Estimation effects on stop-loss premiums under dependence[*]

Willem Albers[†] and Wilbert C.M. Kallenberg[‡]

Abstract

Even a small amount of dependence in large insurance portfolios can lead to huge errors in relevant risk measures, such as stop-loss premiums. This has been shown in a model where the majority consists of ordinary claims and a small fraction of special claims. The special claims are dependent in the sense that a whole group is exposed to damage. In this model, the parameters have to be estimated. The effect of the estimation step is studied here. The estimation error is dominated by the part of the parameters related to the special claims, because by their nature we do not have many observations of them. Although the estimation error in this way is restricted to a few parameters, it turns out that it may be quite substantial. Upper and lower confidence bounds are given for the stop-loss premium, thus protecting against the estimation effect.

1   Introduction. A well-known risk measure for large insurance portfolios is the so called stop-loss premium $E(S-a)^+ = E\{\max(0, S-a)\}$, where S denotes the sum of

[†]*Mailing Address:* Department of Applied Mathematics, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

[‡]*Mailing Address:* Department of Applied Mathematics, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. *E–mail:* `w.c.m.kallenberg@math.utwente.nl`.

the individual claims during a given reference period and a is called the retention. The classical model takes S as a sum of independent terms. This is often not realistic. On the other side of the spectrum, the assumption of comonotonicity produces astronomical effects due to its strong form of dependence. In practice, the dependence will be at a much lower level. However, it has been shown in Albers [1], Reijnen et al. [6] and Albers et al. [2] that even small dependencies can lead to huge errors in relevant risk measures, such as stop-loss premiums. Attributing on average a fraction of merely 1%-5% of the total claim amount to a common risk part turns out to already allow increases of stop-loss premiums by 200%-600%, when dealing with normally distributed claim size distributions, or even up to 50000% for more realistic skewed claim size distributions; see Albers [1] and Reijnen et al. [6]. Therefore, this small fraction of dependence should certainly not be ignored. On the other hand, complete comonotonicity seems to be too much. In fact, on the scale independent-comonotone the model with a (small) common risk part is still close to the independent end-point. For a more detailed discussion on this topic we refer to Reijnen et al. [6], pp. 247-249.

The previous results were obtained in a rather simple model. A more general and flexible model has been presented in Albers et al. [2]. The model makes a distinction between "ordinary" claims, where independence may be assumed, and a small fraction of "special" claims, where dependence appears in the form that a whole group is exposed to damage, due to a special cause (such as an epidemic, an accident, a hurricane etc.). The model is general in the sense that it allows groups of varying sizes, which moreover may overlap and on the other hand do not have to span the whole portfolio. It is flexible, in the sense that it does not require information which is and will remain unavailable from the data. For example, it sometimes may not be easy to identify those individuals who are exposed to a special cause, but did not file a claim. In fact, the model only needs the realized number of special claims.

As usual in stochastic models parameters appear which have to be estimated. Replacing the unknown model parameters by their estimated counterparts obtained from the data, will result in estimation errors. Just as with ignoring the dependence effect, it is too optimistic to act as if the estimation errors are negligible, unless we have a large number of observations. This topic, the effect of the estimation step, is exactly the issue which is addressed in the present paper.

In Section 2 the model is introduced. It turns out that the model is too complicated

to allow an exact evaluation of the estimation effects in such a way that transparent conclusions can be drawn. Therefore, we use some approximations. The accuracy of these approximations have been settled in Lukocius [4]. Two aspects play a role when considering the effect of the estimation step. Obviously, in the first place the accuracy of the estimators, but secondly, also the fluctuation of the stop-loss premium as function of the parameters. The set of parameters may be divided into two parts, those concerning the ordinary claims and those who are inserted in particular for the special causes. For the first part we have a lot of data and these parameters can be estimated very accurately. Due to their nature, special causes do not appear very often and hence estimation of the parameters linked up with the common risk part is much less accurate. As remarked before, their influence on the final outcome, even when a rather small part is due to a common risk, is quite large and hence estimation of the parameters connected with the special causes is the most important issue.

In Section 3 the needed structure of the observations to obtain estimators is given and the estimators based on them are derived. The fluctuations of the stop-loss premium are discussed in Section 4. The behavior of the estimators is the subject of Section 5. Asymptotic normality of the estimators, with respect to the expected total number of claims tending to infinity, is derived. The results of Sections 4 and 5 clearly show that the estimation effect is dominated by the part of the parameters related to the special causes. This is one of the main conclusions of the paper, implying that we only have to worry about that part of the estimation procedure, which simplifies matters. At the same time it is shown that the influence of these remaining estimators in general will be substantial. Hence, the estimation step cannot be ignored. That is the second main conclusion of the paper. In Section 6 it is shown how we can protect against the estimation error. Confidence bounds are derived for that purpose.

The paper is written in such a way that it can be extended in an easy way to other risk measures as for instance the value at risk, since in the theory no special properties of the stop-loss premium are used. Therefore, this part of the paper can be easily generalized with appropriate modifications when other risk measures are applied. Obviously, this does not hold for the numerical calculations, as presented in the tables and figures, where the particular form of the (accurate approximation of the) stop-loss premium, given in the Appendix, is explicitly used.

2   The model. The model is a so called collective model and consists of two parts, the ordinary claims and the special claims, where whole groups are involved. Examples are man and wife both insured in the same portfolio, carpoolers using a collective company insurance, catastrophes like hurricanes or floods hitting numerous insured at the same time. For more details we refer to Albers et al. [2], where the relation with the individual model is given and the impact of the model parameters is discussed, but see also Remark 2.1. Here we mainly restrict attention to a brief description of the model.

We use the following notations

$$N : \text{number of the ordinary claims,}$$
$$C_i : i^{th} \text{ claim size of the ordinary claims,}$$
$$H : \text{number of groups,}$$
$$G_k : k^{th} \text{ group size,}$$
$$D_{jk} : j^{th} \text{ claim size in } k^{th} \text{ group.}$$

The total sum of claims is given by

$$S = \sum_{i=1}^{N} C_i + \sum_{k=1}^{H} \sum_{j=1}^{G_k} D_{jk}. \tag{2.1}$$

Here we clearly see the two parts. The first sum concerns the ordinary claims, the second sum refers to the special claims. They occur groupswise, thus representing dependence in the total claim size. The occurrence of a special claim does not result in a single claim, but in a lot of claims together. So, in this part comonotonicity appears: the whole group has damage.

We assume that $C_1, C_2, \ldots, N, H, G_1, G_2, \ldots, D_{11}, D_{12}, \ldots$ are independent random variables. The name 'dependence model' does not come from the dependence of the claim sizes, but from the clustering of claims in time or space or whatever. As an illustrative example Lukocius [5] simulates a flu epidemic inside a large company, considering several departments as potential places of the mutual infection. The payments which people receive during their illness period can be considered as claims and the sum of all these claims then is modeled as S. The groups of a mutual infection (people which got infection from each other) are considered as groups of a common risk, producing the special claims, while claims from people which got the infection independently or suffer from other types of illness fall in the category of ordinary claims.

All the C_i and D_{jk} have the same distribution and also the G_k have a common distribution. Of course, it is of interest to consider the general case, where the distribution of the C's and that of the D's are different, but we really want to keep the number of additional parameters (above that of the independence model) limited. Contacts with practitioners indicate that otherwise the model becomes quickly too complicated for practical implementation. Hence, the present model may still be a simplification of reality, but it will be much less so than the (included) classical independence model (corresponding to $\varepsilon = 0$), because employing more parameters in principle guarantees a better fit to reality. (Recall the remark, attributed to Tukey: "All models are wrong, but some are more wrong than others.")

The supposed distributions of the random variables are as follows. Here P denotes the Poisson distribution and $\mu_G = EG$.

$$C_i, D_{jk} : \text{Gamma, inverse Gaussian or lognormal}$$
$$N : P(\lambda(1 - \varepsilon))$$
$$H : P\left(\frac{\varepsilon\lambda}{\mu_G}\right)$$
$$G_k : P(L) \text{ with } L : \text{Gamma or inverse Gaussian.}$$

The idea is that a fraction ε of λ, the total expected number of claims, is due to special causes. As ε typically will be (very) small, this clearly shows that the dependence part is really small in terms of the fraction of total expected number of claims. Nevertheless, this may lead to a huge total claim amount, with major consequences for the stop-loss premiums. Since special claims do not occur that often, a pretty high aggregation level is needed. The assumption, therefore, that all special claims lead to similar group sizes, seems rather awkward. Hence G_k, the number of realized claims in the kth group, follows an overdispersed Poisson distribution.

To obtain independence of $H, G_1, G_2, \ldots$, the following assumptions are sufficient: take $H, L_1, L_2, \ldots$ independent, let $G_1 | H = h, L_1 = l_1, L_2 = l_2, \ldots$, $G_2 | H = h, L_1 = l_1, L_2 = l_2, \ldots$ be independent and assume that the distribution of $G_k | H = h, L_1 = l_1, L_2 = l_2, \ldots$ depends only on l_k. Then it is easily seen that

$$P(H = h, G_1 = g_1, \ldots, G_h = g_h) = P(H = h)P(G_1 = g_1) \ldots P(G_h = g_h).$$

So, essentially, we first select an L_i, and given its outcome l_i we subsequently let G_i follow

a Poisson distribution with parameter l_i, thus allowing more variation in the group size than in case of a Poisson distribution with a fixed parameter.

Remark 2.1. As stated before, the dependence comes in, because a whole group of claims accumulates together. To get some additional feeling for the area the present model does cover we translate (2.1) to a corresponding individual model. Consider a large portfolio with m insured. The portfolio is divided into h groups, each of group size g. The jth insured in the ith group has, just like everybody else, a claim probability $(1-\varepsilon)\,q$ for an ordinary claim. Let $X_{ij} = 1$ denote that the jth insured in the ith group has an ordinary claim and otherwise $X_{ij} = 0$. Then the first term of the total claim amount S is given by

$$\sum_{i=1}^{h}\sum_{j=1}^{g}X_{ij}C_{ij}$$

with $P(X_{ij} = 1) = 1 - P(X_{ij} = 0) = (1-\varepsilon)\,q$ and C_{ij} the claim amount of an ordinary claim. This part of the model is in fact nothing else than the usual independence model. But in addition to it, the whole i^{th} group may be hit all together, due to a special cause, in which case each member of the group has damage. Here we clearly see the dependence: if one member of the group has damage due to a special cause, all the others of the group have a claim as well. Denoting $V_i = 1$ when the ith group has been hit and 0 otherwise, the second term of S is written as

$$\sum_{i=1}^{h}\sum_{j=1}^{g}V_iD_{ij}$$

with $P(V_i = 1) = 1 - P(V_i = 0) = \varepsilon q$ and D_{ij} the claim amount of the jth insured in the ith group in case of a special claim. Consider two members of the same group, say the jth and j^{*th} member of group 1. Their contribution to the total claim amount due to special causes is: V_1D_{1j} and $V_1D_{1j^*}$. Clearly, their claims V_1D_{1j} and $V_1D_{1j^*}$ are positively dependent, since they have V_1 in common. The number $N = \sum_{i=1}^{h}\sum_{j=1}^{g}X_{ij}$ of ordinary claims has a binomial distribution with parameters $m = hg$ and $(1-\varepsilon)\,q$ (for short: $Bin(m,(1-\varepsilon)\,q)$). Similarly, the number $H = \sum_{i=1}^{h}V_i$ of groups that have been hit is $Bin(h,\varepsilon q)$ with $h = m/g$. Writing $\lambda = mq$ and replacing $Bin(m,(1-\varepsilon)\,q)$ and $Bin(h,\varepsilon q)$ by $P\left(\lambda(1-\varepsilon)\right)$ and $P\left(\lambda\varepsilon/g\right)$, respectively, where we have used that $h\varepsilon q = m\varepsilon q/g = \lambda\varepsilon/g$, gives the collective model

$$S = \sum_{i=1}^{N}C_i + \sum_{k=1}^{H}\sum_{j=1}^{g}D_{jk}.$$

To allow groups of varying sizes, which moreover may overlap and on the other hand do not have to span the whole portfolio, g is replaced in (2.1) by the random variable G_k, the number of realized claims in the kth group. In this way a more general and flexible model is obtained. For more details we refer to Albers et al. [3].

The choices of the distributions of N, H and G is already discussed in Remark 2.1. Let us now concentrate on that of C and L and on the range of parameters for all the distributions. There are quite a few claim size distributions available in literature. We largely follow Reijnen et al. [6] and consider for the distribution of C the widely-used gamma, inverse-Gaussian and lognormal families. A prototype distribution for L is the gamma distribution. The simulation experiment in Lukocius [5] shows that indeed this distribution performs nicely. A second choice that proves to be quite suitable is the inverse Gaussian distribution. A third choice is the lognormal family. However, this turns out to be too extreme: huge cumulants result and the tails really seem too heavy to adequately model the mixing aspect of G.

Let the standard deviation of a random variable be denoted by σ and let $\gamma = \sigma/\mu$ be its coefficient of variation. The range of parameters that is of interest is given by

$$\lambda \geq 400, \varepsilon \leq 0.05, 5 \leq \mu_G = \mu_L \leq 20, 0.05 \leq \gamma_C \leq 2.5 \tag{2.2}$$

$$\gamma_L \leq 1.5 \text{ for } L : \text{Gamma}, \gamma_L \leq 2.5 \text{ for } L : \text{inverse Gaussian}.$$

Let us now discuss this choice briefly. For more detailed information about the choice of the range of parameters we refer to Albers et al. [2], Section 5. As written in the Introduction the model is too complicated to allow an exact evaluation of the estimation effects in such a way that transparent conclusions can be drawn. Therefore, we use some approximations. Obviously, these approximations should be sufficiently accurate. Therefore, a value of $\lambda \geq 400$ seems to be minimally required, because otherwise the events of interest will be encountered only very rarely. For instance, when $\lambda = 100$ and $\varepsilon = 0.02$, the expected number of special claims is merely 2. If we take $\mu_G = 10$, the expected number of such groups would only be 0.2. This really seems to be too small. Because a small fraction of dependence can create big problems already, we restrict attention to $\varepsilon \leq 0.05$. The lumpiness aspect is already present in the model, studied in Reijnen et al. [6]. So, we simply take the same range for $\mu_G = \mu_L$ as in that paper. The choice of the range of γ_C is based on the work of Reijnen et al. [6], where the skewness of C played an important role in the rule of thumb, which provides an accurate approximation. The extensive numerical

study in Chapter 3 of Lukocius [5] shows that when L follows a gamma distribution $\gamma_L \leq 1.5$ works fine and when L follows an inverse Gaussian distribution even $\gamma_L \leq 2.5$ is fine here.

Remark 2.2. The group size G has expectation μ_G, which in the range of parameters of interest varies between 5 and 20. Hence, G will as a rule be at least equal to 2. However, a value of G equal to 1 is possible. In that case we do not really have a group and it will not be recognized as such. Therefore, one might argue that we should restrict attention to distributions of G starting with 2. For most of the theory developed here this will cause no problem: the results continue to hold for general G. In view of that we will often give the results for this general setting, using the parametrization μ_G, γ_G instead of μ_L, γ_L (see also Remark 3.1). By definition of G the relation between the two forms of parametrization is simply given by

$$\mu_G = E\left(E(G|L)\right) = \mu_L,$$

$$\gamma_G^2 = var\left(\frac{G}{\mu_L}\right) = \mu_L^{-2}\{var(E(G|L)) + E(var(G|L))\} = \mu_L^{-2}\{var(L) + EL\} = \gamma_L^2 + \mu_L^{-1}.$$

On the other hand, in practice we do not have to worry about the restriction, because a value of G equal to 1 will occur only rarely and we may ignore it without making large mistakes.

Remark 2.3. Many other generalizations of the model than the one already mentioned (different distributions for the C's and D's) can be easily thought of. To give but a few examples: the D_{ij} can have different distributions for varying i, all kind of dependencies can exist between the random variables involved, e.g. positive correlation between the G_i and the D_{ij}, the distributions of N and H do not necessarily have to be Poisson etc. However, as explained before, we really want to keep the number of additional parameters (above that of the independence model) limited. Therefore, we do not work out this kind of generalizations in the present paper.

3   Observations and estimators. The basic data are for each individual the pairs (X_i, Y_i) with X_i the claim amount and Y_i the group code, 0 for the independent (ordinary) claim and $1, 2, \ldots$ for the various dependent claims (due to a common risk). From the

observed basic data (x_i, y_i) we can deduce

$$n : \text{the number of independent claims}$$

$$c_1, ..., c_n : \text{the claim amounts for the independent claims}$$

$$h : \text{the number of group codes for the dependent claims}$$

$$g_1, ..., g_h : \text{the group sizes}$$

$$d_{11}, ..., d_{g_h h} : \text{the claim amounts for the dependent claims.}$$

It will typically not be enough to have these data for one year, we usually will need data from several years $t = 1, ..., u$, say. The reason for that is the scarcity of special claims. To get reasonable estimates of ε, μ_G and γ_G we need data from an extended period. The estimators will be based on $N_t, C_{1t}, \ldots, C_{N_t t}, H_t, G_{1t}, \ldots, G_{H_t t}, D_{11t}, \ldots, D_{G_{H_t} H_t t}$, for $t = 1, ..., u$.

For the observed data $n_t, c_{1t}, ..., c_{n_t t}, h_t, g_{1t}, \ldots, g_{h_t t}, d_{11t}, \ldots, d_{g_{h_t} h_t t}$, with $t = 1, ..., u$, the likelihood equals

$$\prod_{t=1}^{u} \left[P\left(N = n_t\right) \left\{ \prod_{i=1}^{n_t} f_C(c_{it}) \right\} P\left(H = h_t\right) \left\{ \prod_{k=1}^{h_t} P(G = g_{kt}) \right\} \left\{ \prod_{k=1}^{h_t} \prod_{j=1}^{g_{kt}} f_C(d_{jkt}) \right\} \right].$$

Using

$$P\left(N = n_t\right) = \frac{\exp\left\{-\lambda(1 - \varepsilon)\right\} \left[\lambda(1 - \varepsilon)\right]^{n_t}}{n_t!},$$

$$P\left(H = h_t\right) = \frac{\exp\{-\varepsilon\lambda\mu_G^{-1}\}(\varepsilon\lambda\mu_G^{-1})^{h_t}}{h_t!},$$

the likelihood can be written as

$$\exp(-\theta)\theta^{n_{tot}+h_{tot}} p^{h_{tot}} (1 - p)^{n_{tot}} \times \left\{ \prod_{t=1}^{u} \prod_{k=1}^{h_t} P(G = g_{kt}) \right\}$$

$$\times \prod_{t=1}^{u} \left[\left\{ \prod_{i=1}^{n_t} f_C(c_{it}) \right\} \left\{ \prod_{k=1}^{h_t} \prod_{j=1}^{g_{kt}} f_C(d_{jkt}) \right\} \right] \times \prod_{t=1}^{u} \frac{1}{n_t! h_t! u^{n_t + h_t}}$$

with

$$\theta = \theta(\lambda, \varepsilon, \mu_G) = u\lambda(1 - \varepsilon + \varepsilon\mu_G^{-1}),$$

$$p = p(\varepsilon, \mu_G) = \frac{\varepsilon\mu_G^{-1}}{1 - \varepsilon + \varepsilon\mu_G^{-1}},$$

$$n_{tot} = \sum_{t=1}^{u} n_t \text{ and } h_{tot} = \sum_{t=1}^{u} h_t.$$

For short we will often write n and h instead of n_{tot} and h_{tot}. Maximizing the likelihood w.r.t. λ for given ε, μ_G gives $\widehat{\theta} = n + h$ and hence

$$\widehat{\lambda} = \widehat{\lambda}(\varepsilon, \mu_G) = \frac{n + h}{u(1 - \varepsilon + \varepsilon \mu_G^{-1})}. \tag{3.1}$$

Inserting it and noting that $\exp(-\widehat{\theta})\widehat{\theta}^{n+h}$ does not depend on (ε, μ_G), the likelihood is maximized w.r.t. ε for given μ_G by taking $\widehat{p} = h/(n+h)$ and hence

$$\widehat{\varepsilon} = \widehat{\varepsilon}(\mu_G) = \frac{h}{h + n\mu_G^{-1}}. \tag{3.2}$$

Inserting this and noting that $\widehat{p}^h(1 - \widehat{p})^n$ does not depend on μ_G, it is seen that we end up with the likelihood of the G's times the likelihood of the C's and D's. This means that we can proceed with estimating the parameters of the distribution of G using only the G-observations and, separately, estimating the parameters of the distribution of C using the C- and D-observations.

Taking for L the gamma-distribution, it follows that G has a negative binomial distribution. Although in general the number of observations from this negative binomial distribution, $\sum_{t=1}^{u} H_t$, will be not very large, the expectation of G is as a rule not small, say between 5 and 20. Under these circumstances, Saha and Paul [7] show that moment estimators are a good alternative to maximum likelihood estimators.

Both when L has a gamma distribution and when L has an inverse Gaussian distribution, G has a distribution with two parameters. Moment estimators do not depend on the parametrization. It is convenient to take as parametrization for G its expectation μ_G and its coefficient of variation γ_G (see also Remarks 2.2 and 3.1). The moment estimates of the expectation and coefficient of variation are

$$\widehat{\mu}_G = \overline{g} = \frac{1}{h} \sum_{t=1}^{u} \sum_{k=1}^{h_t} g_{kt},$$

$$\widehat{\gamma}_G = \frac{\sqrt{\overline{g^2} - \overline{g}^2}}{\overline{g}} \quad \text{with} \quad \overline{g^2} = \frac{1}{h} \sum_{t=1}^{u} \sum_{k=1}^{h_t} g_{kt}^2.$$

Inserting $\widehat{\mu}_G$ in $\widehat{\varepsilon}$, see (3.2), and writing $g_{tot} = \sum_{t=1}^{u} \sum_{k=1}^{h_t} g_{kt}$, yields

$$\widehat{\varepsilon} = \frac{h}{h + n\overline{g}^{-1}} = \frac{h\overline{g}}{h\overline{g} + n} = \frac{g_{tot}}{g_{tot} + n_{tot}}, \tag{3.3}$$

which as the observed fraction special claims indeed is the "natural" estimate of ε. Inserting $\widehat{\varepsilon} = h\overline{g}/(h\overline{g} + n), \widehat{\mu}_G = \overline{g}$ in $\widehat{\lambda}$, see (3.1), moreover gives

$$\widehat{\lambda} = \frac{h\overline{g} + n}{u} = \frac{g_{tot} + n_{tot}}{u},$$

which as the observed total number of claims divided by the number of years also is the "natural" estimate of λ. Writing

$$\overline{h} = \frac{\sum_{t=1}^{u} h_t}{u} = \frac{h}{u}, \overline{n} = \frac{\sum_{t=1}^{u} n_t}{u} = \frac{n}{u},$$

we may also write

$$\widehat{\lambda} = \overline{h}\overline{g} + \overline{n}.$$

For the estimation of the two parameters of the distribution of C we have many observations at our disposal. Hence here we clearly can use moment estimators as well. As parametrization we once more take the expectation μ_C and the coefficient of variation γ_C. This leads to

$$\widehat{\mu}_C = \overline{c + d} = \frac{\sum_{t=1}^{u} \sum_{i=1}^{n_t} c_{it} + \sum_{t=1}^{u} \sum_{k=1}^{h_t} \sum_{j=1}^{g_{kt}} d_{jkt}}{n_{tot} + g_{tot}},$$

$$\widehat{\gamma}_C = \frac{\sqrt{\overline{c^2 + d^2} - \overline{c + d}^2}}{\overline{c + d}} \text{ with } \overline{c^2 + d^2} = \frac{\sum_{t=1}^{u} \sum_{i=1}^{n_t} c_{it}^2 + \sum_{t=1}^{u} \sum_{k=1}^{h_t} \sum_{j=1}^{g_{kt}} d_{jkt}^2}{n_{tot} + g_{tot}}.$$

Summarizing: our estimators are

$$\widehat{\mu}_C = \overline{C + D}, \widehat{\gamma}_C = \frac{\sqrt{\overline{C^2 + D^2} - \overline{C + D}^2}}{\overline{C + D}},$$

$$\widehat{\mu}_G = \overline{G}, \widehat{\gamma}_G = \frac{\sqrt{\overline{G^2} - \overline{G}^2}}{\overline{G}},$$

$$\widehat{\varepsilon} = \frac{G_{tot}}{G_{tot} + N_{tot}}, \widehat{\lambda} = \frac{G_{tot} + N_{tot}}{u}.$$

Remark 3.1. Obviously, we can replace the parameters μ_G, γ_G and its estimators $\widehat{\mu}_G, \widehat{\gamma}_G$ by the parameters μ_L, γ_L and the corresponding estimators $\widehat{\mu}_L, \widehat{\gamma}_L$. Because $\mu_G = \mu_L$ and $\sigma_G^2 = \mu_L + \sigma_L^2$, implying that $\gamma_L = \mu_G^{-1} \sqrt{\sigma_G^2 - \mu_G}$, we get

$$\widehat{\mu}_L = \overline{G},$$

$$\widehat{\gamma}_L = \frac{\sqrt{\overline{G^2} - \overline{G}^2 - \overline{G}}}{\overline{G}}. \tag{3.4}$$

As long as γ_L is not equal to 0 or close to 0, there is no problem with $\widehat{\gamma}_L$. However, when $\gamma_L = 0$ (or close to 0) it may easily happen that $\overline{G^2} - \overline{G}^2 - \overline{G} < 0$ and hence a problem arises with application of (3.4). Note that the case $\gamma_L = 0$ corresponds to a fixed parameter of the Poisson distribution of G, a situation which we also want to take into account. In view of the problems with (3.4), indeed it is more convenient to use the parametrization μ_G, γ_G (see also Remark 2.2).

4 Behavior of $E(S - a)^+$.

The influence of the estimators on $E(S - a)^+$ depends on the behavior of $E(S - a)^+$ as a function of the parameters $\mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon, \lambda$ as well as on the accuracy of the estimators. For instance, if $E(S - a)^+$ is a flat function of the parameters $\mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon, \lambda$ and the estimators are accurate, the small changes due to estimation will have not much effect. So, these two points have to be considered: how is the fluctuation of $E(S - a)^+$ and how accurate are the estimators.

Obviously, the retention a is not just a given number, but will depend on $\mu_S = ES$ and $\sigma_S = \sqrt{var(S)}$: the larger μ_S and σ_S, the larger retention a will be chosen. Defining k by $a = \mu_S + k\sigma_S$, or

$$k = \frac{a - \mu_S}{\sigma_S},$$

we will assume that k is chosen in advance, determining the retention a in "standard units". That means that in our approach k does not depend on the parameters, while a does depend on the parameters $\mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon, \lambda$ through μ_S and σ_S.

In order to get insight into the fluctuation of

$$E(S - a)^+ = \sigma_S E \left(\frac{S - \mu_S}{\sigma_S} - k \right)^+$$

we have to simplify $\sigma_S E(\sigma_S^{-1}(S - \mu_S) - k)^+$ somewhat, because otherwise no conclusions can be drawn. We apply two simplifications. In the first place, $\sigma_S E(\sigma_S^{-1}(S - \mu_S) - k)^+$ is replaced by an approximation, which is simpler, but still sufficiently accurate in the region where we are interested in, see (2.2). This approximation, $SLPapp$, say, concerns the $Gamma - Inverse\ Gaussian\ (G - IG)$ approximation. For a short description of this approximation see the Appendix. That this approximation is indeed accurate in the region considered is shown in the extensive numerical study carried out in Lukocius [4].

Since even then the resulting function is rather complicated, we apply in addition a one step Taylor expansion on the approximation around the true value $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0)$

Table 1: Accuracy of approximation $SLPapp1$.

$(\mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon, \lambda, k)$	γ_L	$SLPapp$	$SLPapp1$	rel. error	abs. error
$(100000, 0.5, 10, 0.6, 0.05, 400, 0)$	0.51	1089184	1131613	0.04	42429
$(110000, 0.3, 12, 1, 0.04, 450, 1)$	0.96	339776	332509	0.02	7267
$(90000, 0.9, 18, 0.7, 0.05, 450, 2)$	0.66	64051	67969	0.06	3918
$(150000, 0.2, 10, 1.1, 0.02, 400, 3)$	1.05	13180	15544	0.18	2364
$(70000, 1, 20, 1, 0.03, 400, 0)$	0.97	957230	1009965	0.06	52735
$(120000, 0.1, 10, 0.6, 0.03, 450, 1)$	0.51	275809	272302	0.01	3508
$(200000, 0.8, 20, 0.5, 0.04, 400, 2)$	0.45	114474	115798	0.01	1324
$(150000, 0.5, 10, 1.1, 0.05, 400, 3)$	1.05	18330	18904	0.03	575

of the parameters. We call this function $SLPapp1$, which is given by

$$SLPapp1\left(\mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon, \lambda\right) = SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right) \qquad (4.1)$$
$$+ \left(\mu_C - \mu_{C0}\right)\frac{\partial}{\partial\mu_C}SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right)$$
$$+ \cdots + \left(\lambda - \lambda_0\right)\frac{\partial}{\partial\lambda}SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right).$$

Table 1 gives an impression of the accuracy of $SLPapp1$. Here C and L each have a (different) gamma-distribution and for the true value of the parameters we have the following representative choice: $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$, implying $\gamma_{L0} = 0.76$. We have $SLPapp(100000, 0.7, 15, 0.8, 0.03, 400) = 1164042, 292282, 56003,$ 9086 for $k = 0, 1, 2, 3$, respectively as our starting values. For convenience also the value of $\gamma_L = \sqrt{\gamma_G^2 - \mu_L^{-1}}$ is given.

This table indicates that the approximation by $SLPapp1$ is sufficiently accurate to proceed with. Note that

$$SLPapp1\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right) = SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right)$$

and hence Table 1 gives also interesting information on the error in

$$SLPapp(\widehat{\mu}_C, \widehat{\gamma}_C, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \widehat{\lambda}) - SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right)$$

due to replacing $SLPapp$ by $SLPapp1$. Hence, further on we concentrate on $SLPapp1$.

Table 2: Coefficients of $SLPapp1$ at $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$ for $k = 0, 1, 2, 3$.

k	$\frac{\partial}{\partial \mu_C} SLPapp$	$\frac{\partial}{\partial \gamma_C} SLPapp$	$\frac{\partial}{\partial \mu_G} SLPapp$	$\frac{\partial}{\partial \gamma_G} SLPapp$	$\frac{\partial}{\partial \varepsilon} SLPapp$	$\frac{\partial}{\partial \lambda} SLPapp$
0	11.6404	$3.8817\mu_{C0}$	$0.1047\mu_{C0}$	$1.3173\mu_{C0}$	$61.9452\mu_{C0}$	$0.0150\mu_{C0}$
1	2.9228	$0.7076\mu_{C0}$	$0.0632\mu_{C0}$	$1.0253\mu_{C0}$	$21.6362\mu_{C0}$	$0.0032\mu_{C0}$
2	0.5600	$-0.0210\mu_{C0}$	$0.0343\mu_{C0}$	$0.6532\mu_{C0}$	$6.4573\mu_{C0}$	$0.0003\mu_{C0}$
3	0.0909	$-0.0459\mu_{C0}$	$0.0116\mu_{C0}$	$0.2336\mu_{C0}$	$1.5790\mu_{C0}$	$-0.0001\mu_{C0}$

The fluctuation of $SLPapp1$ is determined by the coefficients

$$\frac{\partial}{\partial \mu_C} SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0), \ldots, \frac{\partial}{\partial \lambda} SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0).$$

To get some impression about the order of magnitude of these coefficients we have calculated them at $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$ (again for C and L each having a (different) gamma-distribution and for $k = 0, 1, 2, 3$). The results are given in Table 2.

In view of the very small coefficients and the fact that λ is large it seems better to write the term

$$(\lambda - \lambda_0) \frac{\partial}{\partial \lambda} SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0)$$

as

$$\frac{\lambda - \lambda_0}{\lambda_0} \lambda_0 \frac{\partial}{\partial \lambda} SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0).$$

Indeed, in the theory which will be presented next we perform asymptotics for $\lambda \to \infty$ and the appropriate quantity to consider then is $(\lambda - \lambda_0)/\lambda_0$, see Theorems 5.1 and 5.2. A similar remark applies to ε (giving rather large coefficients) and hence we will consider $(\varepsilon - \varepsilon_0)/\varepsilon_0$.

5   Behavior of the estimators.

We study the behavior of the estimators

$$\widehat{\mu}_C, \widehat{\gamma}_C, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \widehat{\lambda}.$$

These are functions of the vector

$$\left(\overline{C+D}, \overline{C^2+D^2}, \overline{G}, \overline{G^2}, \overline{H}, \overline{N} \right).$$

The following theorem gives the limiting distribution of this vector. The skewness of a random variable X is denoted by $\kappa_{3X} = \sigma^{-3} E(X - \mu)^3$ and its kurtosis by $\kappa_{4X} = \sigma^{-4} E(X - \mu)^4 - 3$.

Remark 5.1. Theorems 5.1, 5.2, 5.3 and 6.1 continue to hold for other distributions of C and G as well, provided that their fourth moments are finite.

Remark 5.2. In the following theorems we assume that $\lambda \to \infty$. That seems to be the natural way, because λ is the total expected number of claims, that is the expected number of observations. The other parameters are assumed to be fixed. At first sight it might seem curious that μ_C is called fixed, while in applications it is very large, for example 100000. However, this parameter is essentially a dummy parameter (although it should be estimated!), see also Section 6. We investigate the effect of the estimation in a relative sense, so to say in μ_C-units and therefore it can be considered as fixed.

Theorem 5.1. *Assume that $\lambda \to \infty$ and that $u, \mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon$ are fixed. Let*

$$X_{1\lambda} = \left\{ \frac{\overline{C+D}}{\mu_C} - 1 \right\} \frac{\sqrt{u\lambda}}{\gamma_C},$$

$$X_{2\lambda} = \left\{ \frac{\overline{C^2 + D^2}}{\mu_C^2} - \left(1 + \gamma_C^2\right) \right\} \frac{\sqrt{u\lambda}}{\gamma_C},$$

$$X_{3\lambda} = \left\{ \frac{\overline{G}}{\mu_G} - 1 \right\} \sqrt{\frac{\varepsilon u \lambda}{\mu_G}},$$

$$X_{4\lambda} = \left\{ \frac{\overline{G^2}}{\mu_G} - \mu_G \left(1 + \gamma_G^2\right) \right\} \sqrt{\frac{\varepsilon u \lambda}{\mu_G}}$$

$$X_{5\lambda} = \left\{ \frac{\overline{H} \mu_G}{\varepsilon \lambda} - 1 \right\} \sqrt{\frac{\varepsilon u \lambda}{\mu_G}},$$

$$X_{6\lambda} = \left\{ \frac{\overline{N}}{\lambda(1 - \varepsilon)} - 1 \right\} \sqrt{u\lambda(1 - \varepsilon)}.$$

Then, as $\lambda \to \infty$,

$$(X_{1\lambda}, X_{2\lambda}, X_{3\lambda}, X_{4\lambda}, X_{5\lambda}, X_{6\lambda}) \to (U_1, U_2, U_3, U_4, U_5, U_6)$$

with

$$(U_1, U_2) \sim N\left(0, 0, \begin{matrix} 1 & 2 + \gamma_C \kappa_{3C} \\ 2 + \gamma_C \kappa_{3C} & \gamma_C^2 (\kappa_{4C} + 2) + 4\gamma_C \kappa_{3C} + 4 \end{matrix}\right),$$

$$(U_3, U_4) \sim N\left(0, 0, \begin{matrix} \gamma_G^2 & \mu_G \gamma_G^2 (2 + \gamma_G \kappa_{3G}) \\ \mu_G \gamma_G^2 (2 + \gamma_G \kappa_{3G}) & \mu_G^2 \gamma_G^2 \{\gamma_G^2 (\kappa_{4G} + 2) + 4\gamma_G \kappa_{3G} + 4\} \end{matrix}\right),$$

$$U_5 \sim N(0, 1), U_6 \sim N(0, 1)$$

and $(U_1, U_2), (U_3, U_4), U_5, U_6$ *independent.*

Proof. The proof follows from standard asymptotic normality of random sums, see e.g. Corollary 1 in Teicher [8], and direct calculation of the involved moments. For instance,

$$cov\left(\frac{C}{\mu_C \gamma_C}, \frac{C^2}{\mu_C^2 \gamma_C}\right) = \frac{EC^3 - \mu_C EC^2}{\mu_C^3 \gamma_C^2} = \frac{\kappa_{3C} \gamma_C^3 \mu_C^3 + 3\mu_C^3(\gamma_C^2 + 1) - 2\mu_C^3 - \mu_C^3(\gamma_C^2 + 1)}{\mu_C^3 \gamma_C^2}$$

$$= \kappa_{3C} \gamma_C + 2.$$

The role of "n" is played by λ. The "inflation" of the covariance terms due to different limiting values of the (random) numbers of terms in the sums does not appear here, since the nonzero covariances have the same number of terms. For example, both $\overline{C + D}$ and $\overline{C^2 + D^2}$ have as number of terms $N_{tot} + G_{tot}$.

Obviously, N, having a $P(\lambda(1 - \varepsilon))$-distribution can be considered as a sum of λ independent random variables, each having a $P(1 - \varepsilon)$-distribution, and similarly for H. □

Remark 5.3. Theorem 5.1 can be applied to $G : P(L)$ with parametrization μ_L, γ_L (provided that the fourth moment of L is finite). We rewrite $X_{3\lambda}$ and $X_{4\lambda}$ as

$$X_{3\lambda} = \left\{\frac{\overline{G}}{\mu_L} - 1\right\} \sqrt{\frac{\varepsilon u \lambda}{\mu_L}},$$

$$X_{4\lambda} = \left\{\frac{\overline{G^2}}{\mu_L} - \mu_L (1 + \gamma_L^2) - 1\right\} \sqrt{\frac{\varepsilon u \lambda}{\mu_L}}$$

and use formulas like

$$\gamma_G^2 = \gamma_L^2 + \mu_L^{-1}.$$

We get asymptotic normality with

$$(U_3, U_4) \sim N\left(0, 0, \begin{matrix} \gamma_L^2 + \mu_L^{-1} & \begin{matrix} \mu_L \gamma_L^2 (2 + \gamma_L \kappa_{3L}) \\ + 2 + 3\gamma_L^2 + \mu_L^{-1} \end{matrix} \\ \begin{matrix} \mu_L \gamma_L^2 (2 + \gamma_L \kappa_{3L}) \\ + 2 + 3\gamma_L^2 + \mu_L^{-1} \end{matrix} & \begin{matrix} \mu_L^2 \gamma_L^2 \{\gamma_L^2 (\kappa_{4L} + 2) + 4\gamma_L \kappa_{3L} + 4\} \\ + 2\mu_L(3\gamma_L^3 \kappa_{3L} + 8\gamma_L^2 + 2) + 6 + 7\gamma_L^2 + \mu_L^{-1} \end{matrix} \end{matrix}\right).$$

Obviously, in $X_{5\lambda}$ we can replace μ_G by μ_L.

We are interested in $SLPapp1$, which is a linear combination of $\widehat{\mu}_C, ..., \widehat{\lambda}$. The next theorem gives the limiting distribution of such functions.

Theorem 5.2. *Assume that $\lambda \to \infty$ and that $u, \mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon$ are fixed. Let $c_1, ..., c_6$ be deterministic functions of $\mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon$ and λ. Define*

$$Z_1 = c_1 \frac{\widehat{\mu}_C - \mu_C}{\mu_C} + c_2 \left(\widehat{\gamma}_C - \gamma_C\right),$$

$$Z_2 = c_3 \left(\widehat{\mu}_G - \mu_G\right) \sqrt{\varepsilon} + c_4 \left(\widehat{\gamma}_G - \gamma_G\right) \sqrt{\varepsilon} + c_5 \left(\frac{\widehat{\varepsilon} - \varepsilon}{\varepsilon}\right) \sqrt{\varepsilon} + c_6 \frac{\widehat{\lambda} - \lambda}{\lambda}$$

Then, as $\lambda \to \infty$,

$$\left(\frac{Z_1}{\tau_1}, \frac{Z_2}{\tau_2}\right) \sqrt{u\lambda} \to (V_1, V_2) \tag{5.1}$$

with V_1, V_2 independent and $V_1, V_2 \sim N(0,1)$ with

$$\tau_1^2 = \gamma_C^2 \{c_1^2 + c_1 c_2 (\kappa_{3C} - 2\gamma_C) + c_2^2(\gamma_C^2 + \frac{1}{4}\kappa_{4C} + \frac{1}{2} - \gamma_C \kappa_{3C})\}$$

and

$$\begin{aligned}
\tau_2^2 = {} & c_3^2 \mu_G^3 \gamma_G^2 \\
& + c_4^2 \mu_G \gamma_G^2 \left(\gamma_G^2 - \gamma_G \kappa_{3G} + \frac{1}{4}\kappa_{4G} + \frac{1}{2}\right) \\
& + c_5^2 (1 - \varepsilon) \{\mu_G (1 - \varepsilon) (1 + \gamma_G^2) + \varepsilon\} \\
& + c_6^2 \{\mu_G \varepsilon (1 + \gamma_G^2) + 1 - \varepsilon\} \\
& + c_3 c_4 \mu_G^2 \gamma_G^2 (\kappa_{3G} - 2\gamma_G) \\
& + 2 c_3 c_5 (1 - \varepsilon) \mu_G^2 \gamma_G^2 \\
& + 2 c_3 c_6 \sqrt{\varepsilon} \mu_G^2 \gamma_G^2 \\
& + c_4 c_5 \mu_G \gamma_G^2 (1 - \varepsilon)(\kappa_{3G} - 2\gamma_G) \\
& + c_4 c_6 \mu_G \gamma_G^2 \sqrt{\varepsilon}(\kappa_{3G} - 2\gamma_G) \\
& + 2 c_5 c_6 \sqrt{\varepsilon} (1 - \varepsilon) \{\mu_G (1 + \gamma_G^2) - 1\}.
\end{aligned}$$

Proof. We have

$$\frac{\widehat{\mu}_C - \mu_C}{\mu_C} \sqrt{u\lambda} = \gamma_C X_{1\lambda} \tag{5.2}$$

and

$$(\widehat{\gamma}_C - \gamma_C)\sqrt{u\lambda} = \left[\frac{\sqrt{1 + \gamma_C^2 + \gamma_C X_{2\lambda}(u\lambda)^{-1/2} - \left\{1 + \gamma_C X_{1\lambda}(u\lambda)^{-1/2}\right\}^2}}{1 + \gamma_C X_{1\lambda}(u\lambda)^{-1/2}} - \gamma_C\right]\sqrt{u\lambda}.$$

It follows from Theorem 5.1 that

$$\sqrt{1 + \gamma_C^2 + \gamma_C X_{2\lambda}(u\lambda)^{-1/2} - \left\{1 + \gamma_C X_{1\lambda}(u\lambda)^{-1/2}\right\}^2}$$
$$= \sqrt{\gamma_C^2 + \gamma_C X_{2\lambda}(u\lambda)^{-1/2} - 2\gamma_C X_{1\lambda}(u\lambda)^{-1/2} + O_P(\lambda^{-1})}$$
$$= \gamma_C + \tfrac{1}{2}X_{2\lambda}(u\lambda)^{-1/2} - X_{1\lambda}(u\lambda)^{-1/2} + O_P(\lambda^{-1})$$

as $\lambda \to \infty$. Hence, we get

$$\frac{\sqrt{1 + \gamma_C^2 + \gamma_C X_{2\lambda}(u\lambda)^{-1/2} - \left\{1 + \gamma_C X_{1\lambda}(u\lambda)^{-1/2}\right\}^2}}{1 + \gamma_C X_{1\lambda}(u\lambda)^{-1/2}} - \gamma_C$$
$$= \frac{\gamma_C + \tfrac{1}{2}X_{2\lambda}(u\lambda)^{-1/2} - X_{1\lambda}(u\lambda)^{-1/2} + O_P(\lambda^{-1}) - \gamma_C - \gamma_C^2 X_{1\lambda}(u\lambda)^{-1/2}}{1 + \gamma_C X_{1\lambda}(u\lambda)^{-1/2}}$$
$$= \tfrac{1}{2}X_{2\lambda}(u\lambda)^{-1/2} - X_{1\lambda}(u\lambda)^{-1/2} - \gamma_C^2 X_{1\lambda}(u\lambda)^{-1/2} + O_P(\lambda^{-1})$$

and thus

$$(\widehat{\gamma}_C - \gamma_C)\sqrt{u\lambda} = \tfrac{1}{2}X_{2\lambda} - \left(1 + \gamma_C^2\right)X_{1\lambda} + O_P\left(\lambda^{-1/2}\right) \tag{5.3}$$

as $\lambda \to \infty$.

Next we show that $|c_1/\tau_1|$ and $|c_2/\tau_1|$ are bounded above as functions of λ. Let U_1 and U_2 as given in Theorem 5.1 and $X = \gamma_C U_1, Y = \tfrac{1}{2}U_2 - (1 + \gamma_C^2)U_1$. Then we have $\tau_1^2 = var(c_1 X + c_2 Y)$ and hence $\tau_1^2 \geq \{1 - \rho^2(X,Y)\}\max\{var(c_1 X), var(c_2 Y)\}$. Because X and Y do not depend on λ and therefore also $var(X), var(Y)$ and $\rho(X,Y)$ do not depend on λ, the boundedness of $|c_1/\tau_1|$ and $|c_2/\tau_1|$ immediately follows.

Combination of (5.2) and (5.3) and application of Theorem 5.1 gives

$$\left\{c_1 \frac{\widehat{\mu}_C - \mu_C}{\mu_C} + c_2(\widehat{\gamma}_C - \gamma_C)\right\}\tau_1^{-1}\sqrt{u\lambda} \to V_1.$$

We have

$$(\widehat{\mu}_G - \mu_G)\sqrt{\varepsilon u\lambda} = \mu_G^{3/2}X_{3\lambda} \tag{5.4}$$

and

$$\overline{G^2} - \overline{G}^2 = \mu_G^2 \left(1 + \gamma_G^2\right) + \mu_G^{3/2} \left(\varepsilon u\lambda\right)^{-1/2} X_{4\lambda} - \left(\mu_G + \mu_G^{3/2} \left(\varepsilon u\lambda\right)^{-1/2} X_{3\lambda}\right)^2.$$

It follows from Theorem 5.1 that

$$\sqrt{\overline{G^2} - \overline{G}^2}$$
$$= \sqrt{\mu_G^2 \gamma_G^2 + \mu_G^{3/2} \left(\varepsilon u\lambda\right)^{-1/2} X_{4\lambda} - 2\mu_G^{5/2} \left(\varepsilon u\lambda\right)^{-1/2} X_{3\lambda} + O_P\left(\lambda^{-1}\right)}$$
$$= \mu_G \gamma_G + \tfrac{1}{2}\mu_G^{1/2} \gamma_G^{-1} \left(\varepsilon u\lambda\right)^{-1/2} X_{4\lambda} - \mu_G^{3/2} \gamma_G^{-1} \left(\varepsilon u\lambda\right)^{-1/2} X_{3\lambda} + O_P\left(\lambda^{-1}\right)$$

and thus

$$\left(\widehat{\gamma}_G - \gamma_G\right) \sqrt{\varepsilon u\lambda} = \left\{ \frac{\sqrt{\overline{G^2} - \overline{G}^2}}{\mu_G + \mu_G^{3/2} \left(\varepsilon u\lambda\right)^{-1/2} X_{3\lambda}} - \gamma_G \right\} \sqrt{\varepsilon u\lambda}$$
$$= \tfrac{1}{2}\mu_G^{-1/2} \gamma_G^{-1} \{ X_{4\lambda} - 2\mu_G \left(1 + \gamma_G^2\right) X_{3\lambda} \} + O_P\left(\lambda^{-1/2}\right)$$

as $\lambda \to \infty$. It is seen, cf. e.g. (3.3) that

$$\widehat{\varepsilon} = \frac{\overline{HG}}{\overline{HG} + \overline{N}}.$$

By Theorem 5.1 we get

$$\overline{HG} = \varepsilon\lambda \left\{ 1 + \mu_G^{1/2} \left(\varepsilon u\lambda\right)^{-1/2} X_{5\lambda} \right\} \left\{ 1 + \mu_G^{1/2} \left(\varepsilon u\lambda\right)^{-1/2} X_{3\lambda} \right\} \qquad (5.5)$$
$$= \varepsilon\lambda \left\{ 1 + \mu_G^{1/2} \left(\varepsilon u\lambda\right)^{-1/2} \left(X_{5\lambda} + X_{3\lambda}\right) + O_P\left(\lambda^{-1}\right) \right\}.$$

Together with

$$\overline{N} = \lambda\left(1 - \varepsilon\right) \left[1 + \{(1 - \varepsilon)u\lambda\}^{-1/2} X_{6\lambda} \right] \qquad (5.6)$$

this leads to

$$\widehat{\varepsilon} = \frac{\varepsilon \left\{ 1 + \mu_G^{1/2} \left(\varepsilon u\lambda\right)^{-1/2} \left(X_{5\lambda} + X_{3\lambda}\right) + O_P\left(\lambda^{-1}\right) \right\}}{\varepsilon \left\{ 1 + \mu_G^{1/2} \left(\varepsilon u\lambda\right)^{-1/2} \left(X_{5\lambda} + X_{3\lambda}\right) + O_P\left(\lambda^{-1}\right) \right\} + (1 - \varepsilon) \left[1 + \{(1 - \varepsilon)u\lambda\}^{-1/2} X_{6\lambda} \right]}$$
$$= \varepsilon + (1 - \varepsilon)\varepsilon^{1/2}\mu_G^{1/2} \left(u\lambda\right)^{-1/2} \left(X_{5\lambda} + X_{3\lambda}\right) - \varepsilon\left(1 - \varepsilon\right)^{1/2} \left(u\lambda\right)^{-1/2} X_{6\lambda} + O_P\left(\lambda^{-1}\right)$$

and thus

$$\left(\frac{\widehat{\varepsilon} - \varepsilon}{\varepsilon} \right) \sqrt{\varepsilon u\lambda} = (1 - \varepsilon)\mu_G^{1/2} \left(X_{5\lambda} + X_{3\lambda}\right) - \varepsilon^{1/2}\left(1 - \varepsilon\right)^{1/2} X_{6\lambda} + O_P\left(\lambda^{-1/2}\right) \qquad (5.7)$$

as $\lambda \to \infty$. Finally, we have

$$\frac{\widehat{\lambda} - \lambda}{\lambda} \sqrt{u\lambda} = \left(\frac{\overline{HG} + \overline{N}}{\lambda} - 1\right) \sqrt{u\lambda}.$$

In view of (5.5) and (5.6) we get

$$\frac{\overline{HG} + \overline{N}}{\lambda}$$
$$= \varepsilon \left\{ 1 + \mu_G^{1/2} (\varepsilon u\lambda)^{-1/2} (X_{5\lambda} + X_{3\lambda}) + O_P\left(\lambda^{-1}\right) \right\} + (1 - \varepsilon) \left[1 + \{(1 - \varepsilon) u\lambda\}^{-1/2} X_{6\lambda} \right]$$
$$= 1 + (\varepsilon \mu_G)^{1/2} (u\lambda)^{-1/2} (X_{5\lambda} + X_{3\lambda}) + (1 - \varepsilon)^{1/2} (u\lambda)^{-1/2} X_{6\lambda} + O_P\left(\lambda^{-1}\right)$$

and hence

$$\frac{\widehat{\lambda} - \lambda}{\lambda} \sqrt{u\lambda} = (\varepsilon \mu_G)^{1/2} (X_{5\lambda} + X_{3\lambda}) + (1 - \varepsilon)^{1/2} X_{6\lambda} + O_P\left(\lambda^{-1/2}\right) \tag{5.8}$$

as $\lambda \to \infty$.

By a similar argument as before it follows that $|c_3/\tau_2|, ..., |c_6/\tau_2|$ are bounded above as functions of λ. Note that τ_2^2 is of the form $var\left(c_3 X_1 + ... + c_6 X_4\right)$ and thus $\tau_2^2 \geq (1 - \rho_i^{*2}) var\left(c_{2+i} X_i\right)$, $i = 1, ..., 4$, where ρ_i^{*2} is the multiple correlation coefficient of X_i with the other X_j's, which does not depend on λ.

Combination of (5.4)–(5.8) and application of Theorem 5.1 gives

$$\left\{ c_3 \left(\widehat{\mu}_G - \mu_G\right) \sqrt{\varepsilon} + c_4 \left(\widehat{\gamma}_G - \gamma_G\right) \sqrt{\varepsilon} + c_5 \left(\frac{\widehat{\varepsilon} - \varepsilon}{\varepsilon}\right) \sqrt{\varepsilon} + c_6 \frac{\widehat{\lambda} - \lambda}{\lambda} \right\} \tau_2^{-1} \sqrt{u\lambda} \to V_2.$$

The asymptotic independence of $c_1 \left(\widehat{\mu}_C - \mu_C\right)/\mu_C + c_2 \left(\widehat{\gamma}_C - \gamma_C\right)$ and $c_3 \left(\widehat{\mu}_G - \mu_G\right) \sqrt{\varepsilon}$ $+ c_4 \left(\widehat{\gamma}_G - \gamma_G\right) \sqrt{\varepsilon} + c_5 \left(\widehat{\varepsilon} - \varepsilon\right)/\sqrt{\varepsilon} + c_6 \left(\widehat{\lambda} - \lambda\right)/\lambda$ completes the proof. $\square$

Next we apply Theorem 5.2 in order to get an idea of the impact of the estimators on $SLPapp1$. The error due to estimation, divided by μ_{C0}, equals, cf. (4.1),

$$\mu_{C0}^{-1}\{SLPapp1\left(\widehat{\mu}_C, \widehat{\gamma}_C, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \widehat{\lambda}\right) - SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right)\}$$
$$= \left(\frac{\widehat{\mu}_C - \mu_{C0}}{\mu_{C0}}\right) \frac{\partial}{\partial \mu_C} SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right)$$
$$+ (\widehat{\gamma}_C - \gamma_C) \mu_{C0}^{-1} \frac{\partial}{\partial \gamma_C} SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right)$$
$$+ \cdots + \left(\frac{\widehat{\lambda} - \lambda_0}{\lambda_0}\right) \lambda_0 \mu_{C0}^{-1} \frac{\partial}{\partial \lambda} SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right).$$

The asymptotic distribution of $\mu_{C0}^{-1}\{SLPapp1\left(\widehat{\mu}_C, \widehat{\gamma}_C, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \widehat{\lambda}\right) - SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0},$ $\gamma_{G0}, \varepsilon_0, \lambda_0)\}\sqrt{u\lambda_0}$ is obtained by application of Theorem 5.2 with

$$c_1 = \frac{\partial}{\partial\mu_C}SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right),$$

$$c_2 = \mu_{C0}^{-1}\frac{\partial}{\partial\gamma_C}SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right),$$

$$c_3 = \varepsilon_0^{-1/2}\mu_{C0}^{-1}\frac{\partial}{\partial\mu_G}SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right),$$

$$c_4 = \varepsilon_0^{-1/2}\mu_{C0}^{-1}\frac{\partial}{\partial\gamma_G}SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right),$$

$$c_5 = \varepsilon_0^{1/2}\mu_{C0}^{-1}\frac{\partial}{\partial\varepsilon}SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right),$$

$$c_6 = \lambda_0\mu_{C0}^{-1}\frac{\partial}{\partial\lambda}SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0).$$

The result is a normal distribution with expectation 0 and variance $\tau_1^2 + \tau_2^2$. Hence, this variance gives an idea of the error due to estimation.

As an example we calculate τ_1^2 and τ_2^2 for $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8,$ $0.03, 400)$ and $k = (a - \mu_S)/\sigma_S = 1$ (again with C and L each having a (different) gamma-distribution). Note that $SLPapp(100000, 0.7, 15, 0.8, 0.03, 400) = 292282$ in that case (see Section 4). The values of $c_1, \ldots, c_6$ are easily obtained from Table 2. We get (for the gamma distribution it holds that $\kappa_3 = 2\gamma$ and hence the coefficient of $c_1 c_2$ equals 0)

$$c_1^2\gamma_{C0}^2 = 4.19$$
$$c_1 c_2 \gamma_{C0}^2(\kappa_{3C0} - 2\gamma_{C0}) = 0$$
$$c_2^2\gamma_{C0}^2(\gamma_{C0}^2 + \frac{1}{4}\kappa_{4C0} + \frac{1}{2} - \gamma_{C0}\kappa_{3C0}) = 0.18$$

and therefore

$$\tau_1^2 = 4.37.$$

Using that L has a gamma-distribution, direct calculation (see also (A7)) gives $\kappa_{3G} = 2\gamma_G - \mu_G^{-1}\gamma_G^{-1}$ and $\kappa_{4G} = 6\gamma_G^2 - 6\mu_G^{-1} + \mu_G^{-2}\gamma_G^{-2}$. We obtain

$$c_3^2 \mu_{G0}^3 \gamma_{G0}^2 = 287.43 \tag{5.9}$$

$$\frac{1}{2} c_4^2 \left(\mu_{G0} \gamma_{G0}^4 - \gamma_{G0}^2 + \frac{1}{2} \mu_{G0}^{-1} + \mu_{G0} \gamma_{G0}^2 \right) = 265.21$$

$$c_5^2 (1 - \varepsilon_0) \left\{ \mu_{G0} (1 - \varepsilon_0) \left(1 + \gamma_{G0}^2 \right) + \varepsilon_0 \right\} = 325.47$$

$$c_6^2 \left\{ \mu_{G0} \varepsilon_0 \left(1 + \gamma_{G0}^2 \right) + 1 - \varepsilon_0 \right\} = 2.84$$

$$- c_3 c_4 \mu_{G0} \gamma_{G0} = -25.91$$

$$2 c_3 c_5 (1 - \varepsilon_0) \mu_{G0}^2 \gamma_{G0}^2 = 381.90$$

$$2 c_3 c_6 \sqrt{\varepsilon_0} \mu_{G0}^2 \gamma_{G0}^2 = 23.46$$

$$- c_4 c_5 \gamma_{G0} (1 - \varepsilon_0) = -17.21$$

$$- c_4 c_6 \gamma_{G0} \sqrt{\varepsilon_0} = -1.06$$

$$2 c_5 c_6 \sqrt{\varepsilon_0} (1 - \varepsilon_0) \left\{ \mu_{G0} \left(1 + \gamma_{G0}^2 \right) - 1 \right\} = 38.31$$

and hence

$$\tau_2^2 = 1280.43.$$

This example is really illuminating. It is clearly seen that the contribution of estimating μ_C and γ_C is not very high: τ_1^2 is much smaller than τ_2^2. The reason is that we have a lot of observations for estimating μ_C and γ_C. Typical values for u and λ are values like 7 and 400, respectively. That means about 2800 observations to estimate the parameters of the common distribution of the C_i and D_{jk}. Due to this large number of observations, these estimators are very accurate. Similarly, estimating λ gives also a not very high contribution to the variance $\tau_1^2 + \tau_2^2$. That is seen from the various terms contributing to τ_2^2. The terms in which estimating λ is involved, that is the terms where c_6 appears, are much smaller than the other terms.

    This leads to the following

Conclusion. The estimation error is dominated by the estimation of the parameters related to the common risk, that is by estimating μ_G, γ_G and ε. Therefore, the parameters of the distribution of the C_i and D_{jk}, μ_C and γ_C, and also λ can in fact considered to be known.

Remark 5.4. Theorem 5.2 can be applied to $G : P(L)$ with parametrization μ_L, γ_L (provided that the fourth moment of L is finite), replacing $c_3 (\widehat{\mu}_G - \mu_G) \sqrt{\varepsilon} + c_4 (\widehat{\gamma}_G - \gamma_G) \sqrt{\varepsilon}$

by
$$c_3 \left(\widehat{\mu}_L - \mu_L\right) \sqrt{\varepsilon} + c_4 \left(\widehat{\gamma}_L - \gamma_L\right) \sqrt{\varepsilon} \text{ and } \tau_2^2 \text{ by}$$

$$
\begin{aligned}
\tau_2^2 = {} & c_3^2 \left(\mu_L^3 \gamma_L^2 + \mu_L^2\right) \\
& + c_4^2 \left\{ \mu_L \gamma_L^2 \left(\gamma_L^2 - \gamma_L \kappa_{3L} + \frac{1}{4}\kappa_{4L} + \frac{1}{2}\right) - \gamma_L^2 + \gamma_L \kappa_{3L} + 1 + \frac{1}{2}\mu_L^{-1}\left(1 + \gamma_L^{-2}\right) \right\} \\
& + c_5^2 \left(1 - \varepsilon\right) \left\{ \mu_L \left(1 - \varepsilon\right)\left(1 + \gamma_L^2\right) + 1 \right\} \\
& + c_6^2 \left\{ \mu_L \varepsilon \left(1 + \gamma_L^2\right) + 1 \right\} \\
& + c_3 c_4 \mu_L^2 \gamma_L^2 (\kappa_{3L} - 2\gamma_L) \\
& + 2 c_3 c_5 \left(1 - \varepsilon\right) \left(\mu_L^2 \gamma_L^2 + \mu_L\right) \\
& + 2 c_3 c_6 \sqrt{\varepsilon} \left(\mu_L^2 \gamma_L^2 + \mu_L\right) \\
& + c_4 c_5 \mu_L \gamma_L^2 (1 - \varepsilon)(\kappa_{3L} - 2\gamma_L) \\
& + c_4 c_6 \mu_L \gamma_L^2 \sqrt{\varepsilon}(\kappa_{3L} - 2\gamma_L) \\
& + 2 c_5 c_6 \mu_L \sqrt{\varepsilon} \left(1 - \varepsilon\right) \left(1 + \gamma_L^2\right).
\end{aligned}
$$

So, in the sequel μ_C, γ_C and λ are assumed to be known, while $\varepsilon, \mu_G = \mu_L$ and γ_G or γ_L are estimated by

$$\widehat{\varepsilon} = \frac{G_{tot}}{G_{tot} + N_{tot}},$$

$$\widehat{\mu}_G = \widehat{\mu}_L = \overline{G},$$

$$\widehat{\gamma}_G = \frac{\sqrt{\overline{G^2} - \overline{G}^2}}{\overline{G}}, \widehat{\gamma}_L = \frac{\sqrt{\overline{G^2} - \overline{G}^2 - \overline{G}}}{\overline{G}}$$

with

$$G_{tot} = \sum_{t=1}^{u} \sum_{k=1}^{h_t} G_{kt}, H_{tot} = \sum_{t=1}^{u} H_t, N_{tot} = \sum_{t=1}^{u} N_t,$$

$$\overline{G} = \frac{1}{H_{tot}} \sum_{t=1}^{u} \sum_{k=1}^{h_t} G_{kt}, \overline{G^2} = \frac{1}{H_{tot}} \sum_{t=1}^{u} \sum_{k=1}^{h_t} G_{kt}^2.$$

Writing $SLP(\widehat{\mu}_C, \widehat{\gamma}_C, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \widehat{\lambda})$ for the estimator of the stop-loss premium $E(S - a)^+$, we now have the following result.

Theorem 5.3. *Let* $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0)$ *be the true value of the parameters. Then*

$$SLP(\widehat{\mu}_C, \widehat{\gamma}_C, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \widehat{\lambda}) \approx SLPapp1(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0)$$

and

$$\mu_{C0}^{-1}\{SLPapp1(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0) - SLPapp1(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0)\} \tau^{-1}\sqrt{u\lambda_0\varepsilon_0} \to V$$

as $\lambda_0 \to \infty$*, with* $V \sim N(0,1)$*, in which*

$$\tau^2 = c_3^2\mu_G^3\gamma_G^2 \tag{5.10}$$
$$+ c_4^2\mu_G\gamma_G^2\left(\gamma_G^2 - \gamma_G\kappa_{3G} + \frac{1}{4}\kappa_{4G} + \frac{1}{2}\right)$$
$$+ c_5^2(1-\varepsilon)\{\mu_G(1-\varepsilon)(1+\gamma_G^2) + \varepsilon\}$$
$$+ c_3c_4\mu_G^2\gamma_G^2(\kappa_{3G} - 2\gamma_G)$$
$$+ 2c_3c_5(1-\varepsilon)\mu_G^2\gamma_G^2$$
$$+ c_4c_5\mu_G\gamma_G^2(1-\varepsilon)(\kappa_{3G} - 2\gamma_G),$$

where

$$c_3 = \mu_{C0}^{-1}\frac{\partial}{\partial\mu_G}SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0), \tag{5.11}$$
$$c_4 = \mu_{C0}^{-1}\frac{\partial}{\partial\gamma_G}SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0),$$
$$c_5 = \varepsilon_0\mu_{C0}^{-1}\frac{\partial}{\partial\varepsilon}SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0).$$

Proof. The limiting result follows directly from Theorem 5.2, because

$$\mu_{C0}^{-1}\{SLPapp1(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0) - SLPapp1(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0)\}$$
$$= c_3(\widehat{\mu}_G - \mu_{G0}) + c_4(\widehat{\gamma}_G - \gamma_{G0}) + c_5\frac{\widehat{\varepsilon} - \varepsilon_0}{\varepsilon_0}$$

with c_3, c_4, c_5 given by (5.11). (Note that here we have used in the formulation of the theorem $\sqrt{u\lambda_0\varepsilon_0}$ instead of $\sqrt{u\lambda_0}$, because the expected number of special claims equals $u\lambda_0\varepsilon_0$.) □

Remark 5.5. Theorem 5.3 can be applied to $G : P(L)$ with parametrization μ_L, γ_L (provided that the fourth moment of L is finite), replacing $SLP(\widehat{\mu}_C, \widehat{\gamma}_C, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \widehat{\lambda})$, $SLPapp1($

$\mu_{C0}, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0)$ and $SLPapp1\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right)$ by $SLP(\widehat{\mu}_C, \widehat{\gamma}_C, \widehat{\mu}_L, \widehat{\gamma}_L, \widehat{\varepsilon}, \widehat{\lambda})$, $SLPapp1\left(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_L, \widehat{\gamma}_L, \widehat{\varepsilon}, \lambda_0\right)$ and $SLPapp1\left(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \varepsilon_0, \lambda_0\right)$, respectively, and τ^2 by

$$
\begin{aligned}
\tau^2 = {} & c_3^2\left(\mu_{L0}^3\gamma_{L0}^2 + \mu_{L0}^2\right) && \text{(5.12)}\\
& + c_4^2\left\{\mu_{L0}\gamma_{L0}^2\left(\gamma_{L0}^2 - \gamma_{L0}\kappa_{3L0} + \frac{1}{4}\kappa_{4L0} + \frac{1}{2}\right) - \gamma_{L0}^2 + \gamma_{L0}\kappa_{3L0} + 1 + \frac{1}{2}\mu_{L0}^{-1}\left(1 + \gamma_{L0}^{-2}\right)\right\}\\
& + c_5^2\left(1 - \varepsilon_0\right)\left\{\mu_{L0}\left(1 - \varepsilon_0\right)\left(1 + \gamma_{L0}^2\right) + 1\right\}\\
& + c_3c_4\mu_{L0}^2\gamma_{L0}^2(\kappa_{3L0} - 2\gamma_{L0})\\
& + 2c_3c_5\left(1 - \varepsilon_0\right)\left(\mu_{L0}^2\gamma_{L0}^2 + \mu_{L0}\right)\\
& + c_4c_5\mu_{L0}\gamma_{L0}^2(1 - \varepsilon_0)(\kappa_{3L0} - 2\gamma_{L0}),
\end{aligned}
$$

where

$$
\begin{aligned}
c_3 &= \mu_{C0}^{-1}\frac{\partial}{\partial\mu_L}SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \varepsilon_0, \lambda_0\right), && \text{(5.13)}\\
c_4 &= \mu_{C0}^{-1}\frac{\partial}{\partial\gamma_L}SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \varepsilon_0, \lambda_0\right),\\
c_5 &= \varepsilon_0\mu_{C0}^{-1}\frac{\partial}{\partial\varepsilon}SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \varepsilon_0, \lambda_0\right).
\end{aligned}
$$

6   Effect of estimation, protection. Having established readily applicable formulas for the estimation effects, we investigate the impact of the estimation on the stop-loss premium $E(S - a)^+$. We start with an example. Let the true values of the parameters be equal to $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$ and $k = (a - \mu_S)/\sigma_S = 1$. Let C and L each have a (different) gamma-distribution. As we have seen, see Section 4, $SLPapp(100000, 0.7, 15, 0.8, 0.03, 400) = 292282$ in that case. We may for example ask: what is the probability that $SLPapp1\left(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0\right)$ is smaller than 200000, that is an error of more than 92282? We apply Theorem 5.3. Direct calculation gives $\tau^2 = 36.51$ and hence, with Φ the standard normal distribution function and noting that $10^{-5}(200000 - 292282)(36.51)^{-1/2}\sqrt{12} = -0.53$, we obtain

$$
\begin{aligned}
& P(SLPapp1\left(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0\right) < 200000)\\
={} & P(10^{-5}\{SLPapp1\left(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0\right) - 292282\}(36.51)^{-1/2}\sqrt{12u} < -0.53\sqrt{u})\\
\approx{} & \Phi(-0.53\sqrt{u}).
\end{aligned}
$$

Taking only one year, that is $u = 1$, we see that with a probability as large as 30% we get an estimated value smaller than 200000, while in fact it should have been 292282. This makes clear that indeed one year is not enough. The reason for this is of course that in one year the expected number of groups is (in this case) only $\varepsilon\lambda/\mu_G = 12/15 = 0.8$. This makes the estimation of μ_G, γ_G and ε very inaccurate. When taking $u = 7$, the probability reduces from 30% to 8%.

We see from the example that the estimation effect may be considerable and we may want to protect ourselves against the estimation error, in the sense of confidence bounds for $SLPapp$. The following theorem deals with such a protection.

Theorem 6.1. *Let $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0)$ be the true value of the parameters. Then*

$$\lim_{\lambda_0 \to \infty} P(SLPapp\,(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0) < UB\,(\alpha)) = 1 - \alpha,$$

$$\lim_{\lambda_0 \to \infty} P(SLPapp\,(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0) > LB\,(\alpha)) = 1 - \alpha,$$

$$\lim_{\lambda_0 \to \infty} P(LB(\alpha/2) < SLPapp\,(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0) < UB\,(\alpha/2)) = 1 - \alpha$$

with

$$UB\,(\alpha) = SLPapp1\,(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0) + \Phi^{-1}(1 - \alpha)(\widehat{\varepsilon}u\lambda_0)^{-1/2}\widehat{\tau}\mu_{C0},$$

$$LB\,(\alpha) = SLPapp1\,(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0) - \Phi^{-1}(1 - \alpha)(\widehat{\varepsilon}u\lambda_0)^{-1/2}\widehat{\tau}\mu_{C0},$$

where $\widehat{\tau} = \sqrt{\widehat{\tau^2}}$ and $\widehat{\tau}^2$ is given in (5.10) and (5.11) with $\mu_{G0}, \gamma_{G0}, \varepsilon_0$ replaced by their estimators $\widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}$ (also in $c_3, c_4, c_5, \kappa_{3G0}$ and κ_{4G0}).

Proof. It is easily seen, cf. e.g. Theorem 5.2, that $\widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}$ are consistent estimators of $\mu_G, \gamma_G, \varepsilon$. Writing $\widehat{c}_i$ for c_i with $\mu_{G0}, \gamma_{G0}, \varepsilon_0$ replaced by their estimators $\widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}$, it can be shown (we omit the details, but see Lukocius [5], Chapter 7 for more explanation) that $\widehat{c}_i/c_i \to^P 1$ as $\lambda_0 \to \infty$ and moreover, that the c_i are of the same order of magnitude (that is of exact order $\lambda_0^{1/2}$) for $i = 3, 4, 5$ and hence $\widehat{\tau}/\tau \to^P 1$ as $\lambda_0 \to \infty$. Application of Theorem 5.3 therefore yields

$$(\widehat{\tau}\mu_{C0})^{-1}\left\{SLPapp1\,(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0) - SLPapp1\,(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0)\right\}\sqrt{\widehat{\varepsilon}u\lambda_0}$$
$$\to U$$

with $U \sim N(0,1)$. This implies, writing temporarily $\widehat{S} = SLPapp1\left(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0\right)$ and noting that $SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right) = SLPapp1\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right)$,

$$
\begin{aligned}
&P(SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right) < UB\left(\alpha\right)) \\
=&P(SLPapp1\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right) < \widehat{S} + \Phi^{-1}(1-\alpha)(\widehat{\varepsilon}u\lambda_0)^{-1/2}\widehat{\tau}\mu_{C0}) \\
=&P((\widehat{\tau}\mu_{C0})^{-1}\left\{\widehat{S} - SLPapp1\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right)\right\}\sqrt{\widehat{\varepsilon}u\lambda_0} > -\Phi^{-1}(1-\alpha)) \\
\rightarrow&P(U > -\Phi^{-1}(1-\alpha)) = 1 - \alpha,
\end{aligned}
$$

thus giving the first result. The other statements are obtained in a similar way. $\qquad\square$

Remark 6.1. Theorem 6.1 can be applied to $G : P(L)$ with parametrization μ_L, γ_L (provided that the fourth moment of L is finite), replacing $SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0\right)$ and $SLPapp1\left(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0\right)$ by $SLPapp\left(\mu_{C0}, \gamma_{C0}, \mu_{L0}, \gamma_{L0}, \varepsilon_0, \lambda_0\right)$ and $SLPapp1$ $\left(\mu_{C0}, \gamma_{C0}, \widehat{\mu}_L, \widehat{\gamma}_L, \widehat{\varepsilon}, \lambda_0\right)$, respectively, and $\widehat{\tau}^2$ by the estimated version of (5.12) and (5.13).

Remark 6.2. One may ask why the estimation error is quite substantial. Is it due to the model construction, or the use of the maximum likelihood estimators, or the structure of the stop-loss premium, or the use of the $G - IG$ approximation, or is it due to the further Taylor expansion error? As mentioned before, the estimation error is dominated by the part of the parameters related to the special claims, because by their nature we do not have many observations of them. So, that is the main reason. It is well-known that the more observations, in general the more accurate the estimation. This is so to say an explanation on the most general level. Going somewhat deeper into it, we may distinguish two aspects: the function of the parameters, that have to be estimated (in our case the stop-loss premium) and the accuracy of the estimators of the parameters. If the function is very flat, errors due to estimation may be not very large. With respect to this aspect, obviously the structure of the stop-loss premium comes in. The fluctuation of the stop-loss premium as function of the parameters is expressed by its first order derivatives. These are studied in Section 4. Obviously, the stop-loss premium, being a function of S, is determined by the model construction and hence the model construction plays a role. For instance, in the far more simple model assuming only independent claims, the estimation error will be much less, because the estimation error is dominated by the part of the parameters related to the special claims and they are not present in the independence model. The use of the $G - IG$ approximation and the further Taylor expansion are not important. That the
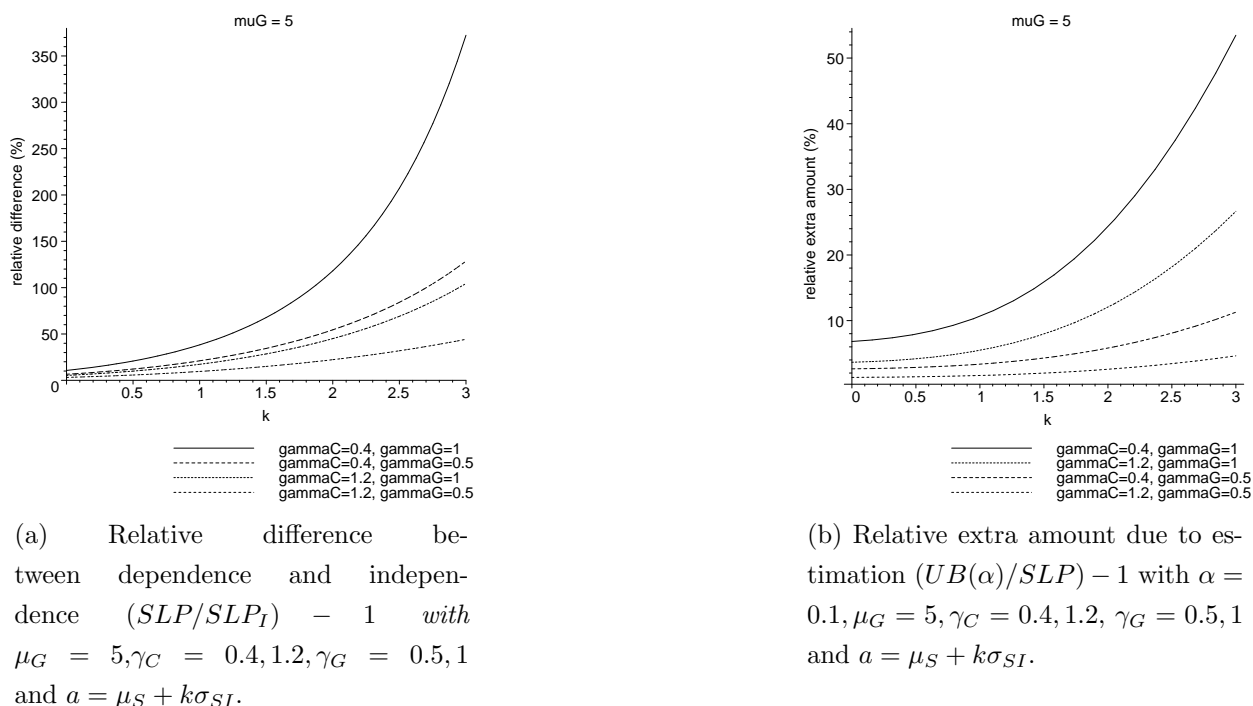
(a) Relative difference between dependence and independence $(SLP/SLP_I) - 1$ *with* $\mu_G = 5, \gamma_C = 0.4, 1.2, \gamma_G = 0.5, 1$ and $a = \mu_S + k\sigma_{SI}$.

(b) Relative extra amount due to estimation $(UB(\alpha)/SLP) - 1$ with $\alpha = 0.1, \mu_G = 5, \gamma_C = 0.4, 1.2, \gamma_G = 0.5, 1$ and $a = \mu_S + k\sigma_{SI}$.

Figure 1

$G - IG$ approximation is accurate was already shown in Lukocius [4]; that also the one step Taylor expansion is accurate is shown in Section 4. The accuracy of the estimators of the parameters is established in Theorems 5.1 and 5.2. There it has been shown, that indeed the part of the parameters related to the special claims are dominating. All estimators, used in the paper are "natural" estimators of the corresponding parameters and therefore the use of the maximum likelihood estimators seems to be not that important. It is very nice that the method of maximum likelihood leads to "natural" estimators, but the fact that indeed we get such "natural" estimators is more important.

Remember that the contribution of estimating μ_C, γ_C and λ is very small compared to that of estimating μ_G, γ_G and ε. Therefore, we assume in Theorem 6.1 again $\mu_{C0}, \gamma_{C0}, \lambda_0$ to be known. Obviously, in practice one should insert the estimators $\widehat{\mu}_C, \widehat{\gamma}_C$ and $\widehat{\lambda}$ in the upper and lower bounds $UB(\alpha)$ and $LB(\alpha)$.

In Figures 1–3 some examples are presented of the extra amount due to the protection against estimation and the effect of dependence in these situations. Figures 1a–3a show the relative difference between the independent case and the dependence one. Let C and
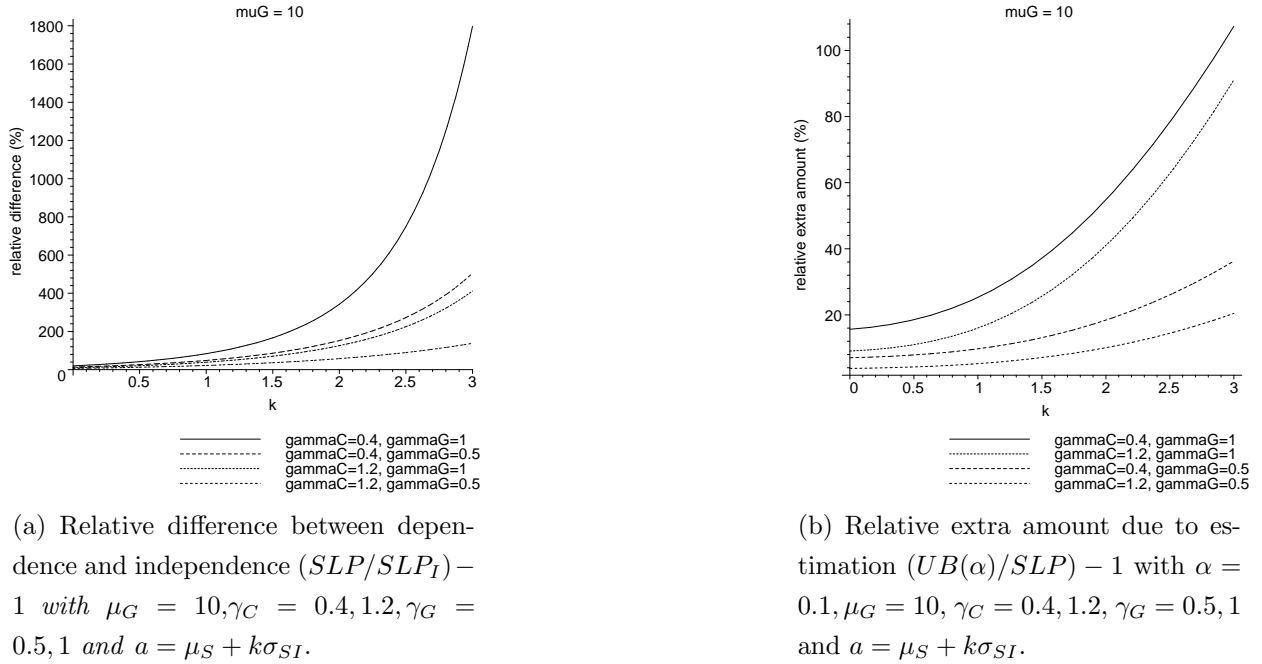
(a) Relative difference between dependence and independence $(SLP/SLP_I) - 1$ with $\mu_G = 10, \gamma_C = 0.4, 1.2, \gamma_G = 0.5, 1$ and $a = \mu_S + k\sigma_{SI}$.

(b) Relative extra amount due to estimation $(UB(\alpha)/SLP) - 1$ with $\alpha = 0.1, \mu_G = 10, \gamma_C = 0.4, 1.2, \gamma_G = 0.5, 1$ and $a = \mu_S + k\sigma_{SI}$.

Figure 2

L each have a (different) gamma-distribution. We take $\gamma_{C0} = 0.4$ or $1.2, \mu_{G0} = 5, 10$ or $15, \gamma_{G0} = 0.5$ or $1, \varepsilon_0 = 0.03$ and $\lambda_0 = 400$. Obviously the independence case is obtained by taking $\varepsilon = 0$. The relative difference is defined by

$$\frac{SLPapp\,(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0) - SLPapp\,(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, 0, \lambda_0)}{SLPapp\,(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, 0, \lambda_0)}$$
$$= \frac{SLP}{SLP_I} - 1,$$

where SLP denotes the (approximated) stop-loss premium $SLPapp\,(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0)$ under dependence and $SLP_I = SLPapp\,(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, 0, \lambda_0)$, the (approximated) stop-loss premium under independence. For a fair comparison we take both for the independence model and the dependence one the same retentions
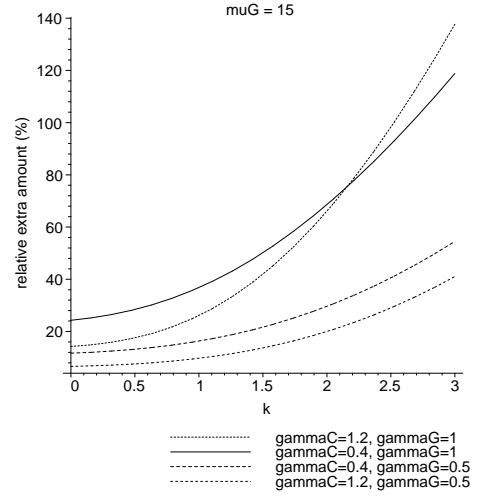
$$a = \mu_S + k\sigma_{SI}$$

with $k = 0, \ldots, 3$ and $\sigma_{SI} = \mu_C \sqrt{\lambda\,(1 + \gamma_C^2)}$, the standard deviation of S for the independence model (see also the Appendix).

Figures 1b–3b show the extra amount due to protection against estimation, also mea-
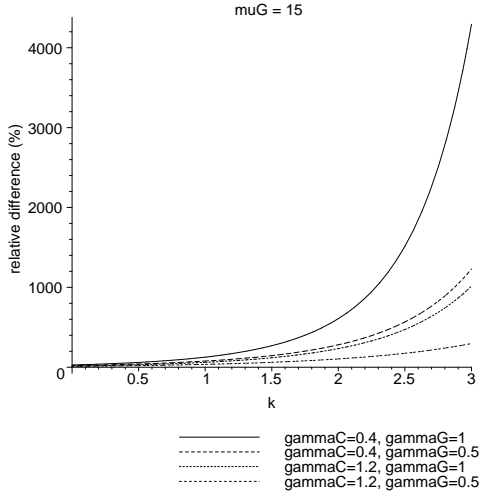
(a) Relative difference between dependence and independence $(SLP/SLP_I) - 1$ $with$ $\mu_G = 15, \gamma_C = 0.4, 1.2, \gamma_G = 0.5, 1$ and $a = \mu_S + k\sigma_{SI}$.

(b) Relative extra amount due to estimation $(UB(\alpha)/SLP) - 1$ with $\alpha = 0.1$, $\mu_G = 15$, $\gamma_C = 0.4, 1.2$, $\gamma_G = 0.5, 1$ and $a = \mu_S + k\sigma_{SI}$.

Figure 3

sured in a relative way by taking

$$\frac{UB\,(\alpha) - SLPapp\,(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0)}{SLPapp\,(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0)}$$

with in $UB\,(\alpha)$ the estimators $\widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}$ and $\widehat{\tau}$ replaced by $\mu_G, \gamma_G, \varepsilon_0$ and $\sqrt{\tau^2}$, respectively. We take $\alpha = 0.1$ and $u = 7$. It is easily seen (see also at the end of this section) that both measures do not depend on μ_{C0}.

Note that the order of the displayed cases is slightly different in the figures a and b: for instance, for $\mu_{G0} = 5$ (Figures 1a, b) the relative difference between dependence and independence is higher for $\gamma_{C0} = 0.4, \gamma_{G0} = 0.5$ than for $\gamma_{C0} = 1.2, \gamma_{G0} = 1$, while their order w.r.t. the relative extra amount due to protection against estimation is reversed.

Figures 1–3 affirm that ignoring dependence may lead to very large errors (up to 4294% in Figure 3). But also the additional step due to protection against estimation is large (up to 138% in Figure 3). A numerical example may illustrate this. Consider again the example with true values of the parameters being equal to $(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0) = (100000, 0.7, 15, 0.8, 0.03, 400)$. Take $k = 1$ and hence $a = \mu_S + \sigma_{SI} = 4 \times 10^7 + 2561250 = 42561250$. If we ignore the dependence structure we get $SLPapp(100000, 0.7, 15, 0.8, 0, 400) =$

211277. If we take into account the dependence without protection against estimation we get $SLPapp(100000, 0.7, 15, 0.8, 0.03, 400) = 382006$. If we add the protection (taking $\widehat{\mu}_G = \mu_{G0} = 15, \widehat{\gamma}_G = \gamma_{G0} = 0.8, \widehat{\varepsilon} = \varepsilon_0 = 0.03, \widehat{\tau} = \sqrt{\tau^2}$) we get $UB(0.1) = 476596$.

The upper and lower bounds $UB(\alpha)$ and $LB(\alpha)$ contain the term $\widehat{\tau}\mu_{C0}$. As this quantity is the less transparent part of $UB(\alpha)$ and $LB(\alpha)$, we will discuss it now. It is seen in the Appendix that

$$SLPapp(\mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon, \lambda) = \mu_C SLPapp(1, \gamma_C, \mu_G, \gamma_G, \varepsilon, \lambda).$$

In view of (5.11) this implies

$$c_3 = \frac{\partial}{\partial \mu_G} SLPapp(1, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0),$$

$$c_4 = \frac{\partial}{\partial \gamma_G} SLPapp(1, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0),$$

$$c_5 = \varepsilon_0 \frac{\partial}{\partial \varepsilon} SLPapp(1, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0).$$

Therefore, see (4.1), using

$$SLPapp1(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0) = SLPapp(\mu_{C0}, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0)$$
$$= \mu_{C0} SLPapp(1, \gamma_{C0}, \mu_{G0}, \gamma_{G0}, \varepsilon_0, \lambda_0),$$

we get

$$UB(\alpha) = \mu_{C0}\{SLPapp1(1, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0) + \Phi^{-1}(1-\alpha)(\widehat{\varepsilon}u\lambda_0)^{-1/2}\widehat{\tau}\},$$
$$LB(\alpha) = \mu_{C0}\{SLPapp1(1, \gamma_{C0}, \widehat{\mu}_G, \widehat{\gamma}_G, \widehat{\varepsilon}, \lambda_0) - \Phi^{-1}(1-\alpha)(\widehat{\varepsilon}u\lambda_0)^{-1/2}\widehat{\tau}\}.$$

So, we see that indeed μ_{C0} is a kind of dummy parameter.

In the special case with L having a gamma distribution,

$$\kappa_{3G} = 2\gamma_G - \mu_G^{-1}\gamma_G^{-1}, \kappa_{4G} = 6\gamma_G^2 - 6\mu_G^{-1} + \mu_G^{-2}\gamma_G^{-2}$$
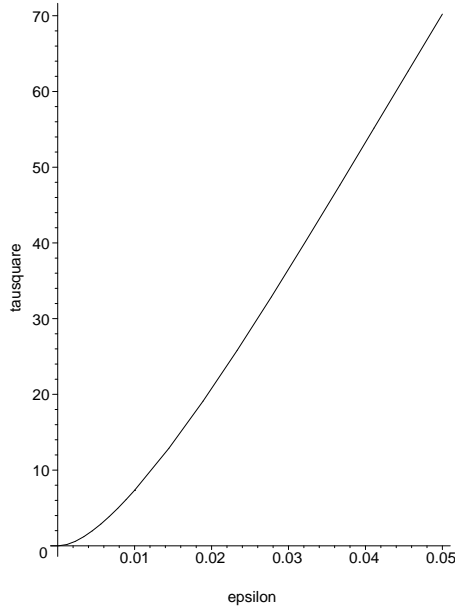
Figure 4: Behavior of $\tau^2(\mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon, \lambda, k)$ as $\varepsilon \to 0$ with $(\mu_C, \gamma_C, \mu_G, \gamma_G, \lambda) = (100000, 0.7, 15, 0.8, 400)$ and $k = (a - \mu_S)/\sigma_S = 1$.

and thus τ^2 reduces to

$$\tau^2 = c_3^2 \mu_G^3 \gamma_G^2 \tag{6.1}$$
$$+ \frac{1}{2}c_4^2 \left(\mu_G \gamma_G^4 - \gamma_G^2 + \frac{1}{2}\mu_G^{-1} + \mu_G \gamma_G^2 \right)$$
$$+ c_5^2 (1 - \varepsilon) \left\{ \mu_G (1 - \varepsilon) \left(1 + \gamma_G^2 \right) + \varepsilon \right\}$$
$$- c_3 c_4 \mu_G \gamma_G$$
$$+ 2 c_3 c_5 (1 - \varepsilon) \mu_G^2 \gamma_G^2$$
$$- c_4 c_5 \gamma_G (1 - \varepsilon).$$

For illustrative purposes we show the behavior of τ^2 in (6.1) as a function of ε (with $(\mu_C, \gamma_C, \mu_G, \gamma_G, \lambda, k) = (100000, 0.7, 15, 0.8, 400, 1)$ keeping fixed). Note that c_3, c_4, c_5 depend on ε in a complicated way. It is clearly seen in Figure 4 that τ^2 tends to 0 if $\varepsilon \to 0$.

Appendix. Approximations.   Here we present three approximations: the gamma approximation, the *Inverse Gaussian* (IG) approximation and the *Gamma − Inverse Gaussian* $(G − IG)$ approximation. For the parameter range and distributions under consideration (see Section 2) the $G − IG$ approximation works well and is best among the three

approximations, see Lukocius [4] for more details. Therefore, the $G - IG$ approximation is recommended. Note that one has to be careful with extending this conclusion outside the parameter range or for other distributions than considered here.

Gamma approximation

A shifted gamma distribution is fitted such that the first three cumulants coincide with those of S. The density of the gamma distribution with parameters α and β (for short: $Gamma(\alpha, \beta)$) is given by

$$f_G(x; \alpha, \beta) = \frac{x^{\alpha-1} \beta^\alpha e^{-\beta x}}{\Gamma(\alpha)}.$$

We approximate S by T such that $T - x_0$ is $Gamma(\alpha, \beta)$, where x_0, α and β are selected such that the first three cumulants of T and S coincide. This is achieved by taking

$$\alpha = \left(\frac{2}{\kappa_{3S}}\right)^2, \beta = \frac{2}{\sigma_S \kappa_{3S}} \text{ and } x_0 = \mu_S - \frac{2\sigma_S}{\kappa_{3S}}.$$

Noting that $a = \mu_S + k\sigma_S$, it leads to the approximation

$$E_G(S - a)^+ = \sigma_S E_G \left(\frac{S - \mu_S}{\sigma_S} - k\right)^+$$

$$= \sigma_S \left\{\frac{2}{\kappa_{3S}} \overline{F}_G \left(k + \frac{2}{\kappa_{3S}}; \frac{4}{\kappa_{3S}^2} + 1, \frac{2}{\kappa_{3S}}\right) - \left(k + \frac{2}{\kappa_{3S}}\right) \overline{F}_G \left(k + \frac{2}{\kappa_{3S}}; \frac{4}{\kappa_{3S}^2}, \frac{2}{\kappa_{3S}}\right)\right\},$$

where

$$\overline{F}_G(x; \alpha, \beta) = 1 - F_G(x; \alpha, \beta)$$

and where $F_G(x; \alpha, \beta)$ is the distribution function of the gamma distribution with parameters α and β.

IG approximation

The density of the IG-distribution with parameters α and β (for short: $IG(\alpha, \beta)$) is given by

$$f_{IG}(x; \alpha, \beta) = \alpha (2\pi\beta)^{-1/2} x^{-3/2} \exp\left\{-\frac{(\alpha - \beta x)^2}{2\beta x}\right\}.$$

For the IG approximation (see Chaubey et al. [3]) we approximate S by T such that $T - x_0$ is $IG(\alpha, \beta)$, where x_0, α and β are selected such that the first three cumulants of T and S

coincide. This is achieved by taking

$$\alpha = \left(\frac{3}{\kappa_{3S}}\right)^2, \beta = \frac{3}{\sigma_S \kappa_{3S}} \text{ and } x_0 = \mu_S - \frac{3\sigma_S}{\kappa_{3S}}.$$

Noting that $a = \mu_S + k\sigma_S$, it leads to the approximation

$$E_{IG}(S-a)^+ = \sigma_S E \left(\frac{S-\mu_S}{\sigma_S} - k\right)^+ = \sigma_S \int_k^\infty \frac{x-k}{\sqrt{2\pi\left(1+\frac{1}{3}x\kappa_{3S}\right)^3}} \exp\left[-\frac{x^2}{2\left(1+\frac{1}{3}x\kappa_{3S}\right)}\right] dx.$$

Using

$$\frac{d}{dx}\left\{\Phi\left(\frac{x}{\sqrt{1+tx}}\right) - \exp\left(\frac{2}{t^2}\right)\Phi\left(\frac{x+\frac{2}{t}}{\sqrt{1+tx}}\right)\right\} = \frac{1}{\sqrt{2\pi(1+tx)^3}}\exp\left[-\frac{x^2}{2(1+tx)}\right],$$

$$\frac{d}{dx}\left\{\frac{2}{t}\exp\left(\frac{2}{t^2}\right)\Phi\left(\frac{x+\frac{2}{t}}{\sqrt{1+tx}}\right)\right\} = \frac{x}{\sqrt{2\pi(1+tx)^3}}\exp\left[-\frac{x^2}{2(1+tx)}\right],$$

we obtain

$$E_{IG}(S-a)^+ = \sigma_S \left\{\left(k+\frac{6}{\kappa_{3S}}\right)\exp\left(\frac{18}{\kappa_{3S}^2}\right)\Phi\left(\frac{-k-\frac{6}{\kappa_{3S}}}{\sqrt{1+\frac{1}{3}k\kappa_{3S}}}\right) - k\Phi\left(\frac{-k}{\sqrt{1+\frac{1}{3}k\kappa_{3S}}}\right)\right\}.$$

$G - IG$ approximation

The $G - IG$ approximation is a combination of the gamma approximation and the IG approximation. Each of these approximations only uses the first three cumulants. A mixing parameter w can be chosen such that the kurtosis of S is fitted as well. The mixing parameter turns out to be

$$w = w(\kappa_{3S}, \kappa_{4S}) = \frac{\frac{5}{3}\kappa_{3S}^2 - \kappa_{4S}}{\frac{5}{3}\kappa_{3S}^2 - \frac{3}{2}\kappa_{3S}^2} = 10 - \frac{6\kappa_{4S}}{\kappa_{3S}^2}.$$

Hence, the $G - IG$ approximation gives

$$E_{G-IG}(S-a)^+ = w(\kappa_{3S}, \kappa_{4S})E_G(S-a)^+ + [1 - w(\kappa_{3S}, \kappa_{4S})]E_{IG}(S-a)^+.$$

Remark A.1. In order that the weight $w(\kappa_{3S}, \kappa_{4S})$ in the $G - IG$ approximation lies between 0 and 1 we should assume $\frac{3}{2}\kappa_{3S}^2 \leq \kappa_{4S} \leq \frac{5}{3}\kappa_{3S}^2$. Unfortunately, often this condition

is not satisfied. However, we may use nevertheless the $G - IG$ approximation (with w not in $(0, 1)$) and simply consider it as an approximation. On the interval in which we are interested ($s > a = \mu_S + k\sigma_S$ with $0 \leq k \leq 3$), often

$$w(\kappa_{3S}, \kappa_{4S}) f_G \left(s - \mu_S + \frac{2\sigma_S}{\kappa_{3S}}; \left(\frac{2}{\kappa_{3S}} \right)^2, \frac{2}{\sigma_S \kappa_{3S}} \right)$$

$$+ \left[1 - w(\kappa_{3S}, \kappa_{4S}) \right] f_{IG} \left(s - \mu_S + \frac{3\sigma_S}{\kappa_{3S}}; \left(\frac{3}{\kappa_{3S}} \right)^2, \frac{3}{\sigma_S \kappa_{3S}} \right)$$

behaves like a density. That is, it is positive on this interval. (In principle, in that case we could even extend it to a density, but note that we should also keep the first four moments of the approximation and those of S equal to each other and that makes it a little bit nasty; therefore we do not bother and consider it simply as an approximation.)

Next we present formulas for $\mu_S, \sigma_S, \kappa_{3S}$ and κ_{4S}.

Formulas for $\mu_S, \sigma_S, \kappa_{3S}$ and κ_{4S}

So far, the approximations are in terms of σ_S, κ_{3S} and κ_{4S}. It remains to link these quantities to the basic parameters $\mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon$ and λ. We start with expressions in case of general C, G (with finite fourth moment), adding for the the sake of completeness also μ_S:

$$\mu_S = \lambda \mu_C, \tag{A1}$$

$$\sigma_S / (\sqrt{\lambda} \mu_C) = \sqrt{1 + \gamma_C^2 - \varepsilon + \varepsilon(1 + \gamma_G^2) \mu_G},$$

$$\kappa_{3S}^* / (\lambda \mu_C^3) = 1 + 3\gamma_C^2 + \kappa_{3C}\gamma_C^3 - \varepsilon(1 + 3\gamma_C^2) + 3\varepsilon\gamma_C^2(1 + \gamma_G^2)\mu_G + \varepsilon(1 + 3\gamma_G^2 + \kappa_{3G}\gamma_G^3)\mu_G^2,$$

$$\kappa_{4S}^* / (\lambda \mu_C^4) = 1 + 6\gamma_C^2 + 4\kappa_{3C}\gamma_C^3 + (\kappa_{4C} + 3)\gamma_C^4$$

$$- \varepsilon(1 + 6\gamma_C^2 + 4\kappa_{3C}\gamma_C^3 + 3\gamma_C^4)$$

$$+ \varepsilon(1 + \gamma_G^2)(4\kappa_{3C}\gamma_C^3 + 3\gamma_C^4)\mu_G$$

$$+ 6\varepsilon\gamma_C^2(1 + 3\gamma_G^2 + \kappa_{3G}\gamma_G^3)\mu_G^2$$

$$+ \varepsilon\{1 + 6\gamma_G^2 + 4\kappa_{3G}\gamma_G^3 + (\kappa_{4G} + 3)\gamma_G^4\}\mu_G^3,$$

$$\kappa_{3S} = \kappa_{3S}^* / \sigma_S^3,$$

$$\kappa_{4S} = \kappa_{4S}^* / \sigma_S^4.$$

So, $SLPapp(\mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon, \lambda)$ is obtained by inserting σ_S, κ_{3S} and κ_{4S} from (A1) into $E_{G-IG}(S - a)^+$. It is easily seen that κ_{3S} and κ_{4S} do not depend on μ_C. Moreover,

$E_{G-IG}(S-a)^+$ is of the form $\sigma_S h(\kappa_{3S}, \kappa_{4S})$ and σ_S is of the form $\mu_C h^*(\gamma_C, \mu_G, \gamma_G, \varepsilon, \lambda)$. Hence, we get

$$SLPapp\,(\mu_C, \gamma_C, \mu_G, \gamma_G, \varepsilon, \lambda) = \mu_C SLPapp\,(1, \gamma_C, \mu_G, \gamma_G, \varepsilon, \lambda)\,.$$

Assuming additionally $G_k : P(L)$, we obtain

$$\mu_S = \lambda \mu_C, \tag{A2}$$

$$\sigma_S/(\sqrt{\lambda}\mu_C) = \sqrt{1 + \gamma_C^2 + \varepsilon(1 + \gamma_L^2)\mu_L},$$

$$\kappa_{3S}^*/(\lambda\mu_C^3) = 1 + 3\gamma_C^2 + \kappa_{3C}\gamma_C^3 + 3\varepsilon(1 + \gamma_C^2)(1 + \gamma_L^2)\mu_L + \varepsilon(1 + 3\gamma_L^2 + \kappa_{3L}\gamma_L^3)\mu_L^2,$$

$$\kappa_{4S}^*/(\lambda\mu_C^4) = 1 + 6\gamma_C^2 + 4\kappa_{3C}\gamma_C^3 + (\kappa_{4C} + 3)\gamma_C^4$$
$$+ \varepsilon\{4(1 + 3\gamma_C^2 + \kappa_{3C}\gamma_C^3) + 3(1 + \gamma_C^2)^2\}(1 + \gamma_L^2)\mu_L$$
$$+ 6\varepsilon(1 + \gamma_C^2)(1 + 3\gamma_L^2 + \kappa_{3L}\gamma_L^3)\mu_L^2$$
$$+ \varepsilon\{1 + 6\gamma_L^2 + 4\kappa_{3L}\gamma_L^3 + (\kappa_{4L} + 3)\gamma_L^4\}\mu_L^3,$$

$$\kappa_{3S} = \kappa_{3S}^*/\sigma_S^3,$$

$$\kappa_{4S} = \kappa_{4S}^*/\sigma_S^4.$$

Hence, $SLPapp\,(\mu_C, \gamma_C, \mu_L, \gamma_L, \varepsilon, \lambda)$ is obtained by inserting σ_S, κ_{3S} and κ_{4S} from (A2) into $E_{G-IG}(S-a)^+$.

In the particular case that C has a gamma distribution we get $\kappa_{3C} = 2\gamma_C$ and $\kappa_{4C} = 6\gamma_C^2$, implying

$$1 + 3\gamma_C^2 + \kappa_{3C}\gamma_C^3 = (1 + \gamma_C^2)(1 + 2\gamma_C^2) \tag{A3}$$

and

$$1 + 6\gamma_C^2 + 4\kappa_{3C}\gamma_C^3 + (\kappa_{4C} + 3)\gamma_C^4 = (1 + \gamma_C^2)(1 + 2\gamma_C^2)(1 + 3\gamma_C^2). \tag{A4}$$

When C has an Inverse Gaussian distribution we get $\kappa_{3C} = 3\gamma_C$ and $\kappa_{4C} = 15\gamma_C^2$, implying

$$1 + 3\gamma_C^2 + \gamma_C^3\kappa_{3C} = 1 + 3\gamma_C^2 + 3\gamma_C^4$$

and

$$1 + 6\gamma_C^2 + 4\gamma_C^3\kappa_{3C} + \gamma_C^4(\kappa_{4C} + 3) = 1 + 6\gamma_C^2 + 15\gamma_C^4 + 15\gamma_C^6.$$

When C has a lognormal distribution we get $\kappa_{3C} = \gamma_C(3 + \gamma_C^2)$ and $\kappa_{4C} = \gamma_C^2(16 + 15\gamma_C^2 + 6\gamma_C^4 + \gamma_C^6)$, implying

$$1 + 3\gamma_C^2 + \gamma_C^3\kappa_{3C} = (1 + \gamma_C^2)^3 \tag{A5}$$

and

$$1 + 6\gamma_C^2 + 4\gamma_C^3\kappa_{3C} + \gamma_C^4(\kappa_{4C} + 3) = (1 + \gamma_C^2)^6. \tag{A6}$$

Remark A.2. Noting that

$$1 + 3\gamma_C^2 + \kappa_{3C}\gamma_C^3 = \mu_C^{-3}EC^3,$$
$$1 + 6\gamma_C^2 + 4\kappa_{3C}\gamma_C^3 + (\kappa_{4C} + 3)\gamma_C^4 = \mu_C^{-4}EC^4$$

and that in case of a gamma distribution we have for $j = 1, 2, ...$

$$\mu_C^{-j}EC^j = \prod_{i=1}^{j}(1 + i\gamma_C^2),$$

while for the lognormal distribution we get for $j = 1, 2, ...$

$$\mu_C^{-j}EC^j = (1 + \gamma_C^2)^{j(j-1)/2},$$

the expressions (A3)–(A6) are easily seen.

Obviously, similar expressions hold for L, having a gamma or an Inverse Gaussian distribution. In particular, when L has a gamma distribution, we obtain

$$\kappa_{3G} = 2\gamma_G - \mu_G^{-1}\gamma_G^{-1}, \tag{A7}$$
$$\kappa_{4G} = 6\gamma_G^2 - 6\mu_G^{-1} + \mu_G^{-2}\gamma_G^{-2}.$$

When C and L have a gamma distribution, we obtain by combination of (A1) and (A7)

$$\kappa_{3S}^*/(\lambda\mu_C^3) = (1 + \gamma_C^2)(1 + 2\gamma_C^2) + \varepsilon\mu_G^2(1 + \gamma_G^2)(1 + 2\gamma_G^2)$$
$$- \varepsilon(1 + 3\gamma_C^2) + \varepsilon\{3\gamma_C^2(1 + \gamma_G^2) - \gamma_G^2\}\mu_G$$

and

$$\kappa_{4S}^*/(\lambda\mu_C^4) = (1 + \gamma_C^2)(1 + 2\gamma_C^2)(1 + 3\gamma_C^2)$$
$$- \varepsilon(1 + 6\gamma_C^2 + 11\gamma_C^4)$$
$$+ \varepsilon\{(1 + \gamma_G^2)11\gamma_C^4 - 6\gamma_C^2\gamma_G^2 + \gamma_G^2\}\mu_G$$
$$+ 2\varepsilon\{3\gamma_C^2(1 + \gamma_G^2)(1 + 2\gamma_G^2) - \gamma_G^2(2 + 3\gamma_G^2)\}\mu_G^2$$
$$+ \varepsilon\{(1 + \gamma_G^2)(1 + 2\gamma_G^2)(1 + 3\gamma_G^2)\}\mu_G^3.$$

References

[1] Albers, W. (1999). Stop-loss premiums under dependence. *Insurance Math. Econom.* **24** 173–185.

[2] Albers, W.; Kallenberg, W. C. M. and Lukocius, V. (2009). A flexible model for actuarial risks under dependence. *Scand. Act. J.* **2** 152–167.

[3] Chaubey, Y. P.; Garrido, J. and Trudeau, S. (1998). On the computation of aggregate claims distributions: some new approximations. *Insurance Math. Econom.* **23** 215–230.

[4] Lukocius, V. (2007). Accuracy of approximations in actuarial overdispersion models. Memorandum, Department of Applied Mathematics, University of Twente, The Netherlands.

[5] Lukocius, V. (2008). Statistical analysis of dependence effects on insurance portfolios. Ph.D. thesis, University Twente, The Netherlands.

[6] Reijnen, R.; Albers, W. and Kallenberg, W. C. M. (2005). Approximations for stop-loss reinsurance premiums. *Insurance Math. Econom.* **36** 237–250.

[7] Saha, K. K. and Paul, S.R. (2005). Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. *Biometrics* **61** 179–185.

[8] Teicher, H. (1965). On random sums of random vectors. *Ann. Math. Statist.* **36** 1450–1458.