Memorandum 1812 (November 2006), Department of Applied Mathematics, University of Twente, Enschede. ISSN 0169-2690
Available from: http://www.math.utwente.nl/publications

Small dependencies and large actuarial risks *

Willem Albers, Wilbert C.M. Kallenberg and Viktor Lukocius

Department of Applied Mathematics
University of Twente
P.O. Box 217, 7500 AE Enschede
The Netherlands

Abstract Methods for computing risk measures such as stop-loss premiums tacitly assume independence of the underlying individual risks. From earlier studies it is already known that this assumption can lead to huge errors even when only small dependencies occur. In the present paper a general model is developed, which covers what happens in practice in a realistic way. Moreover, it is also flexible, in the sense that it allows application in practice. Approximations are presented which are both accurate and transparent and the results obtained are illustrated through some explicit examples.

Keywords and phrases: Aggregate claims; overdispersion; cumulants; dependence; tail events

2000 Mathematics Subject Classification: 62E17, 62P05

1 Introduction and motivation

In this paper we consider large insurance portfolios and focus on the tail behavior of the distribution of S, the sum of the individual claims during a given reference period. A quantity of interest in this respect for example is the stop-loss premium $E(S-a)^+$ for some retention a (here $(x-a)^+$ stands for $\max(0, x-a)$). Typically it is assumed that S is a sum of independent terms, although it is clearly recognized that dependencies do occur in practice (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 quite some time, such effects of dependence have been studied using the notion of comonotonicity (see e.g. Dhaene et al. (2002a,b)). These studies clearly demonstrate that the effects of dependence on stop-loss can be astronomical. However, one could subsequently argue that such strong dependencies will rarely occur in practice and hence that in most cases one still gets away with ignoring dependence. The idea is that as long as dependence effects are 'small on the average', one can hope that the effects on relevant

*This research was supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economic Affairs.

risks, such as stop-loss, will be small as well.

Precisely this optimism has been scrutinized in Albers (1999), using local dependence models, and shown to be entirely unfounded and misleading. 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 a factor (and not a percentage!) between 2 and 6 as compared to the corresponding independent case. Hence there is ample cause to take small dependence effects seriously and to figure out what can be done about it in actuarial practice. This will be the topic of the present paper.

First we will briefly discuss what has been achieved so far. A major complication one immediately encounters when attacking this problem is the fact that the distribution of S is often rather intractable. This is already the case under independence and obviously it will be even more of a problem once dependence has to be taken into account as well. Hence there is need for approximations to the distribution of S which are both accurate and transparent. The need for the first property is obvious; the second is vital in getting a good overall picture of how risks like the stop-loss premium actually depend on the underlying parameters. In Albers (1999) Edgeworth expansions are proposed for this purpose. These well-known refinements of the crude normal approximation, based on the third (and fourth) cumulant(s), work very well in the setup considered there. They are indeed very accurate, especially if the fourth cumulant is used as well. Moreover, they nicely reveal that - and when and how - a bit of dependence strongly affects stop-loss premiums. Hence the goal of that paper has been achieved: crushing the optimistic idea that small dependencies do not matter.

However, Albers (1999) is merely a pilot study and as such not yet suitable for actual application in practice. For example, the underlying claim size distributions were simply assumed to be normal. The first step towards broadening the scope was taken in Reijnen et al. (2005). Here more commonly used claim size distributions were considered: Gamma, inverse Gaussian, lognormal and Pareto. As anticipated, for such skewed distributions, the quality of the Edgeworth expansion as an approximation can no longer be taken for granted and other possibilities need to be investigated. Approximations investigated were: normal power, Gamma, inverse-Gaussian and Gamma/inverse-Gaussian. Just like the Edgeworth expansions, these approximations use, in addition to mean and variance of S, its third (and fourth) cumulant(s). A thorough study over a wide range of underlying claim size distributions and retention values shows that for not too skewed situations the Gamma/inverse-Gaussian is the winner, while otherwise the inverse-Gaussian approximation is best. On this basis, the results from Albers (1999) are reconfirmed and strengthened: small dependencies cause large increases in stop-loss. For the skewed claim size distributions considered here, the increase can even be by a factor 500.

Hence in this second step the approach has been upgraded sufficiently to become suitable for application to the independent case and rather simple dependence models: the underlying claim size distributions are now realistic and adequate approximations are once more available. But it largely remains to deal with the dependence problem. For, also in this respect, Albers (1999) merely offers a simple model, which nicely demonstrates how and why things go wrong, but which is not yet directly applicable to practice. In fact,

the starting point there is the independent case with a large portfolio of size m, in which each individual has a certain small probability q of suffering damage. Dependence is then introduced by subsequently assuming that this portfolio in fact consists of $h = m/g$ groups of fixed and equal size g, where we assume without essential loss of generality that m is a multiple of g. Each of these groups has just a tiny probability of getting exposed to a special cause (such as an epidemic, an accident, a hurricane, etc.).

If this special cause indeed occurs for a given group, the probability of damage for all its members rises from the standard value q to a typically substantially larger value r. Actually, this r may even be 1: the special cause then is a fatal event, leading to sure damage for the whole group involved. Since the probability of the occurrence of a special cause is assumed to be really tiny, it remains possible to set up this model in such a way that merely a fraction ε (between 1% and 5%) of the expected number of claims has to be attributed to this dependent special part. But beyond the mean, this 'ε-fraction idea' no longer holds true. In Albers (1999) it is demonstrated, using the transparency of the approximations, that the relative change of the variance already behaves like εgr, while for the $j^{th}(j = 3, 4)$ cumulant one even gets $\varepsilon(gr)^{j-1}$. Hence it is evident that larger group sizes g in combination with largely increased values r will lead to considerable effects. Moreover, it is also clear that the 'effective' group size gr actually is the quantity that matters. In other words, it seems to make little difference whether g is huge and r not much larger than q, or whether g is not that large but r/q is huge (e.g. $r = 1$), as long as the product gr is kept constant.

In addition to providing a valuable qualitative feeling for the behavior under dependence, the model described above of course also provides the possibility to obtain quantitative results. For each given combination of dependence parameters (ε, g, r), the effect on the stop-loss premium can be adequately approximated. (Hence the aforementioned increase factors between 2 and 6.) However, such results are of limited use, because in practice portfolios will rarely behave in such a nice and orderly fashion. Typically, not everyone will be in one and precisely one special risk group and these groups will also not all have the same size g. More importantly, once a special cause has occurred, it is quite conceivable that it can be determined which claims are due to it and which are not. Note that even this is not trivial: whether home owner's damage is due to having been hit by a hurricane is pretty obvious, but whether an employee's illness is 'individual' or due to an epidemic (e.g. flu), is already less straightforward. At least, in company records for employee illness this distinction is typically not made (yet). But another problem is that it will often be quite hard or outright impossible to figure out what precisely the group size has been for a particular special cause. In other words, to find the size g for a particular group, one would not only need the number of claims due to the special cause defining this group. In addition, it would also be necessary to know how many people were exposed to it, but luckily did not suffer damage (e.g. those employees escaping the flu epidemic that hit their department). Fortunately, the fact that this kind of information will often be out of reach, is not as dramatic as it may seem, since we already noted that the 'effective' group size gr seems to be the quantity that matters, rather than the separate g and r. In fact, in what follows g and r will gradually disappear from sight. These parameters mainly

serve to explain how our final model can be linked to the level of the individual insured.

In the above, the motivation for attacking the problem has been discussed, as well as the results obtained so far. This discussion also helps to make clear what should be done next: we need a more general and more flexible model for small dependence effects under consideration. More general, for example in the sense that the model allows groups of varying sizes, which moreover may overlap and on the other hand do not have to span the whole portfolio. In other words, an insured can be susceptible to $0, 1, 2 \ldots$ special causes, rather than to precisely 1. More flexible, in the sense that the model can be implemented in practice, i.e. it does not require knowledge which is and will remain unavailable from the data. For example, it will not need the group sizes itself, but only the realized numbers of special claims.

Taking these side conditions of generality and flexibility into account, we will propose a suitable model in Section 2 and subsequently analyze it in some detail. We will show how it is related to the more special cases treated in Albers (1999) and Reijnen et al. (2005), but also indicate openings towards further generalization. With the model chosen we thus try to strike a balance between simplicity and ease of application on the one hand and realistic description on the other. Our model will always contain the independence model as a special case and in that sense it will definitely be superior: employing more parameters in principle guarantees a better fit to reality. On the other hand, once the illusion of independence has been abandoned, one might be tempted to go all the way for a 'really realistic' model and as a result keep adding parameters. However, contacts with practitioners indicate that this is quickly bound to become too complicated for practical implementation. Hence the model we propose may still be a simplification of reality, but it will be much less so than the starting point, the independence model. (Recall the remark, attributed to Tukey: "All models are wrong, but some are more wrong than others".)

Just as in the papers quoted, we will need sufficiently accurate approximations to the distribution of S under the present model. Obvious candidates are the approximations from Reijnen et al. (2005). As we remarked earlier, these are all based on the mean and variance of S, together with its third (and fourth) cumulant(s). These quantities are evaluated for the given model in Section 3. Clearly, it is of interest how these three (four) quantities are affected by the dependence parameters from the model considered in the present paper, i.e. to generalize the simple $\varepsilon(gr)^{j-1}$- result from Albers (1999). This aspect is covered in Section 4.

Next, Section 5 is devoted to reporting the results of an extensive numerical study on the accuracy of the approximations. Fortunately, it turns out that the results obtained in Reijnen et al. (2005) more or less continue to hold for the present, considerably more general, model. Just as in that former paper, a simple rule of thumb is presented. Using such a rule, it is easy to determine whether a satisfactory approximation is possible, as well as which choice is best in that case. In addition, it is convincingly demonstrated that within the area where the approximations work well, the dependence effect can indeed already be very large. To this end, some representative numerical examples are given in which stop-loss premiums widely differ when computed either taking the dependence into account (i.e., using the model proposed here) or by simply ignoring such dependence (i.e.

by just using the simple independent model). Finally, in Section 6 a brief summary is given, as well as some indications for further work involving the estimation aspects.

2 The model

First briefly consider the independence model. We have a large portfolio with m insured, each of which has a small probability q to report a claim C, which is a random variable (rv). Hence S can be represented as

$$S = \sum_{j=1}^{m} Y_j C_j, \tag{2.1}$$

where all rv's involved are independent, $P(Y_j = 1) = 1 - P(Y_j = 0) = q$ and C_j is distributed as C, $j = 1, \ldots, m$. For this individual model with identically distributed claims we can also use the representation

$$S = \sum_{i=1}^{N} C_i, \tag{2.2}$$

where N, the total number of claims, is binomial with parameters m and q (for short: $\mathrm{Bin}(m, q)$), the C_i are again independent copies of C, and moreover N and the C_i are independent. From there it is a minor step to a collective model: as m is large and q is small, it makes little difference if we let instead N be Poisson with parameter λ (for short: $P(\lambda)$), where $\lambda = mq$.

In line with (2.2), we now introduce the following dependence model:

$$S = \sum_{i=1}^{N} C_i + \sum_{i=1}^{H} \sum_{j=1}^{G_i} D_{ij}. \tag{2.3}$$

Clearly, more explanation is required here than in the case of (2.2). To begin with, all rv's occurring in (2.3) are once again assumed to be independent. The first term essentially is the same as the whole of (2.2). The only modification is that here we let N be $P((1-\varepsilon)\lambda)$, for some small $\varepsilon > 0$. Hence indeed a fraction ε of the expected number of claims is removed. This we will find back in the second term, as follows. First we have the rv H, which stands for the number of special causes that have occurred during the reference period. Next, the G_i are independent copies of a rv G, and represent the realized numbers of claims due to each of the special causes. Let μ_X denote the expectation of any rv X. We choose H to be $P(\varepsilon\lambda/\mu_G)$, thus indeed obtaining a fraction ε of the expected number of claims in the second term of (2.3). Finally, we have the claim amounts D_{ij}, which are independent copies of a stochastic special claim size D. When comparing models (2.2) and (2.3), it is natural to require the same expected total claim amount. Denoting the claim amount in (2.2) temporarily by $C - IND$, this is realized by requiring

$$\mu_{C-IND} = (1 - \varepsilon)\mu_C + \varepsilon\mu_D. \tag{2.4}$$

In the sequel we shall typically let $D = C$, i.e. the claim amounts are assumed to be identically distributed, whether we have a special cause or not. Consequently, μ_{C-IND} will usually simply equal μ_C. Of course, it is tempting to consider the general case where $D \neq C$, but we really want to keep the number of additional parameters limited (cf. the discussion in the Introduction). Note that many other generalizations can also be easily thought of. To give but a few examples: the D_{ij} can have different distributions for varying i, all kinds of dependencies can exist between the rv's 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. But it seems better to refrain from these generalizations as well for the moment. First we will investigate if and when the present assumptions allow implementation. Once this has been successful, more elaborate schemes can follow.

To get some additional feeling for the area the present model does cover, we shall relate it in passing to the special cases treated before. In analogy to the remark about individual and collective models following (2.2), we can translate matters back for (2.3) as well. We then have:

Remark 2.1 Let for the moment N be Bin$(m, (1-\varepsilon)q)$, H Bin$(h, \varepsilon q/r)$ with $h = m/g$ and G_i Bin(g, r) for all i. Then again $\mu_N = (1-\varepsilon)\lambda$ with $\lambda = mq$, $\mu_H = \varepsilon mq/(gr) = \varepsilon\lambda/\mu_G$, as $\mu_G = gr$. Similarly to (2.1) we can write:

$$S = \sum_{i=1}^{h} \sum_{j=1}^{g} \{X_{ij} C_{ij} + V_i Z_{ij} D_{ij}\}, \tag{2.5}$$

where $P(X_{ij} = 1) = 1 - P(X_{ij} = 0) = (1-\varepsilon)q$, $P(V_i = 1) = 1 - P(V_i = 0) = \varepsilon q/r$ and $P(Z_{ij} = 1) = 1 - P(Z_{ij} = 0) = r$ and the C_{ij} and D_{ij} are the claim amounts. Hence the distribution of S in (2.5) is the same as that of S in (2.3) with now $N = \sum_{i=1}^{h} \sum_{j=1}^{g} X_{ij}$, $H = \sum_{i=1}^{h} V_i$ and $G_i = \sum_{j=1}^{g} Z_{ij}$ having the abovementioned binomial distributions.

Indeed, according to (2.5) the j^{th} insured in the i^{th} group has a claim probability $(1-\varepsilon)q$, just like everybody else. Moreover, in addition he/she has a claim probability r for a special cause damage, provided that $V_i = 1$, i.e. his/her group has been hit. Hence we have now traced back matters to the individual level. Essentially, (2.5) is the model considered in Albers (1999), already briefly sketched in the Introduction. A minor difference of little importance concerns the following aspect. In (2.5), the insured can have both an ordinary and a special claim. One can also choose, as was done in Albers (1999), to model with 'or' rather than with 'and'. Note that this calls for replacing $X_{ij}C_{ij}$ by $(1 - V_i)X_{ij}C_{ij}$ in (2.5), and consequently setting $P(X_{ij} = 1) = 1 - P(X_{ij} = 0) = (1-\varepsilon)q/(1-\varepsilon q/r)$. In that case, a special claim can also contain an ordinary claim.

As concerns the model used in Reijnen et al. (2005), this roughly is (2.5) with the modification that now always $r = 1$. Hence the special cause is fatal: if it hits a group, everybody in it has damage. Thus the G_i are not random but fixed, with value $g\,(= gr)$. Yet another parameter, in addition to ε and g, in that paper is the ratio μ_D/μ_C. This is

not fixed at 1, but can range from 1-3, for example to account for the fact that there the special claim (possibly) also contains the ordinary claim. □

Apparently, it makes little difference whether we let N and H be binomial or Poisson. Likewise, whether G is Bin$(gr, 1)$ (i.e., degenerated at gr), or G is Bin(g, r), also seems to be of minor importance, which suggests that we might as well let G be Poisson too: then G is $P(\mu_G)$ with $\mu_G = gr$. (In Section 3, during the evaluation of the moments and cumulants, this will become explicit, but for the moment we just concentrate on the setup of the model.) Of course, one could argue that this step should have been made at the start and thus the use of r could have been avoided completely. However, as stated before, we feel that it is important to make clear that actually a wide range of possibilities is essentially covered. (This also explains why we had to refer rather extensively to the previous work.) For example, it is clear now that we do cover the situation where people may have been susceptible to a special cause but escaped damage. Moreover, note that in going from the individual to the collective model we have in effect also incorporated the possibility that insured are susceptible to more than just one special cause. There is no fixed link to underlying individuals, so the different G may actually include the same insured.

Consequently, the closest collective analogue to Albers (1999) and Reijnen et al. (2005) generates claims according to three Poisson processes. The main stream is formed by the ordinary claims and has intensity $(1 - \varepsilon)\lambda$. In addition, special events are generated at a rate $\varepsilon\lambda/\mu_G$. For each special event, the third process, with intensity μ_G, provides an outcome. This seems to be the most parsimonious description that still captures the essential features. We have a parameter λ (and no longer a separate large m and small q) which characterizes the ordinary claim process, a parameter ε which specifies the fraction of the claims due to dependence and a parameter μ_G (and no longer a group size g and a special risk probability r) that characterizes the 'lumpiness' of the dependent part. The more lumpy, the more dangerous: cf. the $\varepsilon\mu_G^{j-1}$-result mentioned before.

Nevertheless, while defining (2.3) we did choose N and H to be Poisson, but we did not specify a distribution for G. There is good reason for this restraint, which can be seen as follows. By going to the collective model, we have already removed quite a bit of the rigidity of the previous models. However, the assumption that all special causes lead to similar groups, remains rather awkward. To be realistic, it seems required to allow variation here as well. This can be achieved in a very simple way by requiring that G has the following mixture distribution:

$$\text{given some rv } L = l,\ G \text{ is distributed as } P(l). \tag{2.6}$$

As before, all rv's N, C_i, H, G_i and D_{ij} in (2.3) are assumed to be independent. But now for each G_i, we first select an L_i, and given its outcome l_i we subsequently let G_i be $P(l_i)$. To see that this solution is not only simple, but also make sense, it may once more help to take a brief look at the individual analogue:

Remark 2.2. If we try to introduce more variation into the individual model, matters get rather complicated and thus the transparency is lost to a large extent. On the other hand,

it remains possible to trace back what actually happens to the individual level. In fact, we might consider B categories, and for each $b = 1, \ldots, B$ look at m_b insured, divided into groups of size g_b with increased probability r_b, fraction ε_b, etc. Then an aggregate S will follow by adding a summation over the additional index b in (2.5). The condition about a fraction ε in the dependent part for example will look like $\sum_{b=1}^{B} \varepsilon_b m_b q = \varepsilon m q$. Indeed, this model may be very explicit, but it has also become quite intractable.

However, it nicely indicates the way to follow within the collective framework, as here the very same mechanism can be incorporated quite smoothly. First, switching back from binomial to Poisson leads to H_b that are $P(\lambda_b^*)$ with $\lambda_b^* = \varepsilon_b m_b \lambda/(m\mu_{bG})$, in which $\lambda = mq$ and $\mu_{bG} = g_b r_b$. But this implies that we can also start from a sum process H which is $P(\lambda^*)$, with $\lambda^* = \sum_{b=1}^{B} \lambda_b^*$. Given that an event occurs in H, we subsequently assign it to the b^{th} category with probability $\lambda_b^*/\lambda^* = \{\varepsilon_b m_b/\mu_{bG}\}/\{\sum_{a=1}^{B} \varepsilon_a m_a/\mu_{aG}\}$. But in this way, we actually have defined a rv L for which $P(L = \mu_{bG}) = \lambda_b^*/\lambda^*$, and we can indeed wrap up matters as formulated in (2.6). □

Hence this look at the individual case has made clear that (2.6) is not just a haphazard new proposal for choosing G, after the more simple possibilities of letting G be fixed (Reijnen et al. (2005)), binomial (Albers (1999)) or Poisson (above). It acknowledges the fact that modeling the aforementioned lumpiness of the dependence part may very well require more variation than captured by a single fixed Poisson distribution.

In the special setup in Remark 2.2, the additional variation is achieved by taking a discrete L, distributed over B possible values for μ_G, but in principle any choice for L is feasible, as long as it is nonnegative. Note that we in fact encounter a general phenomenon here, not linked to this particular application area. For this so-called overdispersion problem, a prototype choice is to let L be Gamma distributed with parameters α and β (for short: $Ga(\alpha, \beta)$), i.e.

$$f_L(x) = \beta^\alpha x^{\alpha-1} \frac{\exp(-\beta x)}{\Gamma(\alpha)}. \tag{2.7}$$

Then letting G be $P(l)$ given $L = l$ leads to a unconditional distribution for G which is negative binomial on $\{0, 1, \ldots\}$ with parameters α and $\beta/(\beta+1)$ (for short: NB$(\alpha, \beta/(\beta+1))$).

What is important for us at this point is the following observation. It seems quite reasonable to model N and H as Poisson rv's. Moreover, a Poisson component can be used as well in defining G, but this may not be sufficient. The remaining component of the model is L, and its distribution type is in principle hard to postulate. Roughly speaking, three options exist. The first is to let L be fixed (i.e. $L = \mu_G$) and ignore possible overdispersion. Next, the second is to select some parametric family (e.g. to let L be $Ga(\alpha, \alpha/\mu_G)$) and also estimate the additional parameters from the data. Finally, the third is to assume nothing at all and to estimate all moments and cumulants of L that are required in a nonparametric way. Clearly, the last approach is conceptually superior, but hardest to implement: will there be sufficient data to achieve this? The first approach poses no additional estimation problem, but can easily be too strict to allow a good fit.

Hence the parametric approach might be a good compromise between these two extremes. Note that we do not have L-observations. So, for the estimation of the moments and cumulants we should typically use observations from G.

To obtain some feeling for types of parametric families that might be realistic, we performed some simulation experiments. A company was modeled, consisting of a number of departments. Suitably looking probabilities were chosen for employees falling ill, either individually or by getting flu. Also, probabilities were assigned for infecting other people. Distributions were chosen for the length of absence due to these illnesses and for the damages incurred (in terms of daily salaries). Playing around with the parameter values (performing a sensitivity analysis, to put it more formally) suggested rather heavy tailed types of distributions for G, like a (discrete) Pareto. Another possibility in this respect would be to use (2.6) with $Ga(\alpha, \alpha/\mu_G)$ for L with α small: the probabilities $P(G = j)$ then go down very slowly in j, as the stepwise rate $P(G = j + 1)/P(G = j)$ equals $(\alpha + j)/\{(j + 1)(1 + \alpha/\mu_G)\}$.

Summarizing our model building efforts in this section, we have identified and described the following elements in (2.3). First the parameters: the overall intensity λ, the fraction ε and the lumpiness μ_G. In addition, we have the random aspect L of the group size and, of course, the random claim size amounts C and D, which was already amply discussed in the independent model (cf. Reijnen et al. (2005)).

3 Moments and cumulants

As mentioned in the Introduction, we will need an adequate approximation to the distribution of S under the dependence model from (2.3). All candidates for this purpose (cf. Reijnen et al. (2005)) are based on some (central) moments and cumulants of S. Hence in this section we shall obtain these in terms of λ, ε, μ_G and (suitable characteristics of) L and C. Clearly, the advantage of such explicit expressions will be that the influence of the various components in the model on the moments and cumulants in question is made transparent. First we introduce the notation required. For a single rv X use μ for its mean and σ^2 for its variance. Moreover, for general (central) moments use $\nu_j = EX^j$ and $\mu_j = E(X - \mu)^j$, $j = 1, 2, \ldots$ (hence in particular $\nu_1 = \mu$ and $\mu_2 = \sigma^2$). In addition, let $\gamma = \sigma/\mu$ denote the coefficient of variation (cv). If more than one rv is involved, the respective indices are added, e.g. μ_X, ν_{3X}, etc. Next, the moment generating function Ee^{tX} will be denoted by $M(t)$; as is well-known, $\nu_j = M^{(j)}(0)$, $j = 1, 2, \ldots$ Similarly, let $K^*(t) = \log M(t)$ denote the cumulant generating function, then the j^{th} cumulant $\kappa_j^* = K^{(j)}(0)$. In particular, we have that $\kappa_1^* = \mu$, $\kappa_2^* = \sigma^2$, $\kappa_3^* = \mu_3$, $\kappa_4^* = \mu_4 - 3\sigma^4$. Moreover, let K be the corresponding function for the standardized rv $(X - \mu)/\sigma$, then the standardized cumulants κ_j satisfy $\kappa_1 = 0$, $\kappa_2 = 1$, $\kappa_3 = \mu_3/\sigma^3$, $\kappa_4 = \mu_4/\sigma^4 - 3$.

The approximations proposed in Reijnen et al. (2005) require mean, variance, third (and possibly also fourth) cumulant of S, i.e. κ_{jS}^* for $j = 1, 2, 3, 4$. Consequently, it seems best to just evaluate K_S^*. For the expression in (2.2) it is straightforward that $M_S(t) = M_N(\log M_C(t))$. In other words, we simply have that $K_S^* = K_N^* \circ K_C^*$. Likewise,

in case of our dependence model (2.3) we arrive at

$$K_S^* = K_N^* \circ K_C^* + K_H^* \circ K_G^* \circ K_D^*. \tag{3.1}$$

Note that for any explicit choice for the distributions of N, H, G, C and D, we can directly obtain $\kappa_{jS}^* = K_S^{*(j)}(0)$ for $j = 1 - 4$ from (3.1). For example, we have $\mu_S = \kappa_{1S}^* = \mu_N\mu_C + \mu_H\mu_G\mu_D$ and

$$\sigma_S^2 = \kappa_{2S}^* = \{\sigma_N^2\mu_C^2 + \mu_N\sigma_C^2\} + \{\sigma_H^2\mu_G^2\mu_D^2 + \mu_H\sigma_G^2\mu_D^2 + \mu_H\mu_G\sigma_D^2\}. \tag{3.2}$$

Subsequently, such results can be plugged into a suitable approximation and an equally explicit result can be obtained for each given choice.

Formulas (3.1) and (3.2) hold for quite general N and H. However, for the purpose of obtaining a qualitative overall insight into the behavior of the distribution of S as a function of the underlying model parameters and characteristics, (3.1) is still a bit too general. Consequently, we will now use that we have in fact assumed N and H to be Poisson. Hence, observing that a rv X which is $P(\theta)$ has $K_X(t) = \theta(e^t - 1)$, we arrive for N and H Poisson at

$$K_S^*(t) = (1-\varepsilon)\lambda\{M_C(t) - 1\} + (\varepsilon\lambda/\mu_G)\{M_G(K_D^*(t)) - 1\}. \tag{3.3}$$

From (3.3) we immediately obtain the obvious $\mu_S = (1-\varepsilon)\lambda\mu_C + \varepsilon\lambda\mu_D$, but also that

$$\begin{aligned}
\sigma_S^2 &= (1-\varepsilon)\lambda\nu_{2C} + (\varepsilon\lambda/\mu_G)\{\nu_{2G}\mu_D^2 + \mu_G\sigma_D^2\},\\
\kappa_{3S}^* &= (1-\varepsilon)\lambda\nu_{3C} + (\varepsilon\lambda/\mu_G)\{\nu_{3G}\mu_D^3 + 3\nu_{2G}\mu_D\sigma_D^2 + \mu_G\kappa_{3D}^*\},\\
\kappa_{4S}^* &= (1-\varepsilon)\lambda\nu_{4C} + (\varepsilon\lambda/\mu_G)\{\nu_{4G}\mu_D^4 + 6\nu_{3G}\mu_D^2\sigma_D^2 + \nu_{2G}(4\mu_D\kappa_{3D}^* + 3\sigma_D^4) + \mu_G\kappa_{4D}^*\}.
\end{aligned} \tag{3.4}$$

The expressions in (3.3) and (3.4) are already reasonably tractable. By way of illustration, we shall consider some examples. If G is fixed, i.e. $G = \mu_G = gr$ (cf. Reijnen et al. (2005)), $M_G(K_D^*(t))$ in (3.3) boils down to $\{M_D(t)\}^{\mu_G}$ and the ν_{jG} in (3.4) to μ_G^j. The aforementioned $\varepsilon\mu_G^{j-1}$-effect indeed become visible already in (3.4). When G is $\mathrm{Bin}(g, r)$ instead (cf. Albers (1999)), we get $M_G(K_D^*(t)) = \{1 + r(M_D(t) - 1)\}^{\mu_G/r}$ in (3.3). Supposing that μ_G is substantially larger than 1 - which clearly is the case of interest - the ν_{jG} in (3.4) are dominated by μ_G^j (e.g. $\nu_{2G} = \mu_G^2 + \mu_G(1-r)$). Obviously, for $r = 1$, we are back in the fixed case, while for $r \to 0$ (and $g = \mu_G/r \to \infty$) we get the Poisson case. Here G is P and $M_G(K_D^*(t)) = \exp\{\mu_G(M_D(t) - 1)\}$, while the ν_{jG} in (3.4) are still mainly characterized by μ_G^j (e.g. here $\nu_{2G} = \mu_G^2 + \mu_G$).

As these examples demonstrate, for specific choices of G, we get reasonably explicit results by using (3.3) and (3.4). It just remains to make such a choice as well for C and D in that case. To carry on yet another step for (more) general G, we will resort to the mixture approach we advocated in (2.6). Clearly,

$$E(e^{tG}|L = l) = \exp(l(e^t - 1))$$

and thus $M_G(t) = E(\exp(L(e^t - 1)) = M_L(e^t - 1)$, which in its turn leads to

$$K_S^*(t) = (1-\varepsilon)\lambda\{M_C(t)-1\} + (\varepsilon\lambda/\mu_G)\{M_L(M_D(t)-1)-1\}. \tag{3.5}$$

Under (3.5), the expressions from (3.4) now become

$$\begin{aligned}
\sigma_S^2 &= (1-\varepsilon)\lambda\nu_{2C} + (\varepsilon\lambda/\mu_G)\{\nu_{2L}\mu_D^2 + \mu_G\nu_{2D}\},\\
\kappa_{3S}^* &= (1-\varepsilon)\lambda\nu_{3C} + (\varepsilon\lambda/\mu_G)\{\nu_{3L}\mu_D^3 + 3\nu_{2L}\mu_D\nu_{2D} + \mu_G\nu_{3D}\},\\
\kappa_{4S}^* &= (1-\varepsilon)\lambda\nu_{4C} + (\varepsilon\lambda/\mu_G)\{\nu_{4L}\mu_D^4 + 6\nu_{3L}\mu_D^2\nu_{2D} + \nu_{2L}(4\mu_D\nu_{3D} + 3\nu_{2D}^2) + \mu_G\nu_{4D}\}.
\end{aligned} \tag{3.6}$$

The structure in (3.6) is very similar to that of (3.4), which is not surprising in view of the similarity between (3.3) and (3.5). Basically, a function K^* is replaced by a function $M^* = M - 1$, and hence the corresponding κ_j^*'s become ν_j's. In this connection, note that (3.5) in terms of M^*'s can be written as $K_S^* = (1-\varepsilon)\lambda M_C^* + (\varepsilon\lambda/\mu_G)M_L^* \circ M_D^*$.

The next reduction in complexity is achieved by letting $C = D$, as announced in Section 2. Then the cumulants become really simple:

$$\begin{aligned}
\sigma_S^2 &= \lambda\nu_{2C}\{1 + \varepsilon\nu_{2L}\mu_C^2/(\mu_G\nu_{2C})\},\\
\kappa_{3S}^* &= \lambda\nu_{3C}\{1 + \varepsilon(\nu_{3L}\mu_C^3 + 3\nu_{2L}\mu_C\nu_{2C})/(\mu_G\nu_{3C})\}.\\
\kappa_{4S}^* &= \lambda\nu_{4C}\{1 + \varepsilon(\nu_{4L}\mu_C^4 + 6\nu_{3L}\mu_C^2\nu_{2C} + \nu_{2L}[4\mu_C\nu_{3C} + 3\nu_{2C}^2])/(\mu_G\nu_{4C})\}.
\end{aligned} \tag{3.7}$$

In particular, the terms involving ε explicitly represent the relative increase compared to the independent case, which corresponds to $\varepsilon = 0$. As now moreover $K_S^* = (1-\varepsilon)\lambda M_C^* + (\varepsilon\lambda/\mu_G)M_L^* \circ M_C^*$, we in fact have the general expression

$$\kappa_{jS}^* = \lambda\nu_{jC}\{1+\varepsilon A_j(L,C)\},\ A_j(L,C) = (M_L^* \circ M_C^*)^{(j)}(0)/\{M_L^{*(1)}(0)M_C^{*(j)}(0)\} - 1. \tag{3.8}$$

4 Impact of model parameters

The results on moments and cumulants from Section 3 are in a sufficiently transparent form to allow conclusions on how the underlying parameters affect these outcomes. We shall study such effects in this section in a systematic way, assuming model (2.3) in combination with (2.6) and D in distribution equal to C.

(i) role of λ

The parameter λ represents the overall intensity of the process. Typically, as $\lambda = mq$ becomes larger, all approximations will become more accurate: more claims simply means more information. Technically, this can be seen from the standardized cumulants $\kappa_{jS} = \kappa_{jS}^*/\sigma_S^j$, $j = 3, 4$. For, from (3.8) it is clear that these κ_{jS} are proportional to $\lambda^{1-j/2}$, and thus become small as λ increases.

As the order of magnitude of q seems to be more or less given and invariably quite small, the way to increase λ will be to increase m, i.e. to consider very large portfolios. Hence

the conclusion is that it will really help to consider high aggregation levels. Note that so far we have focused on the quality of the approximations. Another, at least as important, point is how λ will be influencing the dependence effect. The answer is two-fold. From the above it is clear that eventually the standardized skewness and kurtosis terms will vanish. Hence the $\varepsilon\mu_G^{j-1}$-effect will in the long run only play a role for $j = 2$ and in that sense increasing λ is beneficial. On the other hand, as is clear from the expression for σ_S^2 from (3.7), λ is just a multiplicative factor.

(ii) role of ε

Fortunately, about this parameter we can be quite brief, as its role is utterly transparent. It is put into the model in a linear way and that is also the way it comes out at the end in (3.8). (Of course, in the standardized cumulants κ_{jS}, it is slightly more complicated.) Its main role is to make very explicit that out of the average claim amount μ_S indeed only a fraction ε has to be attributed to dependent aspects of the model. Given this side condition, the impact of dependence on tail related aspects of S is subsequently studied.

(iii) role of μ_G

This is what we have called the lumpiness of the dependent part. Clearly, the larger μ_G, the bigger the problems caused by the dependence. In the above, we repeatedly mentioned the $\varepsilon\mu_G^{j-1}$-effect in connection with special cases. In the more general situation of (3.8), this effect continues to play a prominent role as a kind of lower bound. For, in κ_{jS}^* we encounter ν_{jL}/μ_G, which is $\geq \mu_L^j/\mu_G = \mu_G^{j-1}$. More precisely, $\nu_{jL} = E(L - \mu_L + \mu_G)^j$ can obviously be written as a polynomial in μ_G of degree j; the larger the $E(L - \mu_L)^k$, the larger the additional contributions to ν_{jL} beyond μ_G^j.

(iv) role of L

Given the value of μ_G, the rv L in addition describes the so-called overdispersion effect (cf. Section 2), which further aggravates the dependence effect discussed in *(iii)*. This is expressed in the replacement of μ_G^{j-1} by the larger quantity ν_{jL}/μ_G. Consider e.g. $j = 2$, then $\nu_{2L} = \mu_G^2 + \sigma_L^2$ and $\nu_{2L}/\mu_G = \mu_G(1 + \gamma_L^2)$, where γ_L is the coefficient of variation of L. Look once more at the prototype example where L is $Ga(\alpha, \alpha/\mu_G)$. As a $Ga(\alpha, \beta)$ distribution has $\sigma^2 = \alpha/\beta^2$, it follows that $\sigma_L^2 = \mu_G^2/\alpha$ and hence that the increase in this case is expressed by a factor $(1 + \alpha^{-1})$. Of course, as a $Ga(\alpha, \beta)$ distribution has $\nu_j = \Pi_{i=0}^{j-1}((\alpha + i)/\beta) = \nu_1^j \Pi_{i=1}^{j-1}(1 + i/\alpha)$, this can also be seen directly. Using a heavy tailed L, as suggested by the simulations mentioned in Section 2, requires α to be small and hence replacement of μ_G by $\nu_{2L}/\mu_G = \mu_G(1 + \alpha^{-1})$ will have a substantial effect.

Note that by looking at $j = 2$ we have indeed concentrated on the overdispersion aspect. But in a sense this term is somewhat misleading in the present context, as the effects for $j = 3$ or 4 are also of potential importance. In fact, it is quite conceivable that a realistic model will need to allow, be it with tiny probability, the possibility of situations in which a very large expected group size occurs. But that precisely means a very large value of L. Consequently, in addition to a nonnegligible contribution through σ_L^2, we also have to reckon with considerable effects from skewness and kurtosis. To use the $Ga(\alpha, \alpha/\mu_G)$

example once more, we recall that $\nu_j/\mu_G = \mu_G^{j-1}\Pi_{i=1}^{j-1}(1+i/\alpha)$, which indicates a substantial further increase for α small. Another way to see this is to observe that in the Gamma case the standardized cumulants satisfy $\kappa_{3L} = 2\alpha^{-1/2}$ and $\kappa_{4L} = 6\alpha^{-1}$.

(v) role of C

In a sense, the remarks about the impact of C run somewhat parallel to those about L. A major difference of course is that the distribution of C is already important under independence, i.e. for $\varepsilon = 0$. The starting point again is the case of fixed claim amounts, i.e. $C \equiv \mu_C$. However, unlike μ_G, the parameter μ_C is a dummy variable: the problem remains invariant under scale transforms. Hence in illustrative examples, μ_C can freely be set at some representative value. As $\nu_{jC} = \mu_C^j$, the expressions in (3.7) become very simple: in the ε-term, all factors involving a subscript C just drop out.

The next step is the special choice from Albers (1999), where C is assumed to be normal. This means that σ_C^2 is added to the picture, but skewness and kurtosis effects are still absent, as $\kappa_{3C} = \kappa_{4C} = 0$ there. Then the coefficient of variation γ_C suffices to describe the behavior w.r.t. C. We obviously have $\nu_{2C}/\mu_C^2 = 1 + \gamma_C^2$, but also $\nu_{3C}/\mu_C^3 = 1 + 3\gamma_C^2$ and $\nu_{4C}/\mu_C^4 = 1 + 6\gamma_C^2 + 3\gamma_C^4$.

Subsequently, in Reijnen et al. (2005) more realistic choices of C are investigated, for which skewness and kurtosis are positive: Gamma, inverse-Gaussian, lognormal and Pareto. The expressions for the third and fourth moments of C to be used in (3.7) now become

$$\nu_{3C}/\mu_C^3 = 1 + 3\gamma_C^2 + \gamma_C^3\kappa_{3C}, \nu_{4C}/\mu_C^4 = 1 + 6\gamma_C^2 + 4\gamma_C^3\kappa_{3C} + (\kappa_{4C} + 3)\gamma_C^4. \tag{4.1}$$

Typically, the distributions just mentioned are characterized by two parameters. As was argued above, the mean acts as a dummy variable and can be set at an arbitrary illustrative reference value. Hence only one free parameter is left. Quite often it will be feasible to keep using γ_C for this purpose. It has the advantage of being scale invariant, which also is the case for κ_{jC}. Often this will result in relatively simple expressions. For example, if C is $Ga(\alpha, \alpha/\mu_C)$, then $\gamma_C = \alpha^{-1/2}$ and thus $\kappa_{3C} = 2\alpha^{-1/2} = 2\gamma_C$ and $\kappa_{4C} = 6\alpha^{-1} = 6\gamma_C^2$. Next suppose we let C instead be inverse-Gaussian with parameters α and β (for short: $IG(\alpha, \beta)$), i.e.

$$f_C(x) = \alpha(2\pi\beta)^{-1/2}x^{-3/2}\exp\left\{ -\frac{(\alpha - \beta x)^2}{2\beta x} \right\}. \tag{4.2}$$

Then, just as in the Gamma case, $\mu_C = \alpha/\beta$, $\sigma_C^2 = \alpha/\beta^2$, $\gamma_C = \alpha^{-1/2}$ and $\beta = \alpha/\mu_C$ produces the dummy mean value μ_C. But now $\kappa_{3C} = 3\alpha^{-1/2} = 3\gamma_C$ and $\kappa_{4C} = 15\alpha^{-1} = 15\gamma_C^2$, which illustrates that this family models more pronounced skewness and kurtosis than the Gamma family.

Note that similar remarks can be made about the distribution families chosen to model L. Cf. the Gamma example in that case: obviously there we can also write $\kappa_{3L} = 2\alpha^{-1/2} = 2\gamma_L$ and $\kappa_{4L} = 6\alpha^{-1} = 6\gamma_L^2$.

Summarizing, the larger the overall intensity λ of the process, the better it is, as the impact of κ_{jS} for $j = 3, 4$ will decrease in this way. The fraction ε has a straightforward linear effect on the κ^*_{jS} for all $j = 1 - 4$. The average group size ('lumpiness') μ_G has a particularly strong effect, as it shows up in the form of a factor μ_G^{j-1}. This effect is further strengthened by L through overdispersion, as expressed by its coefficient of variation γ_L. Matters do not stop there: beyond overdispersion there are possibly heavy tails effects. These are accounted for through the choice of a suitable family of distributions for L and reflected in how the corresponding κ_{jL} depend on γ_L. A similar account holds for the distribution family selected for the claim amounts C. Only note that this can already be characterized by γ_C, as the average claim amount μ_C plays no role. Hence in terms of parameters we have now lined up:

$$\lambda, \varepsilon, \mu_G, \gamma_L \text{ and } \gamma_C. \tag{4.3}$$

Beyond this, there of course still remain the particular choices of families of distributions for C and L.

5 Numerical study

The main objective of this section is to present a number of typical examples in order to illustrate the impact of ignoring dependence. However, such examples will clearly have to be based on a suitably chosen approximation, preferably from among those considered in Reijnen et al. (2005). Hence, before it makes sense to look at the examples in question, we have to make sure these approximations are sufficiently accurate in the cases considered. To this end, an extensive numerical study has been carried out in Lukocius (2006). For full details, we refer to this report; here we shall present a brief overview of the results obtained.

In doing so, it seems adequate to follow the steps from the previous section. Consequently, we have:

(i) choice of λ

In Albers (1999) the m and q used correspond to values of $\lambda = mq$ between 12 and 30. Already for such small values, the approximations work nicely. But of course, there we only deal with the extremely well-behaved case of normally distributed claim amounts. In Reijnen et al. (2005), we already move on to $\lambda = 80$ in our leading example, thus acknowledging the fact that now more realistic claim size distributions are used. In the present case, through the introduction of L we have added the substantial complication of a mixture distribution for the random group sizes G. But for $\lambda = 80$ and e.g. $\varepsilon = 0.03$, the expected number of special claims is merely 2.4. If we then take the lumpiness $\mu_G = gr = 10$ (cf. Table 4 from Albers (1999)), the expected number of such groups would only be 0.24. This really seems to be too small, especially if we also want to take the mixture effect into account.

Consequently, in the present study we will use $\lambda = 400$. A value of this order of

magnitude seems to be minimally required, because otherwise the events of interest will be encountered only very rarely. As mentioned in Section 4, really large portfolios will thus be required, which is only natural because we are interested in tail phenomena. Clearly, if λ increases even further, the quality of the approximations will only improve.

(ii) choice of ε

Again we can be short here. The intended message is that a small fraction of dependence can create big problems already. In the introduction we mentioned a range of 1 to 5% for ε, so let us fix matters in what follows at $\varepsilon = 0.03$. In view of the more or less linear way in which ε occurs in the formulae, this will be completely representative.

(iii) choice of μ_G

The lumpiness aspect is already present in the previous models, so for μ_G we simply use the values 5, 10 and 20, just as in Reijnen et al. (2005).

(iv) choice of L

Considerable experimentation shows that the prototype example where L is $Ga(\alpha, \alpha/\mu_G)$ indeed performs nicely, in the sense that reasonably looking distributions for G result. The choices for μ_G involved are already given in *(iii)*, so it remains to select a range for α, or equivalently for $\gamma_L = \alpha^{-1/2}$. It turns out that $\gamma_L < 2$ works fine. A second choice that proves to be quite suitable is the inverse-Gaussian family (cf. (4.2)). For this $IG(\alpha, \alpha/\mu_G)$ case once more $\gamma_L = \alpha^{-1/2}$ and even $\gamma_L < 4$ is fine here. A third choice is the lognormal family. However, this turns out to be too extreme for our purposes. Huge cumulants result and the tails really seem too heavy to adequately model the mixing aspect of G.

(v) choice of C

As mentioned in Section 4, the parameter μ_C is a dummy variable, so without loss of generality we fix it at the illustrative value of 10^5, just as in the previous studies. As concerns the choices for C, we largely follow Reijnen et al. (2005). Here we consider the Gamma, inverse-Gaussian and lognormal families. Given μ_C, it remains to indicate ranges for γ_C. In the first two cases we let $\gamma_C \leq 1$ and in the latter $\gamma_C \leq 1.6$.

In the above we have made explicit which range of parameters from (4.3) has been studied, together with the choices of distribution families involved for C and L. Just as in Reijnen et al. (2005), the fortunate conclusion is that the inverse-Gaussian (IG) and/or the Gamma-inverse-Gaussian (G-IG) approximation are very suitable over this, quite wide, region. (For a discussion about the criteria used to judge this, we refer to Reijnen et al. (2005); here we have adapted the same approach.) To be explicit, for the IG case (see Chaubey et al. (1998)), one approximates 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 agree with those of S. As $\mu_T = x_0 + \alpha/\beta$, $\sigma_T = \alpha^{1/2}/\beta$ and $\kappa_{3T} = 3\alpha^{-1/2}$, it readily follows that this will indeed hold for

$$\alpha = \left(\frac{3}{\kappa_{3S}}\right)^2, \ \beta = \frac{3}{\sigma_S \kappa_{3S}} \text{ and } x_0 = \mu_S - \frac{3\sigma_S}{\kappa_{3S}}. \tag{5.1}$$

The resulting T is used to approximate S in all subsequent computations such as stop-loss. In the latter case this e.g. results in the following expression:

$$E_T(S-a)^+ = \sigma_S \int_k^\infty \frac{x-k}{\sqrt{2\pi(1+\frac{1}{3}x\kappa_{3S})^3}} \exp\left[-\frac{x^2}{2(1+\frac{1}{3}x\kappa_{3S})}\right] dx, \tag{5.2}$$

where $k = (a-\mu_S)/\sigma_S$.

Similarly, in the Gamma case (see e.g. Seal (1977)) one approximates S by T^* such that $T^* - x_0$ is $Ga(\alpha,\beta)$. The only difference in the first three cumulants as compared to the IG case is that here $\kappa_{3T^*} = 2\alpha^{-1/2}$, so it suffices to replace the 3's by 2's in the numerator of (5.1):

$$\alpha = \left(\frac{2}{\kappa_{3S}}\right)^2,\ \beta = \frac{2}{\sigma_S\kappa_{3S}} \text{ and } x_0 = \mu_S - \frac{2\sigma_S}{\kappa_{3S}}, \tag{5.3}$$

to achieve the desired match between T^* and S. Let $F_G(x;\alpha,\beta)$ denote the distribution function for $Ga(\alpha,\beta)$ (cf. (2.7)) and let $\overline{F}_G = 1 - F_G$, then it follows that

$$E_{T^*}(S-a)^+ =$$

$$\sigma_S\left\{\frac{2}{\kappa_{3S}}\overline{F}_G\left(k+\frac{2}{\kappa_{3S}};\frac{4}{\kappa_{3S}^2}+1,\frac{2}{\kappa_{3S}}\right) - \left(k+\frac{2}{\kappa_{3S}}\right)\overline{F}_G\left(k+\frac{2}{\kappa_{3S}};\frac{4}{\kappa_{3S}^2},\frac{2}{\kappa_{3S}}\right)\right\}. \tag{5.4}$$

The G-IG approximation (see once more Chaubey et al. (1998)) now follows by mixing the densities of T and T^* in such a way that the resulting fourth cumulant agrees with κ_{4S}. As $\kappa_{4T} = 5\kappa_{3T}^2/3$, while $\kappa_{4T^*} = 3\kappa_{3T^*}^2/2$, it follows that in $(1-w)f_{T^*} + wf_T$ one should use

$$w = 10 - \frac{6\kappa_{4S}}{\kappa_{3S}^2}. \tag{5.5}$$

Actually, this also works quite well for values of w outside (0,1); then we simply consider it as an approximation, rather than as a density. Hence the G-IG approximation for the stop-loss is obtained by mixing (5.2) and (5.4) according to the weight given by (5.5). Do note that both the IG and the G-IG approximation are obtained completely through computation: no simulations are involved.

It remains to indicate when the IG approximation should be used and when the G-IG. Just as in Reijnen et al. (2005), a rather simple rule of thumb turns out to be adequate for the present much more general case:

"if $0 \le \kappa_{3C} \le 10$ and $0 \le \kappa_{4S} \le 70$ then use $G-IG$, otherwise use IG". (5.6)

Of course we should keep in mind that (5.6) holds for the region considered. In particular, for this range of parameters and distributions, no situations occur with $\kappa_{3C} > 14$ and $\kappa_{4S} > 70$. Extension to even more extreme situations, involving e.g. more dependence, fewer claims or more skewed distributions, requires care.

As announced at the beginning of this section, once we have verified the accuracy of the

approximations, we can feel free to use these in order to asses the impact of dependence. Note that this is indeed a relatively simple matter: (3.7) provides explicit expressions for the cumulants in terms of the underlying moments, which in their turn are presented in Section 4 as functions of the parameters from (4.3). Hence it remains to plug such cumulants into e.g. (5.1) to arrive at an adequate approximation of the density f_S. Subsequently, this readily produces approximations to measures of tail behavior like VaR or stop-loss $SL = E(S-a)^+$.

For brevity, we shall restrict attention here to SL, which is considered to be one of the foremost quantities to study (cf. Kaas (1993) and Kaas et al. (1994)). Just as in the previous papers, we look at retentions a covering a broad range: $a = \mu_S + k\sigma_S$, with $1 \leq k \leq 3$. We shall evaluate SL for a number of configurations in (4.3) and each time compare the result to the corresponding outcome SL_I for $\varepsilon = 0$, i.e. when dependence is ignored. Note that some good judgment will be needed to interpret the results: as k increases, we move more into the tail and the dependence effect will become more pronounced. Hence the relative error $SL/SL_I - 1$ will definitely increase. On the other hand, for larger k the magnitude of the stop-losses will decrease and hence the absolute error $SL - SL_I$ will decrease. Incidentally, fair comparison of SL and SL_I obviously requires using the same retention a in both. But as we select a by letting k vary in $a = \mu_S + k\sigma_S$, this means that in both instances the same μ_S and σ_S should be applied. We shall always use the values from independence, i.e. $\mu_S = \lambda\mu_C$ and $\sigma_S = \lambda^{1/2}\mu_C(1+\gamma_C^2)^{1/2}$ (cf. (3.7)).

As announced above, we throughout fix λ at 400 and ε at 0.03. As a representative choice for C we take the lognormal family. As always, we let the dummy mean μ_C equal 10^5, while for the coefficient of variation we choose the values $\gamma_C = 0.4$ and $\gamma_C = 1.2$. For L we use the Gamma family, with parameter values such that $\mu_L (= \mu_G) = 5$ or 10 and $\gamma_L = 0$ (corresponding to the case where G is simply Poisson), 0.75 or 1.5. Using the appropriate approximation according to (5.6), we then evaluate for retentions a with $k = 1, 2, 3$ the stop-loss SL and collect the results in Table 5.1. For comparison to the case where dependence is neglected we also include SL_I by adding the choice $(\mu_L, \gamma_L) = (0, 0)$.

Table 5.1. Stop-loss values (in thousands) for various configurations with C Lognormal and L Gamma distributed.

	$\gamma_C = 0.4$			$\gamma_C = 1.2$		
k	1	2	3	1	2	3
(μ_L, γ_L)						
(0,0)	185	20.7	1.15	283	38.5	3.23
(5,0)	222	31.3	2.51	308	45.5	4.27
(10,0)	261	46.0	5.30	334	54.4	5.84
(5,0.75)	245	40.6	4.30	391	71.6	10.3
(10,0.75)	311	72.6	13.7	427	84.9	13.6
(5,1.5)	319	81.4	18.1	431	88.0	14.7
(10,1.5)	439	172	69.7	511	134	32.2

Note that the results from Table 5.1 nicely illustrate the various phenomena described

in general terms above. As the lumpiness μ_G increases, the dependence effect becomes more pronounced. If in addition overdispersion is allowed (i.e. γ_L becomes positive), matters get substantially more serious. Hence this aspect definitely has to be taken into account as well. Increasing γ_C on the other hand appears to moderate the effects somewhat. However, the main conclusion from Table 5.1, based on comparison of the various SL's to the SL_I's in its first row, is of course that ignoring small dependencies in tail related measures indeed can produce huge errors. For $\gamma_C = 0.4$ and $(\mu_L, \gamma_L) = (10, 1.5)$, the relative error $SL/SL_I - 1$ attains a maximal value 61 for $k = 3$, while the absolute error $SL - SL_I$ is most extreme at the value 254 for $k = 1$.

6 Summary and remaining issues

In (2.3) we proposed a general model which takes into account that, besides N ordinary claims, there may also be H groups of special claims, with sizes $G_1, \ldots, G_H$. For N and H the use of a Poisson distribution was adequate, but for G we argued that some overdispersion should be allowed through an additional L (cf. (2.6)). To analyze the resulting model, accurate and transparent approximations were needed, which could again be based on the first three or four cumulants involved. Hence these quantities were obtained in (3.7) and their behavior as functions of the underlying model parameters (cf. (4.3)) was analyzed in Section 4. In the numerical study reported on in Section 5, a wide parameter region was identified where either (or both) the IG (cf. (5.2)) or $G-IG$ approximation (cf. (5.4)) was (were) adequate, with the simple rule of thumb (5.6) for making a final choice. Based on these approximations, in Table 5.1 a number of explicit examples were collected. These clearly demonstrated the severe impact of ignoring even small dependencies.

As mentioned in the introduction, earlier pivot studies usually did not allow direct application in practice. The advantage of the general and flexible model proposed here, would be that it did. And indeed, suppose an insurance company records for a (combination of) portfolio(s) not only the number of claims and the individual claim amounts, but in addition also keeps track for each event of whether it concerns an ordinary claim, or a claim due to special cause nr.1, to special cause nr.2, etc. Then such a dataset would give us not only the realized n and the c_i, but also the realized h and the g_j. In this sense it is definitely true that the present model allows direct implementation into practice, as soon as the minimally required outcomes have indeed been recorded.

However, the description above immediately shows that one major complication still remains: the estimation aspect. Quite often in stochastic model building this aspect is completely neglected. It is simply advised to replace the unknown model parameters by estimated values obtained from the dataset and to carry on as if nothing has happened. The idea is of course that the estimation errors involved will be negligible. But, just as with ignoring the dependence, this optimism could be quite misleading and, when estimation is not based on a large number of observations, lead to an additional effect. For, once again, we are dealing with tail effects and the relative errors involved need not be small at all. Hence for the present we conform to the commonly accepted strategy of ignoring

the estimation aspects. But we intend to address this interesting issue in a forthcoming paper. At the same time, we will pay additional attention to implementation aspects, e.g. by explicitly listing the actual steps needed for the computation of the stop-loss premiums.

References

Albers, W. (1999). Stop-loss premiums under dependence. *Insurance: Mathematics and Economics* **24**, 173-185.

Chaubey, Y.P., Garrido, J. and Trudeau, S. (1998). On the computation of aggregate claims distributions: some new approximations. *Insurance: Mathematics and Economics* **23**, 215-230.

Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R. and Vyncke, D. (2002a). The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics* **31**, 3-33.

Dhaene, J., Denuit, M., Goovaerts, M.J.,Kaas, R. and Vyncke, D. (2002b). The concept of comonotonicity in actuarial science and finance: applications. *Insurance: Mathematics and Economics* **31**, 133-161.

Kaas, R. (1993). How to (and how not to) compute stop-loss premiums in practice. *Insurance: Mathematics and Economics* **13**, 241-254.

Kaas, R., van Heerwaarden, A.E. and Goovaerts, M.J. (1994). Ordering of Actuarial Risks. *Caire Education Series*. Amsterdam

Lukocius, V. (2006). Accuracy of approximations in actuarial overdispersion models. TW-Report ....(To appear).

Reijnen, R., Albers, W. and Kallenberg, W.C.M. (2005). Approximations for stop-loss reinsurance premiums. *Insurance: Mathematics and Economics* **36**, 237-250.

Seal, H.L. (1977). Approximation to risk theory's $F(x,t)$ by means of the gamma distribution. *ASTIN Bull.* **9**, 213-218.