General Insurance Deductible Ratemaking with Applications to the Local Government Property Insurance Fund

Gee Y. Lee *
Principal Investigator
Department of Risk and Insurance
Wisconsin School of Business

Edward W. Frees †
Co-Investigator
Department of Risk and Insurance
Wisconsin School of Business

November 4, 2016

Abstract. Insurance claims have deductibles, which must be considered when pricing for insurance premium. Deductibles may cause censoring and truncation to the observed insurance claims. For this type of data, the regression approach is often used with deductible amount included as an explanatory variable inside a frequency-severity model, so that the resulting coefficient can be used for an assessment of the relativities for deductibles. This approach has the advantage of incorporating the selection effect into deductible ratemaking. On the other hand, standard actuarial textbooks recommend the maximum likelihood approach for estimating parametric loss models, which can be used for calculating the coverage modification amounts due to the deductibles. In this paper, a comprehensive overview of deductible ratemaking is provided, and the pros and cons of various approaches under different parametric models are compared. The regression approach proves to have an advantage in predicting aggregate claims when deductible choices influence the frequency and severity distributions. The maximum likelihood approach becomes necessary for calculating theoretically correct relativities for deductible levels beyond those observed, for each policyholder. For demonstration, loss models are fit to the Wisconsin Local Government Property Insurance Fund data, and examples are provided for the ratemaking of per-loss deductibles offered by the fund. Selected parametric models from the generalized beta family distributions are compared. Models for specific peril types can be combined to improve the ratemaking, and estimation issues for such models under truncation and censoring are discussed.

*Email: gylee2@wisc.edu. We thank Scott Lennox, Erika Schulty, Dale Hall, Steven Siegel, Barbara Scott, and the Project Oversight Group members Robert Winkelman, Michael Ewald, Jean-Marc Fix, Josh Newkirk, Rafi Herzfeld, Andy Ferris, Crispina Caballero, Joe Wurzburger, Julianne Callaway, Tina Liu, who provided support throughout this project.

†Email: jfrees@bus.wisc.edu.

1 Introduction

A deductible is an important feature of an insurance contract. Deductibles influence the number of times the insured will make a claim and will influence the amount that is reimbursed to the insured in the event of an insured loss. In many cases, deductibles may cause insurance claims to be observed with censoring and truncation. These aspects must be addressed when pricing insurance premiums, and the theoretically correct approach can be discussed from the standpoint of actuarial theory.

To formalize our framework for modeling, for each policyholder let N be the underlying frequencies and Y_j be the severities of the claims, independent of each other. Suppose a deductible d is applied, so that the risk-sharing function is defined as

$$g(Y_j; d) = \begin{cases} 0 & Y_j < d \\ Y_j - d & d \le Y_j < \infty \end{cases}$$

The observed, censored and truncated random variable for claim frequencies and severities for each policyholder can be denoted as

$$N_g(d) = \sum_{j=1}^{N} \mathrm{I}(d < Y_j) \qquad \text{(number of claims)}$$

$$Y_{g,j}(d) = \begin{cases} 0 & Y_j < d \\ Y_j - d & d \le Y_j < \infty \end{cases} \qquad \text{(censored severities)}$$

$$Y_{*,j}(d) = Y_j - d \mid d \le Y_j \qquad \text{(truncated severities)}$$

$$S_g(d) = \sum_{j=1}^{N} Y_{g,j}(d) \qquad \text{(aggregate claims)}$$

where $\mathrm{I}(\cdot)$ is an indicator function, taking on the value 1 if the input condition is true, and 0 otherwise. Note that $N_g(d)$ is a summation of Bernoulli random variables. We will be consistent in notating censored random variables with a subscript (g), to remember they have extra zeros below the censoring point. Also, we will denote truncated random variables and their corresponding parameters with a subscript $*$. It is helpful to understand that truncation is basically observing a subset of a full sample, under some truncation mechanism.

The textbook Klugman et al. (2012) shows in detail how coverage modification affects the claim frequency and severity distributions. The advantage of applying parametric loss models for deductible ratemaking is that accurate, theoretically correct deductible rates can be calculated for insurance losses. When covariates are incorporated into the models, deductibles can be priced in a subject-specific manner, which allows a rating engine to be theoretically correct for all of the policyholders within an insurance company. Empirical work using truncated estimation for insurance claims with real data can give practitioners an illustration of the application of loss models for deductible ratemaking.

Although accurate rates can be calculated using such textbook approaches, a practitioner may be interested in the regression approach for deductible ratemaking, by treating the deductible level as an explanatory variable in a regression model, as in Frees and Lee (2016). This intuitive solution is to use the coefficient estimates for log deductible to calculate the relativities for various deductible levels. This approach is taken, for simplicity of implementation and practicality in ratemaking applications. The approach becomes particularly useful when a large number of explanatory variables are used for ratemaking. A practitioner may be interested in learning when to apply truncated estimation techniques and when the regression approach suffices. Hence, in this paper, a detailed analysis of deductible rating approaches is conducted.

For some more motivation for the study, the reader may consider the situation where an analyst would be interested in developing a pricing structure that incorporates deductibles in a disciplined way, and in knowing how to change prices when the deductibles change. Yet only the reported losses above a certain deductible level may be observed. In this circumstance, an actuarial analyst may need an assessment of the price of a particular insurance policy or a portfolio of policies. For these considerations, we believe an overview of the available methods and a comparison of the approaches using empirical applications to be a meaningful contribution to the literature.

2 Literature Review

There is a large literature discussing problems related to deductible ratemaking. Some foundational literature on deductible pricing, exposure rating and coverage modification is summarized in the following subsections. Statistical methods related to censored and truncated estimation have a long history, as does the insurance economics literature, where the deductible choice of policyholders is studied for the assessment of risk preferences. Readers who are interested in the main subject of the paper may skip this section and go directly to Section 3.

Deductible Pricing

A standard reference for deductible pricing in actuarial science is in Brown and Lennox (2015), where the *indicated deductible relativity* for a single loss of an insurance policy is given by the relationship

$$\text{Indicated deductible relativity} = \frac{E[Y_*(d)]}{E[Y]} = 1 - \text{LER}(d),$$

where LER is an abbreviation of *loss elimination ratio*. The indicated deductible relativity provides an assessment of how much an insurance loss cost is reduced by a deductible, from a per-loss perspective, while the loss elimination ratio provides an assessment of how much the covered loss is reduced by introducing a deductible d. If the policy has an upper limit of coverage, u, then the loss elimination ratio is

$$\text{LER}(d) = \frac{\displaystyle\int_0^d y\text{f}_Y(y)dy + d\int_d^u \text{f}_Y(y)dy}{\displaystyle\int_0^u y\text{f}_Y(y)dy}.$$

This principle can be applied to excess-of-loss treaty pricing for per-loss insurance and reinsurance policies, where losses beyond a retention level are covered by a reinsurer. For the frequency-severity framework, it is helpful to use the notations in the following Section 3, to define the relativity of an *aggregate loss* as

$$\text{REL}(d_0, d) = \frac{E[S_g(d)]}{E[S_g(d_0)]} = \frac{E[N]\displaystyle\int_d^u (1-\text{F}_Y(y))dy}{E[N]\displaystyle\int_{d_0}^u (1-\text{F}_Y(y)dy} = \frac{\displaystyle\int_d^u (1-\text{F}_Y(y))dy}{\displaystyle\int_{d_0}^u (1-\text{F}_Y(y))dy}, \tag{1}$$

where d_0 is a base deductible. In the textbook, Brown and Lennox (2015), the *experience rating* approach, and the *exposure rating* approach for reinsurance pricing are introduced in relation to deductible ratemaking. The former uses a company's historical loss experience, for a best predictor of future experiences. In the latter approach, claim severity distributions are based on industry data. The literature has some work related to excess-of-loss layer rating methodologies.

An article by Bernegger (1997) uses the Maxwell-Boltzmann, Bose-Einstein, Fermi-Dirac distributions. Statistical properties of this class of distributions are introduced further in Wu and Cai (1999). To summarize their approach, an insurance company may be given an increasing curve in δ—say, $H(\delta)$, where $0 < \delta < 1$. In this case, $\delta = d/u$ is a normalized deductible in the $[0, 1]$ interval. This curve is differentiated to obtain an expression for the loss distribution.

Several related studies have been interested in the rating of large insurance losses and the excess-of-loss layer rating. For example, Ludwig (1991) provides an overview of the exposure rating approach. Fasen and Kluppelberg (2014) discusses risk processes for large insurance losses without empirical examples. Several seminars, such as White and Mrazek (2004) and White (2005), introduce advanced practical methodologies for exposure rating approaches. A recent article by Chavez-Demoulin et al. (2016) applies extreme value models to operational risk. Some researchers have studied the data misspecification issue under left-truncation, as the study by Gurnecki et al. (2006) has done. These studies provide good motivation for further studies. Our approach is to provide an empirical demonstration of how coverage modification effects can be incorporated into deductible ratemaking. We emphasize that our approach is distinct from existing work, in that our interest is more focused on the experience rating approach, using data from the Local Government Property Insurance Fund, introduced further in Sections 5.2, 5.3 and 8.3.

Coverage Modification

In many cases, the deductible levels correspond to only small values in the lower tail of the claim distribution. Hence, it is often most efficient to use the regression approach with independent explanatory variables, for both the frequencies and the severities of insurance claims. However, for large deductible amounts, there may be motivation to use other approaches.

For a specific class of frequency distributions, called the $(a, b, 0)$ class distributions, the modification to the frequencies, due to deductibles, has been understood quite well. The $(a, b, 0)$ class distributions, summarized in Table 1, are explained in detail by Klugman et al. (2012). These frequency distributions have the property such that a scale in the parameter θ results in the same scale to the mean of the distribution. If the mean of the distribution with parameter θ is given as $E[Y]$, the mean of the distribution with the scaled parameter θv has mean $E[Y]v$. This property can be easily observed by inspecting Table 1.

Table 1: (a, b, 0) Class Distributions

Name	$B(z)$	$B(\theta(z-1))$	Mean
Poisson	e^z	$e^{\theta(z-1)}$	θ
Binomial	$(1+z)^m$	$((1-\theta)+\theta z)^m$	$m\theta$
Geometric	$(1-z)^{-1}$	$(1-\theta(z-1))^{-1}$	θ
Negative Binomial	$(1-z)^{-r}$	$(1-\theta(z-1))^{-r}$	$r\theta$

To understand the effect of left-truncation of the severities on the frequencies, an effect also called *coverage modifications* in the language of Klugman et al. (2012) is introduced. Let $v = 1 - \mathrm{F}_Y(d)$ be the probability that a loss results in a claim (a payment). The probability-generating function for the modified claim count random variable can be obtained by modifying the probability-generating function of the underlying (a, b, 0) class distribution $P(z) = B[\theta(z-1)]$ in the following way:

$$\begin{aligned} P_g(z) &= P(P_I(z)) = P(1 + v(z-1)) \\ &= B(\theta(1 + v(z-1) - 1)) = B(\theta v(z-1)) = P(z; \theta v). \end{aligned} \tag{2}$$

This interesting result uses the probability-generating function $P_I(z) = 1 - v + vz$ of the Bernoulli random variable, which takes on the value 1 when a loss results in a claim. The sum of N Bernoulli random variables has an $(a,\ b,\ 0)$ class primary distribution and a Bernoulli secondary distribution, in which case the probability-generating function for the secondary distribution can be plugged into the probability-generating function of the primary distribution to obtain the resulting compound distribution. This allows expression (2) to be so simple and intuitive.

The two most often used frequency distributions in actuarial science are the Poisson distribution and the negative binomial distribution. The Poisson distribution is often used in practice. Let N^c be the observed counts, excluding the unobserved claims due to truncation, for a policyholder. Note that from (2), we know the underlying frequencies also follow a Poisson distribution, so that

$$N \sim \text{Poisson}\,(\theta) \iff N_g \sim \text{Poisson}\,(v\theta)\,,$$

where $v = \Pr(Y > d)$ is the amount of coverage modification. In practice, the negative binomial is also often used, in order to accommodate for over-dispersion. In this case, a mean parametrization is used, and the underlying frequencies can be retrieved in a similar way:

$$N \sim \text{NB}\,(r, \theta) \iff N_g \sim \text{NB}\,(r, v\theta)\,,$$

where $v = \Pr(Y > d)$. These properties are valid under the assumption that N and v are independent, meaning the factors determining the severity are independent of the number of claims.

There has been little work on how to use an estimated loss distribution for small-deductible ratemaking in conjunction with the loss frequencies. When the loss frequencies and severities are empirically analyzed together, the sampling frame also becomes an issue. Cummings (2001) recommends the use of the generalized linear models (GLM) approach to deductible ratemaking. His presentation discusses the limitation of the method of Guiahi (2005), where the relationship between the claim frequencies and the deductibles is not considered. For practical reasons, Cummings (2001) recommends using standard GLM models with deductibles as an independent explanatory variable.

Truncated Estimation

There is a vast literature on censored and truncated data modeling in statistics. However, most of these papers focus on the estimation problem, where the goal is to obtain estimates of parameters for a specified distribution from censored or truncated data. Kaplan and Meier (1958) introduced the product limit estimator for censored and truncated data. There have been a number of follow-up studies, including Woodroofe (1985) and Lai and Ying (1991). Kalbfleisch and Prentice (2002) provides treatment of modeling for censored and truncated data for survival models. Finkelstein and Wolfe (1985) take a semi-parametric approach for interval censored failure time data. There have also been a number of studies focusing on estimation problems for particular distributions: Barr and Sherrill (1999) on the truncated normal, Aban et al. (2006) on the truncated Pareto distribution, and Chapman (1956) on the truncated gamma distribution. Discrete data with zero truncation is discussed in Plackett (1953) and Klugman et al. (2012). Recently, Verbelen and Claeskens (2014) applied multivariate Erlang mixture models to censored and truncated data.

Although the literature is vast, a combined estimation of frequencies and severities together is rarely found in the statistical estimation literature. The estimation of censored frequencies under a deductible to the severities distribution is treated theoretically in actuarial textbooks, but more empirical studies seem to be needed.

Insurance Economics

There is, however, another vast but separate literature where the selection effect mentioned in Cummings (2001) is studied in relation to policyholder behavior in the market. The selection effect occurs when specific deductible choices are correlated with the loss profile of a policyholder. The problem of deductible choice is important in insurance economics and risk management, as it is a crucial vehicle for sorting out the adverse-selection and moral-hazard problems in practice, so it is important to study their effects from an academic standpoint. The pricing of a deductible is an interesting problem, and the precise psychological effect of a deductible choice is a problem under active research. Moreover, the deductible choice problem serves as a framework for understanding economic decisions under uncertainty, which can be applied more broadly to problems in society. For this reason, traditional economics textbooks cover the deductible choice problem in depth. There is a vast literature in which deductibles are studied in order to understand the risk preference of policyholders and the presence of adverse selection in insurance markets.

In economics and risk management, articles such as Rothschild and Stiglitz (1976) and Halek and Eisenhauer (2001) have been standard references for the need of deductibles for mitigating adverse selection in insurance markets with hidden information. In economics and behavioral economics, deductible choices of policyholders have been used to study the risk preference of decision makers. Treatment of this literature can be found in Mas-Colell et al. (1995), Koszegi and Rabin (2006), and Sydnor (2010). Econometric approaches for measuring the preference of decision makers in a lab setting have been an active topic of research. For example, Holt and Laury (2002) provide standard procedures for measuring risk preferences in a lab environment. Recently, there is interest in extending these studies into real-world problems, through empirical studies such as Cohen and Einav (2007) and Einav et al. (2012). In particular, Sydnor (2010) illustrates how deductible choices are often unexplained by standard economic theory, and more sophisticated models may be necessary.

3 Theory of Coverage Modification

This section provides useful theoretical results for deductible ratemaking. The proofs are organized in the appendix (see Section 8.1). Similar results can be found in Klugman et al. (2012), Gray and Pitts (2012), Tse (2009), and Bahnemann (2015); however, here the results are simplified and condensed into more general forms, with an emphasis on small deductible changes. The results apply to any distribution Y, without continuity or the existence of a distribution required. The results for frequencies apply to any count random variable N, not just the $(a, b, 0)$ class distributions.

Let us begin by assuming the censored random variable is observed. From an empirical standpoint, the sample size on which estimation can be performed necessarily gets smaller for the claim severities, as Y is truncated to $Y_*(d)$. The first theorem provides a general expression for the difference in expected aggregate claims, under two different deductibles.

Theorem 3.1. *Let N be any count random variable, and Y any random variable, each with finite first moments. If N and Y are independent, then for deductibles $d_1 < d_2$, we have*

$$E[S_g(d_1) - S_g(d_2)] = E[N]E\left[Y_g(d_1) - Y_g(d_2)\right].$$

Theorem 3.1 provides an expression for the difference between two aggregate claims means, under two different deductibles. When the loss severities have a parametric loss distribution, the following corollaries allow a modeler to calculate the mean average claim and relativities for a given deductible d, relative to a base deductible d_0.

Corollary 3.1.1. *If Y has distribution function F_Y, then for $d_2 > d_1$ we have*

$$E[S_g(d_1) - S_g(d_2)] = E[N]\int_{d_1}^{d_2} (1 - F_Y(y))\, dy.$$

Corollary 3.1.2. *Using the notation for relativity REL as defined in equation (1), we have*

$$E[S_g(d_2)] = E[S_g(d_1)] \times REL(d_1, d_2) = E[S_g(d_1)] \times \frac{\displaystyle\int_{d_2}^{\infty} (1 - F_Y(y))\, dy}{\displaystyle\int_{d_1}^{\infty} (1 - F_Y(y))\, dy},$$

where d_1 may be considered as a base deductible.

This motivates the concept of the *relativity* for aggregate claims, as explained in Section 2. Hence, the mean aggregate claim is modified by an amount, depending on d. Given some loss distribution F_Y, the next theorem allows a modeler to recover the underlying loss frequencies, from observed claim frequencies.

Theorem 3.2. *Let N be any count random variable, and Y have distribution function F_Y, each with finite first moments. If Y is independent of N, then $N_g(d)$ satisfies*

$$E\left[N_g(d)\right] = E\left[N\right] \cdot (1 - F_Y(d)).$$

Theorem 3.2 provides an expression for the mean of the observed, censored frequency distribution, in terms of the underlying loss distribution parameters, under deductible d. When the mean frequency is parametrized using a log-link, for regression purposes, parameters for the underlying loss N can be obtained by a regression, using `offset` $= \ln(1 - \mathrm{F}_Y(d))$. Given the loss distributions, the following formulas provide the theoretical marginal changes in the means, under a small deductible change.

Corollary 3.2.1. *If $E[Y^c(d)]$ is differentiable at d, then*

$$\frac{\partial}{\partial d} E[Y_g(d)] = -1 + F_Y(d).$$

If F_Y is differentiable, then

$$\frac{\partial}{\partial d} E[N_g(d)] = -E[N] \cdot f_Y(d).$$

Corollary 3.2.2. *Let N be any count random variable, and Y any random variable, each with finite first moments. If N and Y are independent, then $E[S_g(d)]$ satisfies*

$$\frac{\partial}{\partial d}E[S_g(d)] = -E[N] \cdot (1 - F_Y(d)) \, .$$

Theorem 3.3 (Truncated Severity Modification)**.** *Let Y be any random variable with density f_Y and distribution F_Y. Then, $Y_*(d)$ satisfies*

$$\frac{\partial}{\partial d}E[Y_*(d)] = \frac{f_Y(d)}{1 - F_Y(d)}E[Y_*(d)] - 1$$

4 Approaches to Deductible Ratemaking

In this section, we provide an overview of different empirical approaches to deductible ratemaking and how they can be applied in our framework. The two general approaches are the regression approach and the maximum likelihood approach with truncated estimation methods. First, the sampling frame is formalized. For rating purposes, we assume the following variables are observed:

$$\{N_{g,i}(d_i), \boldsymbol{x}_i, d_i\} \, ,$$

where $\boldsymbol{x}_i$ is a set of explanatory variables including coverage amounts u_i, which could not be adjusted, while d_i are the deductible choices, which can be adjusted by either the policyholder or the insurance company. The coverage amounts u_i are used as the upper-limit amounts, and these are assumed not adjustable by the policyholder or the insurance company. In other data sets, coinsurance amounts also may be observed. Note that the number of losses N_i are realized prior to the loss amounts. For each loss N_i, the amounts y_{ij} are realized, and

$$y_{*,ij}(d_i) = y_{ij} - d_i | y_{ij} > d_i$$

are observed for each loss $j = 1, \ldots N_i$. Hence, the estimation assumes a claims data set and a policyholder data set. If the numbers of observations in these two data sets are considered independent, then standard asymptotic theory could be used for standard error estimates. In many cases, $y_{*,ij}$ may be observed, while y_{ij} is unobserved.

4.1 Maximum Likelihood Approach to Deductible Ratemaking

The maximum likelihood approach is a direct application of the theory, outlined in Section 3. We provide an overview of how the theory in Section 3 and similar results in Gray and Pitts (2012), Tse (2009) and Bahnemann (2015) can be empirically applied to real data. The rating procedure is summarized into four simple steps. The most difficult part is the estimation step, which requires statistical estimation methods for censored and truncated loss distributions. Subsequent steps are simple and straightforward.

Rating Procedure

1. Obtain $\mathrm{F}_Y(y)$ using statistical estimation. This involves censored and truncated estimation methods, as described in Section 8.4 (see Section 3 for related theorems).

2. Obtain $E[N]$, using Section 3 theory. Specifically, use $\ln(1-\mathrm{F}_Y(d))$ as offset and $E[N_g(d)]$ as response in a regression, as described in Section 8.4.
3. To calculate the rates for any d_{new}, obtain $E[Y_g(d_{\text{new}})] = \int_0^{d_{\text{new}}} (1-\mathrm{F}_Y(y))\,dy$, using the estimated loss model from Step 2 and numerical integration.
4. Calculate the new $E[S_g(d_{\text{new}})]$ using, $E[S_g(d_{\text{new}})] = E[N] \cdot E[Y_g(d_{\text{new}})]$ (see Corollary 3.1.2).

Details of the likelihood functions used for the statistical estimation in the first step, as well as the estimated coefficients from available real data, are shown in the Appendix, Section 8.4.

There are two complications to consider. First, for the frequency model parameter estimates, the standard errors become amplified by the modeling error and estimation error from the severity distribution, because of the coverage modification.

Second, if the number of observed claims is considered random, then the size of the claims data depends on the realization of the claim frequencies and severities for each policyholder, and hence the sampling frame becomes complicated. It is possible to show that for any confidence level, it is possible to find a large enough size for the policyholder sample so that the sample size of the claim data set is ensured to be large enough with any desired confidence. There is a large literature on large-sample theory for the validity of the fixed sample size for the claims data set, given a sequential stopping rule. Anscombe (1952) provides a proof for this result. Siegmund (1985) provides an overview of sequential analysis. In particular, Anscombe (1952) shows that for a sequence of proper random variables taking positive integer values N_r, the sequence of statistics based on N_r observations satisfies convergence and uniform continuity in probability. Assuming the sampling frame described above, application of the rating formulas using the maximum likelihood approach would use the above steps.

4.2 Regression Approach to Deductible Ratemaking

The regression approach is to use GLM models with a log deductible covariate. In practice, it is common to assume the observed after-deductible claims follow a gamma distribution or a Pareto distribution. Let Y_* be the observed claims. Then common practice is to parameterize the mean of the gamma distribution, using explanatory variables $\boldsymbol{x}_i$ and coefficients $\boldsymbol{\beta}$:

$$E[Y_*(d_i)] = \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right).$$

In this setup, deductibles may be incorporated into the model. If $\ln d_i = \texttt{lnDeduct}_i$ is included as an explanatory variable, then its coefficient, β_d, would satisfy

$$\frac{\partial E[Y_*(d_i)]}{\partial d} = \frac{\partial \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)}{\partial d} = \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)\beta_d \frac{\partial \ln d_i}{\partial d} = E[Y_*(d_i)]\frac{\beta_d}{d_i}.$$

Hence, for a single policy i, the coefficient β_d can be considered as the deductible *elasticity* of the observed mean:

$$\beta_d = \frac{\partial E[Y_*(d_i)]/E[Y_*(d_i)]}{\partial d/d_i}.$$

In econometrics, elasticity is a term used to denote the percentage change in a variable, in response to a percentage change in an explanatory variable. However, defining a single quantity β_d for a population of policies can be done in many different ways. When the sample of observed deductibles is not uniform, or when the deductible choice distribution is correlated with the response variable, the coefficient β_d may reflect this. For this paper, assessing how well a calculated β_d summarizes the relativity is best done using graphical approaches explained in Section 4.3. Analogously, if used in the regression for $N_g(d_i)$, then the corresponding coefficient γ_d would have the interpretation

$$\gamma_d = \frac{\partial E[N_g(d_i)]/E[N_g(d_i)]}{\partial d/d_i}.$$

In actuarial science, the *pure premium approach* is sometimes used to model the aggregate claims directly, using a compound distribution, such as the *Tweedie* distribution. In this case, a similar approach can be used by including $\ln d_i$ as an explanatory variable in the regression for the aggregate claims. For an overview of the pure premium approach, the reader may refer to Frees (2014) or Shi (2016). In this case, the coefficient ξ_d for $\ln d_i$ would have the interpretation

$$\xi_d = \frac{\partial E[S_g(d_i)]/E[S_g(d_i)]}{\partial d/d_i}.$$

One may compare the preceding derivations with using $\ln u_i = \texttt{LnCoverage}_i$ as an explanatory variable, in order to incorporate the coverage amounts into the regression model, as larger coverage amounts should typically result in higher claims. To ensure the interpretability of the coefficients for log coverage amounts, it is often recommended to use an alternative, *exposure (offset)* approach. According to Frees et al. (2015), in actuarial science *exposures (offsets)* are used to calibrate the size of potential outcome variables. In this case, the mean can be assumed to vary proportionally with an amount E. In this case, the coefficient for $\ln E$ is restricted to be 1 and included in the model as an *offset*. With this convention, we have

$$\mu = E \cdot \exp(\mathbf{x}'\boldsymbol{\beta}) = \exp\left(\mathbf{x}'\boldsymbol{\beta} + \ln E\right).$$

One can consider using $E = u - d$ as the offset amount for the regression, when the deductible amounts are to be incorporated as well. This way, when a policyholder selects a larger coverage amount (upper limit of policy), higher insurance premiums would be charged. Similarly, higher-deductible choices would naturally lead to a discount of the premium. The problem with applying this approach to deductibles is that the precise effects of deductible changes are not considered this way. In most cases, for a unit of deductible change, the scale change of the loss may be different from a unit. Note that in reality, the effect of a deductible change may vary, depending on the loss distributions of the policyholders.

4.3 Relativity Calculation

Further insight can be obtained by comparing the relativities for each approach. Given a base deductible d_0, for the regression approach, we have

$$\mathrm{REL}_{\mathrm{REG}}(d_0, d) = \frac{\exp\left(\boldsymbol{x}'\boldsymbol{\beta} + \beta_d \ln d\right) \cdot \exp\left(\boldsymbol{x}'\boldsymbol{\gamma} + \gamma_d \ln d\right)}{\exp\left(\boldsymbol{x}'\boldsymbol{\beta} + \beta_d \ln d_0\right) \cdot \exp\left(\boldsymbol{x}'\boldsymbol{\gamma} + \gamma_d \ln d\right)} = \left(\frac{d}{d_0}\right)^{\beta_d + \gamma_d}, \tag{3}$$

where we assume $\beta_d + \gamma_d < 0$. When the *pure premium approach* is used with a Tweedie distribution, the relativity given base deductible d_0 would be calculated in a similar way:

$$\mathrm{REL}_{\mathrm{REG}}(d_0, d) = \frac{E[S^c(d)]}{E[S^c(d_0)]} = \frac{\exp\left(\boldsymbol{x}'\boldsymbol{\xi} + \xi_d \ln d\right)}{\exp\left(\boldsymbol{x}'\boldsymbol{\xi} + \xi_d \ln d_0\right)} = \left(\frac{d}{d_0}\right)^{\xi_d}, \tag{4}$$

where we assume $\xi_d < 0$. For a general link function $\eta : (0, \infty) \to (-\infty, \infty)$, the relativity curve becomes

$$\mathrm{REL}_{\mathrm{REG}}(d_0, d) = \frac{\eta^{-1}\left(\boldsymbol{x}'\boldsymbol{\xi} + \xi_d \cdot \eta(d)\right)}{\eta^{-1}\left(\boldsymbol{x}'\boldsymbol{\xi} + \xi_d \cdot \eta(d_0)\right)},$$

where ξ_d is the coefficient for the covariate $\eta(d_i)$ in the regression. Different link functions may result in different shapes of relativity curves, yet in this paper we focus on analyzing the log link. The true relativity is

$$\mathrm{REL}_{\mathrm{CM}}(d_0, d) = \frac{\int_d^u (1 - \mathrm{F}_Y(y))dy}{\int_{d_0}^u (1 - \mathrm{F}_Y(y))dy}, \tag{5}$$

where $d > d_0$. For a Pareto model, if $\alpha > 1$, we have

$$\mathrm{REL}_{\mathrm{CM}}(d_0, d) = \frac{\int_d^u \left(\frac{\lambda}{\lambda + y}\right)^{\alpha} dy}{\int_{d_0}^u \left(\frac{\lambda}{\lambda + y}\right)^{\alpha} dy} = \frac{(\lambda + d)^{-\alpha+1} - (\lambda + u)^{-\alpha+1}}{(\lambda + d_0)^{-\alpha+1} - (\lambda + u)^{-\alpha+1}}, \tag{6}$$

and taking the upper bound u to infinity gives

$$\mathrm{REL}_{\mathrm{CM}}(d_0, d) \to \left(\frac{\lambda + d}{\lambda + d_0}\right)^{-\alpha+1} \quad \text{as } u \to \infty. \tag{7}$$

The reader may compare (4) and (6), and note the similarity, as both are decreasing functions in d, with the true relativity curve depending on the shape parameter of the distribution. In particular, observe that in equation (6), as $\lambda \to 0$ and $u \to \infty$, with $\alpha > 1$, the relativity curve for the Pareto model becomes identical to that of the regression approach, with $\xi_d = -\alpha + 1$ in equation (4). To compare the performance of the regression approach, one may plot the relativities for deductibles in the range $[d_0, u]$ and compare the fit. Note that in parametric models with covariates, the true relativity curve becomes subject specific through the distribution parameters, whereas the regression approach would allow for subject-specific variation through interaction terms with the deductible covariate. The performance of the regression approach depends on how well the relativity curve approximates the true relativity of a given policyholder.

5 Applications

5.1 Simulation

To assess the performance of the regression approach, claims were generated synthetically, using parameters similar to the Local Government Property Insurance Fund (LGPIF) building and contents claims (explained in more detail in Section 5.2); the lightning peril type is used for demonstration. The coefficient estimate results for the severity models are shown in Section 8.4. Claims were synthetically generated using parameters similar to those found from estimation. This way,

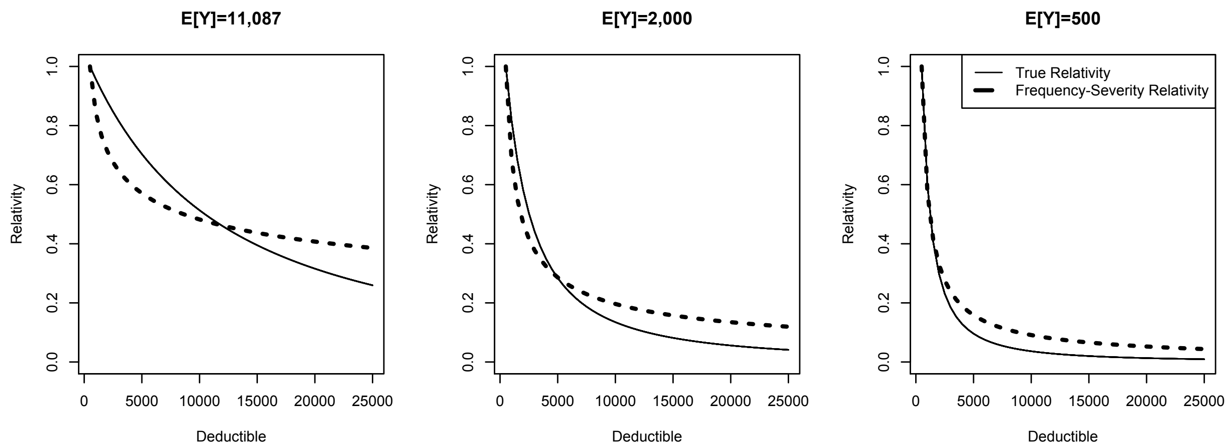

Figure 1: Relativities From Regression Using $\ln(d_i)$, for $E[Y] = 11087$, $E[Y] = 2000$, and $E[Y] = 500$

the deductible level can be adjusted to observe the potential effect on the relativity. $B = 10,000$ policies were generated using a Pareto distribution, with $E[Y] = 11,087$ and $\alpha = 2.553$. The claim frequency mean $E[N] = 1$ was used. Deductibles were synthetically generated by

- Generate d_i from a multinomial distribution over {500, 1000, 2500, 5000, 10000, 25000}, each with probabilities (0.461, 0.215, 0.118, 0.095, 0.011, 0.100), for $i = 1, \ldots, B$ policyholders.

These numbers were used so that the deductible distribution resembles the LGPIF data. Then the regression approach is used to estimate the elasticities. The results are $\beta_d + \gamma_d = -0.2434289$ for the deductible elasticity using the Poisson and gamma family. The relativities are then calculated for deductible levels ranging between the base 500 and 25,000, using the regression coefficients, and the true relativity curve is shown for comparison in Figure 1.

From Figure 1, the reader may observe that the regression approach using $\ln(d_i)$ approximates the true relativity curve in the best possible way, with deviations due to the nature of the log link. In the first panel, notice that for small relativity values, there is a small discrepancy between the dotted and solid lines. If the curve is dilated to the right, eventually the regression approach results in larger and larger deviations from the true relativity, as the curve deviates more from the solid line. For example, if a deductible of 1,000,000 is selected for the reinsurance retention level, the error in the relativity would be substantial when the regression approach is used. The regression approach using $\ln d$ as the explanatory variable results in a curve that is steeper than the true relativity curve for small deductibles, and that flattens out eventually, due to the nature of the link function. In subsequent panels, the scale parameter λ is increased, showing how the regression approximation becomes closer to the true relativity curve when λ approaches zero. In general, we find that the regression approach would be suitable for moderate-size problems, where the scale parameter λ is moderate in size.

In Figure 2, instead of $\ln(d_i)$, we attempt to use $\ln(d_i+\lambda)$ for the explanatory variable. This way, the regression approach becomes identical to the true relativity curve, and in all three panels, the curve fits almost exactly. This suggests several valuable insights. First, we learn that the regression is an approximation to the relativity curve using one parameter only, or in other words, without a scale parameter. This scale parameter is not known in advance without performing maximum

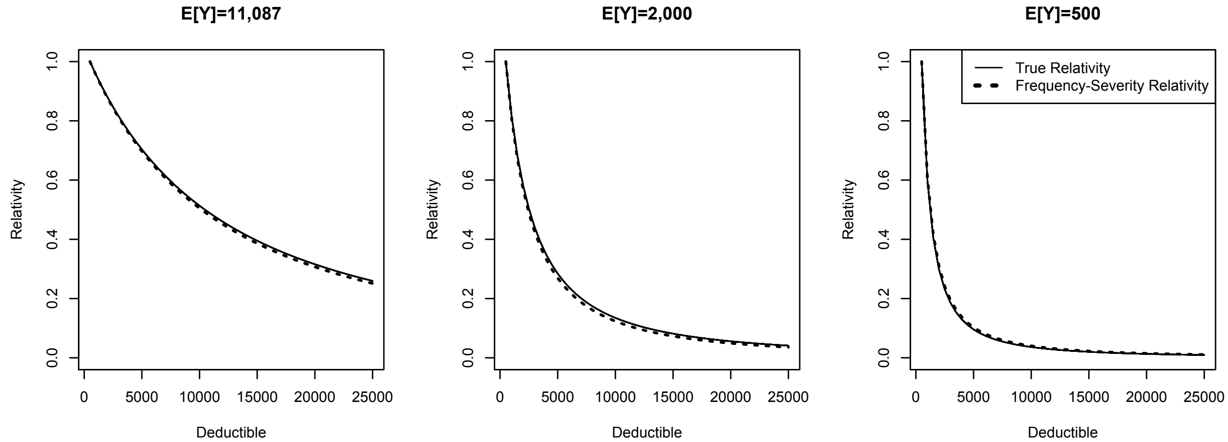

Figure 2: Relativities From Regression Using $\ln(d_i + \lambda)$, for $E[Y] = 11087$, $E[Y] = 2000$, and $E[Y] = 500$

likelihood. Also, we learn that for specific cases, regression may perform well—specifically, when the problem is of moderate size.

5.2 Relativities

Next, we compare the different approaches using real data. Figure 3 shows a histogram of the lightning losses and the fitted distributions using truncated estimation for the exponential, gamma and Pareto models. The reader may observe that the exponential and gamma model fits look about the same, whereas the Pareto model fit is slightly better. Models are fit for various peril types, as shown in the Appendix, Section 8.3.

The regression approach is implemented using a standard `glm` software package with Poisson and gamma families. Log deductibles are included as covariates in each regression model, by peril type. To compare the single relativity obtained from the regression approach, we calculate single relativities for the gamma and Pareto maximum likelihood approaches by

$$\text{REL}_m(d_0, d) = \frac{E\left[\sum_i S_{g,i,m}(d)\right]}{E\left[\sum_i S_{g,i,m}(d_0)\right]}$$

where $S_{g,i,m}(d)$ indicates the aggregate claims for peril type m, for policyholder i, under deductible d. This single quantity allows for a comparison of the relativity obtained from the maximum likelihood approach and the regression approach.

The regression approach coefficient estimates for the $\ln(d)$ explanatory variable are shown in Table 2. These quantities are used to calculate the relativities for the regression approach, by peril type. Relativities for the maximum likelihood approach are calculated using the estimated models in Table 12, in the Appendix.

In Figure 4, the peril types are categorized into nine categories, as explained in more detail in the Appendix. In the figure, relativities for the regression approach and those obtained from

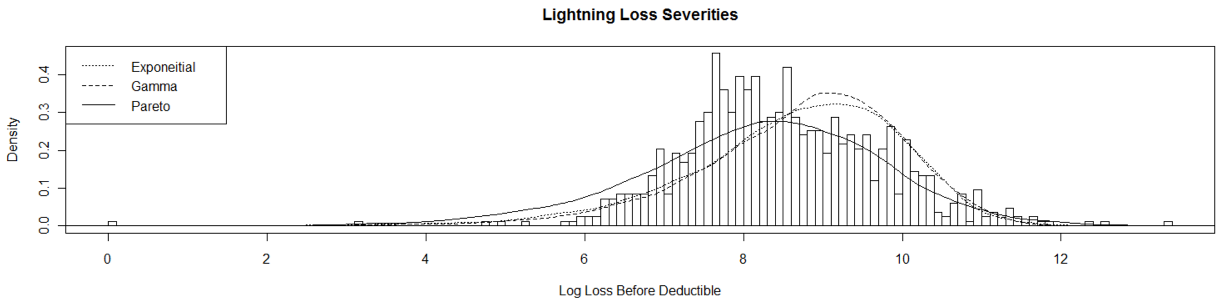

Figure 3: Density of Lightning Losses

Table 2: Coefficients from Regression Approach

	Poisson (γ_d)	Gamma (β_d)	Sum ($\gamma_d + \beta_d$)	Adjustment	$1-\alpha$**
Fire	-0.407	0.228	-0.179		-0.012
Vandalism	-1.335	0.981	-0.354		-0.357
Lightning	-0.822	0.489	-0.332		-0.874
Wind	-0.567	0.259	-0.308		-0.242
Hail*	-0.202	0.594	0.391	-0.202	-0.543
Vehicle	-1.125	0.429	-0.697		-2.924
Water (Non-weather)*	-0.579	0.714	0.135	-0.579	-0.127
Water (Weather)*	-0.375	0.949	0.574	-0.375	-0.000
Misc.*	-0.734	0.716	-0.019	-0.734	-0.063

*Adjustment made for parameter interpretability.
**Shape parameter of Pareto model is shown for reference.

maximum likelihood are overlayed, allowing for comparisons. The numeric values of the relativities are shown in Table 3.

In each panel of Table 3, the relativities are shown in the column for each deductible level. The reader may compare the first panel, showing the regression approach, with the second and third panels, showing the gamma model and Pareto model, using truncated estimation and the rating formulas in Section 3. The leftmost column is the relativity for the 1,000 deductible level. The relativities for the next deductible level, 2,500, are lower as expected, and so on. The lowest relativity indicates the ratio of the aggregate claims under 50,000 to the aggregate claims under the base deductible, 500. Hence, the single quantities in the tables allow for a comparison of the relativity levels for the single elasticity obtained from the regression approach and the maximum likelihood approach.

The reader may observe that regression provides lower relativities in general. In particular, the relativities in the regression approach are somewhat more uniform over different deductible levels, except for very small deductible levels, when compared with the second panel, which uses the same distributional assumption with a maximum likelihood approach. This is due to the nature of the log link. As we will see in subsequent sections, the performance of the aggregate claim prediction is unaffected by this phenomenon. The Pareto distribution has a heavier tail and in general results in higher relativities. The calculations are shown with holdout sample claims for year 2011, while the models were fit using data for years 2006–2010.

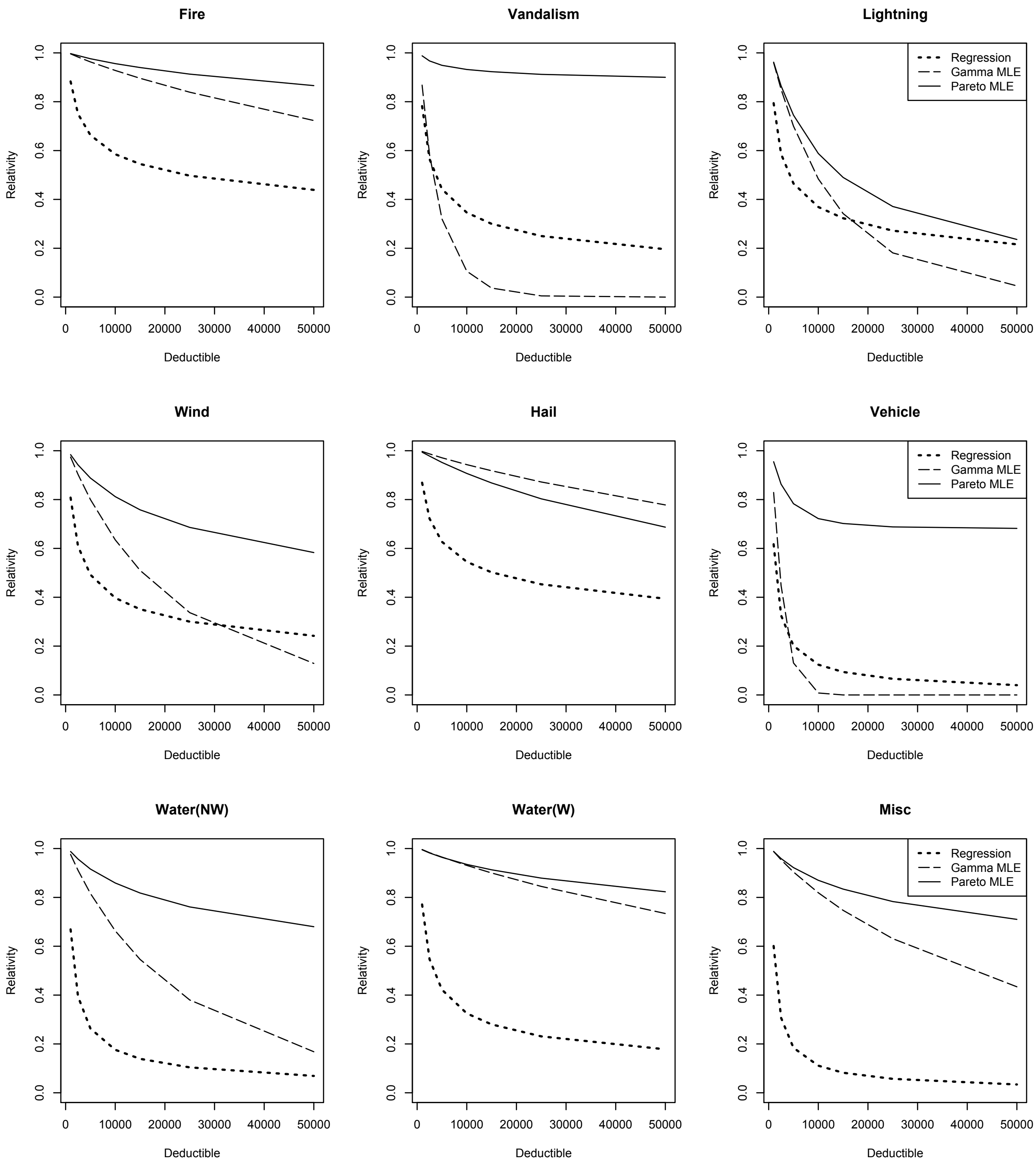

Figure 4: Plot of Relativities for Regression Approach and Selected MLE Approaches

We have been able to make a few observations while implementing each approach. First, when the maximum likelihood approach is used with the gamma distribution assumption, some of the policyholders resulted in relativity values of zero. This shows that although the maximum likelihood approach has the flexibility of providing policyholder-specific relativities by varying the parametrization, sophisticated distributional assumptions, such as a long-tail, Pareto distribution, would be needed for the relativities to be interpretable. In contrast, the regression approach

Table 3: Comparison of Relativities for Regression Approach and Selected MLE Approaches

Regression Approach

Deductible	1,000	2,500	5,000	10,000	15,000	25,000	50,000
Fire	0.883	0.750	0.663	0.585	0.545	0.497	0.439
Vandalism	0.782	0.565	0.442	0.346	0.300	0.250	0.196
Lightning	0.794	0.586	0.465	0.369	0.323	0.272	0.216
Wind	0.808	0.609	0.492	0.397	0.351	0.300	0.242
Hail*	0.869	0.722	0.627	0.545	0.502	0.453	0.394
Vehicle	0.617	0.326	0.201	0.124	0.094	0.066	0.040
Water (Non-weather)*	0.669	0.394	0.263	0.176	0.139	0.104	0.069
Water (Weather)*	0.771	0.547	0.422	0.325	0.280	0.231	0.178
Misc.*	0.601	0.307	0.184	0.111	0.082	0.057	0.034

*lnDeduct has been included only in the frequency regression, for interpretability.

Poisson-gamma MLE

Deductible	1,000	2,500	5,000	10,000	15,000	25,000	50,000
Fire	0.996	0.983	0.963	0.928	0.896	0.839	0.723
Vandalism	0.868	0.583	0.320	0.106	0.037	0.005	0.000
Lightning	0.960	0.851	0.700	0.484	0.342	0.181	0.046
Wind	0.974	0.903	0.800	0.635	0.510	0.337	0.129
Hail	0.997	0.987	0.971	0.943	0.918	0.872	0.778
Vehicle	0.829	0.448	0.131	0.008	0.000	0.000	0.000
Water (Non-weather)	0.976	0.911	0.816	0.663	0.546	0.380	0.168
Water (Weather)	0.996	0.983	0.965	0.931	0.900	0.845	0.734
Misc.	0.988	0.954	0.904	0.819	0.747	0.631	0.434

Poisson-Pareto MLE

Deductible	1,000	2,500	5,000	10,000	15,000	25,000	50,000
Fire	0.997	0.989	0.976	0.956	0.940	0.913	0.866
Vandalism	0.988	0.967	0.949	0.932	0.923	0.912	0.900
Lightning	0.962	0.865	0.745	0.588	0.490	0.371	0.236
Wind	0.984	0.942	0.888	0.812	0.758	0.686	0.583
Hail	0.994	0.978	0.952	0.907	0.868	0.803	0.687
Vehicle	0.955	0.863	0.783	0.722	0.702	0.688	0.682
Water (Non-weather)	0.988	0.957	0.916	0.859	0.818	0.761	0.680
Water (Weather)	0.995	0.982	0.964	0.935	0.913	0.879	0.823
Misc.	0.988	0.959	0.922	0.870	0.834	0.783	0.710

provides a single relativity value, which allows for easier interpretation. When compared with the empirical relativities, shown in the last panel, the regression approach, although not perfect, seems to provide reasonable single-value relativities for an analyst to use.

In particular, observe that the maximum likelihood experienced difficulty in assessing the vandalism relativities, because claims in this category are influenced substantially by the deductible, as claim sizes are small. Regression seems to provide more reasonable relativities for this category. The reader may observe this in Table 2.

Second, the severity model for the regression approach sometimes provided coefficients that could not be interpreted. This situation is illustrated in Table 2. Specifically, hail, water and miscellaneous peril types resulted in β_d values too high. In these cases, the regression model must be fixed, with β_d omitted from the severity model. In Section 5.3, a similar situation was observed for the aggregate claims model. The reader may see the coefficients in Table 15.

5.3 Aggregate Claims Prediction

How well does each method, including the regression approach, perform in aggregate claims prediction? We compare the regression approach with the maximum likelihood rating approach for total aggregate claims prediction for an entire line. From this study, we demonstrate that, although the regression approach provides smaller-than-reasonable relativities for some policies, it performs quite well in terms of total aggregate claims prediction. We expect the regression approach to provide a practical solution for applications with a large number of explanatory variables, where the aggregate claims prediction is of primary interest. For a demonstration, we compare the results of three different approaches:

(A) Regression approach, using Poisson-gamma, with the `lnDeduct` covariate
(B) Maximum likelihood approach, using Poisson-gamma truncated estimation
(C) Maximum likelihood approach, using Poisson-*GB2* truncated estimation

We did not implement a regression approach for the Poisson-*GB2* model, because comparing the three cases listed would suffice in demonstrating the most commonly encountered assumptions in practice. Advantages and disadvantages of regression and maximum likelihood can be illustrated by comparing these three common modeling assumptions. The estimation results are shown in the Appendix. Specifically, Tables 15, 16, 17 and 18 each show the results from the regression approach and the truncated estimation results for gamma, Poisson, and *GB2* models. Using these coefficient estimates, the total aggregate claims are predicted for different deductible levels. For the regression approach, the log deductible amount is multiplied by the coefficient estimate –0.737 shown in Table 15. The coefficient estimates from truncated estimation are used to calculate the aggregate claims, and the aggregate claims are compared with predictions from the regression approach.

Table 4: Out-of-Sample (2011) Performance of Each Approach

	Aggregate Claims	Pearson Correlation with Claims	Spearman Correlation with Claims
(A) Poisson-gamma regression	16,170,966	0.2231	0.3922
(B) Poisson-gamma MLE	11,464,929	0.3358	0.3847
(C) Poisson-*GB2* MLE	20,976,735	0.4157	0.4025
Claims	19,036,189	1.0000	1.0000

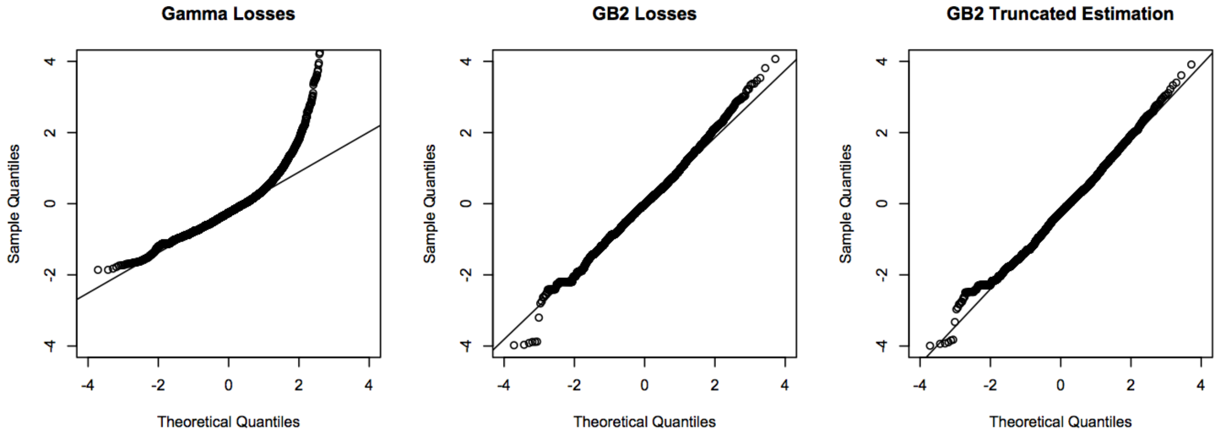

Figure 5: Comparison of Severity Distributions. Panel 1: gamma losses fit; Panel 2: *GB2* losses fit; Panel 3: *GB2* truncated estimation (Kolmogorov-Smirnov test statistics (p-values) 0.199 (0.000), 0.025 (0.003), 0.103 (0.000))

According to Table 4, the regression approach was effective in identifying the ranking of the policyholders (as could be seen from the Spearman correlation), as well as predicting the amount of aggregate claims, compared with the maximum likelihood approach using the same distribution assumption. The aggregate claim amount for the regression approach, 16,170,966, is in fact closer to the empirical claims than 11,464,929. This shows, for many practical rating problems with small data sets, that the regression approach may provide a good assessment for the actual aggregate claims. That is perhaps because the coefficients β_d and γ_d incorporate the deductible selection effect into the rating.

Theoretically, it is possible to incorporate the deductible selection effect into the Poisson-*GB2* model. For example, one may consider fitting a separate severity model for various deductible levels. A similar approach could be used for the mixture approach as well. If a selection effect exists, then the severity model coefficient estimates would provide different parameters for each deductible level. Another potential way to incorporate the selection effect is to use dependence modeling, considering the dependence between deductible levels and the loss severity. This approach is left as future work.

According to Table 4, long-tail loss models, such as the *GB2* model with truncated statistical estimation, and coverage modification theory can improve the deductible rating, as the third row of the table shows. Figure 5 provides some insight into why the prediction improves. The left panel shows the Q-Q plot when the *GB2* model is fit to the underlying loss distribution, assuming it is observed. The LGPIF data records both the underlying loss and the deductible amounts, which allows this figure to be shown as a comparison. The first panel shows that the fit of the gamma distribution suffers for the lower and upper tails of the claims distribution. The middle panel shows that the *GB2* fit is better, and the third panel shows that truncated estimation has recovered the underlying distribution quite well. In general, a better fit of the underlying loss model would result in a better assessment of the coverage modification of the deductibles, for either small values or large values of deductibles.

During our analysis, we performed comparisons of claim scores obtained from different models, for various hypothetical deductible levels. For example, the aggregate claims can be predicted for increasing deductible levels, and applied to all policyholders throughout the property fund. The predicted aggregate claims can then be compared with the hypothetical empirical observed claims, which can be obtained by applying the hypothetical deductible level to the underlying losses. In general, our analyses have shown that, for large deductibles, the *GB2*-01NB model performs best. Defining how large a deductible level is required, for the maximum likelihood approach to become necessary, would be an interesting research question for future studies. The regression approach may be a good method for predicting aggregate claims, for moderate-size data and small deductibles. We find that when accurate deductible relativities are of interest for large losses, more elaborate methods, such as the maximum likelihood approach, are needed with truncated estimation. In particular, if subject-specific deductible relativities are needed or when an excess of loss layers is to be priced, then the maximum likelihood approach would be necessary. In contrast, when a single relativity value is desirable, the regression approach may turn out to be useful.

5.4 Comparison of Frequency Models

The assumption of a *GB2* severity distribution influences the truncation of the underlying frequency distribution. This section compares different frequency model assumptions. We compare the Poisson model and 01-inflated Poisson model by fitting the two distributions to the underlying loss frequencies and then attempting to estimate the same parameters using the censored frequency observations. Details of the estimation issue for 01-inflated count models under deductible influence are covered in the Appendix, Section 8.5. For an introduction to 01-inflated count distributions, the reader may refer to Frees et al. (2015). Briefly speaking, the 01-inflated distribution has a latent variable I, following a multinomial, so that N satisfies

$$N \sim \begin{cases} 0 & I = 0 \\ 1 & I = 1 \\ N_\lambda & I = 2, \end{cases}$$

where N_λ is the secondary Poisson distribution. With this, the probability mass function of N is

$$\mathrm{f}_N(n) = \pi_0 I_{\{n=0\}} + \pi_1 I_{\{n=1\}} + \pi_2 \mathrm{P}_\lambda(n), \tag{8}$$

where $\mathrm{P}_\lambda(n) = \Pr(N_\lambda = n)$. Suppose censored frequencies $N_{\lambda,g}(d)$ are observed. Let π_0, π_1 and $\pi_2 = 1 - \pi_0 - \pi_1$ be the multinomial primary distribution probabilities, and let $\lambda = \exp(\boldsymbol{x}'\boldsymbol{\beta})$ be the mean parametrization of the secondary Poisson distribution. Table 5 shows that the 01-inflated Poisson model fit is better than that of the Poisson model, when the underlying losses are observed.

Table 5: Comparison of Predicted Counts Using Validation Sample

Count	(1) Poisson Underlying	(2) 01-Poisson Underlying	(3) Poisson Censored Estimation	(4) 01-Poisson Censored Estimation	Empirical Counts (2011)
0	703	768	614	603	766
1	194	195	231	236	187
2	84	54	108	109	57
3	42	29	56	57	28
4	23	16	31	32	17
5	14	10	18	19	14
6	8	6	11	12	4
7	6	4	7	7	8
8	4	3	5	5	2
9	3	2	3	4	2
10	2	2	2	2	2
11	2	1	2	2	1
12	1	1	1	1	0
13	1	1	1	1	1
14	1	0	1	1	0
15	1	0	1	1	1
16	1	0	0	0	1
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0

Note: Each column attempts to predict the underlying loss frequencies. Loss count categories 0–19 are shown for illustration. Notice that (3) and (4) overpredict the 0-losses and 1-loss categories and underpredict the 2-losses category. This is as expected, and the reader may understand the reason from the fit of the severity distribution below the deductible. Compare the Q-Q plots in Figure 5.

This situation is shown in the first panel, where the 01-Poisson with underlying frequencies in general fit the empirical counts better, in each count category. For example, the number of zero observations is predicted to be 768, which is close to the empirical 766. The situation is different when a deductible causes censoring. In terms of the predicted counts, the 01-Poisson does not perform better than the Poisson. This is because prediction of the observations below the deductible become difficult, as a censoring is in place. This motivates using a basic model for the counts when censoring is in place. Hence, the Poisson model is used in Section 5.3, which also allows using the $\ln(1 - v_i)$ offset technique for estimation. Table 19 shows the coefficient estimates.

6 Conclusion

Summary

Loss models are built on positive responses. For this reason, log links or other such link functions $\eta : (0, \infty) \to (-\infty, \infty)$ are required to implement a regression approach. As a result, when the deductible amount is included as a covariate, the resulting relativities become an approximation of the theoretically true relativities, and hence are not perfect. Work-around approaches are possible by using d as the explanatory variable or using interaction terms. The regression approach to the ratemaking of per-loss deductibles is hence simple and widely used, as it proves to be valuable when aggregate claims prediction is the interest.

This paper provides an overview of the rating of deductibles for per-loss insurance deductibles. Our contribution is the empirical application of textbook methods for coverage modification to real data, and the generalization of coverage modification theory to loss variables without a continuous distribution. We also provide a comparison of the regression approach to deductible rating, with the maximum likelihood approach. Empirical data are used to calculate aggregate claim amounts under the influence of a deductible. If deductible choices are not uniform or are correlated with the loss distribution, then the rating may become more complicated.

To summarize our work, we have proposed a comprehensive overview of deductible ratemaking using real data. For cases where small deductibles are applied and the aggregate claim amounts are the primary interest, the deductible amounts may be used as covariates in the scale parametrization of a regression model. The reader may compare equations (3) and (6) to understand how the regression approach provides a reasonable approximation of the true relativity curve, and the meaning of the shape parameter of a distribution in relation to the curvature of the relativity curve. This approach is not suitable when deductibles are large or when the precise relativities are of interest for large losses, as shown in Figure 1. When large deductibles are to be priced, the maximum likelihood approach is recommended. For example, excess of loss layers may be priced better by fitting an underlying loss model. Yet, as Table 4 shows, when different deductible-rating approaches were used to predict the aggregate claims, the regression approach performed reasonably well, outperforming the maximum likelihood approach using an identical distributional assumption, in terms of predicted aggregate claim amounts. Our explanation is that regression utilized the deductible selection effect within the data. Specifically, the deductible elasticity of the claim frequency may likely have included effects due to deductible selection.

Future Work

For future studies, it may be interesting to apply more elaborated estimation procedures for the case when the observed data contain only the average severities. Further studies on estimation issues for compound distributions also may be of interest. The effect of an aggregate deductible may also be a natural extension, where an additional layer of deductible is applied to the losses. In addition, the claims may be classified using more elaborated techniques, to allow for efficient peril-specific loss models for deductible ratemaking, using regression and maximum likelihood. The influence of classification on deductible rates may be potentially interesting to study further.

Selection Effect

Deductible selection effects also may be of interest to study further. For example, separate models may be imposed for each deductible choice and tested for significance in the difference of coefficients. The null hypothesis of no deductible selection effect would indicate deductible selection effects are not present in the data, whereas evidence of different coefficients may indicate a selection effect. One may consider null and alternative hypotheses based on the Akaike Information Criterion (AIC). If the fit improvement from more than one pair of parameters is significant enough, then a deductible selection effect would be proven. Because the deductible selection effect relates to a vast economic literature in endogenous selection biases, we consider this to be potentially interesting future work.

7 Acknowledgments

This study has been generously funded by the Society of Actuaries (SOA, http://www.soa.org) through the General Insurance Research Studies grant. We thank Scott Lennox, Erika Schulty, Dale Hall, Steven Siegel, Barbara Scott, and the project oversight group members Robert Winkelman, Michael Ewald, Jean-Marc Fix, Josh Newkirk, Rafi Herzfeld, Andy Ferris, Crispina Caballero, Joe Wurzburger, Julianne Callaway and Tina Liu for their valuable insights regarding this work. We especially thank Robert Winkelman for suggestions on future work on deductible selection effects. We would also like to thank the Wisconsin Office of the Commissioner of Insurance.

References

Aban, Inmaculada B., Mark M. Meerschaert, and Anna K. Panorska (2006). "Parameter Estimation for the Truncated Pareto Distribution," *Journal of the American Statistical Association*, 101(473): 270–277.

Anscombe, F. J. (1952). "Large-sample Theory of Sequential Estimation," *Mathematical Proceedings of the Cambridge Philosophical Society*, 48: 600–607.

Bahnemann, David (2015). *Distributions for Actuaries.* CAS Monograph Series, Number 2.

Barr, Donald R., and Todd Sherrill (1999). "Mean and Variance of Truncated Normal Distributions," *American Statistician*, 53(4): 357–361.

Bernegger, Stefan (1997). "The Swiss Re Exposure Curves and the MBBEFD Distribution Class," *ASTIN Bulletin*, 27: 99–111.

Brown, Robert L., and W. Scott Lennox (2015). *Ratemaking and Loss Reserving for Property and Casualty Insurance.* ACTEX Publications.

Chapman, Douglas G. (1956). "Estimating the Parameters of a Truncated Gamma Distribution," *Annals of Mathematical Statistics*, 27(2): 498–506.

Chavez-Demoulin, Valerie, Paul Embrechts, and Marius Hofert (2016). "An Extreme Value Approach for Modeling Operational Risk Losses Depending on Covariates," *The Journal of Risk and Insurance*, 83(3): 735–776.

Cohen, Alma, and Liran Einav (2007). "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, 97(3): 745–788.

Cummings, David (2001). "Practical GLM Analysis of Homeowners," *Presentation, State Farm Insurance Companies.*

Einav, Liran, Amy Finkelstein, Iuliana Pascu, and Mark R. Cullen (2012). "How General Are Risk Preferences? Choices Under Uncertainty in Different Domains," *American Economic Review*, 102(6): 2606–2638.

Fasen, Vicky, and Claudia Kluppelberg (2014). "Large Insurance Losses Distributions," *Wiley StatsRef: Statistics Reference Online.*

Finkelstein, D. M., and R. A. Wolfe (1985). "A Semiparametric Model for Regression Analysis of Interval-Censored Failure Time Data," *Biometrics*, 41(4): 933–945.

Frees, Edward W. (2014). "Frequency and Severity Models," in Edward W. Frees, Glenn Meyers, and Richard A. Derrig, eds. *Predictive Modeling Applications in Actuarial Science*, Cambridge: Cambridge University Press.

Frees, Edward W., and Gee Y. Lee (2016). "Rating Endorsements using Generalized Linear Models," *Variance*, 10(1).

Frees, Edward W., Gee Lee, and Lu Yang (2015). "Multivariate Frequency-Severity Regression Models in Insurance," *Risks*, 4(1).

Frees, Edward W., and Emiliano A. Valdez (2008). "Hierarchical Insurance Claims Modeling," *Journal of the American Statistical Association*, 103(484): 1457–1469.

Fu, Luyang, and Xianfang Liu (forthcoming). "Finite Mixture Model and Workers Compensation Large Loss Regression Analysis," in Edward W. Frees, Glenn Meyers, and Richard A. Derrig, eds. *Predictive Modeling Applications in Actuarial Science, Vol. II*, Cambridge: Cambridge University Press.

Gray, Roger J., and Susan M. Pitts (2012). *Risk Modelling in General Insurance: From Principles to Practice*, Cambridge: Cambridge University Press.

Guiahi, Farrokh (2005). "Fitting Loss Distributions with Emphasis on Rating Variables," *CAS Specialty Seminar.*

Gurnecki, Krzysztof, Grzegorz Kukla, and Rafal Weron (2006). "Modelling Catastrophe Claims With Left-Truncated Severity Distributions," *Computational Statistics*, 21: 537–555.

Halek, Martin, and Joseph Eisenhauer (2001). "Demography of Risk Aversion," *Journal of Risk and Insurance*, 68(1): 1–24.

Holt, Charles A., and Susan K. Laury (2002). "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), p. p.1644.

Kalbfleisch, John D., and Ross L. Prentice (2002). *The Statistical Analysis of Failure Time Data, Second Edition*, New Jersey: John Wiley & Sons, Inc.

Kaplan, E. L., and P. Meier (1958). "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53: 457–481.

Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot (2012). *Loss Models: From Data to Decisions, Fourth Edition*, New Jersey: John Wiley & Sons.

Koszegi, Botond, and Matthew Rabin (2006). "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121(4): 1133–1165.

Lai, Tze Leung, and Zhiliang Ying (1991). "Estimating a Distribution Function With Truncated and Censored Data," *Annals of Statistics*, 19(1): 417–442.

Ludwig, Steven J. (1991). "An Exposure Rating Approach to Pricing Property Excess-of-Loss Reinsurance," *Casualty Actuarial Society*, 78.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry Green (1995). *Microeconomic Theory, First Edition*, Oxford: Oxford University Press.

Plackett, R. L. (1953). "The Truncated Poisson Distribution," *International Biometric Society*, 9(4): 485–488.

Rosenberg, Marjorie A., Richard W. Andrews, and Peter Lenk (1999). "A Hierarchical Bayesian Model for Predicting the Rate of Nonacceptable In-Patient Hospital Utilization," *Journal of Business and Economic Statistics*, 17(1): 1–8.

Rothschild, Michael, and Joseph Stiglitz (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90(4): 629–649.

Salvadori, Gianfausto, and Carlo De Michele (2007). *Extremes in Nature: An Approach Using Copulas, First Edition*, Dordrecht: Springer.

Shi, Peng (2016). "Insurance Ratemaking Using a Copula-Based Multivariate Tweedie Model," *Scandinavian Actuarial Journal*, 2016(3): 198–215.

Siegmund, David (1985). "The Sequential Probability Ratio Test," in D. Brillinger, S. Fienberg, J. Gani, J. Hartigan, and K. Krickeberg, eds. *Sequential Analysis: Tests and Confidence Intervals*, New York: Springer Science and Business Media.

Sun, Jiafeng, Edward W. Frees, and Marjorie A. Rosenberg (2008). "Heavy-Tailed Longitudinal Data Modeling Using Copulas," *Insurance: Mathematics and Economics*, 42(2): 817–830.

Sydnor, Justin (2010). "(Over)insuring Modest Risks," *American Economic Journal: Applied Economics*, 2(4): 177–199.

Tse, Yiu-Kuen (2009). *Nonlife Actuarial Models: Theory, Methods and Evaluation, First Edition*, Cambridge: Cambridge University Press.

Verbelen, R., and G. Claeskens (2014). "Multivariate Mixtures of Erlangs for Density Estimation Under Censoring and Truncation," *Social Science Research Network*.

White, Steve (2005). "Property Ratemaking — an Advanced Approach: Exposure Rating," *CAS Seminar on Reinsurance*.

White, Steve, and Kari Mrazek (2004). "Advanced Exposure Rating: Beyond the Basics," *CAS Seminar on Reinsurance*.

Woodroofe, M. (1985). "Estimating a Distribution Function With Truncated Data," *Annals of Statistics*, 13(1): 163–177.

Wu, S.Q., and X. Cai (1999). "Maxwell-Boltzmann, Bose-Einstein, Fermi-Dirac Statistical Entropies in a D-Dimensional Stationary Axisymmetry Space-Time," *Cornell University Library*.

Yuan, Ming, and Yi Lin (2006). "Model Selection and Estimation in Regression With Grouped Variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68: 49–67.

Zhang, Zhengjun, Chunming Zhang, and Qiurong Cui (2008). "Quotient Correlation: A Sample-Based Alternative to Pearsons Correlation," *Annals of Statistics*, 36: 1007–1030.

8 Appendix

The appendix of the paper covers some more details of the approaches. First, it presents proofs of the theory in Section 3. Then it provides an overview of parametric loss models, with an emphasis on the generalized beta family distributions. The relativities for each peril and the coefficient estimation results are shown for selected distributions.

8.1 Proof of Theorems

This section provides proofs for the results in Section 3, for the coverage modifications in frequency and severity first moments.

Proof of Theorem 3.1 *Let N be any count random variable, and Y any random variable, each with finite first moments. If N and Y are independent, then for deductibles $d_1 < d_2$, we have*

$$E[S_g(d_1) - S_g(d_2)] = E[N]E\left[Y_g(d_1) - Y_g(d_2)\right].$$

Proof. We have

$$E[S_g(d)] = E\left[\sum_{i=1}^{N} Y_{g,i}(d)\right] = E\left[E\left[\sum_{i=1}^{N} Y_{g,i}(d)\middle|N\right]\right] = E\left[NE\left[Y_g(d)\right]\right] = E\left[N\right]E\left[Y_g(d)\right],$$

and the theorem follows directly. □

Proof of Corollary 3.1.1 If Y has distribution function F_Y, then

$$E[S_g(d_1) - S_g(d_2)] = E[N]\int_{d_1}^{d_2} (1 - \mathrm{F}_Y(y))\, dy.$$

Proof. It suffices to provide an expression for $Y_g(d)$. We have

$$E[Y_g(d)] = \int_d^{\infty} (y-d)d\mathrm{F}_Y(y) = (y-d)(1-F_Y(y))\Big|_d^{\infty} + \int_d^{\infty} (1 - \mathrm{F}_Y(y))\, dy = \int_d^{\infty} (1 - \mathrm{F}_Y(y))\, dy.$$

In particular, we have

$$E[Y_g(d_1) - Y_g(d_2)] = \int_{d_1}^{d_2} (1 - \mathrm{F}_Y(y))\, dy.$$

Note, $E[Y_g(d)]$ is the partial expectation of Y, given $Y > d$. As a special case, the following provides an expression for the modified aggregate claims under any new per-loss deductible level:

$$E[S_g(d)] = E[S] - E[N] \cdot \int_0^{d} (1 - \mathrm{F}_Y(y))\, dy.$$

□

For an expression for the influence on severities, the following corollary states that a unit deductible change results in less than a unit decrease in the severity mean.

Proof of Corollary 3.2.1 If $E[Y_g(d)]$ is differentiable, then

$$\frac{\partial}{\partial d}E[Y_g(d)] = -1 + \mathrm{F}_Y(d). \tag{9}$$

Proof. We have

$$\frac{\partial}{\partial d}E[Y_g(d)] = \lim_{d_2 \to d} \frac{E\left[Y_g(d_2) - Y_g(d)\right]}{d_2 - d} = \lim_{d_2 \to d} \frac{-1}{d_2 - d} \int_d^{d_2} (1 - \mathrm{F}_Y(y))\, dy = -1 + \mathrm{F}_Y(d).$$

An alternative proof assumes a density for Y. Let f_Y be the density, and we have

$$E[Y_g(d)] = E\left[(Y - d) \cdot \mathrm{I}(Y > d)\right] = \int_d^{\infty} (y - d)\mathrm{f}_Y(y) dy = \int_d^{\infty} y\mathrm{f}_Y(y) dy - d(1 - \mathrm{F}_Y(d)).$$

Differentiation gives

$$\frac{\partial}{\partial d}\left[\int_d^{\infty} y\mathrm{f}_Y(y) dy - d \cdot (1 - \mathrm{F}_Y(d))\right] = -d \cdot \mathrm{f}_Y(d) - (1 - \mathrm{F}_Y(d)) + d \cdot \mathrm{f}_Y(d) = -1 + \mathrm{F}_Y(d).$$

□

The next theorem provides an expression for the modified frequencies, under the influence of a deductible d. This result is needed when recovering the underlying loss frequencies from the observed claim frequencies.

Proof of Theorem 3.2 Let N be any count random variable, and Y have distribution function F_Y, each with finite first moments. If Y is independent of N, then $N_g(d)$ satisfies

$$E\left[N_g(d)\right] = E\left[N\right] \cdot (1 - \mathrm{F}_Y(d)). \tag{10}$$

Proof. We have

$$E\left[N_g(d)\right] = E\left[\sum_{i=1}^{N} \mathrm{I}(y > d)\right] = E\left[E\left[\sum_{i=1}^{N} \mathrm{I}(y > d) \middle| N\right]\right] = E\left[N \cdot P(y > d)\right] = E\left[N\right] \cdot (1 - \mathrm{F}_Y(d)).$$

If F_Y is differentiable, then the rate of change of the frequencies can be obtained by

$$\frac{\partial E[N_g(d)]}{\partial d} = E[N] \cdot \frac{\partial\left[1 - \mathrm{F}_Y(d)\right]}{\partial d} = -E[N] \cdot \mathrm{f}_Y(d).$$

□

This provides a framework for understanding the increase in zero probability. For example, in a count regression model, one may be interested in recovering the underlying parameters of the loss frequencies, given an observed count random variable $N_g(d)$. The theorem states that the frequency change is proportional to $v = (1 - \mathrm{F}_Y(d))$. The rate of change of the aggregate claims in response to a unit deductible change can be obtained in a similar way. In general, the overall effect of a deductible change on the expected aggregate claim, using any count random variable and any severity distribution, can be obtained by differentiation. For the truncated claims observations, we provide the following proof.

Proof of Theorem 3.3 Let Y be any random variable with density f_Y and distribution F_Y. Then $Y_*(d)$ satisfies

$$\frac{\partial E[Y_*(d)]}{\partial d} = \frac{\mathrm{f}_Y(d)}{1 - \mathrm{F}_Y(d)} E[Y_*(d)] - 1. \tag{11}$$

Proof. Using the Libnitz rule, we have

$$\begin{aligned}
\frac{\partial E[Y_*(d)]}{\partial d} &= \frac{\partial E[Y - d|Y > d]}{\partial d} \\
&= \frac{\partial E[Y|Y > d]}{\partial d} - \frac{\partial}{\partial d}\{d|Y > d\} \\
&= \frac{\partial}{\partial d}\int_d^\infty \frac{y\mathrm{f}_Y(y)}{1-\mathrm{F}_Y(d)}dy - \frac{\partial}{\partial d}\{d|Y > d\} \\
&= \int_d^\infty \frac{\partial}{\partial d}\frac{y\mathrm{f}_Y(y)}{1-\mathrm{F}_Y(d)}dy - \frac{d\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)} - \frac{\partial}{\partial d}\{d|Y > d\} \\
&= \frac{\mathrm{f}_Y(d)}{(1-\mathrm{F}_Y(d))^2}\int_d^\infty y\mathrm{f}_Y(y)dy - \frac{d\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)} - 1 \\
&= \frac{\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)}E[Y|Y > d] - \frac{\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)}d - 1 \\
&= \frac{\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)}\mathrm{TCE}_Y(d) - \frac{\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)}d - 1 \\
&= \frac{\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)}(\mathrm{TCE}_Y(d) - d) - \frac{\partial}{\partial d}\{d|Y > d\} \\
&= \frac{\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)}E[Y_*] - 1,
\end{aligned}$$

where the last term 1 is obtained by differentiating $\{d|Y > d\}$ with respect to d. Here, TCE is used to denote the tail conditional expectation of a severity distribution. □

8.2 Parametric Models

This section reviews the parametric models for general insurance loss severities. In each of the following models, a continuous distribution of loss severity, given a vector of covariates $\boldsymbol{x}$, needs to be specified. Commonly used distributions in lifetime distributions of survival analysis are exponential, Weibull, gamma distributions. In survival analysis, different aging rates are modeled by allowing the location and scale parameters to be functions of covariates. Common practice is to replace the location parameter with $\boldsymbol{x}'\boldsymbol{\beta}$. In Frees et al. (2015), the average severities Y after averaging over policy-years is used for the response. For this, it is common to assume a parametric model with a logarithmic link for parameter interpretability.

Generalized Beta (GB) Family

The generalized beta random variable Y has the density

$$\mathrm{f}_Y(y; a, b, c, \alpha_1, \alpha_2) = \frac{|a| y^{a\cdot\alpha_1 - 1}(1-(1-c)(y/b)^a)^{\alpha_2 - 1}}{b^{a\cdot\alpha_1} B(\alpha_1, \alpha_2)(1 + c(y/b)^a)^{\alpha_1+\alpha_2}},$$

where $0 < c < 1$, and $b, \alpha_1, \alpha_2 > 0$. Here, $B(\alpha_1, \alpha_2)$ is the beta function. The generalized beta family contains many familiar distributions as special cases: *GB1*, *GB2*, gamma, generalized gamma, Weibull, Burr type 3, Burr type 12, Dagum, log-normal, Lomax, *F*, Rayleigh, chi-square, half-normal, half-Student-t, exponential and log-logistic. Klugman et al. (2012) provides an introductory overview of the generalized beta family distributions. Each special case is obtained by restricting the parameter of the distribution to a specific value or taking the limiting case of the parameter. Some special cases have been more popular than others in the loss-modeling context. The GB family is defined for response values between

$$0 < y^a < \frac{b^a}{1-c}.$$

Yet limiting cases are defined for arbitrarily large y values.

Gamma

In practice, often the underlying losses or the truncated per-payment variables are assumed to follow a gamma distribution, defined for $0 < y < \infty$. This special distribution is obtained by taking $b = \alpha_2^{1/a} s$ and $a = c = 1$, and by letting $\alpha_2 \to \infty$, for the parameters of a random variable in the GB family. The resulting density is

$$\mathrm{f}_Y(y; a = 1, b = \alpha_2^{1/a} s, \alpha_1) = \frac{y^{\alpha_1 - 1}\exp(-y/s)}{s^{\alpha_1}\Gamma(\alpha_1)},$$

where α_1 is a shape parameter and s is a scale parameter. For practitioners, the α_1 parameter can be considered to be where the hump of the distribution happens. As the α_1 parameter gets smaller, the hump goes farther to the right. For regression, the scale can be parametrized by

$$s = \mu/\alpha_1 = \exp(\boldsymbol{x}'\boldsymbol{\beta})/\alpha_1.$$

The limited loss variable for the gamma has some nice properties, such as this:

$$E[Y \wedge d] = \frac{s\Gamma(\alpha_1 + 1)}{\Gamma(\alpha_1)}\Gamma(\alpha_1 + 1; d/s) + d\left[1 - \Gamma(\alpha_1; d/s)\right].$$

Exponential

The gamma model can be further simplified by restricting the shape parameter to $\alpha_1 = 1$. In this case, the density reduces to

$$\mathrm{f}_Y(y; s) = \frac{1}{s} \exp\left(-\frac{y}{s}\right).$$

In basic probability courses, the use of the gamma distribution is often motivated by showing that a sum of α_1 exponential random variables becomes the gamma distribution. Because of its simplicity and treatment in early probability courses, the exponential seems to be used in practice quite often. For advanced applications, the tails of both the gamma and exponential distributions have been considered to be too thin, meaning they underestimate the small probability of large losses.

For this reason, actuarial researchers had begun to apply heavy-tail distributions to insurance loss data. The *GB* family provides some flexible specific cases. For practitioners, these topics become a stretch assignment, because familiar regression software, such as the `glm` or `actuar` package in the R programming environment, does not provide a one-line application of these models in a regression context. A simple R routine, with practical heavy-tail regression modeling capability under censoring and truncation, may eventually become available.

Generalized Gamma (*GG*)

When the condition $a = 1$ is relaxed from the gamma, the resulting distribution is called the generalized gamma (*GG*) distribution. That is, $b = \alpha_2^{1/a} s$, $c = 1$, and $\alpha_2 \to \infty$. The density is

$$\mathrm{f}_Y(y; a, b = \alpha_2^{1/a} s, \alpha_1, \alpha_2) = \frac{|a| y^{a \cdot \alpha_1 - 1} e^{-(y/s)^a}}{s^a \Gamma(\alpha_1)}.$$

The *GG* distribution is defined for $0 < y < \infty$. Comparing with the gamma, we see that the exponent a allows for a flexible modeling of the tail behavior of the distribution. From here, if we further restrict $\alpha_1 = a$, then the resulting distribution is the Weibull distribution:

$$\mathrm{f}_Y(y; a, b = \alpha_2^{1/a} s, \alpha_1 = a, \alpha_2) = \frac{a}{s}\left(\frac{y}{s}\right)^{a-1} e^{-(y/s)^a},$$

which is introduced in early probability courses for its simple form of cumulative distribution function, $\mathrm{F}_Y(y; a, s) = 1 - e^{(y/s)^a}$.

Generalized Beta of First Kind (*GB1*)

Restricting other parameters from the *GB* family results in the *GB1* distribution. This is when the restriction $c = 0$ is imposed. The density is

$$\mathrm{f}_Y(y; a, b, \alpha_1, \alpha_2) = \frac{|a| y^{a \cdot \alpha_1 - 1} \left(1 - (y/b)^a\right)^{\alpha_2 - 1}}{b^{a \cdot \alpha_1} B(\alpha_1, \alpha_2)}.$$

The *GB1* response variable is restricted to be between $0 < y^a < b^a$. The *GB1* distribution also contains the *GG* distribution as a special case, by allowing $b = \alpha_2^{1/a}$ and $\alpha_2 \to \infty$. From the *GB1*, restricting $a = 1$ results in the beta distribution of first kind (*B1*). Finally, the *GB1* distribution has the Pareto distribution as a special case, with $a = -1$ and $\alpha_2 = 1$. Introducing a new parameter λ, the density becomes

$$\mathrm{f}_Y(y; \lambda, \alpha_1) = \frac{\alpha_1 \lambda^{\alpha_1}}{(\lambda + y)^{\alpha_1 + 1}}.$$

From a practitioner's perspective, the Pareto distribution is a useful special case, allowing for simple modeling of tail behavior for response variables defined over $0 < y < \infty$. It has the nice property that

$$E[Y] = \frac{\lambda}{\alpha_1 - 1} \quad \text{and} \quad E[Y - d|Y > d] = \frac{\lambda + d}{\alpha_1 - 1}.$$

In fact, if $Y \sim Pa(\alpha_1, \lambda)$, then the truncated random variable follows $Y - d|Y > d \sim Pa(\alpha_1, \lambda + d)$. See Gray and Pitts (2012) for a proof. In practice, one may think of α_1 as a parameter in charge of the thickness of the tail. Smaller values of α_1 result in thicker tails for the distribution. When $\alpha_1 < 1$, the distribution does not have a finite mean, which is often the case for heavy-tail distributions.

Generalized Beta of Second Kind (*GB2*)

The *GB2* distribution is obtained by restricting $c = 1$ from the *GB* family. The resulting density provides a flexible class of distributions for insurance loss modeling. The density is

$$\mathrm{f}_Y(y; a, b, \alpha_1, \alpha_2) = \frac{a(y/b)^{a \cdot \alpha_1}}{yB(\alpha_1, \alpha_2)[1 + (y/b)^a]^{\alpha_1 + \alpha_2}},$$

and is defined for $0 < y < \infty$. In Frees et al. (2015), and Sun et al. (2008), the density is reparametrized, so that the definition of $GB2(\sigma, \mu, \alpha_1, \alpha_2)$ is

$$\mathrm{f}_Y(y; \mu, \sigma, \alpha_1, \alpha_2) = \begin{cases} \dfrac{[\exp(z)]^{\alpha_1}}{y\sigma B(\alpha_1, \alpha_2)[1 + \exp(z)]^{\alpha_1 + \alpha_2}} & \text{for } y > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where the parametrization for regression is $z = \frac{\ln(y) - \mu}{\sigma}$, with $\mu = \boldsymbol{x}'\boldsymbol{\beta}$. According to Frees et al. (2015) and Sun et al. (2008), if $-\alpha_1 < \sigma < \alpha_2$, then the first moment can be obtained using the following formula:

$$\mathrm{E}\left[y|\mathbf{x}\right] = \frac{B(\alpha_1 + \sigma, \alpha_2 - \sigma)}{B(\alpha_1, \alpha_2)} \exp(\mathbf{x}'\boldsymbol{\beta}), \qquad \text{where } B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}.$$

When the value of α_2 is small compared with σ, the tail of the distribution becomes thick, and the first moment does not exist. In these cases, for ratemaking purposes, a practitioner must use direct integration to find the mean of the limited loss variable. Specifically,

$$E[y \wedge u|\boldsymbol{x}] = \int_{-\infty}^{u} y \cdot \mathrm{f}_Y(y)dy = \int_{-\infty}^{u} (1 - \mathrm{F}_Y(y))\, dy,$$

where $\mathrm{F}_Y(y)$ is the cumulative distribution function for the *GB2*, and u is the policy limit, or in other words, the coverage amount. The *GB2* distribution contains a number of special cases. The special case where $\alpha_1 = 1$ is called the Burr type 12 distribution and has been extensively used in heavy-tail loss modeling in the literature. Other special cases of the *GB2* are *GG*, Weibull, gamma, Burr type 3, Dagum, log-normal, Lomax, F, Rayleigh, chi-square, half-normal, half-Student-t, exponential, log-logistic. The Lomax distribution is essentially a Pareto distribution with a different parametrization.

Beta Distributions

The beta distribution is a name for the special case of the *GB* family, where the restriction is $a = 1$. The beta distribution has as a special case, the first and second kind, each corresponding to restrictions $c = 0$ and $c = 1$. The beta distribution is defined for $0 < y < b/(1 - c)$; however, the second kind is defined for $0 < y < \infty$. Because of its support, insurance loss modeling seems to have focused on using the second kind—or in other words, the *GB2* class distributions—more than other specific cases.

Other Distributions (*EGB*, *GPD*, *GEV*)

New random variables can be created by transformations. For example, the *GB* family distribution can be transformed by taking the transformation $Z = \ln Y$, resulting in the exponential generalized beta (*EGB*) distributions. The *EGB* has several popular specific cases, such as the Gompertz or Gumbell distribution. Recently, some researchers consider the generalized Pareto distributions (*GPD*) and generalized extreme-value distributions (*GEV*) for extreme-value modeling, such as the minimum or maximum of a sample of random varaibles. For interested readers, Salvadori and Michele (2007) provides an overview of the extreme-value approach to distributions, and Chavez-Demoulin et al. (2016) and Zhang et al. (2008) provides some applications of extreme-value theory. Although references are provided for these sources, in this paper we would like to focus on the *GB* family distributions and their applications. Extreme-value distributions may be useful for applications such as excess of loss layers ratemaking, which is not of interest in this paper. We believe that the *GB* family distributions provide a flexible class of parametric models for insurance ratemaking, so we illustrate our approaches using specific cases of the *GB* family.

8.3 LGPIF Data Summary

This section applies our rating procedure to real data, with a full set of explanatory variables. The Wisconsin Office of the Insurance Commissioner administers the Local Government Property Insurance Fund (LGPIF). The LGPIF was established to provide property insurance for local government entities that include counties, cities, towns, villages, school districts and library boards. The fund insures local government property such as government buildings, schools, libraries and motor vehicles, within over a thousand entities, which pay approximately $25 million in premiums each year. In effect, the LGPIF acts as a stand-alone insurance company for local government entities. Table 6 summarizes the number of building and contents (BC) policies in force.

Table 6: Summary Statistics of BC (Primary Coverage) Claims

Year	Average Deductible	Loss Frequency	Claim Frequency	Loss Total	Claim Total	Number of Policyholders
2006	3,048	0.735	0.525	20,313,812	18,161,172	1,158
2007	3,233	0.926	0.611	17,230,457	15,261,868	1,142
2008	3,412	0.747	0.518	11,060,356	9,160,440	1,129
2009	3,517	0.925	0.443	11,047,677	8,774,310	1,113
2010	3,599	1.089	0.633	36,659,296	33,328,603	1,113

The LGPIF data is ideal for this study, because the underlying losses and claims are both recorded in the data server. In many practical situations, the former may not be available. Hence, our goal in this study is to assume the former variable is unobserved and to test our result using the observed, empirical underlying losses, which various hypothetical deductible levels can be applied to and compared against. Table 7 shows a summary of the frequency and severity of claims by deductible choice, for 2006–2010. There are a number of instances with a high deductible level, say $100,000$, which implies that this data set may be studied in relation to risk retention problems in the reinsurance context.

Table 7: Summary Statistics of BC Claims by Deductible

Deductible	Avg. Loss Frequency	Avg. Claim Frequency	Average Loss	Average Claim	Number of Observations
500	0.628	0.621	6,197	5,884	2,674
1,000	0.668	0.641	6,808	6,147	1,067
2,500	0.539	0.506	12,923	11,610	686
5,000	0.606	0.362	39,229	36,987	716
10,000	0.378	0.196	13,044	10,692	209
15,000	0.672	0.224	22,426	17,615	67
25,000	5.973	0.202	34,679	21,654	183
50,000	17.290	1.355	530,867	411,317	31
75,000	7.400	0.000	44,897	0	5
100,000	0.294	0.235	486,350	459,880	17

Table 8 shows the explanatory variables in the policyholder data, given that an entity has purchased BC coverage. Table 10 shows a summary of both the underlying loss data, which usually isn't observable, and the claims data for those losses above the chosen deductible, which is observable in most common practices.

The observed claims are summarized in Table 10, allowing for a comparison of `LossBeforeDeductible` and `LossAfterDeductible` in both cases. Because Table 9 is conditional on `LossAfterDeductible` > 0, the minimum value for `LossAfterDeductible` is zero in Table 10 (losses), whereas it is positive in Table 9 (claims).

Table 8: BC Policies, 2006–2010

	Min.	*Median*	*Mean*	*Max.*	*N*
`CoverageBC`	8,937	11,310,000	37,190,000	2,445,000,000	5,655
Log(`CoverageBC`)	-4.718	2.426	2.128	7.802	
`DeductBC`	500	1,000	3,356	100,000	

Table 9: BC Claims, 2006–2010

	Min.	*Median*	*Mean*	*Max.*	*N*
`LossBeforeDeduct`	504	4094	29,920	12,920,000	3,089
After Deduct	4	2,982	27,420	12,920,000	

Table 10: BC Losses, 2006–2010

	Min.	*Median*	*Mean*	*Max.*	*N*
`LossBeforeDeduct`	1	2,243	19,300	12,920,000	4,285
After Deduct	0	750	16,970	12,920,000	

In Frees et al. (2015), the average severities are used for the modeling, whereas here the claims are directly used without the averaging. In this case, it is helpful to address some additional aspects of the data, for parameter interpretability.

Table 11 summarizes the losses and claims with respect to the three peril type categories. The building and contents coverage has subcoverages, each of which consists of different peril types, which could be clustered into different categories. The property fund classifies the claims into three categories by default. Here they have been manually recategorized into nine broad categories: fire, vandalism, lightning, wind, hail, vehicle, water damages (weather and non-weather) and other perils. Because the scale of the loss distribution is highly dependent on the peril type, it is worthwhile to consider specific peril type categories and fit loss distributions for each. Hence, the question is, Given a claim, how can we accurately classify it into one of these categories? For example, given a claim, could we classify it into either a vandalism claim or other? We have done some preliminary analyses using basic discrete choice models, and what we learned is that more categories result in higher standard errors in the coefficients of the peril type model. For this, Rosenberg et al. (1999) and Yuan and Lin (2006) are some related articles. The claim classification should use explanatory variables of the policyholder, instead of any properties of the claim that is unknown at the instance of the claim. In this case, discrete categories may be easier to conceptualize. Classifying claims using continuous mixture models is left as future work.

When the log coverage is plotted with the average severities of claims, usually a positive correlation can be observed. However, when the underlying losses are plotted without the averaging over policy-year observations, the variation in the response variable is larger. Hence, when a single severity model is used with the coverage amount as an explanatory variable, interpretable coefficients may not be obtained without categorizing the claims into peril type categories. For this reason, we are interested in considering the different severity distributions with respect to various

Table 11: Peril Types of BC Losses

Peril	Average Loss	N	Prob.
Fire	87,168	172	0.034
Vandalism, Theft, Etc.	2,084	1,774	0.355
Lightning	11,087	832	0.167
Wind	18,125	296	0.059
Hail	145,488	76	0.015
Damage by Vehicle	3,905	852	0.171
Water (Weather)	80,432	426	0.085
Water (Non-Weather)	23,974	202	0.040
Misc.	29,150	362	0.073

peril types. The next section will show the coefficient estimation procedure and results in more detail for selected models.

8.4 Coefficient Estimation

The general situation is that the jth observed claim for policyholder i is forced to be in the interval $(0, \infty)$, due to left truncation point d_i. We indicate those claim observations above the deductible by using $j(\varsigma)$, where the indices ς (varsigma) take on values $1, \dots N_g(d)$. To specify the likelihood, we consider modeling the following observed variables:

$$\begin{aligned} N_{g,i}(d) &= \sum_{j=1}^{N_i} I(Y_{ij} > d_i), \\ Y_{*,i,j(\varsigma;i)}(d) &= \sum_{\varsigma=1}^{N_i^c} Y_{i,j(\varsigma)} - d_i | Y_{i,j(\varsigma;i)} > d_i, \\ j(\varsigma; i) &= \text{ index of } i\text{th loss above } d_i,\ \varsigma = 1, \dots, N_{g,i}, \\ i(\iota) &= \text{ index of } j\text{th positive } Y_{*,i,j(\varsigma;i)}(d), \end{aligned}$$

Severity. Here, the likelihood for severities is specified. In most practical situations in actuarial science, upper-tail truncation rarely happens, and we are interested in ordinary left truncation only. Then, the likelihood becomes

$$\begin{aligned} L_{Y|M} = &\prod_{M_\iota = m} \prod_{\varsigma=1}^{n_{g,i(\iota)}} \frac{\mathrm{f}_{Y|M}\left(y_{*,i(\iota),j(\varsigma;i)} + d_{i(\iota)}\right)}{1 - \mathrm{F}_{Y|M}\left(d_{i(\iota)}\right)} \cdot \mathrm{I}\left(y_{*,i(\iota),j(\varsigma;i)} < u_{i(\iota)} - d_{i(\iota)}\right) \\ &+ \prod_{M_\iota = m} \prod_{\varsigma=1}^{n_{g,i(\iota)}} \frac{1 - \mathrm{F}_{Y|M}\left(u_{i(\iota)}\right)}{1 - \mathrm{F}_{Y|M}\left(d_{i(\iota)}\right)} \cdot \mathrm{I}\left(y_{*,i(\iota),j(\varsigma;i)} = u_{i(\iota)} - d_{i(\iota)}\right), \end{aligned}$$

where y_* is used to denote a realization of $Y_*(d)$, and the second term will be nonzero if there is right-censoring due to a policy limit $u_{i(\iota)}$. This provides the likelihood of the conditional severity distribution. Coefficients have been estimated for the exponential, gamma and Pareto distributions, for each peril type separately. Results are shown in Table 12.

Table 12: Exponential, Gamma, Pareto Model Coefficient Estimates

		Exponential		**Gamma**		**Pareto**	
Fire		*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
	(Intercept)	9.484	0.246	9.524	0.385	11.628	1.029
	Coverage	0.455	0.060	0.447	0.096	0.571	0.093
	shape			0.423	0.041	1.012	0.012
	log L	1,723		1,674		1,626	
Vandalism		*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
	(Intercept)	7.713	0.112	7.782	0.001	8.806	0.305
	Coverage	0.116	0.025	0.016	0.001	-0.110	0.041
	shape			0.454	0.001	1.357	0.131
	log L	4,543		4,435		4,324	
Lightning		*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
	(Intercept)	8.028	0.105	8.032	0.101	8.419	0.177
	Coverage	0.324	0.028	0.325	0.027	0.226	0.040
	shape			1.079	0.051	1.874	0.201
	log L	7,346		7,345		7,213	
Wind		*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
	(Intercept)	8.829	0.139	8.860	0.171	9.258	0.607
	Coverage	0.250	0.035	0.235	0.044	0.257	0.066
	shape			0.715	0.058	1.242	0.177
	log L	2,645		2,635		2,552	
Hail		*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
	(Intercept)	9.658	0.181	9.672	0.239	10.481	0.550
	Coverage	0.595	0.051	0.593	0.068	0.276	0.091
	shape			0.584	0.085	1.543	0.436
	log L	857		849		827	
Vehicle		*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
	(Intercept)	7.917	0.117	7.921	0.075	8.123	0.149
	Coverage	0.049	0.029	0.087	0.018	-0.021	0.036
	shape			2.217	0.103	3.924	0.622
	log L	5,753		5,645		5,690	
Water (NW)		*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
	(Intercept)	8.058	0.201	8.072	0.246	10.167	1.460
	Coverage	0.400	0.043	0.386	0.053	0.137	0.088
	shape			0.707	0.070	1.127	0.204
	log L	1,705		1,698		1,653	
Water (W)		*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
	(Intercept)	9.493	0.180	9.516	0.300	12.953	0.438
	Coverage	0.428	0.042	0.428	0.071	0.270	0.059
	shape			0.375	0.023	1.007	0.002
	log L	4,310		4,141		3,910	
Misc.		*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
	(Intercept)	8.959	0.148	9.023	0.230	10.561	1.662
	Coverage	0.351	0.036	0.333	0.057	0.174	0.069
	shape			0.436	0.031	1.063	0.112
	log L	3,075		2,990		2,817	

The fit of these models can be assessed by looking at the Q-Q plots. From Figures 6, 7, 8, 9, 10, 11, 12, 12, 13, 14, the reader may see that the Pareto model fits best for most of the peril types, demonstrating the long-tail nature of the claim severities.

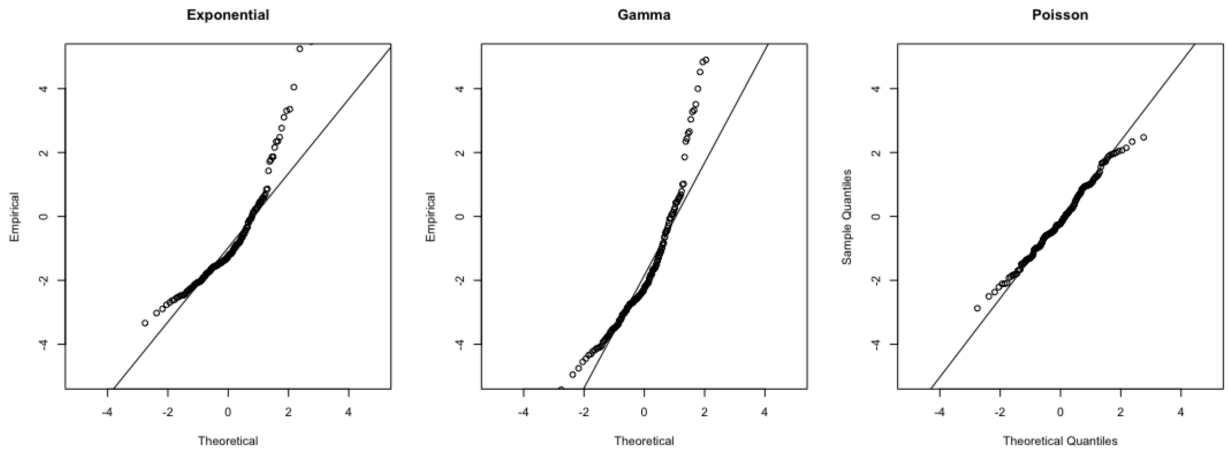

Figure 6: Fire Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.442(0.000), 0.594(0.000), 0.094(0.097)

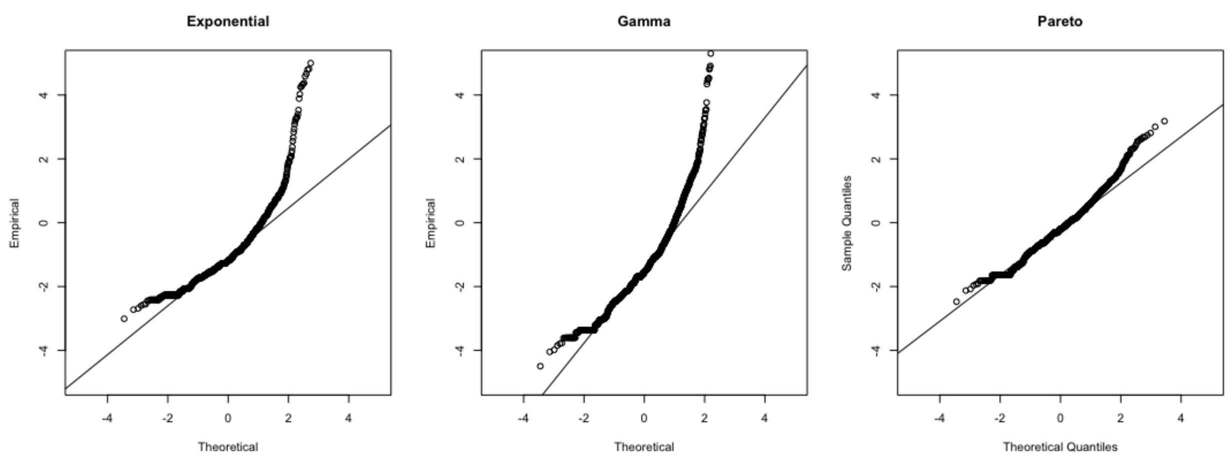

Figure 7: Vandalism Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.472(0.000), 0.520(0.000), 0.143(0.000)

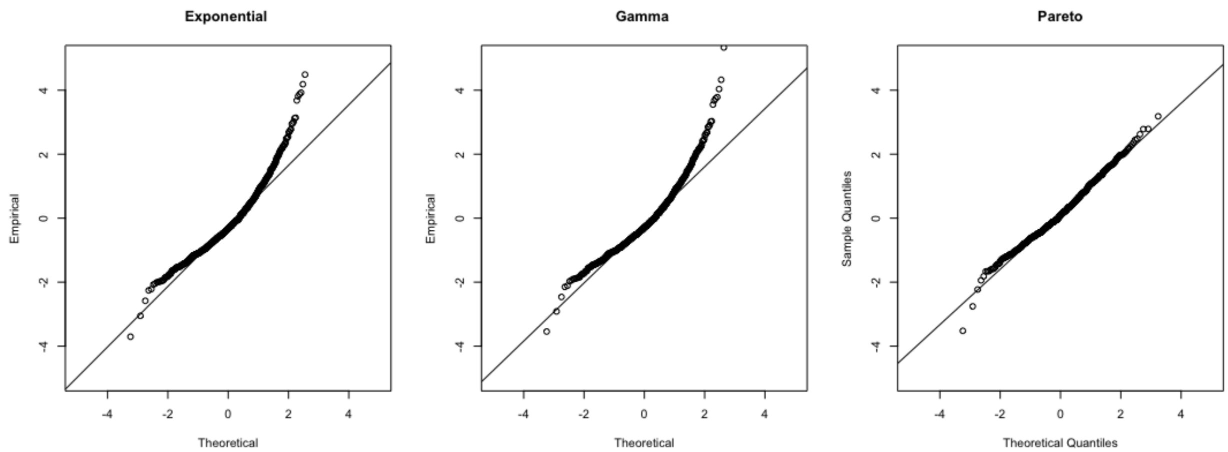

Figure 8: Lightning Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.148(0.000), 0.138(0.000), 0.105(0.000)

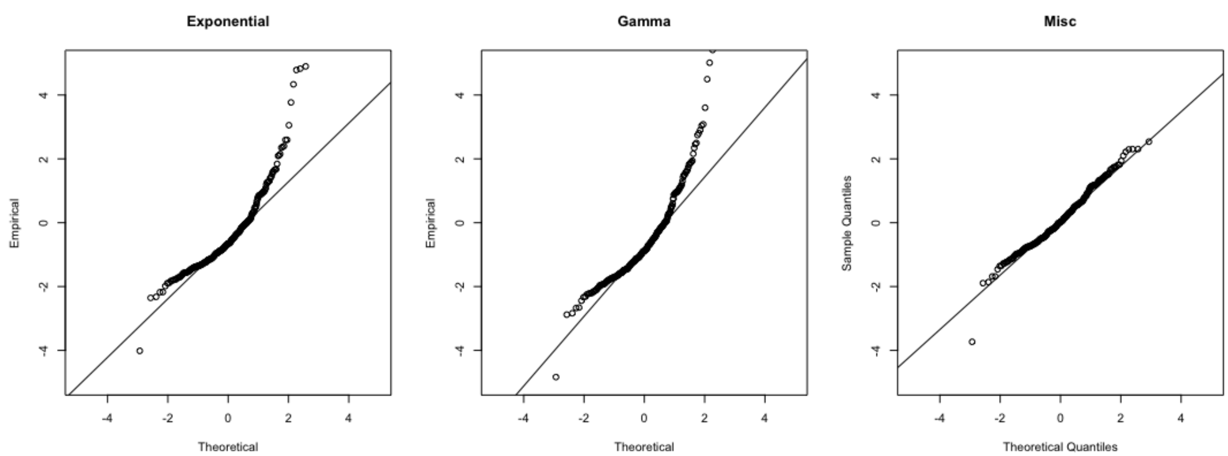

Figure 9: Wind Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.262(0.000), 0.328(0.000), 0.095(0.010)

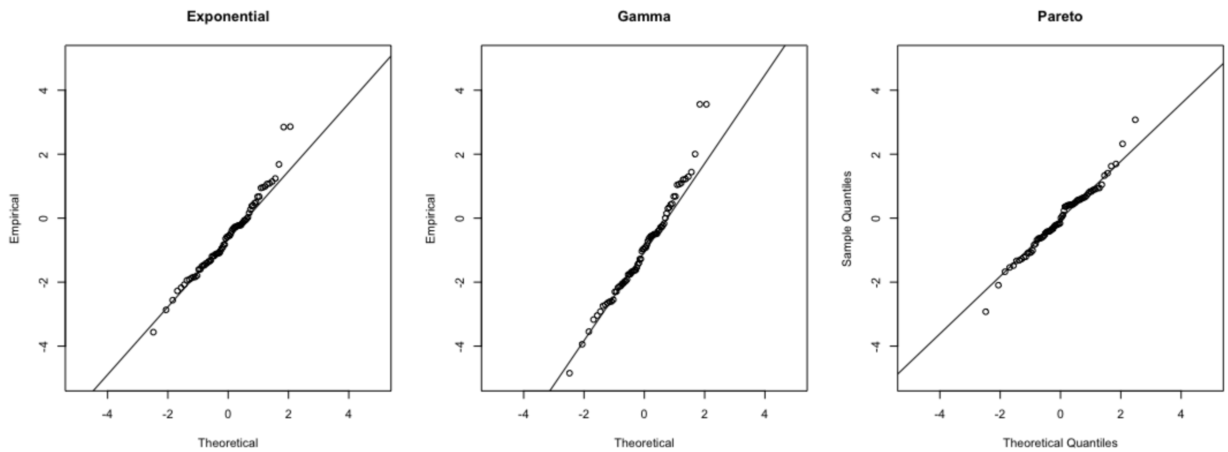

Figure 10: Hail Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.257(0.000), 0.359(0.000), 0.088(0.565)

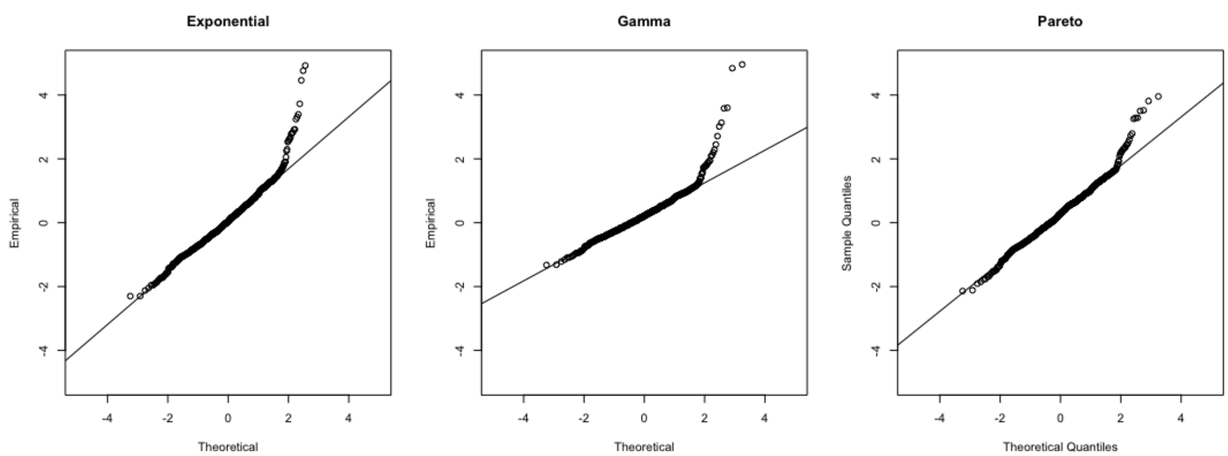

Figure 11: Vehicle Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.084(0.000), 0.246(0.000), 0.159(0.000)

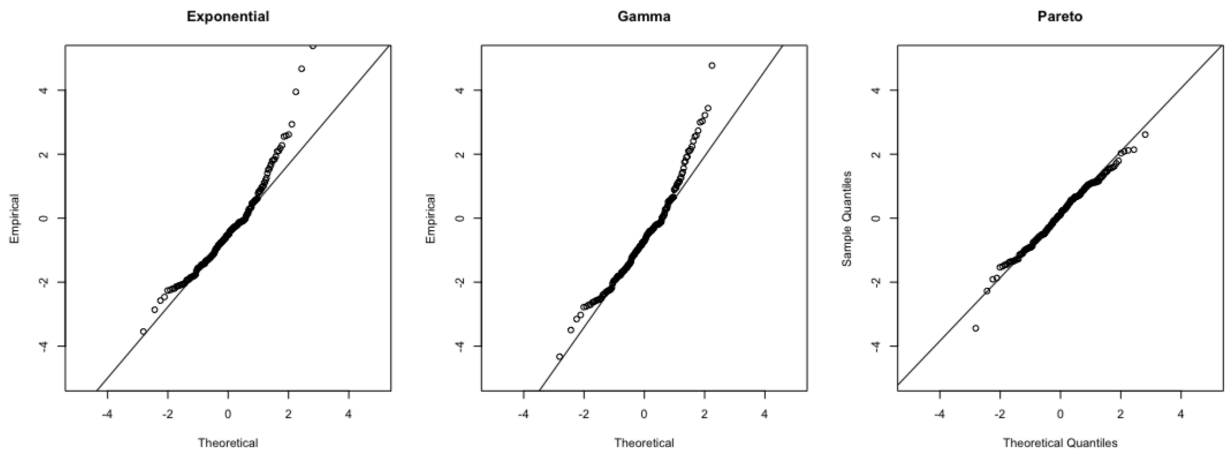

Figure 12: Water (Non-Weather) Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.230(0.000), 0.267(0.000), 0.083(0.128)

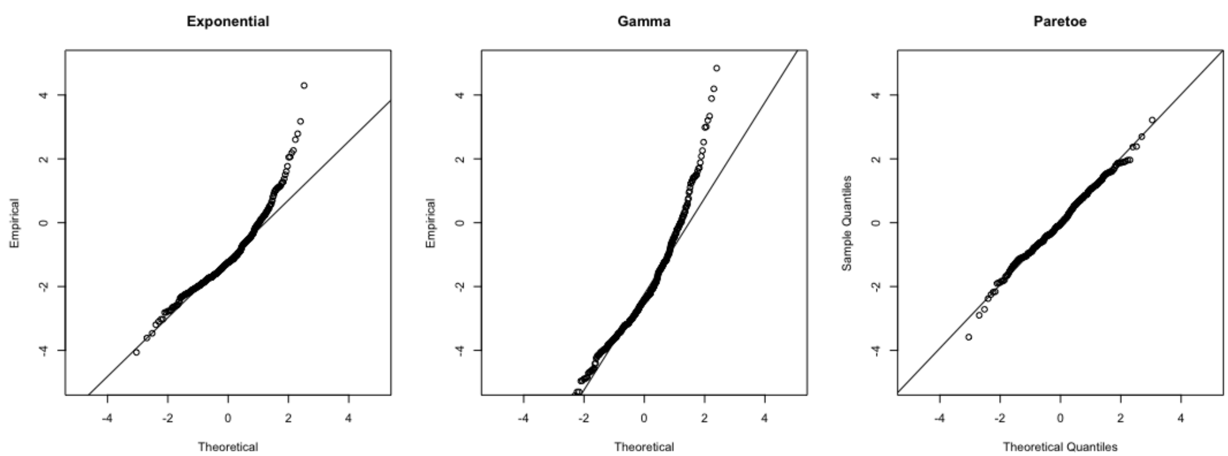

Figure 13: Water (Weather) Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.120(0.000), 0.648(0.000), 0.044(0.377)

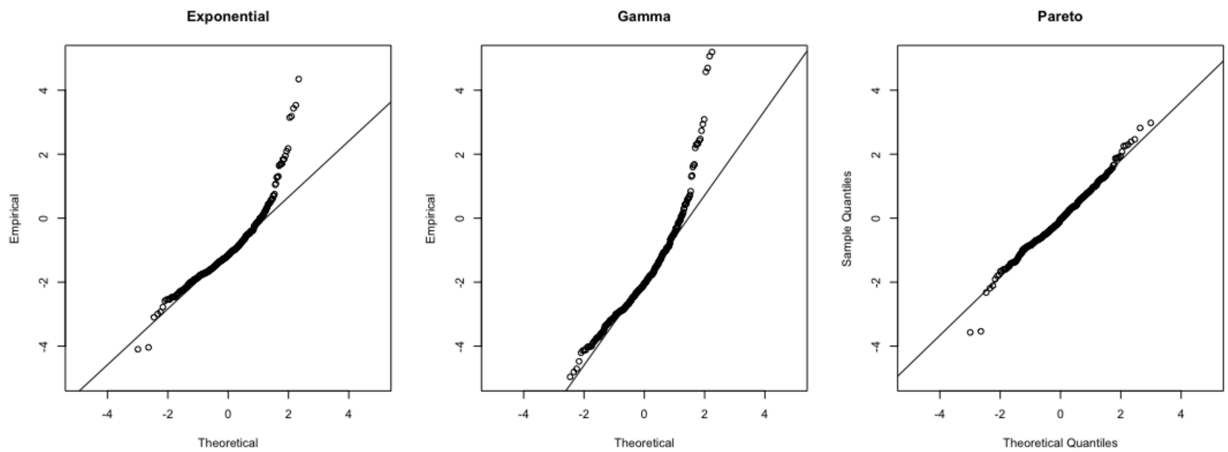

Figure 14: Miscellaneous Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.453(0.000), 0.610(0.000), 0.059(0.160)

The fit of the models may be visually inspected by plotting the empirical data with a plot of the fitted density for each peril type. For example, Figure 3 shows the fit of the three different models (exponential, gamma and Pareto) on a log scale.

Frequency: Obtaining the underlying frequency parameters considers the coverage modification by the deductible. For regression models with a log link, the quantity $\ln(1 - \mathrm{F}_Y(d_i))$ can be used as an offset in standard regression routines. Specifically, for a mean parametrization

$$E[N_i] = \exp\left(\boldsymbol{x}_i'\boldsymbol{\gamma}\right),$$

consider the observed frequencies, $E[N_{g,i}]$. Section 3 theory can be used, so that

$$E[N_{g,i}] = E[N_i](1 - \mathrm{F}_Y(d_i)) = \exp\left(\boldsymbol{x}_i'\boldsymbol{\gamma}\right)(1 - \mathrm{F}_Y(d_i)) = \exp\left(\boldsymbol{x}_i'\boldsymbol{\gamma} + \texttt{offset}\right).$$

Hence, for most frequency regression models, the offset variable $\texttt{offset} = \ln(1 - \mathrm{F}_Y(d_i))$ can be used, in order to recover the underlying loss frequency distribution parameters. For $(a, b, 0)$ class distributions with log link, summarized in Table 1, this approach would work. In particular, the approach would work for the Poisson regression or the negative binomial regression. This can be extended to a general link function η, where the mean is parametrized by

$$E[N_i] = \eta^{-1}\left(\boldsymbol{x}_i'\boldsymbol{\gamma}\right).$$

Estimation becomes more complicated for zero-inflated models, or zero-one-inflated models. For a treatment of zero-one-inflated models, see Frees et al. (2015). In these models, the modification to each component of the primary and secondary probability mass function should be specified in the likelihood function. The estimation of zero-one-inflated Poisson models is covered in Section 8.5. In this paper, we will use the Poisson frequency model for most illustrations. The coefficient estimates for the Poisson model are shown, using different coverage modification models for the severity part. In general, the log likelihood tends to improve when the log-tail, Pareto model is assumed.

Table 13: Poisson Frequency Model Coefficient Estimates

		Fire		Vandalism		Lightning	
		Coef.	*Std.Err.*	*Coef.*	*Std.Err.*	*Coef.*	*Std.Err.*
Exponential	(Intercept)	-5.366	0.225	-5.082	0.156	-3.400	0.093
	Coverage	0.551	0.055	1.007	0.036	0.484	0.024
	-log L	604		1,466		2,091	
Gamma	(Intercept)	-5.374	0.225	-5.374	0.155	-3.242	0.096
	Coverage	0.556	0.055	1.099	0.036	0.518	0.025
	-log L	603		1,590		1,985	
Pareto	(Intercept)	-5.124	0.234	-5.118	0.154	-3.245	0.097
	Coverage	0.557	0.057	1.112	0.035	0.560	0.025
	-log L	594		1,396		1,949	

		Wind		Hail		Vehicle	
		Coef.	*Std.Err.*	*Coef.*	*Std.Err.*	*Coef.*	*Std.Err.*
Exponential	(Intercept)	-4.389	0.161	-4.681	0.210	-4.155	0.122
	Coverage	0.503	0.042	0.137	0.065	0.847	0.030
	-log L	947		375		1,799	
Gamma	(Intercept)	-4.454	0.166	-4.719	0.209	-3.839	0.121
	Coverage	0.599	0.044	0.141	0.065	0.799	0.030
	-log L	933		376		1,837	
Pareto	(Intercept)	-4.250	0.163	-4.693	0.212	-4.179	0.122
	Coverage	0.525	0.043	0.163	0.066	0.883	0.030
	-log L	937		376		1,786	

		Water (NW)		Water (W)		Misc.	
		Coef.	*Std.Err.*	*Coef.*	*Std.Err.*	*Coef.*	*Std.Err.*
Exponential	(Intercept)	-6.242	0.262	-4.910	0.157	-4.546	0.161
	Coverage	0.867	0.058	0.685	0.036	0.550	0.040
	-log L	601		1,155		998	
Gamma	(Intercept)	-6.652	0.282	-5.245	0.174	-4.710	0.172
	Coverage	1.043	0.063	0.899	0.041	0.697	0.043
	-log L	594		1,142		965	
Pareto	(Intercept)	-6.491	0.277	-4.973	0.166	-4.543	0.169
	Coverage	0.990	0.061	0.776	0.038	0.647	0.042
	-log L	593		1,134		963	

Mixture Approach: For modeling insurance losses with different profiles, depending on specific cases, mixture models have been used in the literature. In practice, exponential mixtures and mixed Pareto approaches have been used. For cases with mixture weights in multiple categories, practitioners have used the term *Pareto soup* model; see White (2005). One motivation for mixture models may be the different peril types, under which the loss severities experience different profiles. The modeling of severities, given a specific peril type, can be performed conditional on each peril type to obtain a set of parameters for each category of claim peril. This approach is taken in our aggregate claims prediction. Hence estimation issues for mixture models under censoring and truncation are discussed in detail here.

Suppose the number of claims N and Y are independent. Let M be the peril type categorical variable, so that there are several peril types with respective loss severities, conditional on the peril type category. Then the conditional density for claim severities can be written as

$$\mathrm{f}_{Y|M}(y|m) = \mathrm{f}_Y(y; \theta_m),$$

where we allow the distribution parameters θ_m to vary over different peril types. The unconditional distribution for the severity can be obtained using the mixture

$$\mathrm{F}_Y(y) = \sum_m \mathrm{F}_{Y|M}(y|m)\mathrm{f}_M(m).$$

In this case, the expected loss severity becomes

$$E[Y] = \sum_m E[Y|M = m]\mathrm{f}_M(m).$$

The underlying peril type probabilities, $\mathrm{f}_M(m)$, are required for this. In practice, fixed probabilities $\mathrm{f}_M(m) = p_m$ may be used. Here, in order to determine the peril type probabilities in a data-driven way, we specify the joint density for the frequencies and severities as

$$\mathrm{f}_{Y,N}(y, n) = \sum_m \mathrm{f}_{Y|M}(y|m) \cdot \mathrm{f}_M(m) \cdot \mathrm{f}_N(n),$$

for each peril type m. For related work, Frees and Valdez (2008) use a discrete choice model, focusing on the dependency among auto insurance claim types, and Fu and Liu (forthcoming) use the EM algorithm for estimation of a finite mixture model. Other approaches may be to impose parametric models for the peril type probabilities. The approach we take is to model each specific peril type and consider the coverage modification for that specific peril only.

Hence, calculation of the total building and contents density requires the peril type probabilities, since the above results are peril dependent. The peril type probabilities may be modeled either for all of the perils or for the reclassified peril type categories. For the Pareto mixture, we tried the former, while for the *GB2* mixture, the reclassified approach has been used with categories shown in Table 14. Other data-driven clusterings may be possible. In this case, the clustering has been manually performed into three arbitrarily chosen categories, using the average loss severities. The advantage of clustering the perils into categories is the reduction in the number of parameters to be estimated, especially when covariates are used for the peril type model.

Table 14: Average Loss and Claim Severity by Peril

Category	Perils	Average Loss	Average Claim	N	Proportion
Low	Vandalism, theft, burglary, damage by vehicle	2,675	1,495	2,626	53%
Medium	Lightning, wind, non-weather water damage, misc.	17,721	15,020	1,692	34%
High	Fire, hail, water damage by weather	89,487	83,617	674	14%
Total		19,496	17,167	4,992	100%

In this paper, a basic discrete choice model is used for the peril type probabilities, whose coefficients are estimated using maximum likelihood:

$$\mathrm{f}_M(m) = \begin{cases} \dfrac{\exp(\boldsymbol{x}'\boldsymbol{\omega}_m)}{1 + \sum_{\varphi \neq m_0} \exp(\boldsymbol{x}'\boldsymbol{\omega}_\varphi)} & \text{for peril type } m \neq m_0 \\[2ex] \dfrac{1}{1 + \sum_{\varphi \neq m_0} \exp(\boldsymbol{x}'\boldsymbol{\omega}_\varphi)} & \text{for base peril type } m = m_0, \end{cases}$$

Here, $\boldsymbol{\omega}_m$ are the regression coefficients of interest. With deductible d, the observed peril type probabilities are altered. The underlying ω_m can be recovered by first calculating

$$1 - \mathrm{F}_{Y|M}(d|m), \quad \text{for each } m,$$

which are simply the coverage modification amounts for each peril type. Using the truncated claims, given $Y > d$, the likelihood for peril type is specified using a multinomial model:

$$L_{M|Y>d} = \Pr(M = m|Y > d) = \frac{\Pr(Y > d|M = m)\mathrm{f}_M(m)}{\sum_{\varphi} \Pr(Y > d|M = \varphi)\mathrm{f}_M(\varphi)}$$

$$= \frac{\left(1 - \mathrm{F}_{Y|M}(d|m)\right) \mathrm{f}_M(m)}{\sum_{\varphi} \left(1 - \mathrm{F}_{Y|M}(d|\varphi)\right) \mathrm{f}_M(\varphi)}. \tag{12}$$

The left side of (12) is an observed quantity from the truncated data. The unknowns are the parameters for the unconditional probabilities $\mathrm{f}_M(m)$ for each m and the conditional severity distribution parameters for each peril type. In a maximum likelihood context, these quantities would need to be included in the likelihood for the severity model. The goal is to estimate the $\boldsymbol{\omega}$ in $\mathrm{f}_M(m; \boldsymbol{\omega})$. For this, each observation of $M|Y > d$ within the truncated claims data is used in the following model for the categorical response M:

$$\begin{aligned} L_{Y,M|Y>d} &= \Pr(Y = y, M = m|Y > d) \\ &= \Pr(Y = y|M = m, Y > d) \cdot \Pr(M = m|Y > d) \\ &= L_{Y|M,Y>d} \cdot L_{M|Y>d}. \end{aligned} \tag{13}$$

Taking the log of (13) and summing, we have

$$\log L_{Y,M|Y>d} = \log L_{Y|M,Y>d} + \log L_{M|Y>d}, \tag{14}$$

where the first term is the log-likelihood for the severity for a specific peril type category, and the second term is the likelihood for the peril type. Note that the first term is the likelihood for the severity for a given observed claim, and the second is a probability weight, which is due to the modification by a deductible d.

Poisson-Gamma Regression (Model A)

This section details the models in the main text, Section 5.3 (the aggregate claims study). Two different regression approaches are compared with the gamma model with truncated estimation. We also compare a *GB2* mixture model with entity types used as predictors for peril type categories. The *GB2* distribution is explained in Sun et al. (2008) and Frees et al. (2015). We begin here by showing the two regression approaches. Table 15 shows the coefficient estimates for the first approach. The frequency model contains `lnDeduct` as an explanatory variable. Here, the peril type model has used a simple classification scheme, where the peril types are categorized as having low, medium and high loss severity, before fitting the model. This was necessary because fitting too many peril type predictors resulted in high standard errors for the peril type model. Different approaches may be used for the reclassification of peril type categories.

Table 15: Poisson-Gamma Regression (Model A)

Gamma			Poisson		
Variable	*Coef.*	*Std. Err.*	*Variable*	*Coef.*	*Std. Err.*
(Intercept)	8.442	0.094	(Intercept)	-7.901	0.170
`CoverageBC`	0.353	0.026	`CoverageBC`	0.889	0.018
			`lnDeductBC`	-0.737	0.020
			`NoClaimCreditBC`	-0.409	0.060
`Type`:City	-0.095	0.105	`Type`:City	-0.068	0.063
`Type`:County	-0.093	0.123	`Type`:County	-0.323	0.075
`Type`:Misc	0.358	0.185	`Type`:Misc	-0.335	0.112
`Type`:School	0.764	0.111	`Type`:School	-0.745	0.065
`Type`:Town	0.115	0.194	`Type`:Town	0.021	0.116
ϕ	2.847				
AIC	64,444		*AIC*	9,054	

Table 16 shows the estimation results when $\ln E = \ln(u - d)$ is used as an offset variable in both the frequency and severity regressions. Because the deductible amounts and coverages are used as offsets, they are excluded from the set of explanatory variables in the regression.

Table 16: Regression Approach With Offset

Gamma			Poisson		
Variable	*Coef.*	*Std. Err.*	*Variable*	*Coef.*	*Std. Err.*
(Intercept)	7.295	0.076	(Intercept)	-2.960	0.046
			`NoClaimCreditBC`	-0.669	0.058
`Type`:City	-1.191	0.093	`Type`:City	-0.852	0.056
`Type`:County	-1.953	0.103	`Type`:County	-1.263	0.062
`Type`:Misc	0.796	0.185	`Type`:Misc	-1.904	0.110
`Type`:School	-0.676	0.095	`Type`:School	-1.599	0.057
`Type`:Town	1.476	0.191	`Type`:Town	0.407	0.114
ϕ	3.196				
AIC	65,037		*AIC*	11,828	

Poisson-Gamma MLE (Model B)

Table 17 shows the coefficient estimates when truncated estimation procedures are used with the Poisson frequency and gamma severity distributions. Note that the AIC for the severity model is high, most likely because of the poor model fit in the upper tail. Estimates of the Poisson coefficient are shown in the second panel, where the recovered coefficients are different from Tables 15 and 16.

Table 17: Poisson-Gamma Maximum Likelihood (Model B)

Gamma			Poisson		
Variable	*Coef.*	*Std. Err.*	*Variable*	*Coef.*	*Std. Err.*
(Intercept)	8.157	0.091	(Intercept)	-1.924	0.059
`CoverageBC`	0.373	0.025	`CoverageBC`	0.591	0.016
			`NoClaimCreditBC`	-0.725	0.059
`Type`:City	-0.114	0.103	`Type`:City	0.118	0.064
`Type`:County	-0.109	0.133	`Type`:County	0.087	0.075
`Type`:Misc	0.395	0.185	`Type`:Misc	-0.635	0.112
`Type`:School	0.743	0.110	`Type`:School	-0.774	0.065
`Type`:Town	0.150	0.196	`Type`:Town	-0.195	0.117
ϕ	0.272	0.007			
AIC	70,768		AIC	10,948	

Poisson-*GB2* MLE (Model C)

Table 18 shows the coefficient estimates for the *GB2* severity model. For the severity model, notice that only the scale parameters are peril dependent, in order to reduce the number of parameters used in the model. Also, the reader can observe that the AIC for the severity model becomes much lower under the *GB2* model, compared with the gamma model in Table 17.

Table 18: *GB2* Maximum Likelihood (Model C)

Term	*Variable*	*Coef.*	*Std. Err.*
Low	(Intercept)	7.235	0.138
Medium	(Intercept)	7.619	0.164
	`CoverageBC`	0.198	0.028
High	(Intercept)	8.332	0.216
	`CoverageBC`	0.250	0.040
$\boldsymbol{\omega}_M$	(Intercept)	0.368	0.102
	`Type`:City	-0.705	0.124
	`Type`:County	0.467	0.142
	`Type`:Misc	-0.191	0.248
	`Type`:School	0.056	0.129
	`Type`:Town	-0.314	0.254
$\boldsymbol{\omega}_H$	(Intercept)	-0.669	0.134
	`Type`:City	-0.605	0.166
	`Type`:County	0.444	0.183
	`Type`:Misc	0.081	0.309
	`Type`:School	0.344	0.164
	`Type`:Town	-0.081	0.323
	σ	1.213	0.154
	α_1	1.653	0.337
	α_2	1.677	0.344
AIC		17,737	

Table 18 shows only the severity distribution coefficients. This could be paired with the Poisson model or more advanced frequency model assumptions such as zero-one inflated models. Comparison of various assumptions and coefficient estimates are shown in Table 19. In Table 19, the estimated frequency parameters differ, depending on whether the underlying losses are observed or unobserved. Estimation of the coefficients in Table 19 is explained in the following Section 8.5.

Table 19: Comparison of Coefficients for Frequency Models

		(1) Poisson Underlying		(2) 01-Poisson Underlying		(3) Poisson Censored Estimation		(4) 01-Poisson Censored Estimation	
		Coef.	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
Poisson	(Intercept)	-2.874	0.054	-1.841	0.098	-1.955	0.060	-1.913	0.066
	CoverageBC	0.993	0.012	0.753	0.019	0.733	0.017	0.734	0.020
	NoClaimCreditBC	-0.668	0.047	-0.289	0.122	-0.574	0.059	-0.587	0.065
	TypeCity	-0.597	0.058	-0.025	0.077	0.055	0.063	0.038	0.064
	TypeCounty	-0.540	0.064	-0.130	0.086	-0.190	0.075	-0.214	0.076
	TypeMisc	-1.884	0.113	-0.388	0.168	-0.494	0.112	-0.542	0.121
	TypeSchool	-0.988	0.056	-1.070	0.083	-0.792	0.065	-0.786	0.067
	TypeTown	0.360	0.113	-0.117	0.144	-0.045	0.117	-0.130	0.124
Zero	(Intercept)			-0.684	0.325			-13.048	9.776
	CoverageBC			0.074	0.071			1.228	1.017
	NoClaimCreditBC			1.147	0.243			-5.461	44.880
One	(Intercept)			-3.507	0.436			-6.400	3.860
	CoverageBC			0.481	0.089			-0.196	0.276
	NoClaimCreditBC			0.900	0.337			2.167	3.894

8.5 01-Inflated Model Estimation

To estimate the 01-inflated Poisson model from censored observations, we define $v = 1 - \mathrm{F}_Y(d)$ for notational convenience, since only those losses above the deductible would be observed as a claim. Then the observed zero probabilities would satisfy

$$
\begin{aligned}
&\Pr(N_{\lambda,g}(d) = 0) \\
&= \pi_0 + \pi_2 \mathrm{P}_\lambda(0) + \pi_1(1-v) + \pi_2 \mathrm{P}_\lambda(1)(1-v) + \pi_2 \mathrm{P}_\lambda(2)(1-v)^2 + \pi_2 \mathrm{P}_\lambda(3)(1-v)^3 + \dots \\
&= \pi_0 + \pi_2 \mathrm{P}_\lambda(0) + \pi_1(1-v) + \pi_2 \left[\lambda e^{-\lambda}(1-v) + \frac{e^{-\lambda}\lambda^2}{2!}(1-v)^2 + \frac{e^{-\lambda}\lambda^3}{3!}(1-v)^3 + \dots\right] \\
&= \pi_0 + \pi_2 \mathrm{P}_\lambda(0) + \pi_1(1-v) + \pi_2 \frac{e^{-\lambda}}{e^{-\lambda(1-v)}} \left[\lambda(1-v)e^{-\lambda(1-v)} + \frac{e^{-\lambda(1-v)}(\lambda(1-v))^2}{2!} + \dots\right] \\
&= \pi_0 + \pi_2 \mathrm{P}_\lambda(0) + \pi_1(1-v) + \pi_2 e^{-\lambda v}(1 - \mathrm{P}_{\lambda(1-v)}(0)) \\
&= \pi_0 + \pi_2 \mathrm{P}_\lambda(0) + \pi_1(1-v) + \pi_2 \mathrm{P}_{\lambda v}(0)(1 - \mathrm{P}_{\lambda(1-v)}(0)),
\end{aligned}
$$

where we use the notation $\mathrm{P}_{\lambda(1-v)}(n)$ to denote the probability of the secondary Poisson distribution with parameter $\lambda(1-v)$ being n.

The probability of one claim to be observed is

$$\begin{aligned}
&\Pr(N_{\lambda,g}(d) = 1) \\
&= \pi_1 + \pi_2 \mathrm{P}_\lambda(1) v \binom{1}{1} + \pi_2 \mathrm{P}_\lambda(2)(1-v) v \binom{2}{1} + \pi_2 \mathrm{P}_\lambda(3)(1-v)^2 v \binom{3}{1} + \dots \\
&= \pi_1 + \pi_2 v \left[\mathrm{P}_\lambda(1) + 2\mathrm{P}_\lambda(2)(1-v) + 3\mathrm{P}_\lambda(3)(1-v)^2 + \dots \right] \\
&= \pi_1 + \pi_2 v \left[\lambda e^{-\lambda} + 2\frac{\lambda^2 e^{-\lambda}}{2!}(1-v) + 3\frac{\lambda^3 e^{-\lambda}}{3!}(1-v)^2 + \dots \right] \\
&= \pi_1 + \pi_2 v \lambda \frac{e^{-\lambda}}{e^{-\lambda(1-v)}} \left[e^{-\lambda(1-v)} + \frac{(\lambda(1-v))e^{-\lambda(1-v)}}{1!} + \frac{(\lambda(1-v))^2 e^{-\lambda(1-v)}}{2!} + \dots \right] \\
&= \pi_1 + \pi_2 \lambda v e^{-\lambda v}(1 - \mathrm{P}_{\lambda(1-v)}(0)) \\
&= \pi_1 + \pi_2 \cdot \mathrm{P}_{\lambda v}(1) \cdot (1 - \mathrm{P}_{\lambda(1-v)}(0)),
\end{aligned}$$

and the probability of n claims being observed is

$$\begin{aligned}
&\Pr(N_{\lambda,g}(d) = n) \\
&= \pi_2 \left[\mathrm{P}_\lambda(n) v^n \binom{n}{0} + \mathrm{P}_\lambda(n+1) v^n (1-v) \binom{n+1}{1} + \mathrm{P}_\lambda(n+2) v^n (1-v)^2 \binom{n+2}{2} + \dots \right] \\
&= \pi_2 v^n \lambda^n e^{-\lambda} \frac{1}{n!} \left[1 + \frac{\lambda(1-v)}{1!} + \frac{(\lambda(1-v))^2}{2!} + \dots \right] \\
&= \pi_2 \frac{(\lambda v)^n e^{-\lambda v}}{n!} \\
&= \pi_2 \Pr\left(N_{\lambda v} = n\right) \qquad \text{for } n \geq 2.
\end{aligned}$$

Using these terms, the log-likelihood for policyholder i can be specified as

$$\begin{aligned}
\log L_i = \log\{ &\Pr(N_{\lambda_i,g}(d_i) = 0) \cdot \mathrm{I}(N_{\lambda_i,g}(d_i) = 0) \\
&+ \Pr(N_{\lambda_i,g}(d_i) = 1) \cdot \mathrm{I}(N_{\lambda_i,g}(d_i) = 1) \\
&+ \Pr(N_{\lambda_i,g}(d_i) = n_{\lambda_i,g}) \cdot \mathrm{I}(N_{\lambda_i,g}(d_i) = n_{\lambda_i,g}(d_i)) \}.
\end{aligned}$$

Because the estimation of 01-inflated models under deductible influence is an interesting application, we have discussed the details. Table 5 provides an assessment of the performance of this model with and without deductible influence.