2}+1>\frac{K}{\psi}. \tag{48}$$

Left hand side contains only characteristics of the initial loss X_0. We may conclude that higher relative efficiency of $\hat{X}$ may be expected in case of high relative variability of X_0. The right hand side is a fraction of complex sums of characteristics of both F_j and I_j for which some straightforward statements cannot be easily claimed, however, the following practical example suggests that the relative efficiency observed may be well below one (see Table 2).

4 PRACTICAL EXAMPLE

The properties of both estimators are illustrated on an example based on real data from motor third party liability bodily claims. Simulation study is presented to accompany the analytical results. All financial values are in EUR. The following 'true' values are assumed:

- Maximum $\omega = 9$ development years.
- Gamma distribution with $\mathrm{E}(X_0) = 6\,704$ and $\mathrm{Var}(X_0) = 125216729$ is assumed for the initial value X_0.
- Gamma distribution is also assumed for all development factors F_j, $j = 1,2, \dots , 9$. Moments of the variables are contained in Table 1.
- Sample size (number of claims) $n = 5\,000$.
- Although probabilities of having no observation in j-th development year $\pi_j^{(0)}$ are negligible as maximum equals $\pi_9^{(0)} = 2.1\times10^{-10}$, we set formally also moments of the external estimates. Intentionally, the values are selected to be of a very low quality. Both the mean and the variance (α and β) of the external estimate is set twice the true values. Note that in the case of one observation, the variance of the estimate would be equal to the variance of F_j, i.e. would be $\beta/2$.

Table 1 Parameters used as the true values in the simulation. $\mathbf{E}(F_j)$ and $\mathbf{Var}(F_j)$ are calculated based on Formulas (3) and (A6)

j	$\mathrm{E}(I_j) = p_j$	$\mathrm{E}(D_j)$	$\mathrm{Var}(D_j)$	$\mathrm{E}(F_j) = \mu_j$	$\mathrm{Var}(F_j) = \sigma_j$
1	0.236	1.27	1.73	1.27	1.73
2	0.198	1.08	0.98	1.38	5.29
3	0.138	0.97	0.31	1.34	7.27
4	0.216	0.84	0.13	1.13	6.25
5	0.124	0.79	0.12	0.89	4.78
6	0.050	0.82	0.09	0.73	3.70
7	0.019	0.83	0.08	0.60	2.87
8	0.014	0.76	0.08	0.46	1.93
9	0.004	0.84	0.08	0.38	1.52

Source: Own construction

The parameters shown in Table 1 imply the moments of the ultimate claim value X. The 'true' mean calculated using Formula (9) is $\mathrm{E}(X) = 7833$ and the 'true' variance calculated using Formula (10) is $\mathrm{Var}(X) = 975670440$. Based on these values and assumptions stated in Section 2.1, random portfolios of n claims were generated. For each such portfolio, both estimators of the ultimate claim value $\tilde{X}$ and $\hat{X}$

were calculated. The simple average of the ultimate claim value is calculated directly $\tilde{X}$ from the values observed. For the detailed estimator $\hat{X}$ all the estimators involved such as initial claim size, probabilities of claim closure for each development year, and development factors for each development years are calculated.

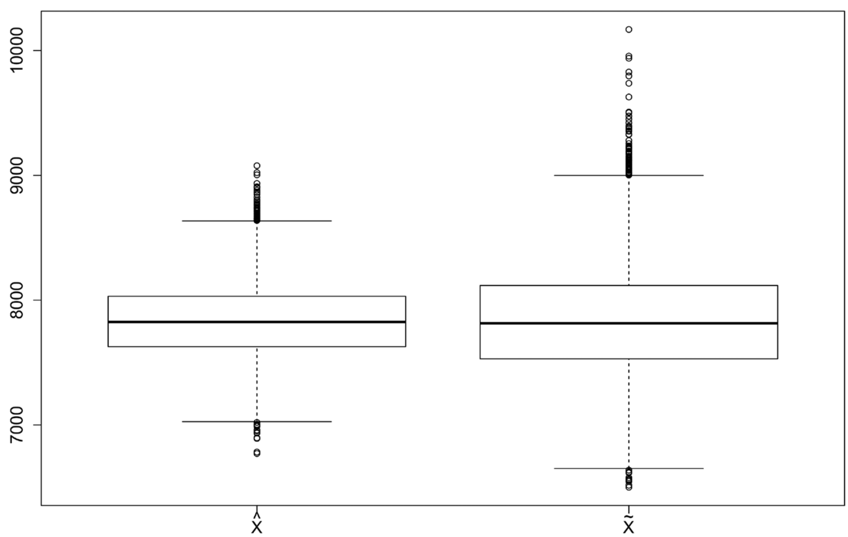
Portfolios were generated randomly 10 000 times and properties of both estimators $\tilde{X}$ and $\hat{X}$ were calculated from the simulations. Both analytic as well as simulated results are for the experiment presented in Table 2. Given the true process follows Assumptions A1–A5, the gain in efficiency using the detailed data to estimate the ultimate claim value is rather high, approximately 55%. The differences in distributions of the estimators are demonstrated in Figure 1. The box plots clearly show smaller variance of the estimator $\hat{X}$. The relative efficiency as a function of the sample size n is plotted in Figure 2.

Table 2 Comparison of the variance of the estimate $\hat{X}$ based on micro data and the simple average $\tilde{X}$

	Analytic result	Simulation
$\psi = \mathrm{Var}(\Sigma\, I_j\, F_j)$	4.729	.
$\phi = \mathrm{Var}(\Sigma\, \hat{p}_j\, \hat{\mu}_j)$	0.001	0.001
$\mathrm{Var}(\hat{X})$ (micro data)	89 293.000	88 638.000
$\mathrm{Var}(\tilde{X})$ (simple avg.)	195 108.000	192 674.000
$\mathrm{Var}(\hat{X})/\mathrm{Var}(\tilde{X})$	**0.458**	0.460
$\lim_{n\to\infty} \mathrm{Var}(\hat{X})/\mathrm{Var}(\tilde{X})$	0.457	.

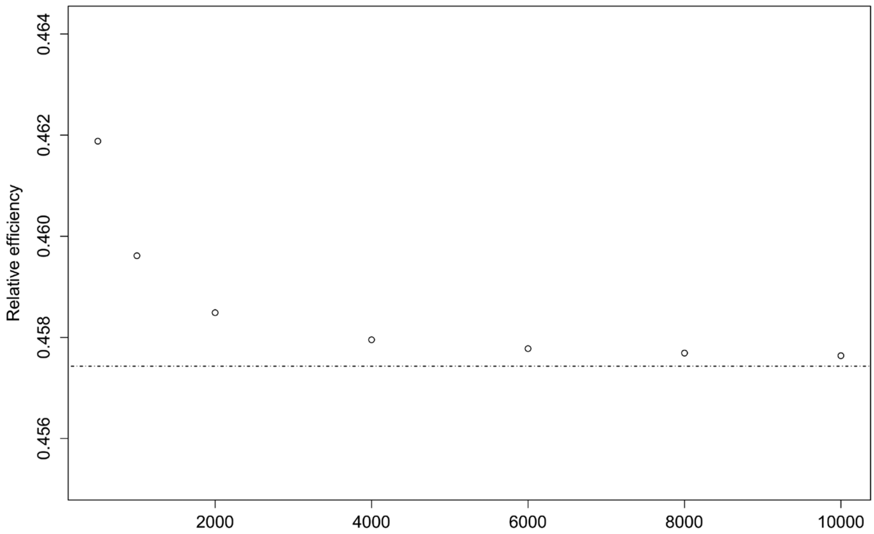
Source: Own construction

Figure 1 Boxplot of the simulated estimates for the simple average $\tilde{X}$ and estimate based on detailed settlement process data $\hat{X}$



Source: Own construction

Figure 2 Relative efficiency as a function of the sample size ***n*** and its asymptote



Source: Own construction

CONCLUSION

Models based on aggregated data are rather common in general insurance. Models based on individual data are generally more complex, requiring more calculations and, consequently, more computer time and human capacities. It is not obvious at first sight if this extra effort actually leads to higher efficiency as the number of parameters involved is usually also higher and hence higher estimation error occurs. In this article, we focused on one particular quantity – the ultimate claim value. A simple average, which would usually be the most common estimator employed, was compared to an estimator based on the more detailed data collected during the whole settlement process. It was shown that the estimator proposed is asymptotically unbiased. The approximate formula for its variance was derived and the difference in efficiency from the simple average was evaluated. The estimator is distribution-free. Although it was not proven that the efficiency of the more complex estimator for the assumed process would always be higher, it can be concluded that rather high gains in efficiency may be achieved. In the example presented, the increase of efficiency is almost 55%.

ACKNOWLEDGEMENTS

The author is grateful for comment received by an anonymous referee which helped to improve the quality of the article. The article was written with the support of the Grant Agency of the Czech Republic to the project no. P404/12/0883 and with the institutional support of the long term conceptual development of science and development of the Faculty of Informatics and Statistics of the University of Economics, Prague.

References

ANTONIO, K. AND PLAT R. Micro-Level Stochastic Loss Reserving for General Insurance. *Scandinavian Actuarial Journal*, 2014, 7, pp. 649–69.

ARJAS, E. The Claims Reserving Problem. In: Non-Life Insurance: Some Structural Ideas, *ASTIN Bulletin*, 1989, 19.
BOHRNSTEDT, G. W. AND GOLDBERGER, A. S. On the Exact Covariance of Products of Random Variables. *Journal of the American Statistical Association*, 1969, 64(328), pp. 1439–1442.
DRIESKENS, D., HENRY, M., WALHIN, J. F., WIELANDTS, J. Stochastic Projection for Large Individual Losses. *Scandinavian Actuarial Journal*, 2012(1), pp. 1–39.
ENGLAND, P. D. AND VERRAL, R. J. Stochastic Claims Reserving in General Insurance. *British Actuarial Journal*, 2002, 8(3), pp. 443–518.
HERBST, T. An Application of Randomly Truncated Data Models in Reserving IBNR Claims. *Insurance: Mathematics and Economics*, 1999, 25, pp. 123–131.
HUANG, J., QIU, C., WU, X., ZHOU, X. An Individual Loss Reserving Model with Independent Reporting and Settlement. *Insurance: Mathematics and Economics*, 2015, 64, pp. 232–245.
HUDECOVÁ, Š. AND PEŠTA, M. Modeling Dependencies in Claims Reserving with GEE. *Insurance: Mathematics and Economics*, 2013, 53(3), pp. 786–794.
KENDALL, M. AND STUART, A. *The Advanced Theory of Statistics. Vol. 1: Distribution Theory.* London: Griffin, 1977.
LARSEN, C. R. An Individual Claims Reserving Model. *ASTIN Bulletin International Actuarial Association*, 2007, 37(1), pp. 113–132.
MURPHY, K. AND MCLENNAN, A. A Method For Projecting Individual Large Claims. *Casualty Actuarial Society Forum*, 2006, pp. 205–36.
NORBERG, R. Prediction Of Outstanding Liabilities in Non-Life Insurance. *ASTIN Bulletin International Actuarial Association*, 1993, 23(1), pp. 95–115.
NORBERG, R. Prediction of Outstanding Liabilities II. Model Variations and Extensions. *ASTIN Bulletin International Actuarial Association*, 1999, 29(1), pp. 5–25.
PEŠTA, M. AND OKHRIN O. Conditional Least Squares and Copulae in Claims Reserving for a Single Line of Business. *Insurance: Mathematics and Economics*, 2014, 56, pp. 28–37.
PIGEON, M., ANTONIO, K., DENUIT, M. Individual Loss Reserving with the Multivariate Skew Normal Framework. *ASTIN Bulletin International Actuarial Association*, 2013, 43(3), pp. 399–428.
PIGEON, M., ANTONIO, K., DENUIT, M. Individual Loss Reserving Using Paid–Incurred Data. *Insurance: Mathematics and Economics*, 2014, 58, pp. 121–131.
TAYLOR, G., MCGUIRE, G., SULLIVAN, J. Individual Claim Loss Reserving Conditioned by Case Estimates. *Annals of Actuarial Science*, 2008, 3, pp. 215–56.

APPENDIX

A1 Some Auxiliary Formulas

A1.1 Variance of product of random variables

Lemma 8. The variance of the product of two finite random variables A and B is:

$$\begin{gathered} Var(AB) = \\ \mathrm{Cov}(A^2, B^2) + (\mathrm{Var}(A) + \mathrm{E}(A))^2(\mathrm{Var}(B) + \mathrm{E}(B))^2 - \\ (\mathrm{Cov}(A, B) - \mathrm{E}(A)\mathrm{E}(B))^2. \end{gathered} \tag{A1}$$

Proof. This formula is derived setting:

$$\mathrm{Var}(AB) = \mathrm{E}(A^2B^2) - \mathrm{E}^2(AB) \tag{A2}$$

and inserting:

$$\mathrm{E}(A^2B^2) = \mathrm{Cov}(A^2, B^2) + \mathrm{E}(A^2)\mathrm{E}(B^2), \tag{A3}$$

and

$$\mathrm{E}^2(AB) = (\mathrm{Cov}(A,B) + \mathrm{E}(A)\mathrm{E}(B))^2. \tag{A4}$$

A1.2 Variance of product of independent random variables

The variance of the product of two finite uncorrelated random variables A and B is:

$$\begin{aligned}\mathrm{Var}(AB) = {} & \mathrm{Cov}(A^2, B^2) + \mathrm{Var}(A)\mathrm{Var}(B) + \mathrm{Var}(A)\mathrm{E}^2(B) \\ & + \mathrm{Var}(B)\mathrm{E}^2(A).\end{aligned} \tag{A5}$$

This formula follows directly from inserting (Cov(A, B) = 0 in Formula (A1) in Appendix. If the variables A and B are independent, also Cov(A^2, B^2) = 0, and Formula (A5) reduces to:

$$\mathrm{Var}(AB) = \mathrm{Var}(A)\mathrm{Var}(B) + \mathrm{Var}(A)\mathrm{E}^2(B) + \mathrm{Var}(B)\mathrm{E}^2(A). \tag{A6}$$

A1.3 Approximate covariance of product of random variables

The approximate formula for the covariance of the product of random variables is stated in (Kendall and Stuart, 1977):

$$\begin{aligned}&\mathrm{Cov}(AB, UV) \approx \\ &\qquad \mathrm{E}(A)\mathrm{E}(U)\mathrm{Cov}(B,V) + \mathrm{E}(A)\mathrm{E}(V)\mathrm{Cov}(B,U) + \\ &\qquad \mathrm{E}(B)\mathrm{E}(U)\mathrm{Cov}(A,V) + \mathrm{E}(B)\mathrm{E}(V)\mathrm{Cov}(A,U).\end{aligned} \tag{A7}$$

The exact formula is stated in (Bohrnstedt and Goldberger, 1969).

A2 Moments of the components and estimators

A2.1 Proof of Lemma 1

Variance of the sum can be written as:

$$\mathrm{Var}\left(\sum_{j=1}^{\omega} I_j F_j\right) = \sum_{j=1}^{\omega} \mathrm{V\,ar}(I_j F_j) + 2\sum_{j<k} \mathrm{C\,ov}(I_j F_j, I_k F_k). \tag{A8}$$

The variance of the product V ar$(I_j F_j)$ contained in the first term can be derived using Formula (A6) and inserting (6) for Var(I_j).

$$\begin{aligned}\mathrm{Var}(I_j F_j) &= \mathrm{Var}(I_j)\mathrm{Var}(F_j) + \mathrm{Var}(I_j)\mathrm{E}^2(F_j) + \mathrm{Var}(F_j)\mathrm{E}^2(I_j) \\ &= p_j \sigma_j^2 + p_j(1 - p_j)\mu_j^2.\end{aligned} \tag{A9}$$

The covariance contained in the second term of (A6) is for $j < k$ using the notation (1):

$$\begin{aligned}&cov(I_j F_j, I_k F_k) = \\ &\qquad \mathrm{E}(I_j I_k)\mathrm{E}(F_j F_j)\mathrm{E}({}_j F_k) - \mathrm{E}(I_j)\mathrm{E}(F_j)\mathrm{E}(I_k)\mathrm{E}(F_j)\mathrm{E}({}_j F_k) = \\ &\qquad \mathrm{E}({}_j F_k)\left[\mathrm{E}(I_j I_k)E(F_j^2) - \mathrm{E}(I_j)\mathrm{E}(I_k)\mathrm{E}^2(F_j)\right].\end{aligned} \tag{A10}$$

As $I_j\, I_k$ is always equal to 0 due to the fact that if I_j=1, I_k must equal to zero and vice versa, $\mathrm{E}(I_j I_k) = 0$ and we can write using (4):

$$\mathrm{Cov}(I_j F_j, I_k F_k) = -\frac{\mu_k}{\mu_j} p_j p_k \mu_j^2 = -\mu_k p_j p_k \mu_j. \tag{A11}$$

If we now insert (A9) and (A11) into (A8), we get Formula (8).

A2.2 Proof of Lemma 2

Implication (18) implies also:

$$\bar{N}_j = 0 \Rightarrow \hat{p}_j = \frac{N_j}{n} = 0 \tag{A12}$$

and therefore $\mathrm{E}(\hat{p}_j | \bar{N}_j = 0) = 0$. Using the iterated expectations on $E(\hat{p}_j)$ gives Formula (26).

For $j < k$ joint probability that $\bar{N}_j > 0 = \bar{N}_k > 0$ equals to probability that $\bar{N}_k > 0$ as the combination $(\overline{N}_j = 0, \overline{N}_k > 0)$ cannot appear due to implication (16). Using the iterated expectations we can write:

$$\mathrm{E}(N_j N_k) = \mathrm{E}(N_j N_k | \bar{N}_j, \bar{N}_k > 0)\left(1 - \pi_k^{(0)}\right). \tag{A13}$$

Inserting this relation and Formula (26) and (14) in the covariance formula, we get:

$$\mathrm{Cov}(\hat{p}_j, \hat{p}_j | \bar{N}_j > 0, \bar{N}_k > 0) = \frac{Cov(N_j, N_k) - \pi_j^{(0)} E(N_j N_k)}{n^2 \left(1 - \pi_j^{(0)}\right)\left(1 - \pi_k^{(0)}\right)}. \tag{A14}$$

Inserting further:

$$\begin{aligned}\mathrm{E}(N_j N_k) &= \mathrm{Cov}(N_j, N_k) + \mathrm{E}(N_j)\mathrm{E}(N_k) \\ &= \mathrm{Cov}(N_j, N_k) - n^2 \mathrm{Cov}(I_j, I_k)\end{aligned} \tag{A15}$$

and using (7) yields Formula (25).

A2.3 Proof of Lemma 3

As F_j and I_j are independent we may also state that $\bar{N}_j$ and F_j are in the case $\bar{N}_j > 0$ independent. Furthermore, given the value of $\bar{N}_j$, observations $F_{j,u}$, $u = 1, \dots, \bar{N}_j$ is series of independent identically distributed variables. Using the iterated expectations the expected value of the estimator equals:

$$\mathrm{E}(\hat{\mu}_j) = \mathrm{E}_{\bar{N}_j > 0} \mathrm{E}(\hat{\mu}_j | \bar{N}_j) = \pi_j^{(0)} \mathrm{E}(\hat{\mu}_j | \bar{N}_j = 0) + \left(1 - \pi_j^{(0)}\right) \mathrm{E}(\hat{\mu}_j | \bar{N}_j > 0), \tag{A16}$$

where $\mathrm{E}_{\bar{N}_j > 0}$ denotes expectation over $\bar{N}_j$ conditional on $I(\bar{N}_j > 0)$. Inserting (31) and:

$$\mathrm{E}(\hat{\mu}_j | \bar{N}_j > 0) = \mu_j \tag{A17}$$

yields Formula (33).

The variance of the estimator $\hat{\mu}_j$ can be derived using the Law of total variance:

$$\mathrm{Var}(\hat{\mu}_j) = \mathrm{Var}_{\bar{N}_j>0}\big(\mathrm{E}(\hat{\mu}_j|\bar{N}_j)\big) + \mathrm{E}_{\bar{N}_j>0}\big(\mathrm{Var}(\hat{\mu}_j|\bar{N}_j)\big). \tag{A18}$$

We can write for first term:

$$\begin{aligned}\mathrm{Var}_{\bar{N}_j>0}\big(\mathrm{E}(\hat{\mu}_j|\bar{N}_j)\big) &= \pi_j^{(0)}\alpha_j^2 + \left(1-\pi_j^{(0)}\right)\mu_j^2 - \left(\pi_j^{(0)}\alpha_j + \left(1-\pi_j^{(0)}\right)\mu_j\right)^2 = \\ &= \pi_j^{(0)}\left(1-\pi_j^{(0)}\right)\left(\alpha_j-\mu_j\right)^2.\end{aligned} \tag{A19}$$

And for the second term:

$$\begin{aligned}&\mathrm{E}_{\bar{N}_j>0}\big(\mathrm{Var}(\hat{\mu}_j|\bar{N}_j)\big) = \\ &\qquad = \pi_j^{(0)}\mathrm{Var}\big(\hat{\mu}_j|\bar{N}_j=0\big) + \left(1-\pi_j^{(0)}\right)\mathrm{E}\big(\mathrm{Var}(\hat{\mu}_j|\bar{N}_j>0)\big).\end{aligned} \tag{A20}$$

Inserting (32) and:

$$\begin{aligned}&\mathrm{E}\left(\mathrm{Var}\left(\hat{\mu}_j\middle|\bar{N}_j>0\right)\right) = \\ &\qquad \mathrm{E}\left(\mathrm{Var}\left(\frac{\sum_{u=1}^{\bar{N}_j}F_{j,u}}{\bar{N}_j} \mid \bar{N}_j>0\right)\right) = \mathrm{E}\left(\frac{1}{\bar{N}_j} \mid \bar{N}_j>0\right)\sigma_j{}^2.\end{aligned} \tag{A21}$$

and using the notation (21) we get Formula (34).

A2.4 Proof of Lemma 4

Let us assume the covariance of the estimators $\mathrm{Cov}(\hat{\mu}_j,\hat{\mu}_k)$ for $j<k$ conditional on some fixed $\bar{N}_j, \bar{N}_k > 0$. The inequality $j<k$ implies $\bar{N}_j \geq \bar{N}_k$ as $\bar{N}_j$ contains all claims contained in $\bar{N}_k$.

$$\begin{aligned}&\mathrm{Cov}(\hat{\mu}_j,\hat{\mu}_k|\bar{N}_j,\bar{N}_k) = \\ &\quad \frac{1}{\bar{N}_j}\frac{1}{\bar{N}_k}\mathrm{E}\left(\sum_{u=1}^{\bar{N}_j}F_{j,u}\sum_{l=1}^{\bar{N}_k}F_{k,l}\right) - \frac{1}{\bar{N}_j}\frac{1}{\bar{N}_k}\mathrm{E}\left(\sum_{u=1}^{\bar{N}_j}F_{j,u}\right)\mathrm{E}\left(\sum_{l=1}^{\bar{N}_k}F_{k,l}\right).\end{aligned} \tag{A22}$$

We may write for the first term:

$$\mathrm{E}\left(\sum_{u=1}^{\bar{N}_j}F_{j,u}\sum_{l=1}^{\bar{N}_k}F_{k,l}\right) = \mathrm{E}\left(\sum_{u=1}^{\bar{N}_j}\sum_{l=1}^{\bar{N}_k}F_{j,u}\,F_{j,l}\;{}_jF_{k,l}\right). \tag{A23}$$

The variables $F_{j,l},\ F_{j,l}^{k}$ are independent. As we assume random sample, we may also state that the variables $F_{j,u}$ and $F_{j,l}$ are independent as long as $u \neq l$. The double sum contains $\min(\bar{N}_j, \bar{N}_k) = \bar{N}_k$ terms for which $u = l$ and $\bar{N}_j\bar{N}_k - \min(\bar{N}_j,\bar{N}_k) = \bar{N}_k$ terms for which $u \neq l$. Therefore we may write:

$$\mathrm{E}\left(\sum_{u=1}^{\bar{N}_j} F_{j,u} \sum_{l=1}^{\bar{N}_k} F_{k,l}\right) = \sum_{u=1}^{\bar{N}_k} \mathrm{E}\left(F_{j,u}^2\right)\mathrm{E}\left(\,_jF_{k,l}\right) + \sum\sum_{u \neq l} \mathrm{E}\left(F_{j,u}\right)\mathrm{E}\left(F_{j,l}\right)\mathrm{E}\left(\,_jF_{k,l}\right) =$$
$$= \bar{N}_k \mathrm{E}\left(F_j^2\right)\mathrm{E}\left(\,_jF_k\right) + \bar{N}_k\bar{N}_j \mathrm{E}\left(F_j\right)^2 \mathrm{E}\left(\,_jF_k\right) - \bar{N}_k \mathrm{E}\left(F_j\right)^2 \mathrm{E}\left(\,_jF_k\right) =$$
$$= \bar{N}_k \mathrm{E}\left(\,_jF_k\right)\mathrm{Var}(F_j) + \bar{N}_k\bar{N}_j \mathrm{E}\left(F_j\right)^2 \mathrm{E}\left(\,_jF_k\right). \tag{A24}$$

The second term of covariance (1) equals:

$$\mathrm{E}\left(\sum_{u=1}^{\bar{N}_j} F_{j,u}\right) \mathrm{E}\left(\sum_{l=1}^{\bar{N}_k} F_{k,l}\right) = \bar{N}_j\bar{N}_k \mathrm{E}\left(F_j\right)^2 \mathrm{E}\left(\,_jF_k\right). \tag{A25}$$

Therefore inserting (A24) and (A25) into (A22) and using (5), we get:

$$\mathrm{Cov}(\hat{\mu}_j, \hat{\mu}_k | \bar{N}_j, \bar{N}_k) = \frac{1}{\bar{N}_j} \frac{\mu_k}{\mu_j} \sigma_j^2 = \frac{1}{\bar{N}_j} \mathrm{Cov}(F_j, F_k). \tag{A26}$$

Taking the expectation over $\bar{N}_j$ conditional on $\bar{N}_j > 0$ and the notation (21), we get Formula (35).

A2.5 Proof of Lemma 5

For any j, $k = 1, \dots \omega$, the conditional expected value of the product $\hat{p}_j\, \hat{\mu}_j$ equals to:

$$E\left(\hat{p}_j\hat{\mu}_k \mid \bar{N}_j > 0, \bar{N}_k > 0\right) =$$
$$\mathrm{E}\left(\frac{N_j}{n} \frac{\sum_{u=1}^{\bar{N}_k} F_{k,u}}{\bar{N}_k} \mid \bar{N}_j > 0, \bar{N}_k > 0\right) =$$
$$\mathrm{E}_{N_j>0,\bar{N}_k>0}\left(\frac{N_j}{n} \mid \bar{N}_j > 0, \bar{N}_k > 0\right) \mathrm{E}\left(F_k \mid \bar{N}_j > 0, \bar{N}_k > 0\right) =$$
$$\mathrm{E}\left(\hat{p}_j | \bar{N}_j > 0, \bar{N}_k > 0\right) \mathrm{E}\left(\hat{\mu}_k | \bar{N}_j > 0, \bar{N}_k > 0\right) \tag{A27}$$

which proves (36).

A3 Moments of Product of the Estimators

A3.1 Proof of Lemma 6

The expected value of the product $\hat{p}_j\, \hat{\mu}_j$ can be expressed as:

$$\mathrm{E}\left(\hat{p}_j\hat{\mu}_j\right) = \pi_j^{(0)} \mathrm{E}\left(\hat{p}_j\hat{\mu}_j | \bar{N}_j = 0\right) + \left(1 - \pi_j^{(0)}\right) \mathrm{E}\left(\hat{p}_j\hat{\mu}_j | \bar{N}_j > 0\right). \tag{A28}$$

Implication (18) implies that the first term equals 0.

The expectation in the second term is:

$$\mathrm{E}\left(\hat{p}_j\hat{\mu}_j|\bar{N}_j>0\right)=\mathrm{EE}\left(\frac{N_j}{n}\frac{\sum_{u=1}^{\bar{N}_j}F_{j,u}}{\bar{N}_j}\mid \bar{N}_j>0\right)= \\ =\mathrm{E}\left(\frac{N_j}{n}\mid \bar{N}_j>0\right)\mathrm{E}\left(F_j\right). \tag{A29}$$

Using (26) and inserting (A29) into (A28), we get Formula (37).

The estimators $\hat{p}_j$ and $\hat{\mu}_j$ are generally dependent. The variance of the $\hat{p}_j\,\hat{\mu}_j$ can be derived using Formula (A5):

$$\mathrm{Var}\left(\hat{p}_j\hat{\mu}_j\right)=\mathrm{Cov}\left(\hat{p}_j^2,\hat{\mu}_j^2\right)+\mathrm{Var}(\hat{p}_j)\mathrm{Var}(\hat{\mu}_j)+\mathrm{Var}(\hat{p}_j)\mathrm{E}^2(\hat{\mu}_j) \\ +\mathrm{Var}(\hat{\mu}_j)\mathrm{E}^2(\hat{p}_j). \tag{A30}$$

The covariance of the squares of the estimators $\mathrm{Cov}\left(\hat{p}_j^2,\hat{\mu}_j^2\right)$ can not be generally derived. Its first order approximation equals to:

$$\mathrm{Cov}\left(\hat{p}_j^2,\hat{\mu}_j^2\right)\approx 4\mathrm{E}\left(\hat{\mu}_j\right)\mathrm{E}\left(\hat{p}_j\right)\mathrm{Cov}\left(\hat{p}_j,\hat{\mu}_j\right)=0. \tag{A31}$$

Lemma 5 states that the estimators $\hat{p}_j$ and $\hat{\mu}_j$ are uncorrelated. Formula for the variance of product of independent estimators (A6) is therefore approximately valid for uncorrelated variables.

$$\mathrm{Var}\left(\hat{p}_j\hat{\mu}_j\right)\approx\mathrm{Var}(\hat{p}_j)\mathrm{Var}(\hat{\mu}_j)+\mathrm{Var}(\hat{p}_j)\mathrm{E}^2(\hat{\mu}_j) \\ +\mathrm{Var}(\hat{\mu}_j)\mathrm{E}^2(\hat{p}_j). \tag{A32}$$

If we collect $\mathrm{Var}(\hat{p}_j)$ and insert the results (24) and (23) we get the Formula (38). The consistency is implied by the fact that $\hat{\mu}_j$ is a consistent estimator and hence both terms has limit zero.

A4 Covariance of Product of the Estimators

Situations of random sample listed in the first column of Table A1 need to be considered. An approximate formula for the covariance conditioning on the first situation, $\bar{N}_j>0$ and $\bar{N}_k>0$, is first derived.

Table A1 Conditional expected values of $\hat{p}_j\,\hat{\mu}_j$ and $\hat{p}_k\,\hat{\mu}_k$ conditioning on different constellations of $\bar{N}_j$ and $\bar{N}_k$

Condition	Probability	$\mathrm{E}(\hat{p}_j\,\hat{\mu}_j)$	$\mathrm{E}(\hat{p}_k\,\hat{\mu}_k)$	$\mathrm{E}(\hat{p}_j\,\hat{\mu}_j,\hat{p}_k\,\hat{\mu}_k)$
$\bar{N}_j>0,\bar{N}_k>0$	$1-\pi_k^{(0)}$	$p_j\,\mu_j$	$p_k\mu_k$	$p_j\mu_jp_k\mu_k$
$\bar{N}_j>0,\bar{N}_j=0$	$\pi_k^{(0)}-\pi_j^{(0)}$	$p_j\,\mu_j$	0	0
$\bar{N}_j=0,\bar{N}_k=0$	$\pi_j^{(0)}$	0	0	0

Source: Own construction

A4.1 Proof of Lemma 7

Using Formula (A7) and Lemma 5, we get:

$$\begin{aligned}&\mathrm{Cov}\left(\hat{p}_j\hat{\mu}_j,\hat{p}_k\hat{\mu}_k|\bar{N}_j,\bar{N}_k>0\right)\approx\\&\qquad \mathrm{E}\left(\hat{p}_j|\bar{N}_j,\bar{N}_k>0\right)\mathrm{E}\left(\hat{p}_k|\bar{N}_j,\bar{N}_k>0\right)\mathrm{Cov}\left(\hat{\mu}_j,\hat{\mu}_k|\bar{N}_j,\bar{N}_k>0\right)+\\&\qquad \mathrm{E}\left(\hat{\mu}_j|\bar{N}_j,\bar{N}_k>0\right)\mathrm{E}\left(\hat{\mu}_k|\bar{N}_j,\bar{N}_k>0\right)\mathrm{Cov}\left(\hat{p}_j,\hat{p}_k|\bar{N}_j,\bar{N}_k>0\right).\end{aligned} \tag{A33}$$

Inserting Formulas (A17), (26), (35) and (25) we get:

$$\begin{aligned}&\mathrm{Cov}\left(\hat{p}_j\hat{\mu}_j,\hat{p}_k\hat{\mu}_k|\bar{N}_j,\bar{N}_k>0\right)\approx\\&\qquad \frac{p_jp_kn_j^{-1}Cov\left(F_j,F_k\right)}{\left(1-\pi_j^{(0)}\right)\left(1-\pi_k^{(0)}\right)}+\frac{\mu_j\mu_kCov\left(I_j,I_k\right)\left(1-n\pi_j^{(0)}\right)}{n\left(1-\pi_k^{(0)}\right)},\end{aligned} \tag{A34}$$

which is Formula (39).

Second column of Table 3 contains probabilities of the three situations. The Law of total covariance is applied:

$$\begin{aligned}&\mathrm{Cov}\left(\hat{p}_j\hat{\mu}_j,\hat{p}_k\hat{\mu}_k\right)\\&\qquad =\ \mathrm{E}\left(\mathrm{Cov}(\hat{p}_j\hat{\mu}_j,\hat{p}_k\hat{\mu}_k|\bar{N}_j,\bar{N}_k)\right)\\&\qquad +\ \mathrm{Cov}\left(\mathrm{E}\left(\hat{p}_j\hat{\mu}_j|\bar{N}_j,\bar{N}_k\right),\mathrm{E}\left(\hat{p}_k\hat{\mu}_k|\bar{N}_j,\bar{N}_k\right)\right).\end{aligned} \tag{A35}$$

Due to implication (A12), multiples containing $\hat{p}_k$ conditioning on $\bar{N}_k=0$ all equal 0. Therefore conditional covariance in the first term in both cases where $\bar{N}_k=0$ equals 0 and we may write:

$$\begin{aligned}&\mathrm{E}\left(\mathrm{Cov}(\hat{p}_j\hat{\mu}_j,\hat{p}_k\hat{\mu}_k|\bar{N}_j,\bar{N}_k)\right)\\&\qquad =\mathrm{Cov}\left(\hat{p}_j\hat{\mu}_j,\hat{p}_k\hat{\mu}_k|\bar{N}_j>0,\bar{N}_k>0\right)\left(1-\pi_k^{(0)}\right).\end{aligned} \tag{A36}$$

Table A1 can also be used to calculate the second term of Formula (14). Based on this table we may write:

$$\begin{aligned}\mathrm{E}_{\bar{N}_j,\bar{N}_k}\mathrm{E}\left(\hat{p}_j\hat{\mu}_j\right)&=\left(1-\pi_k^{(0)}\right)p_j\mu_j+\left(\pi_k^{(0)}-\pi_j^{(0)}\right)p_j\mu_j\\&=p_j\mu_j\left(1-\pi_j^{(0)}\right),\end{aligned} \tag{A37}$$

$$\mathrm{E}_{\bar{N}_j,\bar{N}_k}\mathrm{E}\left(\hat{p}_k\hat{\mu}_k\right)=p_k\mu_k\left(1-\pi_k^{(0)}\right), \tag{A38}$$

$$\mathrm{E}_{\bar{N}_j,\bar{N}_k}\mathrm{E}\left(\hat{p}_j\hat{\mu}_j\hat{p}_k\hat{\mu}_k\right)=p_j\mu_jp_k\mu_k\left(1-\pi_k^{(0)}\right). \tag{A39}$$

The covariance $\mathrm{Cov}\left(\mathrm{E}\left(\hat{p}_j\hat{\mu}_j|\bar{N}_j,\bar{N}_k\right),\mathrm{E}\left(\hat{p}_k\hat{\mu}_k|\bar{N}_j,\bar{N}_k\right)\right)$ is then:

$$
\begin{aligned}
&\operatorname{Cov}\left(\mathrm{E}\left(\hat{p}_j\hat{\mu}_j|\bar{N}_j,\bar{N}_k\right),\mathrm{E}\left(\hat{p}_k\hat{\mu}_k|\bar{N}_j,\bar{N}_k\right)\right)= \\
&\qquad = \mathrm{E}_{\bar{N}_j,\bar{N}_k}\mathrm{E}\left(\hat{p}_j\hat{\mu}_j\hat{p}_k\hat{\mu}_k\right)-\mathrm{E}_{\bar{N}_j,\bar{N}_k}\mathrm{E}\left(\hat{p}_j\hat{\mu}_j\right)\mathrm{E}_{\bar{N}_j,\bar{N}_k}\mathrm{E}(\hat{p}_k\hat{\mu}_k) \\
&\qquad = p_j\mu_jp_k\mu_k\left(1-\pi_k^{(0)}\right)\pi_j^{(0)}.
\end{aligned}
\tag{A40}
$$

Inserting (5) and (7) in (39) we get Formula (40).