On fitting dependent nonhomogeneous loss models to unearned premium risk

Jessup, Sébastien Boucher, Jean-Philippe Pigeon, Mathieu

Chaire Co-operators en analyse des risques actuariels
Département de mathématiques, UQAM.
Montréal, Québec, Canada.

Abstract

Unearned premium, or more particularly the risk associated to it, has only recently received regulatory attention. Unearned losses occur after the evaluation date for policies written before the evaluation date. Given that an inadequate acquisition pattern of premium and approximate modelling of premium liability can lead to an inaccurate reserve around unearned premium risk, an individual nonhomogeneous loss model including cross-coverage dependence is proposed to provide an alternative method of evaluating this risk. Claim occurrence is analysed in terms of both claim seasonality and multiple coverage frequency. Homogeneous and heterogeneous distributions are fitted to marginals. Copulas are fitted to pairs of coverages using rank-based methods and a tail function. This approach is used on a recent Ontario auto database.

Keywords Unearned premium risk, Loss reserving, Predictive Modelling, Dependence

1 Introduction

Non-life insurance companies face high volatility due to the nature of losses they must provide coverage for. Regulation thus requires insurers to maintain funds under solvency constraints to ensure that up to a certain risk level, insureds' claims will not suffer from an insurer's solvency issues. These funds form actuarial reserves, and their risk requires accurate and reliable actuarial models. As such, a significant portion of actuarial literature is focused on modelling reserves and ensuring optimal capital allocation.

There exist very specific guidelines to determine the capital requirement to be maintained by an insurer varying from one country and even one state/province to another. For Property and Casualty (P&C) insurers, also known as non-life insurers, US regulation as defined by the National Association of Insurance Commissioners uses risk-based capital requirements (see Feldblum (1996) for more information), while the Canadian requirement is set forth as the conditional tail expectation (CTE) at a 99% level for insurance risk (Office of the Superintendent of Financial Institutions, 2018a). Different guidelines exist elsewhere in the world. Insurance risk can generally be further broken down into four parts: capital required for unpaid claim liabilities, capital required for premium liabilities, margin required for reinsurance ceded to unregistered reinsurers and catastrophe reserves.

An insured may or may not incur a loss over the duration of their contract. This contract is written on a certain effective date, and obligates the insurer to cover for losses defined in the contract for its specified length, which is usually but not always one year. In the event of a loss, there are often delays between occurrence and reporting dates, and depending on the nature of the claim, it may lead to multiple payments before the file is closed. Actuaries must establish the amount of capital to be maintained as reserves on a specific date called the evaluation date.

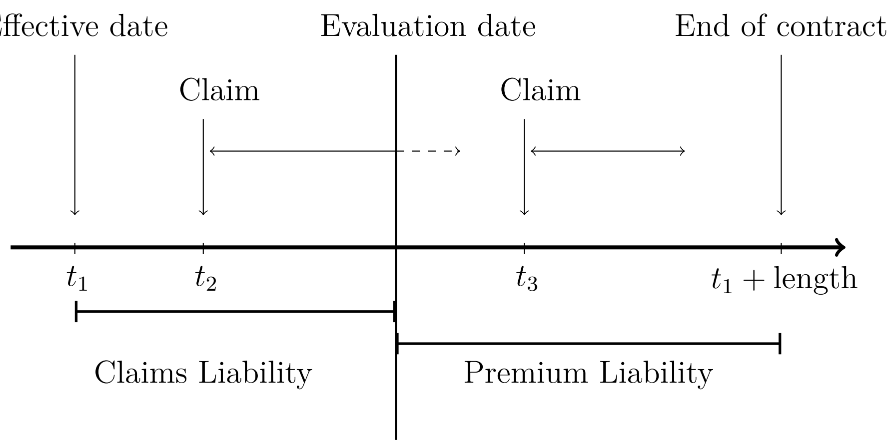


Figure 1 – Split between claims liabilities and premium liabilities

Classical reserving methods usually focus on unpaid claim liabilities. This liability stems from the outstanding loss amount left to be paid from accidents incurred on or before the evaluation date. In Figure 1, the outstanding loss from any claim occurring between the effective date and the evaluation date would fall into unpaid claim liabilities. This includes reported but not settled (RBNS) claims, and evaluation of incurred but not reported (IBNR) claims. There are many existing methods to evaluate such claims, see Wüthrich and Merz (2008) and Friedland (2010) for an extensive discussion of existing methods.

Premium liability is what drives the risk linked to unearned premium, stemming from potential future losses occurring after the evaluation date from contracts effective prior to the evaluation date. Figure 1 gives a visual representation of this separation. In other words, the risk linked to unearned premium is that the unearned premium reserve (UPR) will be insufficient to cover for premium liability.

This type of loss is the focus of our article. More specifically, we use piece-wise constant risk exposure and cross-coverage dependence to evaluate unearned premium risk. Through this, we aim to identify the main drivers of risk relating to unearned premium and determine the role of cross-coverage dependence. Our hypothesis is that risk is driven by loss seasonality, loss distributions, the acquisition pattern of premium, contract subscription patterns, and cross-coverage dependence. In Section 2, we explain what is unearned premium risk and what methods are available to build a model around it. In Section 3, we present the models proposed to evaluate this risk by including dependence between coverages, and apply these models to actual data in Section 4. Finally, we provide concluding remarks and potential adaptations in Section 5.

2 Unearned premium risk

To evaluate the risk linked to unearned premium, there are two main options: to use an aggregated framework (see Subsection 2.1) or to use individual data (see Subsection 2.2). Aggregated methods essentially group all claims based on accident year and the time elapsed since the loss, generally referred to as development period, whereas individual methods keep all claim information on an individual level.

2.1 Traditional methods

Historically speaking, aggregated models have been favoured to calculate reserves due to their ease of use and non-intensive computing time. In this framework, one would use loss development triangles. This tool owes its name to its shape; to build a loss development triangle one must aggregate loss payments by accident year and development period, thus creating a triangle shape. Table 1 illustrates this aggregation. From paid amounts (P), one can project outstanding claim amounts (O) using methods such as the stochastic Chain-Ladder method (Mack, 1993) or the Bornhuetter-Ferguson method (Bornhuetter and Ferguson, 1972).

Table 1 – Loss development triangle with subsequent year

	Development period					
AY	**1**	**2**	**3**	**4**	**5**	**6**
2014	P	P	P	P	P	O
2015	P	P	P	P	O	O
2016	P	P	P	O	O	O
2017	P	P	O	O	O	O
2018	P	O	O	O	O	O
2019	F	F	F	F	F	F

Note: paid claims (P), outstanding claims (O), future claims (F)

Development triangles allowed for a convenient way of aggregating losses and calculate outstanding losses in non-computationally intensive ways. These methods are useful in valuing unpaid claim liabilities, but do not tell us much about the last line of Table 1, which consists of future claims (F). A part of these claims comes from contracts effective on the evaluation date: these make up most of premium liability, which is what we attempt to evaluate through our proposed model.

Loss analysis linked to unearned premium has only recently started receiving attention. As suggested by the Canadian Institute of Actuaries (2014), through triangles one would require a projection of the loss ratio for the following year and an estimation of the UPR to calculate premium liability. One can then evaluate ultimate losses for each year evaluated through methods such as Mack (1993), Bornhuetter and Ferguson (1972), and many others, then compare these losses to the premium written each year. This allows for determining the loss ratio to be multiplied by the UPR. The UPR is usually evaluated by supposing a uniform acquisition of premium, such that it is equal to the sum of written premium multiplied by the remaining fraction of contracts across all contracts.

The aggregated framework idea was reworked to determine the variance of such a method. Li (2010) proposes an extension of Mack's model using the next accident year's expected loss ratio to find an estimate of the prediction error of premium liability without assumptions concerning the underlying loss distribution. Priest (2012) instead proposes a model supposing that losses are driven by three factors, one specific to each accident year, one depending on both accident year and development period, and some random effect. Both Li and Priest rely on the hypothesis that the evaluation of premium liability is an extension of the evaluation of outstanding claims.

Some issues arise with these proposals. Barnett et al. (2005) demonstrates that the development method supposes independence between accident years, however Meyers (2013) suggests that this is in fact not the case as most contracts span more than one accident year. For example, a contract written on July 1^{st}, 2017 for a one year duration will end on June 30^{th}, 2018. Moreover, as most insureds keep the same insurer year after year, an insurer's portfolio will consist of mostly the same risks, which should induce between-year dependence.

Beyond between-year dependence, one should consider between-coverage dependence. Regulatory guidelines for Canadian insurers as provided by the Office of the Superintendent of Financial Institutions (2018b) require insurers to hold reserves for expected outstanding amount for each coverage. This is intuitively affected by dependence between coverages. Literature exists concerning dependence between lines of business when modelling loss reserve through aggregated loss triangles, however this is very limited when considering dependence between coverages, which is most likely due to only recently having the computational capacity to look at this level of loss.

Another problem is that current methods suppose losses occur uniformly throughout the year, which is not necessarily true as explained in Collins and Hu (2003). Factors affecting this assumption include but are not limited to loss seasonality, loss trends (e.g. inflation), legal changes affecting claims, climate change, etc.

2.2 Reserve for unearned premium risk

Any aggregated method is in fact highly dependent on the acquisition pattern of premium, given that losses linked to unearned premium are a subset of total losses, and that premium is the only way of measuring the size of this subset. Evaluation of premium liability would then entirely depend on having the right UPR. Bearing in mind that insurers usually use a uniform acquisition pattern for premium, the following example quickly illustrates how this first option can be problematic.

Example 2.1. Suppose two insureds, A and B, have contracts with an effective date of April 1^{st}. A has an equal chance of incurring a loss any time of the year. Meanwhile, B is twice as likely to have an accident between January and March than the rest of the year, in the sense that B's risk level for the first quarter is 2λ while it is λ the rest of the year. Based on loss exposure, as of January 1^{st}, by separating the year into quarters A will have 1/4 of her risk remaining. B will however have 2/5 of his risk remaining. Under uniform acquisition of premium, both insureds would have an unearned premium of $0.25P$, with P being their premium, but B's unearned premium should be $0.4P$, for a shortfall of $0.15P$.

With Example 2.1 in mind, it becomes clear that loss seasonality may play an important role in the evaluation of unearned premium risk. With the same seasonality, but different effective dates, we would however obtain a somewhat different scenario.

Example 2.2. Suppose two insureds, A and B, have contracts with an effective date of October 1^{st}, with the same seasonality as in Example 2.1. Based on loss exposure, as of January 1^{st}, A will have 3/4 of her risk remaining. B will however have 4/5 of his risk remaining. Under uniform acquisition of premium, both insureds would have an unearned premium of $0.75P$, with P being their premium, but B's unearned premium should be $0.80P$, for a shortfall of $0.05P$.

This example allows us to see that for the same loss seasonality, the effective date at which contracts are written affects the unearned premium risk, and with perfect recognition of seasonality in the acquisition of premium, then there would be no unearned premium risk. Evidently the behaviour of losses also needs to be taken into account to determine P, so these examples allow us to postulate that the main drivers of unearned premium risk are: loss distributions, loss seasonality, contract subscription patterns, and premium acquisition.

2.2.1 Individual loss reserving models

With advances in computational power, we have accessible information about each insured and claim, meaning that instead of only having data concerning the total amount paid for a given accident year and development period, we have data for every insured. This change in data availability allows us to use models such as Antonio and Plat (2014) and Pigeon et al. (2013) for reserving purposes. An approach of this type allows for modelling losses linked to unearned premium directly from previous losses of the same type. This eliminates the issue of depending on premium to evaluate which portion of losses is associated to unearned premium.

We can take the individual model approach one step further by taking into account that insurance is generally separated into multiple coverages. For example, in car insurance in Ontario, accident benefits is a no-fault coverage that pays for an insured's injuries resulting from a car accident, while the third party liability coverage mainly pays for injuries caused by an insured driver to another person (such as the other driver). These coverages are usually valued separately in terms of costs and expected claims. It is however a fairly intuitive leap to see how if a car crash is severe enough to cause injuries, it is likely to do so for both drivers, which leads to dependence in both occurrence of losses between coverages and in loss amounts.

Using reserving approaches, dependence has mostly been studied from a line of business point of view. Different approaches based on copulas, where for example Côté et al. (2016) fits generalised linear models (GLM) to marginal lines of business and selects copulas to capture dependence through rank-based methods, while Cossette et al. (2013a) considers dependence between risks within an insurance portfolio and fits a Farlie-Gumbel-Morgenstern copula to model this dependence.

As such, there exist many individual loss models in reserving that take dependence into account. These models however use past data to model claims that have already occurred; we are interested in future losses, which have not yet occurred, and so are more interested in a pricing approach.

2.2.2 Using a pricing approach

Under a pricing approach, Frees (2008) suggests that considering dependence within an insurance contract is important for pricing purposes. This dependence can take multiple forms, both in claim occurrence and in claim amounts. Abdallah et al. (2016) uses the Sarmanov family of multivariate distributions to build a bivariate claim count model. Frees and Valdez (2008) instead models which coverages are affected when an accident occurs through a multinomial logit model, then use a t-copula to model dependence between losses. Recently, Pechon et al. (2019) takes an approach of combining guarantees (coverages) and policyholders through a multivariate Poisson-mixture to capture dependence.

With this range of approaches to capturing dependence between coverages, we would then want to take into account dependence between coverages in an individual model with appropriate copulas, and model dependence between coverage occurrence. This idea stems from the intuition that if a claim involves multiple coverages, there is likely to be a link between those coverages, and it is worthwhile to investigate how this dependence behaves. We thus want a model capable of capturing loss seasonality, contract subscription patterns, while considering loss dependence between coverages.

3 Modelling approach

We are interested in evaluating the risk linked to unearned premium, which is the potential excess of future losses over the unearned premium reserve. When a claim occurs, one or more coverages can be affected. To this end, as mentioned in the previous section, to model future losses we need to consider claim frequency, loss amount, as well as dependence between coverages.

In regard to frequency, we make a simplifying assumption for the model. A contract can in theory have more than one loss in a year. In actuarial databases, this is however rather infrequent. For example, first looking at a full year, in Shi et al. (2018), Table 1 presents some summary statistics of claim frequency, where we see that 0.35% of people will experience 2 or 3 losses in one year, or in Table 2 of Boucher et al. (2007) where 0.50% of contracts have more than one reported claim. These examples are for the year as a whole; future losses are only those losses that occur after the evaluation date. Multiple future losses in one year are thus very rare. As such, we suppose that a contract can only incur one future loss in a given year without losing much information through an indicator variable. In this way, we avoid having to consider multiple future losses for a single contract by supposing that if there is presence of loss, only one loss occurs. Another important assumption we will use is that we suppose that all contracts are written for a one year term, which is a standard assumption in actuarial models. Lastly, we assume that loss amount is independent from loss occurrence, meaning that the moment when a claim happens during the year has no incidence on loss amount, which is again a standard assumption.

In Subsection 3.1, we define notation used throughout the rest of the paper. Subsection 3.2 presents the model in general form for a certain number C of coverages, as well as explaining how we consider dependence between coverages. Subsection 3.3 gives justification for coverage grouping in our model and Subsection 3.4 explains the reserve for unearned premium risk.

3.1 Notation

The following are necessary definitions used in our model for some k^{th} contract. Figure 2 illustrates the various events related to unearned premium risk.

- N_k is a discrete random variable for the number of losses occurring after the evaluation date (future losses);
- J_k is an indicator random variable for the presence of a future loss as per our hypothesis such that
$$J_k = \begin{cases} 1 & \text{if } N_k > 0 \\ 0 & \text{otherwise;} \end{cases}$$
- $\mathbf{Y}_k = [Y_k^{(1)} \cdots Y_k^{(C)}]$ is a vector of continuous positive random variables for the paid future loss amount for each coverage c, $c = 1, \ldots, C$;
- $\mathbf{I}_k = [I_k^{(1)} \cdots I_k^{(C)}]$ is a vector of indicator variables for the presence of a loss for each coverage c, $c = 1, \ldots, C$;
- $t_k^{(E)}$ is the time since the last evaluation date on the effective date of a contract, where for example a contract written on November 1$^{\text{st}}$ with an evaluation date

of December 31st would have $t_k^{(E)} = 0.8329$ (304 days out of 365 since the last evaluation);

- E_k is the exposure to future risk, which will be further defined in Section 3.3;

- P_k^{UE} is the unearned premium. Based on the acquisition pattern used by an insurer, this is not a random variable as the remaining portion of risk can easily be evaluated at the evaluation date. For example, if we assume a uniform acquisition of premium, then $P_k^{UE} = P_k(1 - t_k^{(E)})$, with P_k the written premium for a k^{th} contract.

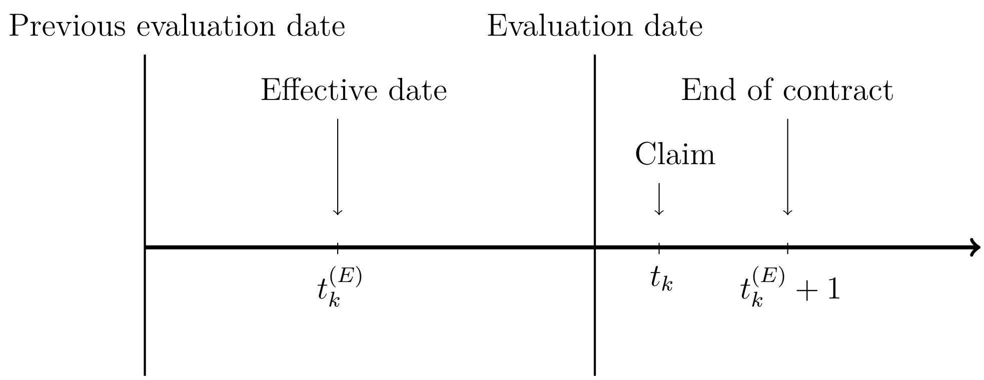


Figure 2 – Illustration of time-related random variables for a one-year contract

3.2 General model in dimension C

Based on our previous definitions, the risk linked to unearned premium, which we call Z, depends on total future losses S^* and the total unearned premium across all n contracts, such that

$$Z = S^* - \sum_{k=1}^{n} P_k^{UE}. \tag{1}$$

As explained in Subsection 3.1, the sum of unearned premium can be calculated at the evaluation date based on an insurer's acquisition pattern. In fact, for the rest of this paper, we will generally suppose that an insurer uses uniform acquisition of premium, such that P_k^{UE} is known at the evaluation date. The model is flexible enough to use other different acquisition patterns, but for illustration, and because it is almost always used in practice, uniform acquisition is supposed. Our interest is thus on modelling S^*. Recall our hypotheses, which are that

- A contract can only incur one future loss in a year;

- Contracts are written for a one-year period;

- Loss amount is independent from loss occurrence.

Grouping theses assumptions, we can then define future loss S_k in terms of $I_k^{(c)}$ and $Y_k^{(c)}$:

$$S_k = \begin{cases} \sum_{c=1}^{C} I_k^{(c)} Y_k^{(c)} & \text{if } J_k = 1, \\ 0 & \text{if } J_k = 0, \end{cases} \tag{2}$$

and subsequently

$$S^* = \sum_{k=1}^{n} S_k.$$

Supposing that the total unearned premium is known at the evaluation date, to model Z we thus need to model frequency N_k, coverage occurrence $\mathbf{I}_k$, and coverage loss amount $\mathbf{Y}_k$.

We include dependence in this model when considering $\mathbf{I}_k$ and $\mathbf{Y}_k$ and consider two cases: independence between coverages and presence of cross-coverage dependence. This choice of consideration follows from our research focus to determine the role of cross-coverage dependence.

For coverage occurrence $\mathbf{I}_k$, we can consider separate occurrence for each coverage or use joint occurrence probabilities. The problem with using an independent occurrence hypothesis is that claims involving more than two coverages seldom happen, and so independent occurrence can increase those probabilities. We choose to keep $\mathbf{I}_k$ fixed across both cases as the empirical distribution of coverage occurrence, we will elaborate on this idea in Section 4.

Then, for both the independent and dependent scenarios, we model $\mathbf{Y}_k$ through copulas, where

$$F_{\mathbf{Y}_k}(\mathbf{y}) = C(F_{Y_k^{(1)}}(y^{(1)}), \cdots, F_{Y_k^{(C)}}(y^{(C)})).$$

in the perspective of considering between-coverage dependence in losses. See Sklar (1973) or a classical textbook on modelling dependence with copulas such as Joe (2014) for more details on copulas. In the independent case, we use the independence copula whereas when taking dependence between loss types we allow for other copulas.

3.3 Seasonality

As introduced in Example 2.1, risk is not always distributed uniformly throughout the year, despite the simplifying assumption used by most insurers, which is the basis of risk linked to unearned premium. In this paper, we use an idea proposed in Verrall and Wüthrich (2016), which is to suppose that the arrival rate at time s for the k^{th} contract $\lambda_0(s)$ is piece-wise constant. In other words, there exists a partition of time through the year $\mathcal{A} = \{A_m\}_{m=1,\ldots,M}$ such that for each $A_m \in \mathcal{A}$, $\lambda_0(s) = \lambda_m$, with $m = 1, \ldots, M$.

Assume N_k follows a non-stationary Poisson process with multiplicative intensity function, similarly to Zhao and Zhou (2010). Let $A_m = [t_{m-1}, t_m)$. Then we can build an exposure function $E_0^{(C)}(t)$ s.t.

$$E_0^{(C)}(t) = \frac{\sum_{m=1}^{M} \frac{\int_{t_{m-1}}^{\min(t_m,t)} \mathbb{I}(t<t_m)\lambda_0(s)ds}{t_m - t_{m-1}}}{\sum_{m=1}^{M}\lambda_m}$$
$$= \frac{\sum_{m=1}^{M} \mathbb{I}(t<t_m)\lambda_m \frac{\min(t_m,t)-t_{m-1}}{t_m-t_{m-1}}}{\sum_{m=1}^{M}\lambda_m}, \tag{3}$$

where each set A_m adds risk up to λ_m, with maximum risk for a full year being $\sum_{m=1}^{M}\lambda_m$. Using this exposure measure, we then have

$$\lambda_k(t|\mathbf{X}_k) = E_0^{(C)}(t)e^{\mathbf{X}_k^T\boldsymbol{\beta}},\ k = 1,\dots,n, \tag{4}$$

with $\mathbf{X}_k$ a $(p\times 1)$ set of covariates and $\boldsymbol{\beta}$ a $(p\times 1)$ corresponding set of predictors obtained through maximum likelihood estimation (MLE), with p the number of covariates. Note that here we implicitly assume that the covariates have no impact on the exposition function.

In Figure 2, consider the previous evaluation date as $t = 0$ and current evaluation date as $t = 1$. Under our previous assumption that a contract has a one year length, that contract's exposure to future risk is equivalent to $[1,t_k^{(E)}+1] \equiv [0,t_k^{(E)}]$, supposing that seasonality does not change from one year to the next. From there, using Equation 4 we trivially obtain

$$\mathbb{E}[N_k|\mathbf{X}_k] = E_0^{(C)}(t_k^{(E)})e^{\mathbf{X}_k^T\boldsymbol{\beta}}. \tag{5}$$

3.4 Reserve linked to unearned premium risk

Proposition 3.1. Let Z be a random variable as defined in Equation 1 for the risk linked to unearned premium. For a portfolio containing n contracts independent and C coverages, the expected value is given by

$$\mathbb{E}\left[Z|t_k^{(E)}\right] = \sum_{k=1}^{n}\left(1-e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\right)\sum_{c=1}^{C}\Pr(\mathbb{I}_k^{(c)}=1)\mathbb{E}\left[Y_k^{(c)}\right] - \sum_{k=1}^{n}P_k^{(UE)}, \tag{6}$$

and the variance is given by

$$\begin{aligned}\mathrm{Var}\left[Z|t_k^{(E)}\right] = &\left(1-e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\right)\left[\sum_{c=1}^{C}\mathbb{E}\left[(I_k^{(c)})^2\right]\mathbb{E}\left[(Y_k^{(c)})^2\right] - \mathbb{E}\left[I_k^{(c)}\right]^2\mathbb{E}\left[Y_k^{(c)}\right]^2\right.\\ &+2\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}\left(\mathbb{E}\left[Y_k^{(i)}Y_k^{(j)}\right]\Pr(I_k^{(i)}=1,I_k^{(j)}=1)\right.\\ &\left.\left.-\mathbb{E}\left[I_k^{(i)}\right]\mathbb{E}\left[I_k^{(j)}\right]\mathbb{E}\left[Y_k^{(i)}\right]\mathbb{E}\left[Y_k^{(j)}\right]\right)\right]\\ &+\left(1-e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\right)e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\left[\sum_{c=1}^{C}\mathbb{E}\left[I_k^{(c)}\right]\mathbb{E}\left[Y_k^{(c)}\right]\right]^2.\end{aligned} \tag{7}$$

Proof. The proof is in Appendix A. □

Moreover, given the distribution of $T^{(E)} \in [0,1]$, a continuous random variable for the effective date of contracts, then

$$\mathbb{E}[Z] = \sum_{k=1}^{n} \left[\sum_{c=1}^{C} \Pr(\mathbb{I}_k^{(c)} = 1) \mathbb{E}\left[Y_k^{(c)}\right] \int_0^1 \left(1 - e^{-\lambda_k(s|\mathbf{X}_k)}\right) f_{T^{(E)}}(s) ds \right] - \sum_{k=1}^{n} P_k^{(UE)}.$$

Then, given the distribution of $T^{(E)}$,

$$\begin{aligned}
\mathrm{Var}[Z] = &\sum_{k=1}^{n} \Bigg(\Bigg[\sum_{c=1}^{C} \mathbb{E}\left[(I_k^{(c)})^2\right] \mathbb{E}\left[(Y_k^{(c)})^2\right] - \mathbb{E}\left[I_k^{(c)}\right]^2 \mathbb{E}\left[Y_k^{(c)}\right]^2 \\
&+2 \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} \left(\mathbb{E}\left[Y_k^{(i)} Y_k^{(j)}\right] \Pr(I_k^{(i)} = 1, I_k^{(j)} = 1) - \mathbb{E}\left[I_k^{(i)}\right] \mathbb{E}\left[I_k^{(j)}\right] \mathbb{E}\left[Y_k^{(i)}\right] \mathbb{E}\left[Y_k^{(j)}\right] \right) \Bigg] \\
&\times \int_0^1 \left(1 - e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\right) f_{T^{(E)}}(s) ds \\
&+ \left[\sum_{c=1}^{C} \mathbb{E}\left[I_k^{(c)}\right] \mathbb{E}\left[Y_k^{(c)}\right] \right]^2 \int_0^1 \left(1 - e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\right) e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)} f_{T^{(E)}}(s) ds \Bigg).
\end{aligned}$$

Knowing the expected value and variance give us some information about the distribution of Z, however as suggested in Kaye (2005), there are multiple ways of measuring risk and allocating capital in general insurance with different risk measures for which we need to know the full distribution of Z. For example, Cossette et al. (2013b) gives possible applications of the Value-at-Risk as the most widely used risk measure in both insurance and finance for capital allocation. As such, we can define the reserve, or allocated capital, R for Z as

$$R = \rho_\alpha(Z),$$

where $\rho_\alpha : \mathcal{L} \to \mathbb{R}$ is a risk measure based on the distribution of $Z \in \mathcal{L}$ such as the Value-at-Risk or the conditional tail expectation at a certain risk level α.

3.4.1 Algorithm to find the reserve for unearned premium risk

The algorithm to simulate the reserve for unearned premium risk is thus as follows:

1. For all contracts, generate a realisation of J_k through N_k and the contract's exposure $E_0^{(C)}(t_k)$, where $J_k = 1$ if $n_k > 0$ and 0 otherwise.

2. If $J_k = 1$, then there is a loss. Generate a realisation of $\mathbf{I}_k$ from the possible loss scenarios based on empirical observations.

3. Generate a realisation of losses $y^{(c)} \sim Y^{(c)}$ for affected coverages. In the independent case, use the independence copula; in the dependent case, use an appropriate copula to capture dependence.

4. Sum losses across all coverages.

5. Calculate the reserve for unearned premium risk as the sum of losses less the total unearned premium.

6. Repeat this procedure a large number of times.

7. Use these results to obtain the predictive distribution of Z.

8. Use appropriate risk measures to determine the reserve for unearned premium risk.

4 Analysis

4.1 Data

We analyse data from a Canadian insurer in Ontario consisting of 132,093 auto contracts with information concerning 45 different coverages with effective dates ranging from December 15, 2015 to December 31, 2018. We choose to focus on five main coverages (Financial Services Commission of Ontario, 2016):

- Accident Benefits (AB), which is a no-fault coverage for benefits that the driver or another insured person may receive if injured or killed in an auto accident (income replacement, medical fees, rehabilitation, etc.);

- Collision (Coll), which covers for material damage to an insured's vehicle when involved in a collision with another object or the insured's vehicle rolls over and for which they are at least partially responsible for;

- Comprehensive (Comp), which covers for damage to an insured's vehicle from perils not linked to a collision with another vehicle, such as hail, theft, vandalism, or hitting a wild animal;

- Direct Compensation Property Damage (DCPD), which is direct compensation covering damage to an insured's vehicle when they are not at fault in an accident, with at least one of the other drivers being insured. It can be seen as the flip side of Collision;

- Third Party Liability (TPL), which can cover for injuries caused to another person or damage to someone's property, as well as protect the insured in the event of lawsuits.

There are 10,423 claims for these five coverages, of which 87.4% involve only one coverage, 11.4% involve two coverages, and 1.2% involve three or more coverages. Due to the limited sample size and out of parsimony, we limit our analysis to two coverages at a time in our analysis of dependence.

The database is built on a transactional basis, where each intervention between the insured and the insurer creates a new line, even if there are no changes. Claim information consists of the date of occurrence, coverages affected, and incurred amount for each coverage.

Using this data, we attempt to address the following research questions:

1. What factors actually drive unearned premium risk?

2. What is the impact of cross-coverage dependence on this risk?

4.2 Fitting the model

We use December 31st as our evaluation date. Due to our focus on the risk linked to unearned premium, this means we only have two years of losses to work with, per se losses occurring in 2017 from contracts effective in 2016, and losses occurring in 2018 from 2017 policies. We choose to fit our model on 2017 losses then observe its accuracy on 2018 losses. This is a limiting factor in our analysis due to the very small number of years available for analysis and the volatile nature of losses from one year to the next.

To get total future loss S^*, we use two models, $S^{(comp)}$ with dimension $C = 1$, and $S^{(crash)}$ with dimension $C = 4$, consisting of respectively only the comprehensive coverage, and the four other coverages, such that

$$S_k^{(comp)} = I^{(comp)} * Y^{(comp)}, \text{ and}$$
$$S_k^{(crash)} = I^{(AB)} * Y^{(AB)} + I^{(coll)} * Y^{(coll)} + I^{(DCPD)} * Y^{(DCPD)} + I^{(TPL)} * Y^{(TPL)},$$

provided $J_k = 1$. The reason behind this choice is the distribution of $\lambda_0(t)$. Figure 3 shows the observed frequency by month for AB, Coll, DCPD and TPL (left) as compared to the frequency for Comp (right). The seasonality for the group of four coverages is very similar, while Comp has a very different seasonality. This is mostly due to the fact that Comp provides coverage for accidents that do not occur while driving (vandalism, hail, theft), whereas the other coverages require accidents occurring while on the road, justifying keeping it separate. Note however that we observe weak seasonality, which is a limiting factor in our analysis of its impact on unearned premium risk.

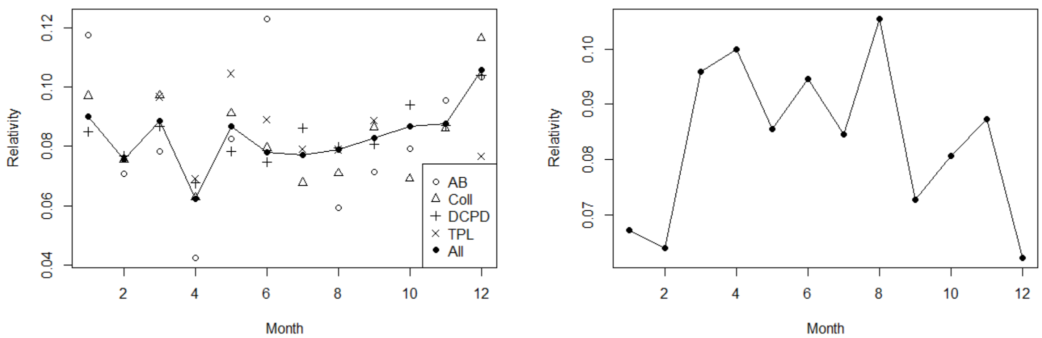


Figure 3 – Relativity of claim occurrence by month for claims linked to car crashes (left) and comprehensive (right)

For both models, we need $\lambda_0(t)$, N_k, $\mathbf{I}_k$ and $\mathbf{Y}_k$. As mentioned in Subsection 3.2, we use two approaches to investigate the impact of cross-coverage dependence on unearned premium risk; one where losses are assumed independent between coverages and one where we consider dependence between coverages.

In both cases, for each model $\lambda_0(t)$ is fitted using the empirical distribution observed in Figure 3. That is, to determine $E_0^{(C)}(t)$ to be used as an exposure measure, we use an empirical approach by using the hypothesis that the partitions in Equation 3 are monthly, meaning that we suppose the risk level is constant through January, but different from February, and so on. This allows us to find $\mathbb{E}[N_k]$ for each contract using Equation 5.

We use the empirical distribution of observed groupings of coverages to simulate the behaviour of coverage occurrence $\mathbf{I}_k$. In the four-coverage case, this does induce some dependence in our independent scenario, but as mentioned previously, the probability of multiple coverages occurring would be over-evaluated if we used independent occurrence of coverages. Using three coverages as an example, in reality we observed 121 out of 10 423 claims involving three or more coverages, or a 1.2% probability, but in an independent model this probability increases to 3.9%. Given that 121 data points is insufficient to model dependence, we restrict our analysis to two coverages per claim, and so imposing a joint empirical distribution allows us to prevent more than two coverages occurring simultaneously. Table 2 lists scenario probabilities across collision-linked events adjusted for the occurrence of only two coverages.

Table 2 – Scenario probabilities of coverages for an accident

	AB	Coll	DCPD	TPL	Probability
1 coverage	1	0	0	0	2.77%
	0	1	0	0	35.30%
	0	0	1	0	40.78%
	0	0	0	1	1.29%
2 coverages	1	1	0	0	3.54%
	1	0	1	0	7.23%
	1	0	0	1	1.89%
	0	1	1	0	3.01%
	0	1	0	1	3.27%
	0	0	1	1	0.94%

Note: Probabilities are adjusted to sum to 100%

Next, in order to model $\mathbf{Y}_k$, we need to select marginal claim distributions for each coverage, and appropriate copulas based on the observed dependence structure. To evaluate the impact of loss distributions on unearned premium risk, we consider two cases for our marginal distributions. First, we select a homogeneous distribution for all losses of a particular coverage, and then we use a generalised linear models (GLM) approach to have a heterogeneous distribution of losses based on individual characteristics.

Although normally we would select among common loss distributions in actuarial literature for potential marginals, see Klugman et al. (2012), we choose to restrict ourselves to the Gamma distribution, optimised using the *actuar* (Dutang et al., 2008) package in R. The motivation behind this choice lies in comparing similar distributions, in this case a Gamma distribution with another Gamma distribution, where in the heterogeneous model

$$\mathbb{E}\left[Y_k^{(c)}\right] = e^{\mathbf{D}_k^T \boldsymbol{\gamma}},\ k = 1, \ldots, \mathrm{n} \tag{8}$$

with $\mathbf{D}_k$ a $(q\times 1)$ set of covariates which can be different from the one used for frequency, and $\boldsymbol{\gamma}_k$ a $(q\times 1)$ set of predictors optimised through maximum likelihood estimation. The parameters obtained for the homogeneous case are in Appendix C while the predictors obtained for the heterogeneous model are in Appendix D.

Then, to determine which pairs of coverages may have dependent loss amounts, we look at Kendall's tau and Spearman's rho (see Kendall (1948)), listed in Table 3, and calculated with the help of the *VGAM* (Yee et al., 2010) package in R. We determine that there is weak dependence between Accident Benefits and DCPD, tail dependence between Accident Benefits and Liability, and strong dependence between Collision and DCPD, as we can observe in Figures 4 to 6. The line observed near 0.44 for accident benefits in Figure 4 stems from policy limits and is not abnormal.

Table 3 – Kendall's Tau and Spearman's Rho for potentially dependent coverages

Coverages	Kendall's Tau	Spearman's Rho
AB, Coll	0.011	0.077
AB, DCPD	**0.053**	**0.097**
AB, Liab	**0.280**	**0.397**
Coll, DCPD	**0.429**	**0.435**
Coll, Liab	0.027	0.043
DCPD, Liab	-0.005	-0.000

For copulas, we use rank plots and a left-right tail function defined as

$$LR(z) = \begin{cases} \Pr(F_X(x) < z | F_Y(y) < z), & \text{if } 0 \leq z < 0.5 \\ \Pr(F_X(x) > z | F_Y(y) > z), & \text{if } 0.5 \leq z < 1, \end{cases}$$

which allows for comparing the curve obtained through the function with a theoretical curve and choosing the closest fit. More information about this function can be found in Venter (2002) and Boucher et al. (2008).

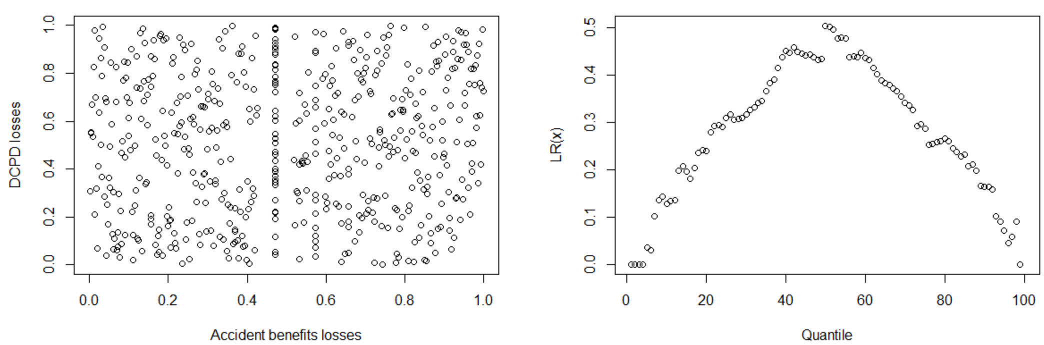
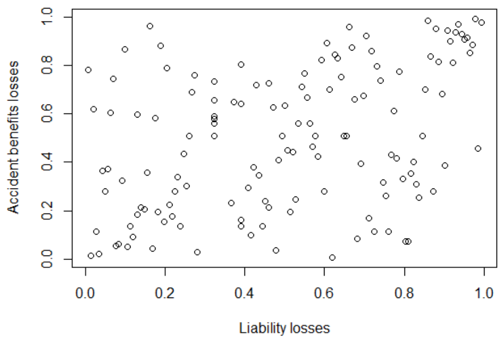


Figure 4 – Rank plot (left) and left-right tail function (right) for Accident Benefits and DCPD



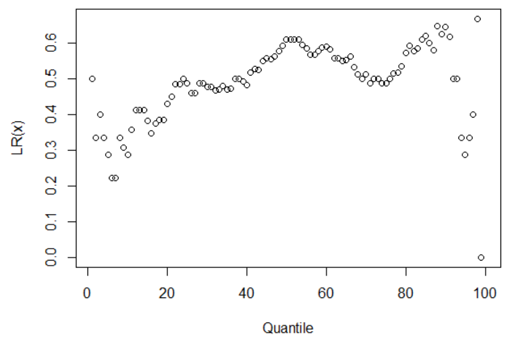


Figure 5 – Rank plot (left) and left-right tail function (right) for Accident Benefits and Liability

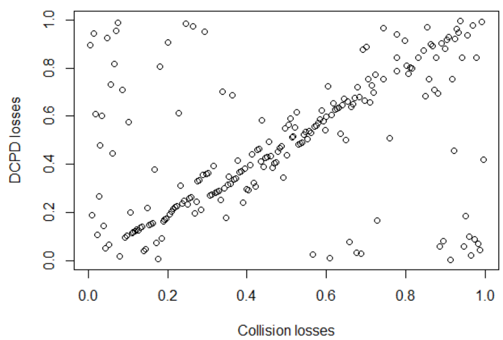


Figure 6 – Rank plot for Collision and DCPD

These figures allow us to select a Frank copula between AB and DCPD, a Gumbel copula between AB and Liability, and we happen to know the exact link between Collision and DCPD. These coverages are in fact two sides of the same coin, where claim cost is split based on percentage of responsibility. We thus model this link through observed frequency of percentage breakdown between 0/100, 25/75, 40/60, and 50/50, with proportions found in Table 4. Note that in 100% at fault or not at fault cases, the accident will be covered fully respectively by Coll (at fault) or DCPD (not at fault).

Table 4 – Collision/DCPD percentage split

	Coll/DCPD split						
	0/100	**25/75**	**40/60**	**50/50**	**60/40**	**75/25**	**100/0**
Proportion	9.25%	3.96%	7.49%	65.20%	4.85%	1.76%	7.49%

4.3 Results

With our previous selections, we can obtain reserve amounts for the unearned premium risk, assuming the insurer uses uniform acquisition of premium. We run 20,000 simulations following the algorithm described in Section 3.4 with the *copula* package in R, allowing us to obtain an unearned premium risk distribution and evaluate different risk measures. The choice of 20,000 is motivated by a balance between sufficient data and computing time. Given that we have an Ontario database, we consider the Financial Services Commission of Ontario (FSCO) guidelines, which follow Canadian federal guidelines, and so we look at the 99^{th} Conditional Tail Expected (CTE) as well as the 99.5^{th} Value-at-Risk (VaR), presented in Table 5 under both homogeneous and heterogeneous distributions. We present what would happen under independent occurrence (that is, independent $I_k^{(c)}$) in Table 6.

Table 5 – Reserve amounts using multiple loss approaches

Approach		**Fitted set (000s)**		**Predictive set (000s)**	
		Ind.	**Dep.**	**Ind.**	**Dep.**
Homogeneous	**Mean**	−2 485	−2 662	−2 284	−2 513
	VaR (99.5%)	2 417	2 175	2 949	2 699
	CTE (99%)	2 622	2 286	3 105	2 861
Heterogeneous	**Mean**	−4 009	−4 389	−4 302	−4 770
	VaR (99.5%)	3 167	2 901	3 189	2 835
	CTE (99%)	3 548	3 299	3 617	3 248
	Variance	5 261 643**		5 835 444**	
Formula (Prop. 3.1)	**Mean**	−4 013	−4 285*	−4 304	−4 634*
	Variance	5 254 415**		6 066 490**	
Observed		−2 952		−8 247	

* The mean is lower for the dependent case because of the added joint distribution of (Coll, DCPD).
** Variance is in millions not thousands.

Table 6 – Reserve amounts in a purely independent heterogeneous model

	Fitted set (000s)	**Predictive set (000s)**
Mean	−3 990	−4 271
VaR (99.5%)	1 100	1 119
CTE (99%)	1 224	1 285

We note that the dependent model always returns a lower reserve than the heterogeneous case. While seemingly counter-intuitive due to the positive dependence, this is expected. In the particular case of the Coll-DCPD relationship, modelling claims involving both these coverages leads to lower amounts than modelling each coverage separately, then adding them up. AB and TPL claims can incur large amounts, but given our choice of the Gamma distribution which does not have a heavy tail, in the absence of large AB-Liab claims, the Coll-DCPD relationship leads to lower simulated reserve amounts.

Bearing in mind the research questions of drivers of unearned premium risk and impact of cross-coverage dependence, these results, despite not being directly intuitive, highlight the importance of properly recognising the relationship between coverages. While one might expect positive dependence to lead to higher loss reserves, it is important to take into account what actually occurs when multiple coverages are implied, as joint losses might behave differently than independent losses. Looking at means as a simple example of this idea, the mean of joint Coll-DCPD claims is much lower than the mean of collision claims added to the mean of DCPD claims, which creates bias in an independent model that can be corrected in an individual model taking dependence into account. We can therefore see that cross-coverage dependence plays a central role in proper evaluation of reserve amounts.

Next, we note that the means are negative; this indicates a surplus. This is to be expected, as an insurer does not only pay losses, but also pays expenses and generally keeps a certain profit margin. One would thus expect an average loss ratio around 70%. While on the fitted data both the homogeneous and heterogeneous models capture this well enough, on the predictive set we see that the homogeneous model has a distribution almost entirely to the right of the observed reserve amount while the heterogeneous model presents a better fit. We thus see that for the same loss distribution (Gamma in this case), using an individual model enables more predictive power, and so the choice of loss model is an important driver of risk.

Finally, we see by comparing Tables 5 and 6 that using independent occurrence of coverages increases Z. This is expected, as explained in Subsection 4.2, as independent occurrence leads to a disproportionate percentage of claims with multiple coverages, which in turn leads to potentially higher total loss. This highlights the importance of respecting the dependence structure in coverage occurrence and not limiting ourselves to dependence in loss distributions. So we can conclude that cross-coverage dependence, both in terms of occurrence and loss amount, is an important element of modelling unearned premium risk.

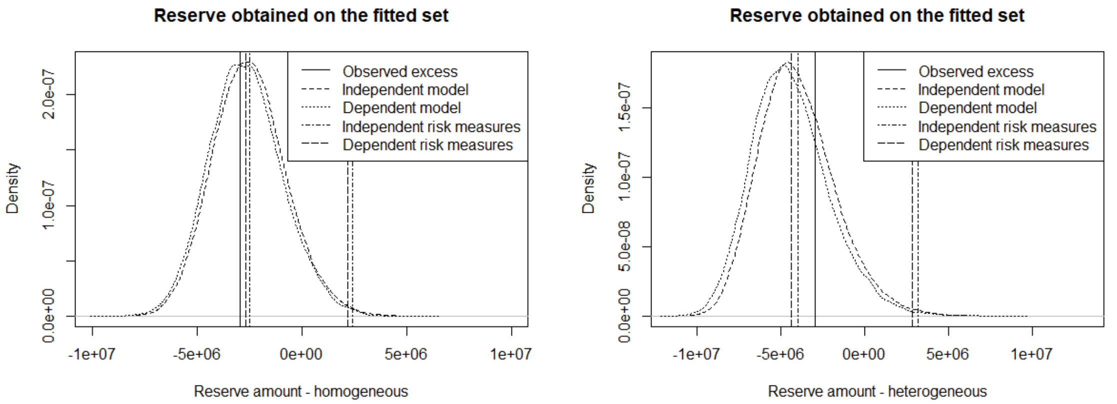


Figure 7 – Reserve amounts for unearned premium risk for fitted data using a homogeneous model (left) and heterogeneous model (right)

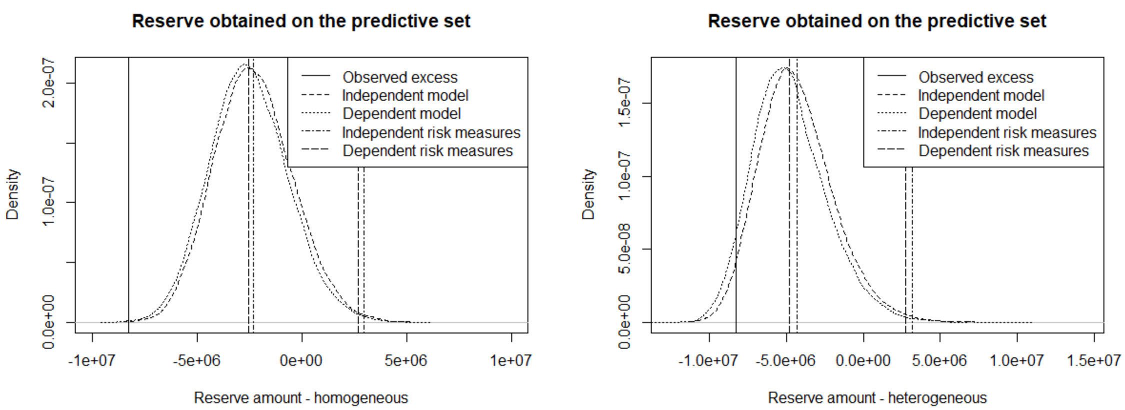


Figure 8 – Reserve amounts for unearned premium risk for predictive set using a homogeneous model (left) and heterogeneous model (right)

In Figures 7 and 8, we compare the distribution of unearned premium risk with and without dependence for both a homogeneous loss model and a heterogeneous one. We only present the models using seasonal exposure instead of uniform exposure because both curves are nearly identical; this is due to weak seasonality in our data and would likely create a larger difference with stronger seasonality. Our model implies that seasonality is in fact one of the main drivers of unearned premium risk, and so in a situation with strong seasonality, acknowledging seasonality in our exposure measure and premium acquisition would become important. The vertical full line is the actual observed amount of future losses minus the actual unearned premium reserve supposing a uniform acquisition pattern. The dotted vertical lines are the mean and 99.5% quantile for the independent case while the long dashes are the same measures for the dependent case. Notice that the dependent model is lower than the independent model in both the fitted dataset and the predictive dataset, as explained earlier in this subsection.

Relating this back to our supposition that loss distribution is one of the main drivers of unearned premium risk, we can deduce from Figures 7 and 8 as well as Table 5 that the homogeneous models vastly over-evaluate the necessary reserves while the heterogeneous models provide an accurate reserve, thus suggesting that using an individual GLM approach is better than using a homogeneous model.

5 Conclusion

In this paper, we analysed the risk linked to unearned premium in Property & Casualty insurance through a non-homogeneous Poisson process and an individual model including between-coverage dependence. That is, we used piece-wise constant seasonality of losses as an exposure base to build a generalised linear model for claim occurrence, and modelled loss amounts through copulas.

Recalling the main questions at the start of Section 4, we can conclude that there are four main drivers to unearned premium risk. Given that exposure to risk in our model is driven by loss seasonality and the effective date of contracts, we see that the timing of losses during the year and when contracts are written, or subscription patterns, are two important drivers of risk.

Next, unearned premium risk being the risk that future losses exceed the unearned premium reserve, loss distribution and premium acquisition are necessarily other main drivers of risk. In terms of loss distribution, our data study clearly indicates that recognising cross-coverage dependence has an impact on projecting losses and so for a more accurate model, it is preferable to include dependence. In fact, we showed that some coverages, such as Collision and DCPD, have strong dependence and should not be considered independently. As such, current models for Claims Liability would also gain in using between-coverage dependence, allowing for a more realistic claims projection. Then, we suggested that the UPR can easily be established at the evaluation date. It is easy to see how recognising seasonality in the acquisition of premium would lead to a better cashflow matching between losses and premium, thus potentially decreasing unearned premium risk.

There are therefore four main drivers to unearned premium risk: seasonality, loss distributions, the acquisition pattern of premium, and the subscription pattern of insureds. Our two other questions are addressed within these four drivers of risk. All things considered, we can conclude that given the advances in computational power, it would be encouraged to stop working with loss development triangles and to use stronger statistical tools to evaluate reserves while including dependence.

Given the high variability of losses in P&C insurance, it would be interesting to have more accident years to work with, as we only had data to work with for two accident years. Having multiple years of data would also enable us to observe the impact of trends on unearned premium risk. Our database unfortunately had fairly weak seasonality and so we could not fully observe its impact on unearned premium risk. It would be interesting to see how risk shifts with stronger presence of seasonality, for example in a heavily snow-prone region where winter generally implies more accidents. Moreover, in the presence of strong seasonality one may suppose that a particular type of accident would be more prevalent (e.g. sliding on ice, storm surges, hail storms), leading to a relation between time of loss and type/amount of loss. It would thus also be interesting to relax the hypothesis of independence between time of loss and loss amount. We may also want to generalise our model to more than one claim, where we use N_k instead of transforming it into a binary variable.

6 Acknowledgements

We would like to thank the Quebec insurance regulator, the Authorité des Marchés Financiers, for its financial support in this project, as well as The Co-Operators for providing data to analyse.

Appendix A - Proof of Proposition 3.1

We seek to evaluate the sum of losses linked to unearned premium. For a loss to fall into this category, it must arrive after the evaluation date but before the end of the contract. Recall that N_k is a discrete random variable for the number of future losses for the k^{th} contract, $k = 1, \dots, n$. Then S_k the future loss across C coverages is

$$S_k = \begin{cases} \sum_{c=1}^{C} \mathbb{1}_k^{(c)} Y_k^{(c)} & \text{if } N_k > 0 \\ 0 & \text{if } N_k = 0. \end{cases}$$

It then follows that

$$\mathbb{E}\left[S_k | t_k^{(E)}\right] = \mathbb{E}\left[\mathbb{E}\left[S_k | N_k, t_k^{(E)}\right] | t_k^{(E)}\right] = \Pr(N_k > 0) \sum_{c=1}^{C} \Pr(\mathbb{1}_k^{(c)} = 1) \mathbb{E}\left[Y_k^{(c)}\right],$$

which makes sense intuitively: the expected future loss is the sum of weighted expected losses by coverage, weighted by the probability of a loss occurring. Now, recalling Equations 4 and 5, we can rewrite the expected value as

$$\mathbb{E}\left[S_k | t_k^{(E)}\right] = \left(1 - e^{-\lambda_k (t_k^{(E)} | \mathbf{X}_k)}\right) \sum_{c=1}^{C} \Pr(\mathbb{1}_k^{(c)} = 1) \mathbb{E}\left[Y_k^{(c)}\right].$$

Straightforwardly, we obtain

$$\mathbb{E}\left[Z | t_k^{(E)}\right] = \sum_{k=1}^{n} \left(1 - e^{-\lambda_k (t_k^{(E)} | \mathbf{X}_k)}\right) \sum_{c=1}^{C} \Pr(\mathbb{1}_k^{(c)} = 1) \mathbb{E}\left[Y_k^{(c)}\right] - \sum_{k=1}^{n} P_k^{(UE)}.$$

The proof for the variance is similar to the one for the expected value. Consider the indicator function

$$J_k = \begin{cases} 1 & \text{if } N_k > 0 \\ 0 & \text{if } N_k = 0, \end{cases}$$

defined in Subsection 3.1. We have

$$\begin{aligned}
\mathrm{Var}[S_k] =& \mathbb{E}[\mathrm{Var}[S_k|J_k]] + \mathrm{Var}[\mathbb{E}[S_k|J_k]] \\
\mathrm{Var}[S_k] =& \Pr(N_k > 0)\mathrm{Var}\left[\mathbf{I}_k^T \mathbf{Y}_k\right] + \Pr(N_k > 0)\Pr(N_k = 0)\left[\mathbb{E}\left[\mathbf{I}_k^T \mathbf{Y}_k\right]\right]^2 \\
=& \Pr(N_k > 0)\left[\sum_{c=1}^{C} \mathrm{Var}\left[I_k^{(c)} Y_k^{(c)}\right] + 2\sum_{i=1}^{C-1}\sum_{j=i+1}^{C} \mathrm{Cov}\left[I_k^{(i)} Y_k^{(i)}, I_k^{(j)} Y_k^{(j)}\right]\right] \\
&+ \Pr(N_k > 0)\Pr(N_k = 0)\left[\sum_{c=1}^{C} \mathbb{E}\left[I_k^{(c)}\right]\mathbb{E}\left[Y_k^{(c)}\right]\right]^2 \\
=& \Pr(N_k > 0)\left[\sum_{c=1}^{C} \mathbb{E}\left[(I_k^{(c)})^2\right]\mathbb{E}\left[(Y_k^{(c)})^2\right] - \mathbb{E}\left[I_k^{(c)}\right]^2 \mathbb{E}\left[Y_k^{(c)}\right]^2\right. \\
&\left. +2\sum_{i=1}^{C-1}\sum_{j=i+1}^{C} \mathbb{E}\left[I_k^{(i)} Y_k^{(i)} I_k^{(j)} Y_k^{(j)}\right] - \mathbb{E}\left[I_k^{(i)} Y_k^{(i)}\right]\mathbb{E}\left[I_k^{(j)} Y_k^{(j)}\right]\right] \\
&+ \Pr(N_k > 0)\Pr(N_k = 0)\left[\sum_{c=1}^{C} \mathbb{E}\left[I_k^{(c)}\right]\mathbb{E}\left[Y_k^{(c)}\right]\right]^2 \\
=& \left(1 - e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\right)\left[\sum_{c=1}^{C} \mathbb{E}\left[(I_k^{(c)})^2\right]\mathbb{E}\left[(Y_k^{(c)})^2\right] - \mathbb{E}\left[I_k^{(c)}\right]^2 \mathbb{E}\left[Y_k^{(c)}\right]^2\right. \\
& +2\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}\left(\mathbb{E}\left[Y_k^{(i)} Y_k^{(j)}\right]\Pr(I_k^{(i)} = 1, I_k^{(j)} = 1)\right. \\
&\left.\left. -\mathbb{E}\left[I_K^{(i)}\right]\mathbb{E}\left[I_k^{(j)}\right]\mathbb{E}\left[Y_k^{(i)}\right]\mathbb{E}\left[Y_k^{(j)}\right]\right)\right] \\
&+ \left(1 - e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\right) e^{-\lambda_k(t_k^{(E)}|\mathbf{X}_k)}\left[\sum_{c=1}^{C} \mathbb{E}\left[I_k^{(c)}\right]\mathbb{E}\left[Y_k^{(c)}\right]\right]^2.
\end{aligned}$$

Appendix B

Table 7 – MLE parameters for all coverages for the Gamma distribution

	AB	**Coll**	**Comp**	**DCPD**	**Coll+DCPD**	**TPL**
Shape	0.50 (0.00)	1.40 (0.07)	0.47 (0.02)	1.41 (0.04)	1.41 (0.12)	0.34 (0.04)
Scale	63 747.38 (7 838.25)	4 460.40 (277.31)	4 254.52 (319.18)	6 043.20 (349.27)	4 838.24 (492.46)	203 320 (41 129.67)

Appendix C

Table 8 – GLM loss predictors obtained by MLE by coverage

	AB		**Coll**		**Comp**		**DCPD**		**TPL**	
Intercept	9.29	(0.59)***	8.35	(0.14)***	6.94	(0.28)***	8.38	(0.14)***	11.03	(1.29)***
Var1	0.26	(0.26)	0.11	(0.07)	0.44	(0.15)**	0.18	(0.07)*	NA	NA
Var2	0.07	(0.34)	0.14	0.10	0.06	(0.24)	-0.00	(0.09)	NA	NA
Var3	0.64	(0.27)*	-0.03	(0.07)	-0.05	(0.16)	-0.10	(0.07)	NA	NA
Var4	-0.56	(0.41)	-0.17	(0.11)	-0.43	(0.21)*	-0.31	(0.11)**	-0.75	0.60
Var5	-0.98	(0.44)*	-0.08	(0.12)	0.08	(0.28)	-0.17	(0.11)	-0.48	0.66
Var6	1.24	(0.85)	0.40	(0.17)*	0.32	(0.37)	0.35	$(0.20)^{+}$	-0.66	(0.41)
Var7	0.44	(0.46)	-0.04	(0.13)	-0.20	(0.24)	0.08	(0.11)	-0.62	(0.90)
Var8	0.74	(0.50)	0.19	(0.12)	0.56	(0.24)*	0.01	0.11	0.59	(0.71)
Var9	0.13	(0.69)	0.34	$(0.17)^{+}$	0.58	$(0.34)^{+}$	-0.03	(0.17)	0.69	(1.19)
Var10	-0.26	(0.27)	0.11	(0.08)	-0.16	(0.16)	0.27	(0.07)***	1.03	(1.37)
Var11	1.44	(0.63)*	-0.10	(0.34)	0.10	0.45	0.05	0.21	-1.07	(0.42)*
Var12	-0.27	(0.43)	0.34	(0.13)	0.74	(0.31)*	0.05	(0.14)	-0.39	(0.79)

Note: Variable names removed for confidentiality purposes. TPL variables limited due to lower data than the other coverages. ***: significant at 0.1%, **: significant at 1%, *: significant at 5%, $^{+}$: signicant at 10%.

References

Abdallah, A., Boucher, J.-P., and Cossette, H. (2016). Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. *Insurance: Mathematics and Economics*, 68:120–133.

Antonio, K. and Plat, R. (2014). Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 2014(7):649–669.

Barnett, G., Zehnwirth, B., and Dubbossarsky, E. (2005). When can accident years be regarded as development years? *Proceedings of the Casualty Actuarial Society*, 92(177):239–256.

Bornhuetter, R. L. and Ferguson, R. E. (1972). The actuary and IBNR. *Proceedings of the casualty actuarial society*, 59(112):181–195.

Boucher, J.-P., Denuit, M., and Guillén, M. (2007). Risk classification for claim counts: a comparative analysis of various zeroinflated mixed poisson and hurdle models. *North American Actuarial Journal*, 11(4):110–131.

Boucher, J.-P., Denuit, M., and Guillen, M. (2008). Models of insurance claim counts with time dependence based on generalization of poisson and negative binomial distributions. *Variance*, 2(1):135–162.

Canadian Institute of Actuaries (2014). Premium liabilities. *Educational Notes*.

Collins, E. and Hu, S. (2003). Practical considerations in valuing premium liabilities. *Institute of Actuaries of Australia 14th General Insurance Seminar*.

Cossette, H., Côté, M.-P., Marceau, E., and Moutanabbir, K. (2013a). Multivariate distribution defined with farlie–gumbel–morgenstern copula and mixed erlang marginals: Aggregation and capital allocation. *Insurance: Mathematics and Economics*, 52(3):560–572.

Cossette, H., Mailhot, M., Marceau, É., and Mesfioui, M. (2013b). Bivariate lower and upper orthant value-at-risk. *European actuarial journal*, 3(2):321–357.

Côté, M.-P., Genest, C., and Abdallah, A. (2016). Rank-based methods for modeling dependence between loss triangles. *European Actuarial Journal*, 6(2):377–408.

Dutang, C., Goulet, V., Pigeon, M., et al. (2008). actuar: An r package for actuarial science. *Journal of Statistical software*, 25(7):1–37.

Feldblum, S. (1996). Naic property/casualty insurance company risk-based capital requirements. *Proceedings of the Casualty Actuarial Society*, 83(159):300–389.

Financial Services Commission of Ontario (2016). What do the coverages mean. url http://www.fsco.gov.on.ca/fr/auto/brochures/pages/what-the-coverages-mean-glossary.aspx. Last visited 2019-05-18.

Frees, E. W. (2008). Dependent insurance risks. *Encyclopedia of Quantitative Risk Analysis and Assessment*, 2.

Frees, E. W. and Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 103(484):1457–1469.

Friedland, J. (2010). Estimating unpaid claims using basic techniques. In *Casualty Actuarial Society*, volume 201.

Joe, H. (2014). *Dependence modeling with copulas*. Chapman and Hall/CRC.

Kaye, P. (2005). Risk measurement in insurance a guide to risk measurement, capital allocation and related decision support issues. *Casualty Actuarial Society Discussion Paper*.

Kendall, M. G. (1948). Rank correlation methods.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss models: from data to decisions*. John Wiley & Sons.

Li, J. (2010). Prediction error of the future claims component of premium liabilities under the loss ratio approach. *Variance*, 4(2):155–169.

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, 23(2):213–225.

Meyers, G. (2013). Correlations between accident years in loss reserve models. *The Actuarial Review*, 40(1):27–28.

Office of the Superintendent of Financial Institutions (2018a). Minimum capital test for federally regulated property and casualty insurance companies. url http://www.osfi-bsif.gc.ca/eng/fi-if/rg-ro/gdn-ort/gl-ld/Pages/mct2018.aspx. Last visited 2019-06-27.

Office of the Superintendent of Financial Institutions (2018b). Own risk and solvency assessment. url http://www.osfi-bsif.gc.ca/eng/fi-if/rg-ro/gdn-ort/gl-ld/Pages/e1918.aspx. Last visited on 20019-08-07.

Pechon, F., Denuit, M., and Trufin, J. (2019). Multivariate modelling of multiple guarantees in motor insurance of a household. *European Actuarial Journal*, pages 1–28.

Pigeon, M., Antonio, K., and Denuit, M. (2013). Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin: The Journal of the IAA*, 43(3):399–428.

Priest, C. (2012). Premium liability correlations and systemic risk. *Australian Actuarial Journal*, 18(1):1–66.

Shi, P., Zhang, W., and Boucher, J.-P. (2018). Dynamic moral hazard: A longitudinal examination of automobile insurance in canada. *Journal of Risk and Insurance*, 85(4):939–958.

Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, 9(6):449–460.

Venter, G. G. (2002). Tails of copulas. *Proceedings of the Casualty Actuarial Society*, 89(171):68–113.

Verrall, R. and Wüthrich, M. (2016). Understanding reporting delay in general insurance. *Risks*, 4(3):25.

Wüthrich, M. V. and Merz, M. (2008). *Stochastic claims reserving methods in insurance*. John Wiley & Sons.

Yee, T. W. et al. (2010). The vgam package for categorical data analysis. *Journal of Statistical Software*, 32(10):1–34.

Zhao, X. and Zhou, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics*, 46(2):290–299.