Stockholms universitet

Model Risk in the Cramér–Lundberg Model: The Impact of Clustered and Dependent Claims

Alvar Mikkola

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2026:12
Matematisk statistik
Maj 2026

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2026:12**
http://www.math.su.se

Model Risk in the Cramér–Lundberg Model: The Impact of Clustered and Dependent Claims

Alvar Mikkola*

May 2026

Abstract

Two insurance models can have the same expected aggregate loss but still describe very different levels of risk. This thesis studies this issue by extending the classical Cramér–Lundberg model to allow claim arrivals to generate several related subclaims. The aim is to understand how clustering and dependence between the number of subclaims and their sizes affect total losses.

We compare the extended model with two alternative models. The first keeps the clustered claim count but removes the dependence in subclaim sizes. The second replaces the clustered count with a Poisson count with the same mean. This makes it possible to separate the effect of clustering from the effect of dependence.

For all three models, formulas for the expected value and variances are derived and compared with Monte Carlo simulations. The results show that the expected loss is the same in all three models, but their coefficients of variability differs. The extended model gives the largest variance, while the adjusted classical model gives the smallest. The difference becomes larger as the group size increases, showing that dependence can affect risk even when the expected loss is unchanged. The results are also consistent with earlier research showing that dependence and clustering can affect insurance risk.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: alvarmikkola@gmail.com. Supervisor: Ola Hössjer and Leo Levenius.

Sammanfattning

Denna uppsats studerar hur beroende mellan skadefrekvens och skadestorlek påverkar den totala skadeutvecklingen i försäkringsmodeller. Utgångspunkten är den klassiska Cramér–Lundberg modellen som utvidgas genom att varje skadeankomst kan ge upphov till flera relaterade delskador. Syftet är att undersöka hur klustring och beroende mellan antalet delskador och deras storlekar påverkar den totala risken.

Den utvidgade modellen jämförs med två alternativa modeller. Den första behåller den klustrade skadeprocessen men tar bort beroendet mellan delskadornas storlekar. Den andra ersätter den klustrade skadeprocessen med en Poissonprocess med samma väntevärde. På detta sätt blir det möjligt att särskilja effekten av klustring från effekten av beroende.

För samtliga modeller härleds uttryck för väntevärde och varians, vilka sedan jämförs med Monte Carlo-simuleringar. Trots att den förväntade totala skadan är densamma i samtliga fall visar resultaten att variationskoefficienten skiljer sig åt. Den utvidgade modellen ger störst varians, medan den justerade klassiska modellen ger minst. Skillnaderna blir större när gruppstorleken ökar, vilket visar att både klustring och beroende kan påverka risknivån även när väntevärdet är oförändrat. Resultaten är också förenliga med tidigare forskning om beroende och klustring i försäkringsmodeller.

Acknowledgements

I would like to thank my supervisors, Ola Hössjer and Leo Levenius, for their guidance and support throughout this thesis. Both have contributed to the development of the thesis idea and provided valuable knowledge in insurance mathematics, as well as helpful feedback during the writing process.

I would also like to acknowledge the use of ChatGPT, which was used for minor grammar adjustments and for code- and LaTeX-related tasks, such as implementing the Monte Carlo simulations and generating the corresponding plots and tables.

Finally, I would like to thank my friends, Zacharias Veiksaar and Emre Kaplaner, for their support during this work, and my friend Simon Salmi for providing helpful feedback on the thesis.

Contents

1 Introduction

A common problem in insurance mathematics is to model the total amount paid in claims over a fixed time period. This total amount depends both on the number of claims occur and the size of the claims. Some time periods may have few claims, while others may involve many claims or unusually large losses. An important question in insurance mathematics is therefore how this randomness affects the total loss over time and the financial position of an insurer.

A widely used model for studying this problem is the classical Cramér–Lundberg model. In this model, claims arrive randomly over time and each claim has a random size, while premium income accumulates at a constant rate. However, the model relies on assumptions that may not always be realistic. In the classical model, claims arrive one by one over time, and each arrival corresponds to one claim with a random size. This means that the model treats claims as separate events and assumes that the claim sizes are independent of when the claims occur.

In practice, this may be too simple. A single event can generate several related claims. For instance, consider a traffic accident. If only one car is involved, the damage may be small, such as a scratch in the paint. In a larger accident involving several cars, there may instead be several claims at the same time, one for each damaged car. The costs may also be higher since a larger accident is likely to cause more serious damage. This means that the number of claims and the sizes of the claims may be dependent rather than independent.

This thesis studies how such dependence affects the aggregate loss. To do this, we extend the classical Cramér–Lundberg model by allowing each arrival to generate several subclaims. The sizes of the subclaims may depend on how many are generated, which allows the model to include both clustering and dependence.

The main focus is on how this dependence changes the variability of the aggregate loss. The goal is to understand how clustering and dependence affect the spread of losses, even when the expected loss remains the same.

2 The classical Cramér–Lundberg model

We use the classical Cramér–Lundberg model as the baseline model in this thesis. Historically, the model is associated with Lundberg's early work on ruin problems and Cramér's later development of these ideas [7, pp. 17–18]. The model describes the capital of an insurance company by combining initial capital, premium income, and random claim payments.

In this section we define the model and state its main assumptions. Later, we will change these assumptions when we introduce dependence between the

number of subclaims and their sizes.

Definition 2.1: Cramér–Lundberg model.

Assume an insurance company starts with initial capital $u \geq 0$ and receives premium income at a constant rate $c > 0$. Let $N(t)$ denote the number of claims reported up to time t, and let X_i be the size of the ith claim. Then, the classical Cramér–Lundberg model is defined by

$$U(t) = u + ct - \sum_{i=1}^{N(t)} X_i, \qquad t \geq 0,$$

where $U(t)$ denotes the insurer's capital at time t, see Jacobs [4, p. 4].

In this thesis, we focus on the aggregate loss term $\sum_{i=1}^{N(t)} X_i$. The terms u and ct are included because they are part of the classical model, but they are not studied further. We complete Definition 2.1 by stating the standard modeling assumptions of the classical Cramér–Lundberg setup. Following Levenius [7, pp. 17–18], we use three assumptions describing the claim arrivals, claim sizes, and their independence.

Assumption 2.2: Claim arrivals.

The claim frequency is modeled by a counting process $N = (N(t))_{t\geq 0}$, which is assumed to be a homogeneous Poisson process with a constant intensity $\lambda > 0$. This process is defined by the sequence of random arrival times $0 \leq T_1 \leq T_2 \ldots$ such that the number of claims at any time t is determined by $N(t) = \sup\{n \geq 1 : T_n \leq t\}$.

Assumption 2.3: Claim sizes.

The costs associated with individual claims are denoted by a sequence of random variables $\{X_i\}_{i\geq 1}$. It is assumed that these variables are independent and identically distributed (i.i.d.) and non-negative, following a common distribution function $F_X(x) = \mathbb{P}(X_k \leq x)$. Furthermore, the claim sizes are assumed to have a finite mean, $\mathbb{E}[X_i] < \infty$, to ensure that the aggregate loss and ruin calculations are mathematically valid and the expected total loss is finite.

Assumption 2.4: Claims independent of sizes.

We assume that the costs of the claims $\{X_i\}_{i\geq 1}$ are independent of the arrival process $N = (N(t))_{t\geq 0}$. This means that the time a claim is reported has no influence on the size of the loss.

3 The extended Cramér–Lundberg model

In the previous section, we defined the classical Cramér–Lundberg model and its main assumptions. We now extend this model by allowing each claim arrival to generate several related subclaims. The main difference is that the subclaim sizes may depend on the number of subclaims generated by the same arrival. This creates dependence between claim frequency and severity.

To include this dependence, we keep the capital process

$$U(t) = u + ct - S(t), \quad t \geq 0,$$

but define a new aggregate claims process $S(t)$.

3.1 General setup and dependence

To model the dependence between the frequency and severity of claims, we introduce a hierarchical structure for the aggregate loss process. Let $N = (N(t))_{t\geq 0}$ be a homogeneous Poisson process with intensity $\lambda > 0$, representing the times $0 \leq T_1 \leq T_2 \leq \dots$ when a claim is reported. At each arrival time T_i, the number of subclaims is given by the random variable $K_i \in \mathbb{N}$. We assume that the sequence $(K_i)_{i\geq 1}$ consists of i.i.d. random variables with a probability distribution $\mathbb{P}(K_i = n) = p_n$ for $n \in \mathbb{N}$.

The dependence in this model is introduced by allowing the distribution of the subclaim sizes, denoted by X_{ij}, to depend on the total number of subclaims K_i within the same arrival. Specifically, conditional on $K_i = n$, the subclaim sizes $X_{i1}, \dots, X_{in}$ are assumed to be i.i.d. random variables with a conditional density $f_{X|K_i=n}$ on $[0, \infty)$. The total claim amount produced at arrival time T_i is then defined as the sum of its subclaims:

$$Z_i := \sum_{j=1}^{K_i} X_{ij},$$

where $Z_i = 0$ if $K_i = 0$. Therefore the aggregate claims process $S(t)$ is defined as the following double sum:

$$S(t) := \sum_{i=1}^{N(t)} Z_i = \sum_{i=1}^{N(t)} \sum_{j=1}^{K_i} X_{ij},$$

where the sequence of pairs $((K_i, (X_{ij})_{=1}^{K_i}))_{i\geq 1}$ is assumed to be i.i.d. and independent of N.

3.2 Model specification and the generic subclaim variable

While the previous section provides a general framework, we must specify the distribution of K_i and X_{ij} in order to do concrete calculations and comparisons.

First, we define the distribution for the number of subclaims. To study the effect of claim clustering, we consider a two-point distribution where each arrival produces either a single subclaim or a cluster of m subclaims. We let $m \geq 2$ be a fixed integer and $0 \leq p \leq 1$, defining the distribution as:

$$\mathbb{P}(K_i = 1) = 1 - p, \quad \mathbb{P}(K_i = m) = p.$$

When $p = 0$, the model reduces to the classical Cramér–Lundberg model where each arrival generates exactly one claim. By varying p and m, we can examine how subclaims increase the variance of the total loss process.

Second, we specify the distribution for the subclaim sizes. To get specific formulas for the moments of the aggregate loss, we assume that given $K_i = n$, each subclaim size X_{ij} follows a gamma distribution:

$$X_{ij} \mid K_i = n \sim \Gamma(\alpha(n), \beta(n)),$$

where $\alpha(n) > 0$ is the shape parameter and $\beta(n) > 0$ is the scale parameter. The gamma distribution is a common choice for insurance losses because it only takes positive values. It is also flexible enough to model many different types of claims while remaining mathematically convenient. Under this assumption, the mean and variance for a single subclaim are defined as:

$$\mathbb{E}[X_{ij} \mid K_i = n] = \alpha(n)\beta(n) \quad \text{and} \quad \mathrm{Var}(X_{ij} \mid K_i = n) = \alpha(n)\beta(n)^2.$$

Thirdly, we introduce a generic subclaim variable Y to ensure a fair comparison between models. This variable represents the size of a typical subclaim across the entire process. Because an arrival with m subclaims contributes more individual claims to the total loss than an arrival with only one subclaim, the distribution of Y cannot be a simple average. Instead we must weight the distribution by the number of subclaims each type of arrival provides. For any measurable set $I \subseteq [0, \infty)$, the probability is defined as:

$$\mathbb{P}(Y \in I) = \frac{(1-p)\mathbb{P}(X_{ij} \in I \mid K_i = 1) + pm\mathbb{P}(X_{ij} \in I \mid K_i = m)}{1 - p + pm},$$

where the denominator $1 - p + pm$ is the expected number of subclaims per arrival, $\mathbb{E}[K_i]$. Under our gamma assumption, the distribution of Y becomes a weighted mixture of two gamma distributions:

$$\mathbb{P}(Y \in I) = \frac{(1-p)\mathbb{P}(\Gamma(\alpha(1)), \beta(1)) \in I) + pm\mathbb{P}(\Gamma(\alpha(m)), \beta(m)) \in I)}{1 - p + pm}.$$

This generic variable Y allows us to define a classical model that has the same expected total number of claims as our extended model, which is necessary when comparing the variance and model risk in Section 5.

3.3 Moment calculations

In this section, we derive the formulas for the expected value and variance of the aggregate loss process $S(t)$. To do this we first define the moment generating function (MGF) for our extended model. The definitions and properties of moment generating functions used below follow [3, Ch. 3].

3.3.1 Moment generating function

> Definition 3.1: Moment generating function
>
> Let X be a real-valued random variable. The moment generating function of X is defined by:
>
> $$M_X(\theta) := \mathbb{E}\left[e^{\theta X}\right].$$
>
> We say that the moment generating function exists if there is a constant $h > 0$ such that $M_X(\theta) < \infty$ for all $|\theta| < h$; see [3, p. 63].

To find the MGF for the aggregate loss $S(t) = \sum_{i=1}^{N(t)} Z_i$, we use the fact that the arrival process $N(t)$ follows a Poisson distribution with intensity λt and that the arrival amounts Z_i are independent and identically distributed and independent of $N(t)$. By conditioning on the number of arrivals $N(t) = k$, we get:

$$M_{S(t)}(\theta) := \mathbb{E}[e^{\theta S(t)}] = \sum_{k=0}^{\infty} \mathbb{P}(N(t) = k)\mathbb{E}[e^{\theta \sum_{i=1}^{k} Z_i} \mid N(t) = k].$$

Because the arrival process is independent of the claim amounts, the conditional expectation simplifies to the product of the individual MGFs:

$$\mathbb{E}[e^{\theta \sum_{i=1}^{k} Z_i} \mid N(t) = k] = \prod_{i=1}^{k} \mathbb{E}[e^{\theta Z_i}] = \big(M_Z(\theta)\big)^k,$$

where $M_Z(\theta) := \mathbb{E}[e^{\theta Z_1}]$ is the MGF of a single arrival. By substituting the Poisson distribution and recognizing the power series for the exponential function, we get the general form for the process:

$$M_{S(t)}(\theta) = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \big(M_Z(\theta)\big)^k = e^{\lambda t (M_Z(\theta) - 1)}.$$

The next step is to determine $M_Z(\theta)$, which represents the total claim amount from a single arrival $Z_i = \sum_{j=1}^{K_i} X_{ij}$. To separate the randomness of the number of subclaims from their individual sizes, we condition on the count $K_i = n$:

$$M_Z(\theta) = \sum_{n=0}^{\infty} \mathbb{P}(K_i = n)\mathbb{E}[e^{\theta \sum_{j=1}^{n} X_{ij}} \mid K_i = n].$$

Given that the subclaim sizes X_{ij} are conditionally i.i.d. for a fixed n, the expectation of the sum becomes the nth power of the subclaim MGF, $M_{X|K_i=n}(\theta)$, which under the gamma assumption exists for $\theta < \frac{1}{\beta(n)}$. This allows us to express the MGF for a single arrival as:

$$M_Z(\theta) = \sum_{n=0}^{\infty} \mathbb{P}(K_i = n)\big(M_{X|K_i=n}(\theta)\big)^n.$$

By combining these results, we find the general MGF for the aggregate claims process:

$$M_{S(t)}(\theta) = e^{\lambda t(\sum_{n=0}^{\infty} p_n (M_{X|K_i=n}(\theta))^n - 1)}.$$

Finally, we apply the gamma assumption that we introduced in Section 3.2. Under this assumption, the MGF for a single subclaim is $M_{X|K_i=n}(\theta) = (1 - \beta(n)\theta)^{-\alpha(n)}$. Substituting this into the general form gives the specific expression for our model:

$$M_{S(t)}(\theta) = e^{\lambda t(\sum_{n=0}^{\infty} p_n (1-\beta(n)\theta)^{-n\alpha(n)} - 1}.$$

With the moment generating function determined, we can now derive formulas for the first two moments of the aggregate claims process $S(t)$.

THEOREM 3.2: MOMENTS FROM THE MOMENT GENERATING FUNCTION

Let $t \geq 0$ be fixed and suppose that the moment generating function $M_{S(t)}(\theta)$ exists in a neighborhood of 0. By Theorem 3.3 in Gut [3, p. 64]

$$\mathbb{E}[S(t)^n] = M_{S(t)}^{(n)}(0), \quad n = 1, 2, \ldots$$

In particular,

$$\mathbb{E}[S(t)] = M'_{S(t)}(0)$$

and

$$\mathrm{Var}(S(t)) = M''_{S(t)}(0) - \big(M'_{S(t)}(0)\big)^2.$$

3.3.2 Expectation

The expected value of the aggregate loss is found by evaluating the first derivative of the moment generating function at zero, as stated in Theorem 3.2. Beginning with the general form for the aggregate process, we have:

$$M_{S(t)}(\theta) = e^{\lambda t(M_Z(\theta)-1)},$$

the first derivative with respect to θ is:

$$M'_{S(t)}(\theta) = \lambda t M'_Z(\theta) e^{\lambda t(M_Z(\theta)-1)}.$$

By evaluating this at $\theta = 0$ and noting that $M_Z(0) = 1$, we find that the expected total loss is proportional to the expected cost of a single arrival:

$$\mathbb{E}[S(t)] = M'_{S(t)}(0) = \lambda t M'_Z(0) = \lambda t \mathbb{E}[Z].$$

To determine $\mathbb{E}[Z]$, we apply the two-point distribution for the number of subclaims K_i, defined in Section 3.2. The moment generating function for a single arrival Z_i is therefore:

$$M_Z(\theta) = (1-p)M_{X|1}(\theta) + p(M_{X|m}(\theta))^m.$$

Differentiating this expression with respect to θ yields:

$$M'_Z(\theta) = (1-p)M'_{X|1}(\theta) + pm(M_{X|m}(\theta))^{m-1}M'_{X|m}(\theta).$$

When evaluated at $\theta = 0$, we use the property that any MGF at zero equals one and its derivative at zero equals the mean, $M'(0) = \mathbb{E}[X]$. The expected cost of a single arrival thus becomes:

$$M'_Z(0) = (1-p)\mathbb{E}[X \mid 1] + pm\mathbb{E}[X \mid m].$$

Finally, by substituting the expected values for the gamma distribution from Section 3.2, we get the final formula for the expected aggregate loss:

$$\mathbb{E}[S(t)] = \lambda t\big((1-p)\alpha(1)\beta(1) + pm\alpha(m)\beta(m)\big).$$

3.3.3 Variance

Based on Theorem 3.2, the variance of the aggregate loss is defined by:

$$\mathrm{Var}(S(t)) = M''_{S(t)}(0) - (M'_{S(t)}(0))^2.$$

Applying the product rule to the first derivative of the aggregate process yields the general expression:

$$M''_{S(t)}(\theta) = [\lambda t M'_Z(\theta))^2 + \lambda t M''_Z(\theta)]e^{\lambda t(M_Z(\theta)-1)}.$$

Evaluating this at $\theta = 0$ gives $M''_{S(t)}(0) = (\lambda t M'_Z(0))^2 + \lambda t M''_Z(0)$. When substituted into the variance formula, the squared terms cancel, simplifying the result to:

$$\mathrm{Var}(S(t)) = \lambda t M''_Z(0).$$

For our two-point model, the second moment of a single arrival $M''_Z(0)$ is derived by differentiating the MGF twice, which becomes:

$$M''_Z(0) = (1-p)\mathbb{E}[X^2 \mid K_i = 1] + p[m(m-1)(\mathbb{E}[X \mid K_i = m])^2 + m\mathbb{E}[X^2 \mid K_i = m]].$$

Finally, we substitute the second moments for the gamma distribution, $\mathbb{E}[X^2 \mid m] = \alpha(m)\beta(m)^2(1+\alpha(m))$, to reach the final formula for the aggregate variance:

$$\mathrm{Var}(S(t)) = \lambda t\big((1-p)\alpha(1)\beta(1)^2(1+\alpha(1)) + pm\alpha(m)\beta(m)^2(1+m\alpha(m))\big).$$

4 Comparison models

In this section we compare the extended model with two related models. The purpose of this comparison is to separate the two ways in which the extended model differs from the classical Cramér–Lundberg model. First, the total number of subclaims is generally not Poisson distributed, so the counting assumption in Assumption 2.2 is no longer satisfied. Second, the subclaim sizes depend on the group size, and therefore the claim sizes are not independent and identically distributed as in the classical model.

The models we use for comparison are as follows:

1. The adjusted classical model $S_{CL}(t)$: This model uses a Poisson process for the number of claims as in the classical Cramér–Lundberg model, and i.i.d. claim sizes with distribution Y.
2. The approximation model $S_{CL}^{appr}(t)$: This model keeps the same number of subclaims as in the extended model, but replaces the original dependent subclaim sizes with i.i.d. random variables having the same distribution as Y.

The reason for introducing these models is that they allow us to study the effect of these two properties separately. The approximation model $S_{CL}^{appr}(t)$ removes the dependence in subclaim sizes but keeps the same counting structure as for the extended model. The adjusted classical model $S_{CL}(t)$ goes one step further by also bringing back the Poisson counting assumption.

To make this precise, we let

$$N^*(t) = \sum_{i=1}^{N(t)} K_i$$

denote the total number of subclaims up to time t, where $N(t)$ is a Poisson process and K_i is the number of subclaims generated at arrival i. Under the two-point distribution, we can also write

$$N^*(t) = N_1(t) + mN_2(t),$$

where $N_1(t)$ and $N_2(t)$ are independent Poisson random variables with parameters $(1-p)\lambda t$ and $p\lambda t$.

We now show that $N^*(t)$ is generally not Poisson. Using the property $\mathrm{Var}(N^*(t)) = \lambda t \mathbb{E}[K^2]$ and the two-point distribution of K: $\mathbb{P}(K=1) = 1-p$ and $\mathbb{P}(K = m) = p$, we get $\mathbb{E}[K^2] = 1 - p + m^2 p$. Hence,

$$\mathrm{Var}(N^*(t)) = \lambda t(1 - p + m^2 p).$$

On the other hand,

$$\mathbb{E}[N^*(t)] = \lambda t(1 - p + mp).$$

Since $m \geq 2$, it follows that $\text{Var}(N^*(t)) > \mathbb{E}[N^*(t)]$. A Poisson random variable has the property that its variance equals its mean, so this shows that $N^*(t)$ is generally not Poisson distributed. Therefore the extended model does not satisfy the counting assumption of the classical Cramér–Lundberg model.

4.1 The approximation model

The first comparison model we define is the approximate model, denoted as $S_{CL}^{appr}(t)$. This model allows us to study the effect of grouped arrivals separately, that is, the extra volatility caused by subclaims arriving in clusters rather than one by one. In this model we keep the group arrival process $N^*(t)$ from our extended model, but we drop the assumption that subclaim sizes depend on the group size. Instead, we assume that all subclaim sizes are independent and follow the distribution of the generic subclaim variable Y we defined in Section 3.2. The model is hence defined as:

$$S_{CL}^{appr}(t) = \sum_{i=1}^{N^*(t)} Y_i,$$

where $N^*(t)$ is the non-Poisson arrival process we defined in the previous section.

4.1.1 Expectation

To find the expected value, we substitute the expressions for the arrival process and the generic subclaim variable Y. From the previous section, we know that $\mathbb{E}[N^*(t)] = \lambda t(1 - p + pm)$. The expected value of the generic subclaim size Y, which is a weighted mixture of gamma distributions, is defined as:

$$\mathbb{E}[Y] = \frac{(1-p)\alpha(1)\beta(1) + pm\alpha(m)\beta(m)}{1 - p + pm}.$$

By multiplying these two terms together, the factor $(1 - p + pm)$ cancels out, leaving:

$$\mathbb{E}[S_{CL}^{appr}(t)] = \lambda t((1-p)\alpha(1)\beta(1) + pm\alpha(m)\beta(m))$$

This result is identical to the expected value of the extended model $S(t)$, which we derived in Section 3.3.1. This is important because it means that any differences in variance or ruin probabilities between the models are due to clustering and dependence, rather than differences in the expected total claim amount.

4.1.2 Variance

To find an explicit formula for the variance, we use the law of total variance for a compound process where the counts and sizes are independent:

$$\text{Var}(S_{CL}^{appr}(t)) = \mathbb{E}[N^*(t)]\text{Var}(Y) + \text{Var}(N^*(t))(\mathbb{E}[Y])^2.$$

Using the identity $\mathrm{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$, we rewrite this as:

$$\mathrm{Var}(S_{CL}^{appr}(t)) = \mathbb{E}[N^*(t)]\mathbb{E}[Y^2] + (\mathbb{E}[Y])^2(\mathrm{Var(N^*(t))} - \mathbb{E}[N^*(t)]).$$

To get the explicit formula, we substitute the moments we derived in Section 3. We find that the difference between the variance and the expected value of the claim count is:

$$\mathrm{Var}(N^*(t)) - \mathbb{E}[N^*(t)] = \lambda t(1 - p + m^2 p) - \lambda t(1 - p + mp) = \lambda t mp(m - 1).$$

Furthermore, under the gamma assumption, the second moment of a subclaim in a group of size n is defined as $\alpha(n)\beta(n)^2(1 + \alpha(n))$. By substituting these into the first term, we get:

$$\mathbb{E}[N^*(t)]\mathbb{E}[Y^2] = \lambda t\big[(1 - p)\alpha(1)\beta(1)^2(1 + \alpha(1)) + pm\alpha(m)\beta(m)^2(1 + \alpha(m))\big].$$

By combining this with the other moments, we arrive at a explicit formula for the variance:

$$\begin{aligned}\mathrm{Var}(S_{CL}^{appr}(t)) \quad = \quad & \lambda t\big[(1 - p)\alpha(1)\beta(1)^2(1 + \alpha(1)) + pm\alpha(m)\beta(m)^2(1 + \alpha(m))\big] \\ & +\lambda t\, mp(m - 1)\left(\mathbb{E}[Y]\right)^2,\end{aligned}$$

where:

$$\mathbb{E}[Y] = \frac{(1 - p)\alpha(1)\beta(1) + pm\alpha(m)\beta(m)}{1 - p + pm}.$$

4.2 The adjusted classical model

The second comparison model is the adjusted classical model, denoted by $S_{CL}(t)$. This model goes one step further than the approximation model by also keeping the Poisson counting assumption of the classical Cramér–Lundberg model. More precisely, we replace the non-Poisson count $N^*(t)$ by a Poisson process with the same mean while still assuming that the claim sizes are i.i.d. with the same distribution as Y. In this way the model satisfies the classical counting approach while still using the same generic subclaim variable Y. We therefore define

$$S_{CL}(t) = \sum_{i=1}^{\tilde{N}(t)} Y_i,$$

where $(\tilde{N}(t))_{t\geq 0}$ is a Poisson process with intensity $\lambda(1 - p + mp)$.

4.2.1 Expectation

To compute the expected value of $S_{CL}(t)$, we first note that:

$$\mathbb{E}[\tilde{N}(t)] = \lambda t(1 - p + mp).$$

It follows that:

$$\mathbb{E}[S_{CL}(t)] = \mathbb{E}[\tilde{N}(t)]\mathbb{E}[Y] = \lambda t(1-p+mp)\frac{(1-p)\alpha(1)\beta(1)+pm\alpha(m)\beta(m)}{1-p+mp},$$

which simplifies to

$$\mathbb{E}[S_{CL}(t)] = \lambda t((1-p)\alpha(1)\beta(1)+pm\alpha(m)\beta(m)).$$

So the adjusted classical model has the same expected aggregate loss as both the extended model and the approximation model.

4.2.2 Variance

Since $\tilde{N}(t)$ is Poisson distributed and the claim sizes are i.i.d., $S_{CL}(t)$ is a compound Poisson model. Its variance is therefore given by:

$$\text{Var}(S_{CL}(t)) = \mathbb{E}[\tilde{N}(t)]\mathbb{E}[Y^2].$$

Using the second moment of Y, we get:

$$\text{Var}(S_{CL}(t)) = \lambda t((1-p)\alpha(1)\beta(1)^2(1+\alpha(1))+pm\alpha(m)\beta(m)^2(1+\alpha(m))).$$

5 Results

In this section we present numerical results based on Monte Carlo simulations for the three aggregate loss models. The purpose is to complement the theoretical results derived earlier in the thesis by studying the shape of the aggregate loss distribution under each model. Because explicit density functions are difficult to derive, we approximate the distributions numerically from simulated samples.

5.1 Theoretical results

Before presenting the simulations, we summarize the exact moment formulas derived in Sections 3 and 4. The purpose is to show what differences should be expected between the three aggregate loss models.

All three models have the same expected aggregate loss:

$$\mathbb{E}[S] = \mathbb{E}[S_{CL}^{appr}(t)] = \mathbb{E}[S_{CL}(t)] = \lambda t((1-p)\alpha(1)\beta(1)+pm\alpha(m)\beta(m)).$$

Hence, the models are compared under the same expected total loss. Any differences between them therefore come from their variance and distributional shape, not from their mean. The variances are given in Table 1.

Table 1: Exact variance formulas for the three aggregate loss models.

Model	Variance
Extended	$\lambda t\Big((1-p)\alpha(1)\beta(1)^2(1+\alpha(1)) + pm\alpha(m)\beta(m)^2(1+m\alpha(m))\Big)$
Approximation	$\lambda t\Big((1-p)\alpha(1)\beta(1)^2(1+\alpha(1)) + pm\alpha(m)\beta(m)^2(1+\alpha(m))\Big) + +\lambda t\, mp(m-1)(E[Y])^2$
Adjusted classical	$\lambda t\Big((1-p)\alpha(1)\beta(1)^2(1+\alpha(1)) + pm\alpha(m)\beta(m)^2(1+\alpha(m))\Big)$

The adjusted classical model has the smallest variance. This model replaces the clustered subclaim count by a Poisson process with the same mean and assumes that the claim sizes are independent with common distribution Y. The approximation model keeps the clustered subclaim count and therefore contains the additional term:

$$\lambda tmp(m-1)(\mathbb{E}[Y])^2.$$

This term is positive whenever $p > 0$ and $m > 1$, so the approximation model has a larger variance than the adjusted classical model.

The extended model differs from the approximation model because the distribution of each subclaim depends on the group size, whereas in the approximation model all subclaims follow the same distribution. In the parameter choices used below, grouped subclaims have both a larger mean and a larger spread than single subclaims. For these parameter values, this gives:

$$\mathrm{Var}(S_{CL}(t)) < \mathrm{Var}(S_{CL}^{appr}(t)) < \mathrm{Var}(S(t)).$$

Since the three models have the same expected aggregate loss, we compare relative spread using the coefficient of variation, defined by:

$$R(X) = \frac{\sqrt{\mathrm{Var}(X)}}{\mathbb{E}[X]}.$$

The formulas in Table 1 suggest that the differences between the models should become larger as the group size m increases, since the additional variance terms depend on m.

5.2 Simulation setup

In the simulations we first set $\lambda = 1$ and $t = 1$, so that $\lambda t = 1$, and for each parameter choice and claim size model, we perform 10000 Monte Carlo simulations. This gives rise to 10000 i.i.d. simulated claim sizes $S_1, \ldots, S_{10000}$, each having the same distribution as $S(t)$, $S_{CL}(t)$ or $S_{CL}^{appr}(t)$ for this particular choice of parameters. We also include an additional set of simulations with

$\lambda = 5$ and $t = 2$, so that $\lambda t = 10$, for one fixed choice of p and several values of m. The purpose of this second setting is to study how the aggregate loss distribution changes when the expected number of claim arrivals is larger.

In the single subclaim case we assume that $X \mid K_i = 1$ follows a gamma distribution with parameters $\alpha(1) = 2$ and $\beta(1) = 0.5$. This gives $\mathbb{E}[X \mid K_i = 1] = 1$, so the mean subclaim size is normalized to 1.

For a gamma-distributed random variable, $R(X)$ is equal to $\frac{1}{\sqrt{\alpha}}$, meaning that the parameter α controls the spread relative to the mean. In the single-subclaim case, where $\alpha(1) = 2$, the ratio is $\frac{1}{\sqrt{2}}$.

For grouped subclaims, we assume that $X \mid K_i = m$ follows a gamma distribution with parameters $\alpha(m) = 1$ and $\beta(m) = 2$. This gives $\mathbb{E}[X \mid K_i = m] = 2$, meaning that grouped subclaims are larger on average than single subclaims. In this case, the coefficient of variation is 1. Hence, grouped subclaims are not only larger on average but also more spread out relative to the mean. This reflects the idea that arrivals generating many subclaims may also be associated with more severe and more uneven losses. In this way, the simulation reflects the dependence between group size and subclaim size in the extended model. If the parameters were the same in both cases, then the subclaim size distribution would not depend on the group size K.

In the simulations, we consider several different parameter choices. In particular, we study the case $m = 2$ together with a small value of p, since this represents a common case where grouped arrivals can occur but are still rare. We also consider larger values of m in order to see how larger groups affect the aggregate loss distribution. All three models are investigated for each choice of parameters.

5.3 Simulations

Tables 2–5 show exact and simulated values of the coefficient of variation of the aggregate loss for the three models and for different choices of p and m. In all cases, the simulated values are very close to the corresponding exact values and the results show a clear difference between the models as the group size m increases.

Table 2: Exact and simulated values of the coefficient of variation for the three models when $p = 0.01$ and $\lambda t = 1$.

	Extended		Approximation		Adjusted classical	
m	Exact	Simulated	Exact	Simulated	Exact	Simulated
1	1.239	1.221	1.239	1.239	1.239	1.247
2	1.275	1.286	1.253	1.226	1.245	1.249
5	1.503	1.574	1.331	1.324	1.260	1.275
10	2.039	2.003	1.540	1.513	1.270	1.258
20	3.076	3.012	2.069	2.121	1.264	1.272
30	3.912	3.969	2.601	2.581	1.240	1.252

For $p = 0.01$ the three models are relatively close when m is small. However, as m increases, the coefficient of variation clearly increases in both the extended model and the approximation model, while it remains almost unchanged in the adjusted classical model. Notably, for $m = 30$, the coefficient of variation is 3.969 in the extended model, 2.581 in the approximation model, and 1.252 in the adjusted classical model. This shows that even when grouped arrivals are rare, large group sizes can have a strong effect on the spread of the aggregate loss in the extended and approximation model.

Table 3: Exact and simulated values of the coefficient of variation for the three models when $p = 0.05$ and $\lambda t = 1$.

	Extended		Approximation		Adjusted classical	
m	Exact	Simulated	Exact	Simulated	Exact	Simulated
1	1.287	1.203	1.287	1.264	1.287	1.287
2	1.409	1.395	1.332	1.345	1.297	1.311
5	1.879	1.866	1.524	1.524	1.276	1.280
10	2.482	2.480	1.889	1.914	1.194	1.220
20	3.133	3.064	2.466	2.451	1.041	1.043
30	3.466	3.528	2.847	2.820	0.928	0.928

For $p = 0.05$ the same pattern appears, but the difference between the models becomes even more visible. The coefficient of variation again increases considerably with m in the extended model and the approximation model. On the other hand, the adjusted classical model stays at a much lower level and even decreases for larger values of m. This is a clear difference from the other two models. For instance, when $m = 30$, the coefficient of variation is 3.528 in the extended model, 2.820 in the approximation model, and 0.928 in the adjusted classical model.

Table 4: Exact and simulated values of the coefficient of variation for the three models when $p = 0.10$ and $\lambda t = 1$.

	Extended		Approximation		Adjusted classical	
m	Exact	Simulated	Exact	Simulated	Exact	Simulated
1	1.333	1.245	1.333	1.337	1.333	1.307
2	1.490	1.522	1.382	1.378	1.321	1.325
5	1.923	1.899	1.582	1.578	1.217	1.220
10	2.322	2.356	1.899	1.926	1.054	1.069
20	2.656	2.708	2.289	2.276	0.850	0.854
30	2.800	2.825	2.500	2.528	0.730	0.727

For $p = 0.10$ the same overall pattern remains. The extended model has the largest coefficient of variation for all values of m, followed by the approximation model while the adjusted classical model has the smallest values. As for $p = 0.05$, the coefficient of variation of the adjusted classical model decreases as m becomes larger, while for the other two models it increases. At $m = 30$, the coefficient of variation is 2.825 in the extended model, 2.528 in the approximation model and 0.727 in the adjusted classical model.

Table 5: Exact and simulated values of the coefficient of variation for the three models when $p = 0.05$ and $\lambda t = 10$.

	Extended		Approximation		Adjusted classical	
m	Exact	Simulated	Exact	Simulated	Exact	Simulated
1	0.407	0.389	0.407	0.403	0.407	0.405
2	0.446	0.446	0.421	0.417	0.410	0.409
5	0.594	0.594	0.482	0.479	0.404	0.396
10	0.785	0.776	0.597	0.602	0.378	0.370
20	0.991	0.992	0.780	0.784	0.329	0.332
30	1.096	1.101	0.900	0.903	0.293	0.293

Table 5 shows the results for $\lambda t = 10$ and $p = 0.05$. As in tables 2-4, the simulated values are very close to the exact values. Compared with the case $\lambda t = 1$, the coefficient of variation is smaller for all three models, but the same overall pattern remains, that is, the coefficient of variation is largest in the extended model, followed by the approximation model and smallest in the adjusted classical model. As m increases, the coefficient of variation increases in the extended and approximation models while it decreases in the adjusted classical model.

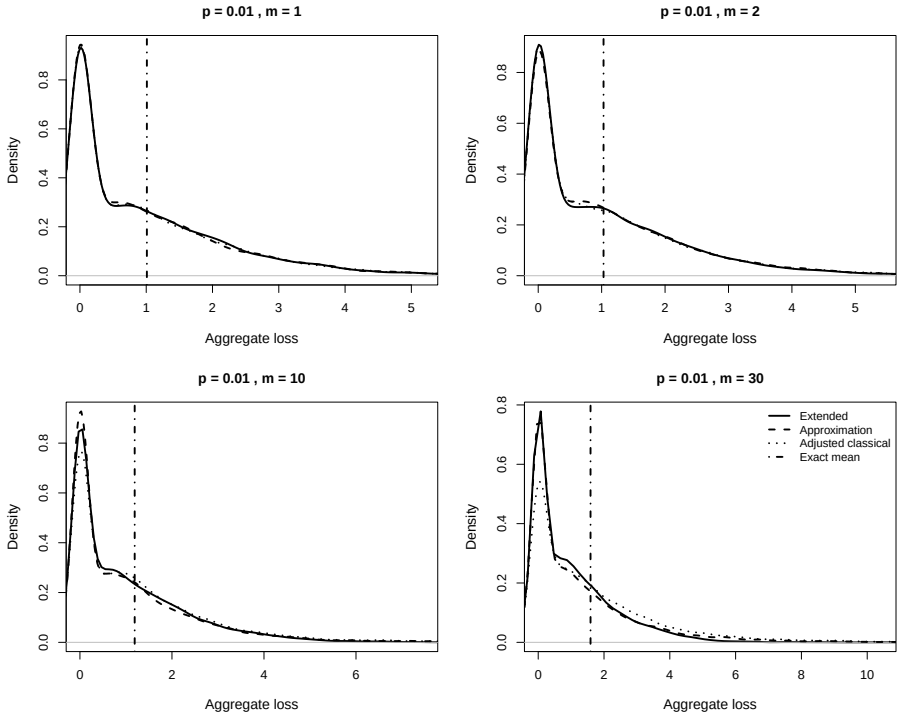

Figure 1: Estimated density functions for the three models when $p = 0.01$ and $\lambda t = 1$. In each subplot, the vertical line shows the mean and the horizontal axis is limited to the 99th percentile of the simulated sample.

Figure 1 shows the estimated density functions for the case $p = 0.01$. For small values of m, the three models are very similar and the differences between the densities are difficult to distinguish. The clearest difference between the densities appears when $m = 30$, where the adjusted classical model has relatively more mass further to the right, while the extended model decreases more quickly. The approximation model lies between the two models.

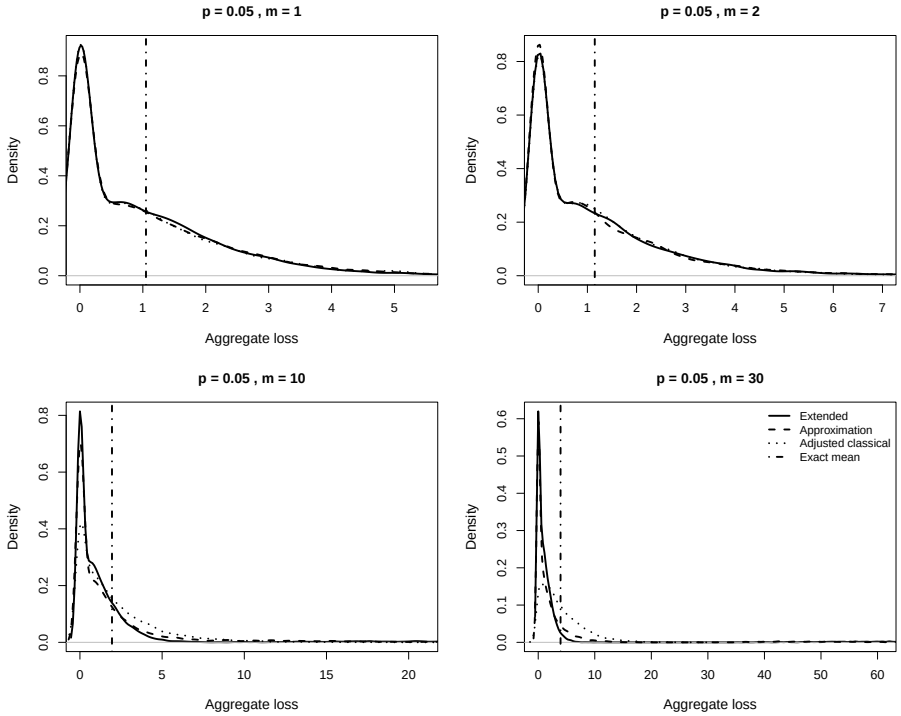

Figure 2: Estimated density functions for the three models when $p = 0.05$ and $\lambda t = 1$. In each subplot, the vertical line shows the mean and the horizontal axis is limited to the 99th percentile of the simulated sample.

Figure 2 shows the estimated density functions for the case $p = 0.05$. Compared with Figure 1, the differences between the three models are more visible, especially for $m = 10$ and $m = 30$. For small values of m, the three densities are still relatively close, whereas for larger values of m the separation becomes clearer. As before the approximation model lies between the extended and adjusted classical models.

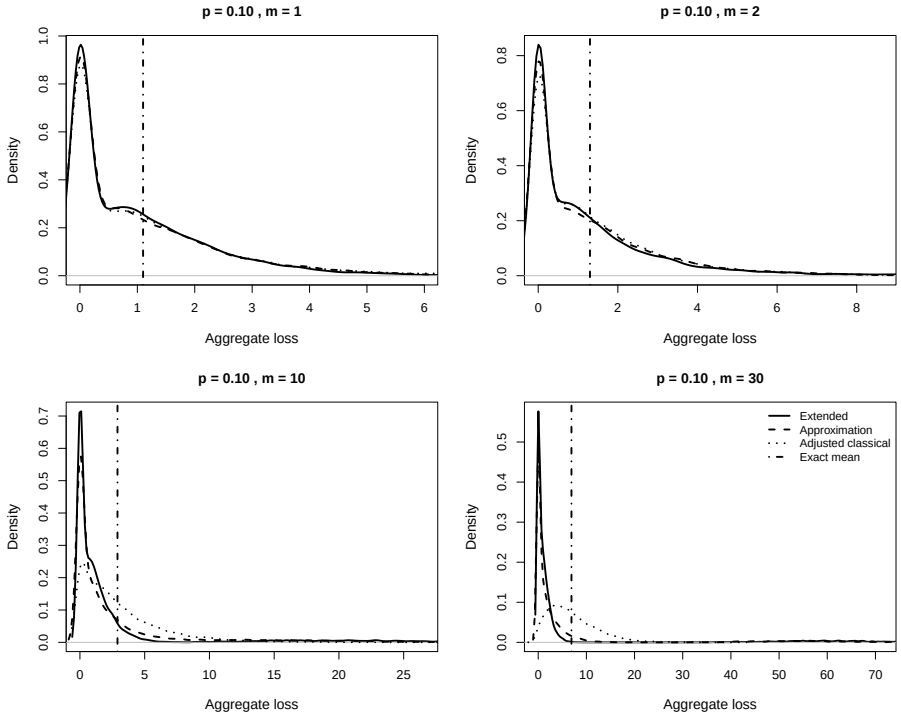

Figure 3: Estimated density functions for the three models when $p = 0.10$ and $\lambda t = 1$. In each subplot, the vertical line shows the mean and the horizontal axis is limited to the 99th percentile of the simulated sample.

Figure 3 shows the estimated density functions for the case $p = 0.10$. Compared with Figures 1 and 2, the differences between the three models are even clearer, especially for $m = 10$ and $m = 30$. Notably, the extended model has a sharper peak near small aggregate losses and then decreases quickly, while the adjusted classical model is more concentrated around the mean. The approximation model again lies between the two models.

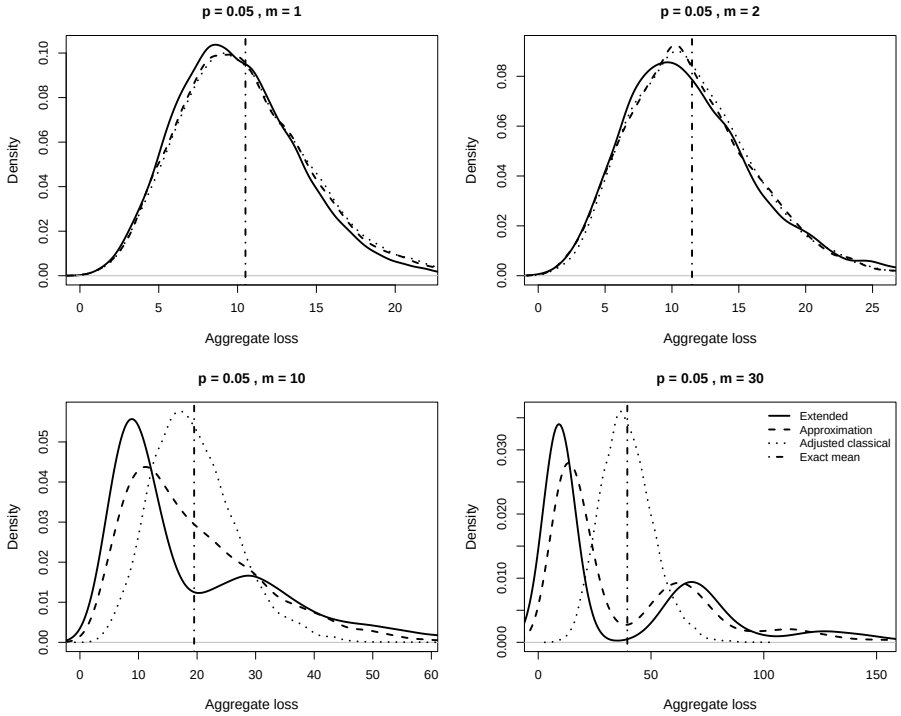

Figure 4: Estimated density functions for the three models when $p = 0.05$ and $\lambda t = 10$. In each subplot, the vertical line shows the mean and the horizontal axis is limited to the 99th percentile of the simulated sample.

Figure 4 shows the estimated density functions for the case $p = 0.05$ and $\lambda t = 10$. Compared with Figures 1–3, the shapes of the densities are now clearly different. For $m = 1$ and $m = 2$, the densities are roughly bell-shaped for all three models. As m increases, however, the extended and approximation models are no longer bell-shaped but instead develop noticeable peaks with additional mass further to the right. The adjusted classical model keeps its shape and remains centered around the mean.

The additional simulations in Appendix 7.1 show the same pattern for $\lambda t = 100$. Compared with $\lambda t = 1$ and $\lambda t = 10$, the coefficient of variation is smaller for all three models, meaning that the aggregate losses are more concentrated around their expected values. The estimated densities also become more bell-shaped, which is consistent with the central limit theorem.

As an additional way of comparing the models, we also use boxplots of the simulated aggregate loss for the cases $\lambda t = 1$ and $\lambda t = 10$. While the density plots show the overall shape of the distributions, the boxplots make it easier to compare the differences in spread and especially in tail behavior across all three

models.

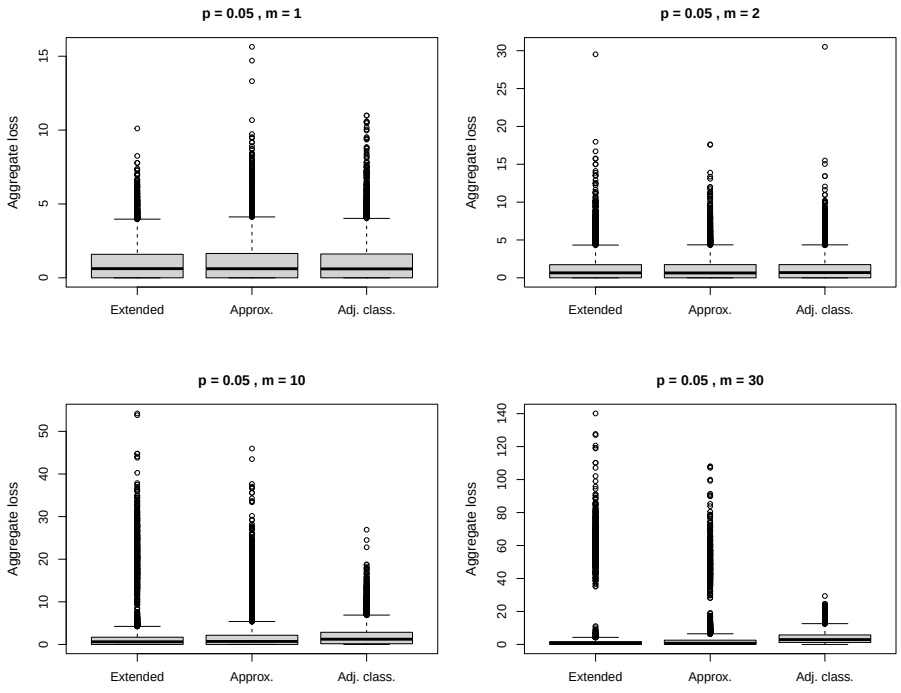

Figure 5: Boxplots of the simulated aggregate loss for the three models when $p = 0.05$ and $\lambda t = 1$ for $m = 1, 2, 10$ and $m = 30$.

Figure 5 shows boxplots of the simulated aggregate loss for the case $p = 0.05$ and $\lambda t = 1$. For $m = 1$ and $m = 2$, the three models are very similar. As m increases, however, the upper-tail observations becomes more apparent. Also, the extended and approximation models remain more concentrated at lower aggregate loss values, while the adjusted classical model appears to have a larger interquartile range.

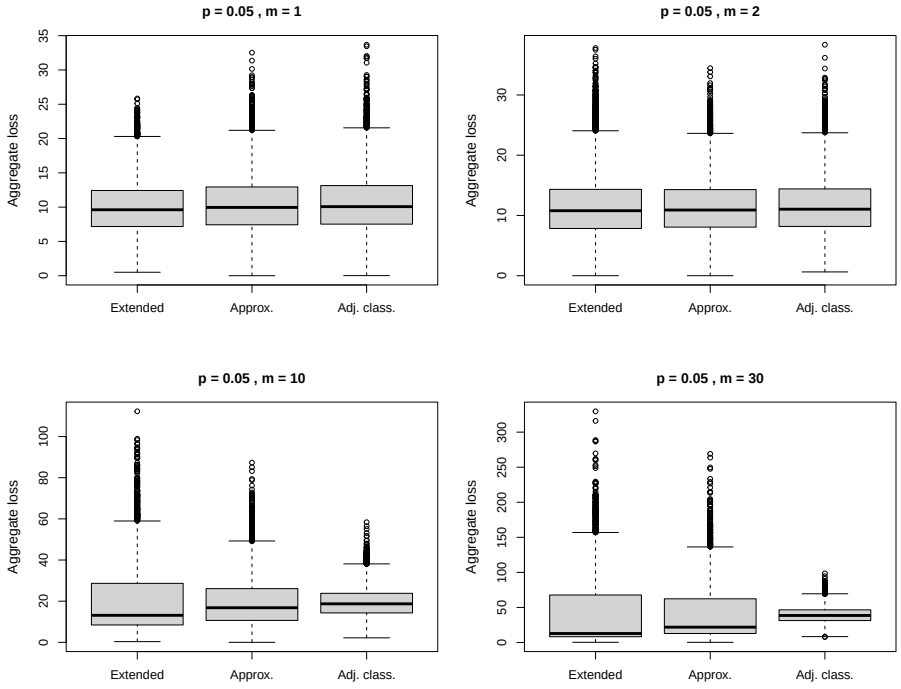

Figure 6: Boxplots of the simulated aggregate loss for the three models when $p = 0.05$ and $\lambda t = 10$ for $m = 1, 2, 10$, and 30.

Figure 6 shows boxplots of the simulated aggregate loss for the case $p = 0.05$ and $\lambda t = 10$. For $m = 1$ and $m = 2$, the three models are still very similar. For $m = 10$, however, the differences become clearer as the extended and approximation models show a larger spread along with more large outlying values, while the adjusted classical model appears more symmetric and concentrated around its median.

6 Discussion

6.1 Main findings

The results show that the dependence between claim frequency and severity increases the variability of the aggregate loss. All three models have the same expected aggregate loss. However, their variances differ. The adjusted classical model has the smallest variance, the approximation model has a higher variance, and the extended model has the largest. This shows that both clustering

and dependence contribute to increased variability and that dependence adds additional variability beyond clustering alone. This effect becomes stronger as the group size m increases. Even when grouped arrivals are rare, for example when $p = 0.01$, large group sizes lead to a much higher coefficient of variation in the extended model compared to the adjusted classical model. This means that the spread of the aggregate loss is strongly affected by the dependence.

6.2 Relation to earlier research

When comparing the results of this thesis with previous research, we see that they are consistent with earlier work. The classical Cramér–Lundberg model is an important standard model, but Mandjes and Boxma describe it as a "gross simplification of reality" since several features that matter in practice are not included [1, p. 1]. Saruan and Effendie also state that the independence assumption does not describe real insurance data well [8, pp. 95-96]. Other papers use more complex models to study similar effects. For example, Hawkes processes can be used to show how one claim can increase the probability of further claims [2]. Lévy subordinators have also been used to study how clustering can affect ruin probabilities and lead to heavier-tailed risk behavior [6]. The results in this thesis support the same general idea, that is, dependence between claim frequency and claim size can increase the variability of the aggregate loss. Therefore, ignoring this dependence may lead to inaccurate estimates of aggregate risk and variability. [5, 9].

6.3 Implications

The implications of the results are mainly related to the role of dependence in the aggregate loss. We see that the classical Cramér–Lundberg model can still be useful when clustering is weak. This is also visible in the simulations where the densities of the three models are very similar for small values of m. In these cases, we can view the classical model as a reasonable approximation.

However, we also see that this changes when the group size m increases. Larger values of m make the effect of dependence more important. In particular, when subclaim sizes depend on the group size, the variability of the aggregate loss increases significantly. This shows that the total loss is affected when claim frequency and claim size are dependent.

For applications, this means that we should consider dependence when claims occur in groups. Ignoring this effect may lead to an underestimation of variability, especially in situations with large claim groups. Therefore models that include dependence are more appropriate when the data contain clustering or large events that generate multiple related claims.

6.4 Limitations

This study uses a simplified model in order to isolate the effects of clustering and dependency. For instance, the number of subclaims is assumed to follow a two-point distribution where each arrival produces either one subclaim or m subclaims. This makes clustering easier to study, but in real data the number of subclaims may vary more.

The choice of gamma distributions for the subclaim sizes is another limitation. The gamma distribution is mathematically convenient and commonly used, however, it may not describe real loss data well, especially for very large losses. Other distributions could yield different results for variability.

Also, the thesis focuses mainly on the expected value, variance and coefficient of variation. These measures are useful when comparing the models, but they do not directly measure the probability of very large losses. Therefore it would be interesting to study tail behavior and ruin probabilities in future work. The density plots provide some visual information about the tails, but they do not provide enough detail to study extreme losses. This was left for future work. It would also be interesting to test more values of λt, especially since the results changed significantly when $\lambda t = 10$. Looking at more values of intensities could show more clearly how the models behave when the expected number of claim arrivals increases.

References

[1] Peter Braunsteins and Michel Mandjes. The cramér–lundberg model with a fluctuating number of clients. *Insurance: Mathematics and Economics*, 112: 1–22, 2023. doi: 10.1016/j.insmatheco.2023.05.007.

[2] Yiling Chen and Baojun Bian. Optimal dividend policy in an insurance company with contagious arrivals of claims. *Mathematical Control and Related Fields*, 11(1):1–22, 2021.

[3] Allan Gut. *An intermediate course in probability*. Springer, 2 edition, 2009.

[4] Tom T. M. M. Jacobs. The Cramér–Lundberg model and copulas, 2021. Bachelor's thesis, Eindhoven University of Technology.

[5] Övgücan Karadağ Erdemir and Meral Sucu. "A comparative study on modeling of dependence between claim severity and frequency with archimedean copulas". *Journal of Statisticians: Statistics and Actuarial Sciences*, 13(1): 18–29, 2020.

[6] Jonathan Klinge and Maren Diane Schmeck. "Asymptotics of ruin probabilities in a subordinated Cramér–Lundberg model". *arXiv*, 2026.

[7] Leo G. Levenius. Den Cramérska assuransmatematiken. Master's thesis, Stockholms universitet, Matematiska institutionen, 2025.

[8] Sandy Salomo Saruan and Adhitya Ronnie Effendie. "A deeper look into an insurance risk model with two types of claims and FGM copula dependency". *Indonesian Actuarial Journal*, 1(2):95–112, 2025.

[9] Peng Shi and Zifeng Zhao. "Regression for copula-linked compound distributions with applications in modeling aggregate insurance claims". *arXiv*, 2021.

Appendix

This appendix contains additional material related to the simulation. First, we present extra simulation results for the case $\lambda t = 100$. These results are included to show how the models behave when the expected number of arrivals is larger. We then present the R code used for the simulations, focusing on the main simulation functions.

A. 1 Additional simulation results

Table 6: Exact and simulated values of the coefficient of variation for the three models when $p = 0.05$ and $\lambda t = 100$.

	Extended		Approximation		Adjusted classical	
m	Exact	Simulated	Exact	Simulated	Exact	Simulated
1	0.129	0.121	0.129	0.126	0.129	0.130
2	0.141	0.141	0.133	0.133	0.130	0.129
5	0.188	0.187	0.152	0.152	0.128	0.127
10	0.248	0.246	0.189	0.190	0.119	0.119
20	0.313	0.311	0.247	0.244	0.104	0.104
30	0.347	0.347	0.285	0.284	0.093	0.093

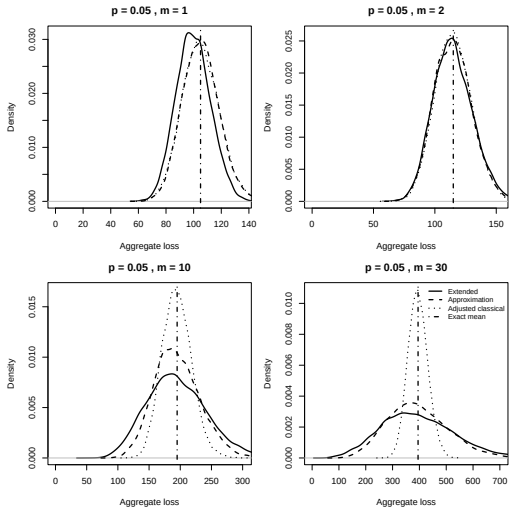

Figure 7: Estimated density functions for the three models when $p = 0.05$ and $\lambda t = 100$. In each subplot, the vertical line shows the exact mean and the horizontal axis is restricted to the 99th percentile of the simulated sample.

A. 2 Simulation code

The R code below gives the simulation functions. The parameter values used in the simulations are stated in Section 5.2.

```
set.seed(12345)

sim_Y <- function(p, m, a1, b1, am, bm) {
  w1 <- (1 - p) / (1 - p + m * p)
  if (runif(1) <= w1) rgamma(1, a1, scale = b1)
  else rgamma(1, am, scale = bm)
}

sim_ext <- function(lambda, t, p, m, a1, b1, am, bm) {
  N <- rpois(1, lambda * t)
  S <- 0
  for (i in seq_len(N)) {
    K <- sample(c(1, m), 1, prob = c(1 - p, p))
    if (K == 1) S <- S + rgamma(1, a1, scale = b1)
    else S <- S + sum(rgamma(m, am, scale = bm))
  }
  S
}

sim_appr <- function(lambda, t, p, m, a1, b1, am, bm) {
  N <- rpois(1, lambda * t)
  K <- sample(c(1, m), N, replace = TRUE, prob = c(1 - p, p))
  sum(replicate(sum(K), sim_Y(p, m, a1, b1, am, bm)))
}

sim_cl <- function(lambda, t, p, m, a1, b1, am, bm) {
  N <- rpois(1, lambda * t * (1 - p + m * p))
  sum(replicate(N, sim_Y(p, m, a1, b1, am, bm)))
}

ext_sample <- replicate(nsim, sim_ext(lambda, t, p, m, a1, b1, am,
    bm))
appr_sample <- replicate(nsim, sim_appr(lambda, t, p, m, a1, b1, am
    , bm))
cl_sample <- replicate(nsim, sim_cl(lambda, t, p, m, a1, b1, am, bm
    ))
```