Rating Endorsements Using Generalized Linear Models

By Edward W. Frees and Gee Lee

ABSTRACT

Insurance policies often contain optional insurance coverages known as endorsements. Because these additional coverages are typically inexpensive relative to primary coverages and data can be sparse (coverages are optional), rating of endorsements is often done in an ad hoc manner after a primary analysis has been conducted. This paper describes a study of the Wisconsin Local Government Property Insurance Fund where it is desirable to have a formal mechanism for rating endorsements. Our goal is to provide prediction algorithms that are transparent and that promote equity among policyholders by determining rates that reflect the appropriate level and amount of uncertainty of each risk. To accommodate potentially conflicting goals of data complexity and algorithmic transparency, we utilize shrinkage techniques to moderate the effects of endorsements with penalized likelihoods. We find that the rating algorithms using shrinkage techniques have a predictive accuracy that are comparable to unbiased generalized linear model techniques and provide relativities for endorsements that are consistent with sound economic, risk management, and actuarial practice.

1. Introduction

It is common for insurance policies to contain optional insurance coverages, often referred to as endorsements or riders. These options may include alternative deductibles and coverage limits and they may also provide extensions to the type of peril (e.g., stolen jewelry in homeowners insurance) covered. Rate manuals provide guidance for the surcharge associated with these optional coverages. For example, Werner and Modlin (2010) describe processes of incorporating endorsement surcharges into rates. For the actuary who uses generalized linear model (GLM) techniques and is charged with developing an associated set of rates, how does one determine surcharges associated with endorsements?

There are several reasonable approaches for addressing this question. One approach is that endorsements form a relatively small fraction of the premium base and so only informal, ad hoc approaches are needed. Actuaries, of course, typically have substantial amounts of experience when ratemaking and this experience can be a guide to setting rates for such a relatively small part of the business. Another approach is to use information from an external agency for this set of relativities, even if GLM techniques are being using in conjunction with company data for the primary set of rates. A third approach, especially for large companies, is to treat endorsements as merely another type of coverage and use GLM techniques to determine this set of prices. This approach requires a substantial amount of data as well as claims that are identified by type of endorsement.

This paper is motivated by a rating study in which none of these approaches are appropriate. Our work makes three contributions. First, we consider the Wisconsin Local Government Property Insurance Fund and describe a process for determining intuitively appealing rates, for a political environment, based on GLM techniques. Second, we provide a detailed case study, so that other analysts may replicate parts of our approach. Through our use of GLM techniques, we provide relativities not only for our primary rating variables, but also for endorsements in a case

when it is not known whether or not a claim is due to an endorsement. Third, we explore the use of shrinkage estimation in ratemaking, and demonstrate that little predictive ability is lost when the base rating variables are left stable.

1.1. Fund description

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, including counties, cities, towns, villages, school districts, and library boards. The fund insures local government property, such as government buildings, schools, libraries, and motor vehicles. The fund covers all property losses except those resulting from flood, earthquake, wear and tear, extremes in temperature, mold, war, nuclear reactions, and embezzlement or theft by an employee.

The fund covers over a thousand local government entities who pay approximately $25 million in premiums each year and receive insurance coverage of about $75 billion. State government buildings are not covered; the LGPIF is for local government entities that have separate budgetary responsibilities and who need insurance to moderate the budget effects of uncertain insurable events. Claims for state government buildings are charged to another state fund that essentially self-insures its properties.

The fund offers three major groups of insurance coverage: building and contents (BC), inland marine (construction equipment), and motor vehicles. For this paper, we focus on BC, as this was the primary motivation for developing the fund; coverage for local government property has been made available by the State of Wisconsin since 1911. However, even within this primary coverage, there are many optional coverages offered, including business interruption and fine arts endorsements.

In effect, the LGPIF acts as a stand-alone insurance company, charging premiums to each local government entity (policyholder) and paying claims when appropriate. Although the LGPIF is not permit-

ted to deny coverage for local government entities, these entities may go onto the open market to secure coverage. Thus, the LGPIF acts as a "residual" market to a certain extent, meaning that other sources of market data may not reflect its experience.

1.2. Determining effective relativities

Although it is government insurance, because the LGPIF essentially acts as a stand-alone insurance company, many of its goals are similar to those of a private insurer. An analysis of LGPIF claims serves as important input for determining rates that the LGPIF charges its policyholders; these rates should reflect the appropriate level and amount of uncertainty of an insurance coverage. Particularly for a public entity such as the LGPIF, the ratemaking process should be transparent and seek to promote equity among policyholders.

Because the LGPIF has a moderate amount of exposure, as will be seen, there is little difficulty in using commonly accepted generalized linear modeling (GLM) techniques to determine rates that are unbiased and transparent for the primary building and contents coverage. However, the usual approaches for handling endorsements were deemed less than satisfactory for three reasons. First, as of this writing (2014–2015), the LGPIF is undergoing a major rate restructuring; due to the political environment, seemingly ad hoc adjustments, even if small, are deemed inappropriate. Second, information from external agencies is expensive and not particularly relevant; the LGPIF is a government entity and acts as a residual market, meaning that there is limited information on comparable risk pools. (See the Association of Government Risk Pools, http://www.agrip.org/, for one set of possible comparables.) Third, LGPIF data for optional coverages is limited, implying that the usual GLM techniques are not suitable for rating the optional coverages, such as endorsements.

To rate endorsements, this paper explores the use of GLM techniques with restrictions on the coefficients through shrinkage using well-known penalized likelihood methods (cf., Brockett, Chuang, and

Pitaktong 2014). Analysts have vague knowledge and impressions about the size and magnitude of these coefficients, stemming from business practice, economic theory, and an understanding of general risk management practice. For example, if x is a binary variable representing the adoption of an alarm system (an "alarm credit"), then the analyst expects the associated coefficient to be negative in the neighborhood of 0 to -10%. That is, if a policyholder manages risk appropriately by introducing alarms, then resulting rates should be at least as low as without the adoption of an alarm system. Estimated alarm credit regression coefficients that are positive are not acceptable for rating purposes.

Compared to the traditional methods of simply including endorsements after the primary analysis has been done, our approach has two main advantages. First, we can use the data to suggest ways of introducing relativities for endorsements in a disciplined manner. Second, because we use GLM techniques, our approach is naturally multivariate and the introduction of endorsements accounts for the presence of other rating variables. Further, as we will see, the shrinkage methods used in this paper have the flexibility to also be used in other situations where the analyst wishes to moderate the effect of unreliable data.

The plan for the rest of the paper follows. We begin in Section 2 by giving more information about the data from the LGPIF as used in this study. Section 3 describes the shrinkage estimation techniques. Sections 4 and 5 describe the results of the model fitting from in-sample and out-of-sample perspectives, respectively. Section 6 provides concluding remarks, and alternative analyses are in the Appendix Section 7.

2. Data

2.1. Fund claims and rating variables

Building and contents is the fundamental coverage underpinning the LGPIF and is the focus of this paper. The claims may be a damage to the base property, content, or other properties covered by endorsements

Table 1. Building and contents claims summary

Year	Average Frequency	Average Severity	Average Coverage	Number of Policyholders
2006	0.951	9,695	32,498,186	1,154
2007	1.167	6,544	35,275,949	1,138
2008	0.974	5,311	37,267,485	1,125
2009	1.219	4,572	40,355,382	1,112
2010	1.241	20,452	41,242,070	1,110
2011	1.036	7,869	42,503,989	1,094

purchased by the policyholder. Hence, the observed claim amounts may vary according to specific terms of the endorsements, selected and purchased by the policyholder. The observed amounts reflect the total end result of each claim; however, the specific contribution by the endorsement is unobserved. Summary statistics of the data show that the average claim varies widely, especially with a high 2010 value due to a single large claim. The total number of policyholders is steadily declining and, conversely, the coverage is steadily increasing. Throughout this section, we summarize the distribution of average severity for policyholders; that is, for each policyholder, we examine total severity divided by the number of claims, i.e., the pure premium or loss cost. In our modeling sections, we appropriately weight by numbers of claims.

Table 1 shows policies beginning in 2006 because there was a shift in claim coding in 2005 so that comparisons with earlier years are not helpful. To mitigate the effect of open claims, we consider policy years prior to 2012. This means we have six years of data, years 2006, . . . , 2011, inclusive. We use a common strategy in predictive modeling where we split our data into a "training" and a "validation" sample.

Specifically, we use years 2006–2010 inclusive (the training sample) to develop our rating factors. Then we apply these factors and 2011 rating variables to predict 2011 claims (the validation sample). Thus, henceforth our summary statistics refer to the 2006–2010 training data. Appendix 7.4 provides an alternative cross-sectional cross-validation.

For the training sample, Table 2 summarizes the distribution of our two continuous outcomes, frequency and claims amount. It is not surprising that the two distributions are right-skewed and correlated with one another. In addition, the table summarizes our continuous rating variables, (building and contents) coverage, and deductible amount. The table also suggests that these variables have right-skewed distributions. Moreover, they will turn out to be useful for predicting claims, as suggested by the positive correlations in Table 2 for coverage and deductible. We use a non-parametric (also known as "Spearman") correlation due to the skewness of the data and the presence of zeros.

Table 3 describes the rating variables considered in this paper. To handle the skewness, we will henceforth focus on logarithmic transformations of cover-

Table 2. Summary of claim frequency and severity, deductibles, and coverages

	Minimum	Median	Average	Maximum	Spearman Correlation	
					Frequency	Claim*
Claim Frequency	0	0	1.109	263	—	–0.065
Claim Severity	0	0	9,292	12,922,21	–0.065	—
Deductible	500	1,000	3,365	100,00	0.041	0.324
Coverage (000's)	8.937	11,354	37,281	2,444,797	0.406	0.243

*The claim correlations are based on 1,679 observations with at least one claim using the average claim (amount divided by frequency).

Table 3. Description of base rating variables

Variable	Description
EntityType	Categorical variable that is one of six types: (Village, City, County, Misc, School, or Town)
LnCoverage	Total building and content coverage, in logarithmic millions of dollars
LnDeduct	Deductible, in logarithmic dollars
NoClaimCredit	Binary variable to indicate no claims in the past two years
Fire5	Binary variable to indicate the fire class is below 5. (The range of fire class is 0:10)

age and deductibles. To get a sense of the relationship between the noncontinuous rating variables and claims, Table 4 relates the claims outcomes to these categorical variables. Table 4 suggests substantial variation in the claim frequency and average severity of the claims by entity type. It also demonstrates higher frequency and severity for the Fire5 variable and the reverse for the NoClaimCredit variable. The relationship for the Fire5 variable is counterintuitive in that one would expect lower claim amounts for those policyholders in areas with better public protection (when the protection code is five or less). Naturally, there are other variables that influence this relationship. We will see that these background variables are accounted for in the subsequent multivariate regression analysis, which yields an intuitive, appealing (negative) sign for the Fire5 variable.

The Appendix (Table 20) shows the claims experience by alarm credit. It underscores the difficulty of examining variables individually. For example, when looking at the experience for all entities, we see that policyholders with no alarm credit have on average lower frequency and severity than policyholders with the highest (15%, with 24/7 monitoring by a fire station or security company) alarm credit. In particular, when we look at the entity type School, the frequency is 0.422 and the severity 25,257 for no alarm credit, whereas for the highest alarm level it is 2.008 and 85,140. This may simply imply that entities with more claims are the ones that are likely to have an alarm system. Summary tables do not examine multivariate effects; for example, Table 4 ignores the effect of size (as we measure through coverage amounts) that affect claims.

2.2. Endorsements

As described in Section 2.1, we do not actually observe claims from an endorsement. For example,

Table 4. Claims summary by entity type, fire class, and no claim credit

Variable	Number of Policies	Claim Frequency	Average Severity
		EntityType	
Village	1,341	0.452	10,645
City	793	1.941	16,924
County	328	4.899	15,453
Misc	609	0.186	43,036
School	1,597	1.434	64,346
Town	971	0.103	19,831
Fire5=0	2,508	0.502	13,935
Fire5=1	3,131	1.596	41,421
NoClaimCredit=0	3,786	1.501	31,365
NoClaimCredit=1	1,853	0.310	30,499
Total	5,639	1.109	31,206

Table 5. Description of endorsements

Variable	Description
Business Interruption	Reimburses an insured for business interruption (lost profits and continuing fixed expenses).
Accounts Receivable	Adds coverage for money owed by its debtors during business interruption due to a covered loss.
Pier and Wharf	Loss of watercraft, by the pressure of ice or water on piers and wharves
Fine Arts	Adds coverage (agreed value) on fine arts, either per item or per exhibit
Golf Course Grounds	Adds coverage to golf course type property such as greens, tees, fairways, etc.
Special Use Animal	Adds coverage for police enforcement animals, such as dogs and horses
Zoo Animals	Adds coverage for zoo animals. Animal mortality is specifically excluded.
Vacancy Permit	Allows claims from covered losses arising from vacant property
Monies and Securities	Adds coverage for monies and securities for loss by theft, disappearance, or destruction (A: loss inside premise, B: loss outside premise).
Monies and Securities (limited term)	Adds limited term coverage for monies and securities
Other Endorsements	Other additional endorsements, including ordinance & law, and extra expenses

if a policyholder purchases a Golf Course Grounds endorsement and has a claim that is from this additional coverage, we are not able to observe this connection with our data. We do observe the additional claim, whether the policyholder has the endorsement, and the amount of coverage under the endorsement. In this sense, endorsements can be treated as another rating variable in our algorithms.

Table 5 describes endorsements, or optional coverages, that are available to LGPIF policyholders. Table 6 summarizes the claims experience by endorsement. Policyholders with the Zoo Animals endorsement experience an average annual claim frequency of 73.9. Presumably, policyholders paying for this extra protection would enjoy higher property claims and so should be charged additional premiums. The

Table 6. Summary of claim frequency and severity by endorsement

Endorsements	Number of Observations	Claim Frequency	Average Severity	Average Endorsement Coverage	Spearman Correlation of Coverage with Frequency	Severity*
Business Interruption	225	6.427	48,612	2,679,595	0.392	0.249
Accounts Receivable	172	5.285	29,743	853,966	0.188	0.097
Pier and Wharf	312	2.510	24,649	245,445	0.067	0.083
Fine Arts	67	13.493	37,896	12,160,956	0.297	0.187
Golf Course Grounds	28	18.000	20,866	237,500	0.749	0.166
Zoo Animals	10	73.900	18,554	1,102,790	0.877	0.462
Special Use Animal	256	5.547	13,127	21,903	0.168	0.073
Vacancy Permit	225	4.902	21,232	1,779,212	0.053	0.316
Monies and Securities (A,B)	2,137	2.000	29,999	58,928	0.255	0.137
Monies and Securities (limited term)	556	1.739	19,811	416,587	0.143	0.091
Other Endorsements	53	4.906	28,245	4,763,019	−0.003	0.334
Total (*All Policies*)	5,639	1.109	17,287			

*The severity correlations are based on observations with at least one claim using the average severity (amount divided by frequency).

most frequently subscribed endorsement is the Monies & Securities, which covers monetary losses by theft, disappearance, or destruction. The average coverage and number of observations are over five years (2006–2010), the in-sample period. For example, the Zoo Animals coverage consists of 10 observations over five years and these were from the Henry Vilas Zoo in Dane County and the Milwaukee County Zoo in Milwaukee County.

Table 6 shows that a policyholder with any type of endorsement has a higher claims frequency compared to the total of all policyholders. Similarly, for most endorsements, policyholders have a higher average severity, with Pier and Wharf, Monies and Securities (limited term), and Other Endorsements being the exceptions. The effect of higher severity seems to be particularly large for certain endorsements, such as Zoo Animals, Golf Course Grounds, and Fine Arts.

To help establish the relationship between endorsements and claims outcomes, Table 6 also shows the average endorsement coverage (the average is over policyholders with some positive coverage). The table summarizes the Spearman correlation of the amount of endorsement coverage versus the frequency and severity of claims observed. It is not surprising that all of these correlations are positive, indicating that more coverage means both a higher frequency and severity of claims. In keeping with our frequency-severity approach to modeling, note that the claims severity correlations are calculated for observations with at least one claim.

3. Claims modeling

As described in Section 1, this paper uses generalized linear models, and following industry norms, employs logarithmic link functions that result in multiplicative relativities. We investigated both the frequency-severity approach as well as the Tweedie ("pure premium") approach. Both models have strengths and weaknesses and, for our data set, predict claims on our holdout sample roughly equally well. See Frees (2014) for a comparison

of these two modeling approaches. This section describes estimation techniques employed and model specifications.

3.1. Shrinkage estimation

3.1.1. Linear model shrinkage

To introduce shrinkage estimation, we first provide a review in the context of the linear model; see, for example, Hastie, Tibshirani, and Friedman (2009) for further information. For notation, assume that y_i is the dependent variable and that $x_{i1}, \ldots, x_{ik}$ is the set of covariates (including coefficients for rating variables and endorsements). Then, the set of shrinkage estimators of $\beta = (\beta_0, \ldots, \beta_k)'$ is determined by minimizing

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{k} x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{k}\beta_j^2. \tag{3.1}$$

Values of λ control the complexity of the model; smaller values mean less shrinkage. At one extreme, a value of $\lambda = 0$ reduces to ordinary least squares. At the other extreme, as λ approaches infinity, β approaches (or, is "shrunk towards") 0, so the data becomes less relevant (has smaller weight) in determining values of β. Note that in equation (3.1) the intercept β_0 is typically not included, as this would make the procedure dependent on the origin of y; to illustrate, subtracting 250 (for example, for a deductible) from each value of y would substantially alter results.

Equivalent to equation (3.1), one could also determine β by minimizing the sum of squares

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{k} x_{ij}\beta_j\right)^2$$

but subject to a constraint of the form $\sum_{j=1}^{k}\beta_j^2 < c$. This formulation is desirable in that one can directly see how the β coefficients are being "shrunk" towards zero (as c becomes small). Thus, shrinkage estimation is a desirable intermediate device between (i) leaving a coefficient in the equation and (ii) removing it completely. Through shrinkage, we can include a rating

variable but shrink its coefficient and hence reduce its effect on the predicted values.

After centering the *y*'s and *x*'s, we can also write the shrinkage estimators in the form

$$\hat{\beta}_{shrink} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}. \qquad (3.2)$$

This equation has two appealing interpretations. First, even in the case when some of the rating variables are collinear so that $\mathbf{X}'\mathbf{X}$ is no longer invertible, the matrix $\mathbf{X}'\mathbf{X} + \lambda\mathbf{I}$ is invertible. Equation (3.2) was first known as a type of "ridge regression" to handle problems of collinearity.

Second, assuming normality of the outcomes and the regression coefficients, one can show that $\hat{\beta}_{shrink}$ represents the posterior mean of β. Through this Bayesian context, one can think about the coefficients β having a distribution (centered about 0) and the analyst is allowed to incorporate his or her belief about the precision of the coefficient through a prior distribution.

3.1.2. Generalized linear model shrinkage

More generally, coefficients may be shrunk towards selected (possibly non-zero) values and we need not shrink all of them. In keeping with common statistical notation, we will make the term $\|\mathbf{R}\beta - \mathbf{r}\|^2$ small, where $\mathbf{R}\beta$ represents sets of linear combinations of regression parameters ($\mathbf{R}$ is known) and $\mathbf{r}$ represents a vector of selected values.

For generalized linear models, the idea behind shrinkage estimation is to make a logarithmic likelihood large subject to requiring $\|\mathbf{R}\beta - \mathbf{r}\|^2$ to be small. This naturally leads to the notion of a penalized likelihood of the form,

$$l(\beta) = \sum_{i=1}^{n} \log f(y_i) - \lambda\|\mathbf{R}\beta - \mathbf{r}\|^2,$$

where $f(\cdot)$ is a density or mass function. For example, for a Poisson distribution with mean $\mu_i = \exp(\mathbf{x}'_i\beta)$, we have $f_i(y_i) = \mu_i^{y_i} e^{-\mu_i}/y_i!$ and

$$l(\beta) = \sum_{i=1}^{n}\{y_i\mathbf{x}'_i\beta - \exp(\mathbf{x}'_i\beta) - \ln(y_i!)\} - \lambda\|\mathbf{R}\beta - \mathbf{r}\|^2.$$

For the application in this paper, we have

$$\mathbf{R} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad \mathbf{r} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

β_1 = coefficients for base rating variables
β_2 = coefficients for endorsements

The shrinkage approach can be understood as a special case of constrained estimation, where the coefficient β is restricted to be within a neighborhood of $\mathbf{r}$. By varying the shape of the constraint region, it is possible to obtain various properties of the resulting coefficient. We provide more details in Appendix Section 14 for the interested reader.

3.2. Offsets and endorsements

Variables described in Tables 3 and 5 were used to calibrate generalized linear models with logarithmic links and estimation methods outlined in Section 3.1. We also used the following offset variable

$$offset = \ln(0.95)AC05 + \ln(0.90)AC10 \\ + \ln(0.85)AC15.$$

Here, AC05 represents a binary variable to indicate the presence of a 5% alarm system, and similarly for AC10 and AC15. This seems a sound practice and so we retain this offset variable in our analysis.

When the LGPIF began capturing alarm system data, premium "credits" in the amount of 5% were given to those with AC05=1, and similarly for the other two categories. Alarm systems at the 5% level mean that automatic smoke alarms exist in some of the main rooms, those at the 10% level mean they exist in all of the main rooms. At the 15% level, facilities are monitored on a 24 hours per day, 7 days per week basis by a police, fire, or security company. The policyholder is eligible for a premium credit of an amount determined by the specified percentage, depending on the alarm credit amount. This section describes how alarm credit is used as an offset in our model. Table 20, in the Appendix, shows a summary of the claims with respect to different alarm credit categories.

Table 6 suggests that not only the presence of an endorsement but also its coverage amount may influence claims outcomes. To capture this, suppose that y_B represents claims from a base coverage with mean $\mu_B = \exp(\mathbf{x}'\boldsymbol{\beta})$. Let y_E be the claims from an endorsement that we assume has mean μ_E. Then, the observed response y has the following mean structure

$$\mu = Ey = \begin{cases} \mu_B = \exp(\mathbf{x}'\boldsymbol{\beta}) \\ \quad \text{endorsement not present} \\ \mu_B + \mu_E = \exp(\mathbf{x}'\boldsymbol{\beta} + \beta_E x_E) \\ \quad \text{endorsement present} \end{cases}.$$

We can readily accommodate this in a GLM structure using an interaction term of the presence of an endorsement with the variable x_E. We use

$$x_E = \ln\left(1 + \frac{Coverage_E}{Coverage_B}\right), \tag{3.3}$$

where $Coverage_E$ and $Coverage_B$ represent amount of coverage for the endorsement and base (building and contents), respectively. With this specification, we have

$$\mu_E = \exp(\mathbf{x}'\boldsymbol{\beta} + \beta_E x_E) - \mu_B$$

$$= \mu_B\left[\left(1 + \frac{Coverage_E}{Coverage_B}\right)^{\beta_E} - 1\right]$$

$$\approx \mu_B\left[\left(1 + \beta_E \frac{Coverage_E}{Coverage_B}\right) - 1\right]$$

$$= \beta_E \times Coverage_E \times \left(\frac{\mu_B}{Coverage_B}\right),$$

using the approximation $(1 + z)^b \approx 1 + bz$. With this, we may think of the appropriate cost of the endorsement μ_E as a factor times the endorsement coverage, rescaled by the overall cost per unit coverage. The factor, β_E, is estimated from the data.

For our data, some of the estimated coefficients associated with endorsement variables were insignificant and negative, making them unacceptable for rating purposes (this would mean that the policy-holder electing the endorsement coverage would pay less premiums than otherwise). In particular, LnAcc RecCovRat, LnPierWharfCovRat, and LnMoney SecCovRat were insignificant and negative, when included in the frequency model. One way to rate these variables is to include them in the severity model as covariates. An alternative is to include a binary variable to indicate having the endorsement, instead of using the log coverage ratios. Hence, we elect to use three indicator variables, AccRec, PierWharf, and MoneySec, in the frequency model.

In our model, another offset was used for Vacancy-Permit. In part, this was because interpretable coefficients could not be obtained for this endorsement variable from the given data, even when included as an indicator and shrinkage applied. Moreover, we had available prior information on the impact of this endorsement from historical precedence where the rate for VacancyPermit had been 0.4 times building rate. Therefore,

$$offset_{VP} = 0.4 \times \ln\left(1 + \frac{Coverage_{VP}}{Coverage_B}\right)$$

was added as an additional offset in the model.

3.3. Advantages of the shrinkage approach

The shrinkage approach provides a framework for controlling the coefficients of the endorsements, restricting them to be small, yet meaningful, values. Using a standard GLM, data-driven approach for rating endorsements can result in coefficients that cannot be interpreted in a meaningful way. For instance, the data may indicate that purchasing ZooAnimals coverage amounts to a seven-fold increase in premium. Applying shrinkage only to endorsements allows the base rating variables to remain at an actuarially fair level, while unreasonable behaviors of the endorsement coefficients are contained.

This approach is simple and flexible, and prevents the endorsement premiums from becoming unfair for those who hold only particular endorsements. For example, charging too much for ZooAnimals may

result in unfair premiums for Milwaukee and Dane County, as these two policyholders are the ones who happen to have a public zoo endorsement coverage. In addition, the method allows for a sound risk management practice in a political setting, as the tuning parameter λ may be selected to accommodate the expectations of the environment in which the relativities are to be used. When the expectations regarding contributions from the endorsements are high, then the tuning parameter may be released to allow for an elevated premium level for the endorsements, while they could be shrunk to a small but still meaningful level when the contributions must remain low. In this process, the base rating variables remain stable, and hence ensure a steady outsample performance.

4. Results from the claims modeling

This section presents results using the frequency-severity approach as it provides more intuitive expressions for our parameter estimates. For comparison, we include the Tweedie model results in the Appendix Section 7.2 using the shrinkage approach with ridge penalty.

4.1. Frequency-severity modeling using shrinkage estimation

Table 7 provides fitting results for claims frequency, using the Poisson model. We incorporated base variables described in Table 3, and selected interaction terms and the offset variables described in Section 3.2. Estimation was conducted using shrinkage techniques in Section 3.1 but shrinking only the endorsement terms, not the base rating variables. For example, in Table 7, the covariate LnBusInterCovRat represents the business interruption endorsement variable given in equation (3.3). Parameter estimates for various values of the shrinkage parameter λ are given. Note that even though our shrinkage focuses on the endorsement variables, parameter estimates for other variables are affected due to the multivariate nature of the regression model.

Table 7 shows that Deductible and the interaction term between NoClaimCredit and lnCoverage dis-

play negative coefficients, as anticipated. It is notable that Fire5 also shows a negative coefficient, in contrast to the relationship suggested by the summary statistics in Table 4. This result is sensible, given that a low fire class represents higher public protection. Also, as anticipated, the coefficients for the endorsements are all positive and significant. The model is estimated with λ increasing from 0 (no shrinkage) to 1,000 (shrinkage). As λ increases, we observe the coefficients shrink towards zero.

Table 8 provides fitting results for claims severity, using the gamma model. Specifically, we used a logarithmic link function with the average claim as the dependent variable and the number of claims as the weight; cf. Frees (2014) for further discussions of this specification. As is common in severity modeling, there were fewer variables that were statistically significant when compared to the frequency model and so the model specification is much simpler. The coefficient for LnCoverage is negative; however, the coefficient in the frequency model is positive, and hence the overall effect is positive and interpretable. As shown in Table 4, cities and counties tend to have smaller average severities, and presumably the effect is due to such entities.

4.2. Parameter interpretation

The parameter estimates provided in Section 4.1 necessarily reflect the complexity of the system. To help interpret them, in this section we focus on a "typical" policyholder whose coverage is at the median of the distribution.

For our dataset, the median (50th percentile) BC coverage was $11.35 million, corresponding to 2.43 (= ln 11.353,57 million) as shown in Table 9. Recall that LnCoverage is the total building and content coverage, in logarithmic millions of dollars.

Using this median coverage, Table 10 provides relativities for the rating factors. Table 10 shows that the entity "School" pays less, and that "City," "County," "Misc," and "Town" pay more, all relative to the reference category "Village." As we apply shrinkage to the endorsements, the relativity estimate for each

Table 7. Poisson frequency model using shrinkage estimation

	λ = 0		λ = 5		λ = 500		λ = 1,000	
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
	Base Rating Variables							
(Intercept)	−1.593	0.127	−1.581	0.127	−1.687	0.127	−1.705	0.127
LnCoverage	0.675	0.046	0.678	0.046	0.712	0.046	0.719	0.046
LnDeduct	−0.061	0.011	−0.063	0.011	−0.046	0.010	−0.042	0.010
TypeCity	0.044	0.144	0.046	0.144	0.105	0.143	0.113	0.143
TypeCounty	−0.445	0.199	−0.565	0.194	−1.468	0.171	−1.632	0.165
TypeMisc	−0.173	0.173	−0.142	0.171	−0.083	0.171	−0.076	0.171
TypeSchool	−5.915	0.173	−5.913	0.173	−5.906	0.174	−5.916	0.174
TypeTown	−0.466	0.160	−0.464	0.160	−0.434	0.162	−0.427	0.162
Fire5=1	−0.179	0.037	−0.182	0.037	−0.182	0.037	−0.182	0.037
NoClaimCredit=1	−0.005	0.105	−0.003	0.105	−0.006	0.105	−0.005	0.105
	Interaction Terms							
LnCoverage* TypeCity	0.040	0.049	0.039	0.049	0.022	0.050	0.019	0.050
LnCoverage*TypeCounty	0.130	0.056	0.158	0.055	0.355	0.051	0.390	0.050
LnCoverage*TypeMisc	−0.251	0.059	−0.258	0.059	−0.281	0.059	−0.287	0.059
LnCoverage*TypeSchool	1.245	0.051	1.245	0.051	1.215	0.052	1.212	0.052
LnCoverage*TypeTown	0.158	0.092	0.156	0.092	0.154	0.093	0.152	0.093
LnCoverage*NoClaimCredit	−0.185	0.026	−0.185	0.026	−0.184	0.026	−0.185	0.026
	Endorsements							
LnBusInterCovRat	0.236	0.047	0.234	0.046	0.072	0.027	0.041	0.020
LnSpecialAnimalCovRat	0.055	0.705	0.009	0.289	0.001	0.032	0.001	0.022
LnZooAnimalCovRat	1.977	0.765	0.259	0.292	0.007	0.032	0.004	0.022
LnFineArtsCovRat	0.298	0.057	0.332	0.053	0.107	0.027	0.064	0.021
LnGolfCourseCovRat	0.915	0.359	0.225	0.274	0.003	0.032	0.001	0.022
LnOtherCovRat	0.300	0.046	0.293	0.045	0.078	0.027	0.044	0.021
	Endorsement Indicators							
AccRec	0.427	0.059	0.421	0.057	0.118	0.028	0.068	0.021
PierWharf	0.407	0.043	0.411	0.042	0.125	0.026	0.073	0.020
MoneySec	0.181	0.031	0.178	0.030	0.113	0.022	0.079	0.018
−2 Log L	−7,393		−7,380		−7,179		−7,190	

entity type is smoothed, reflecting the change in the relativity estimates for the endorsements.

Note that the relativity of School is very small in comparison to other entity types; recall that relativities, like regression coefficients, summarize marginal changes in variables and may not capture all relevant data features. In this case, although 2.43 (ln11.353,57) is the median, it is only at the 11th percentile for

Schools. So, if this example were focused on schools, then we would use a higher coverage amount to reflect the typical school coverage.

Table 10 also shows the relativity estimates for the three endorsement indicators. Note that AccRec, PierWharf, and MoneySec are used as indicators, while other endorsements are used as log coverage ratios in the frequency model. Because we have

Table 8. Gamma severity model for average claim

	Estimate	Standard Error
Base Rating Variables		
(Intercept)	9.385	0.122
LnCoverage	−0.130	0.026
TypeCity	0.474	0.139
TypeCounty	1.107	0.159
TypeMisc	1.656	0.274
TypeSchool	1.038	0.146
TypeTown	0.293	0.290
φ (dispersion)	7.119	

Table 9. Coverage quantiles

Percent	10%	25%	50%	75%	90%	95%
LnCoverage	−0.704	0.785	2.430	3.606	4.487	4.943

fold (7.220) increase in premium without shrinkage, while the effect is significantly mitigated after shrinkage is applied (1.296 with $\lambda = 5$ and as small as 1.004 with $\lambda = 1,000$). The effect of having Golf-CourseCov results in a nearly three-fold (2.497) increase in premium without shrinkage, while the effect is mitigated to 1.252 with $\lambda = 5$ and as small as 1.001 with $\lambda = 1,000$.

A rating engine may be recommended using the relativities shown in Table 10. The final recommendation to the property fund consists of tabulated rating factors, which can be applied to the base premium in a multiplicative manner. The endorsement factors are then applied additively.

not applied shrinkage to the severity model, as λ is increased, the severity model remains the same. The final relativity estimate is obtained by multiplying the exponentiated estimates from the frequency model and the severity model. The reader may observe that having, say, ZooAnimalCov results in a seven-

Table 10. Relativities for base rating variables and endorsements

	$\lambda = 0$	$\lambda = 5$	$\lambda = 500$	$\lambda = 1,000$
Base Rating Variables				
LnCoverage	1.726	1.730	1.789	1.804
LnDeduct	0.941	0.939	0.955	0.959
TypeCity	1.848	1.846	1.881	1.883
TypeCounty	2.658	2.525	1.651	1.527
TypeMisc	2.395	2.431	2.436	2.419
TypeSchool	0.157	0.157	0.147	0.145
TypeTown	1.235	1.231	1.262	1.264
Fire5=1	0.836	0.833	0.833	0.834
NoClaimCredit=1	0.635	0.637	0.636	0.635
Endorsements				
LnBusInterCovRat	1.266	1.263	1.075	1.042
LnAddInsCovRat	1.350	1.340	1.081	1.045
LnSpecialAnimalCovRat	1.056	1.009	1.001	1.001
LnZooAnimalCovRat	7.220	1.296	1.007	1.004
LnFineArtsCovRat	1.347	1.394	1.113	1.066
LnGolfCourseCovRat	2.497	1.252	1.003	1.001
Endorsement Indicators				
AccRec	1.532	1.523	1.125	1.071
PierWharf	1.502	1.508	1.134	1.075
MoneySec	1.199	1.195	1.120	1.082

Hence, the premium is calculated by the following rating formula:

$$\text{Premium} = (\text{BasePremium}) \times (\text{NoClaimFactor})$$
$$\times (\text{AlarmCreditFactor})$$
$$\times (\text{DeductibleFactor})$$
$$+ (\text{EndorsementRates}).$$

The base premium is tabulated for the six different entities (including the base entity, Village), and two different fire classes. This base premium is adjusted for alarm credit, and a no-claims discount factor is applied depending on the policyholder's experience. The deductible factor is selected and multiplied from a tabulated table of eleven deductible categories. Finally, the endorsement factors are added. Note that we could have also included endorsements multiplicatively based on the discussion in Section 3.2. We chose to use additive terms to be consistent with prior LGPIF practice.

5. Out-of-sample performance

The models described in Section 3 with fitted parameter values in Section 4 provide the basis for developing a rating algorithm. With this information, we can generate predictions based on 2011 (out-of-sample) rating variables. To assess the viability of these predictions, we compare them to 2011 out-of-sample claims. We also have available a Premium variable that was generated by an external agency (based on a very expensive process). For another comparison, we also generated scores for the Tweedie model based on the parameter results in the Appendix. This section compares our predictions with held-out claims and this premium score.

Table 11 reports correlations among scores and claims. For both the frequency-severity and the Tweedie model, there were very strong correlations between the scores from the usual unbiased methods without shrinkage (corresponding to $\lambda = 0$) and shrinkage-based scores (corresponding to $\lambda = 1,000$). Note, from Table 11, the outsample correlation for $\lambda = 1,000$ differs only by a little. Because of this strong relationship for these two extreme values of λ, we do not include scores for intermediate values of λ. Moreover, this means that at least for this data set, little predictive ability is lost by using shrinkage methods to give much more intuitively appealing relativities.

We note the strong correlation, nearly 94.29%, between the external agency Premium and the Tweedie model scores, as shown in Figure 1. This suggests that our analysis is able to reproduce (expensive) external agency scores effectively. Table 11 demonstrates that all three scoring approaches, the frequency-severity, the Tweedie, and the external agency premium score, fare about the same in predicting out of sample claims. The frequency-severity model does the best, while the Tweedie model shows the highest correlation with the external agency scores. Note that our frequency-severity scores outperform the external agency scores by a small amount.

To get a better sense of the meaning of these correlations, Figure 2 shows the relationship between our frequency-severity (two-part model, or TPM) score and held out claims. The left-hand panel shows the relationship in terms of dollars and the right-hand panel gives the same data but using logarithmic

Table 11. Spearman correlations among scores and out of sample claims

	Freq-Sev Model		Tweedie		
	$\lambda = 0$	$\lambda = 1,000$	$\lambda = 0$	$\lambda = 1,000$	Premiums
Frequency-Severity Model, $\lambda = 1,000$	0.9938				
Tweedie Model, $\lambda = 0$	0.7594	0.7601			
Tweedie Model, $\lambda = 1,000$	0.7452	0.7512	0.9825		
Out of Sample Premiums	0.7487	0.7527	0.9429	0.9474	
Out of Sample Claims	0.4330	0.4330	0.4201	0.4154	0.4218

Figure 1. Comparison of frequency-severity model scores and Tweedie model scores to external agency premium scores. The Spearman correlation coefficients are 74.87% and 94.29%
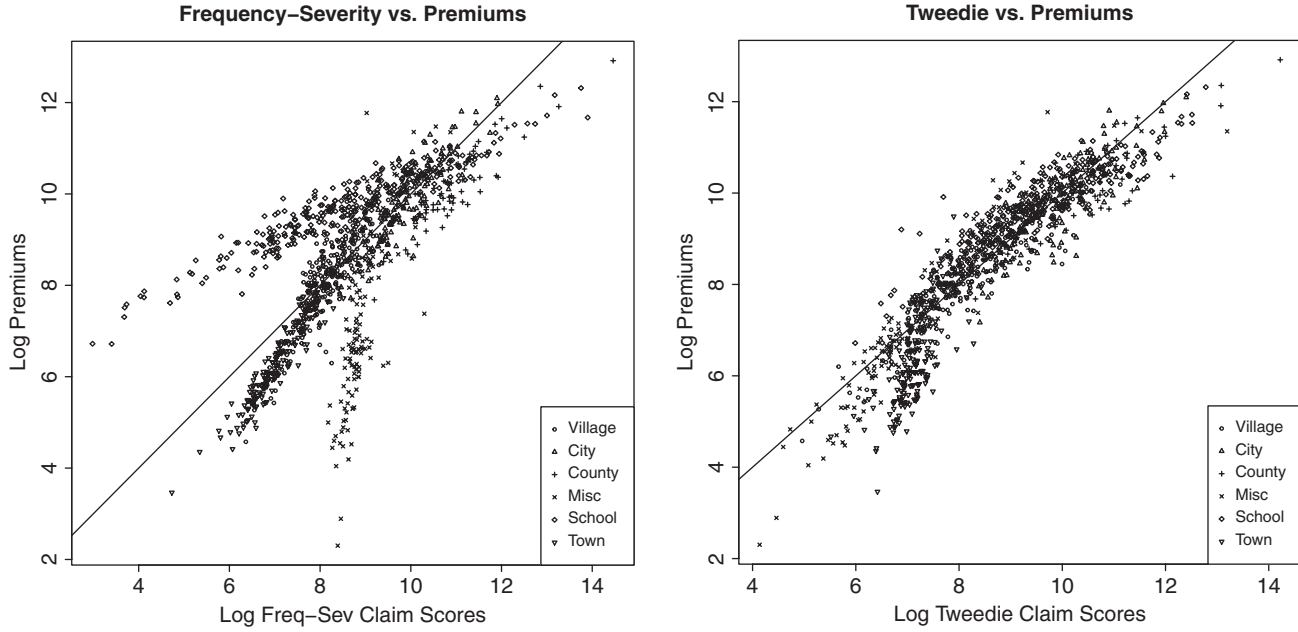

Figure 2. Comparison of frequency-severity scores to out of sample claims for 2011. The Spearman correlation coefficient is 43.30%
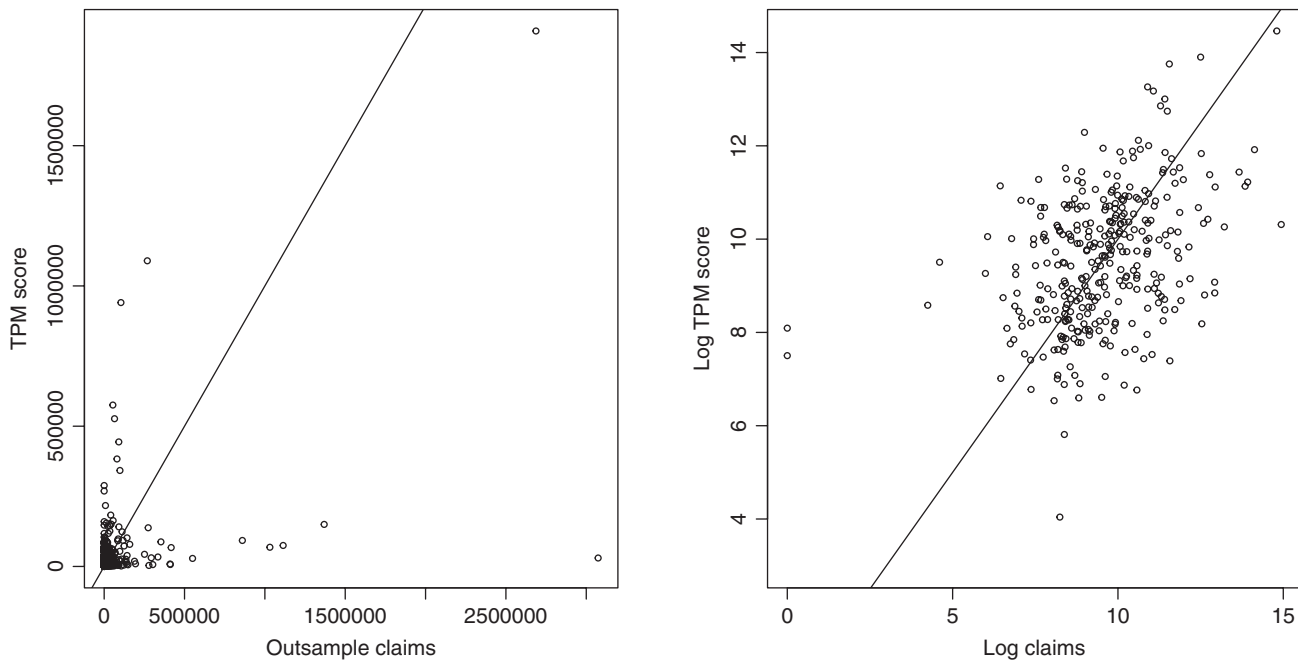
Table 12. Gini indices of predictive claim scores

	Frequency-Severity		Tweedie Model		
	($\lambda = 0$)	($\lambda = 1000$)	($\lambda = 0$)	($\lambda = 1000$)	Premiums
Gini Index	69.66%	69.96%	69.23%	69.77%	72.69%

scaling. For this figure, each plotting symbol corresponds to a policyholder and the overall Spearman correlation is a strong 43.30%.

We believe that our work is fairly typical of analyses of insurance company data. For statistical significance and interpretability of the coefficient estimates for the endorsements, we prefer the frequency-severity approach presented in Section 4.1. However, the Tweedie approach presented in the Appendix uses fewer parameters, and fares evenly when compared to the external agency premium scores. We think both approaches are sensible and the choice will ultimately depend on the actuary who is analyzing and making inferences from the data.

Appendix 7.4 shows an alternative robustness check, using a randomly selected cross-sectional sample of policyholders for out of sample validation. We further check the predictive ability of our claim scores using the Gini index. This is a newer measure developed in Frees, Meyers, and Cummings (2011). For our application, the Gini index is twice the average covariance between the predicted outcome and the rank of the predictor.

Table 12 summarizes these results. By inspecting the Gini indices, we observe only minute differences in the explanatory ability after applying the shrinkage technique. The Gini index is 70.05% using only the base rating variables, 69.66% with the endorsements, and 69.96% using shrinkage estimation. In the same way, the Pure Premium (Tweedie) scores show a 69.74% for the base score, 69.23% with endorsements in the model, and 69.77% using shrinkage estimation. For comparison, the Gini index of the external agency premiums turned out to be 72.69%.

In order to test the significance of the differences among these scores, we use *Theorem 5* of Frees, Meyers and Cummings (2011) which provides standard errors for the difference of two Gini indices. Table 13 shows that the differences among the scores are insignificant. For example, the difference between the frequency-severity score with $\lambda = 1,000$ and the external agency premiums is 0.027; however, the difference is within twice the standard error of the difference statistic. Further, *Corollary 3* in Frees, Meyers and Cummings (2011) established the asymptotic normality of the distribution of the difference

Table 13. Difference in Gini indices among scores. The external agency premiums have a higher Gini index; however, differences are statistically insignificant

	Freq-Sev ($\lambda = 0$)	Freq-Sev ($\lambda = 1,000$)	Tweedie (Base)	Tweedie ($\lambda = 0$)	Tweedie ($\lambda = 1,000$)	Premiums
Freq-Sev (Base)	−0.001 (0.068)	0.002 (0.065)	−0.001 (0.066)	−0.006 (0.064)	−0.000 (0.067)	0.029 (0.057)
Freq-Sev ($\lambda = 0$)		0.003 (0.068)	0.001 (0.069)	−0.004 (0.072)	0.001 (0.069)	0.030 (0.062)
Freq-Sev ($\lambda = 1,000$)			−0.002 (0.068)	−0.007 (0.069)	−0.002 (0.068)	0.027 (0.059)
Tweedie (Base)				−0.005 (0.070)	0.000 (0.039)	0.030 (0.060)
Tweedie ($\lambda = 0$)					0.005 (0.071)	0.035 (0.062)
Tweedie ($\lambda = 1,000$)						0.029 (0.059)

statistic so that we can rely upon the usual normal-based rules for assessing statistical significance.

6. Concluding remarks

There are three main contributions of this paper. First, we have presented a detailed analysis of a government entity, the Wisconsin Local Government Property Insurance Fund. There is little in the literature on government property and casualty actuarial applications and we hope that this application will interest readers. Moreover, the LGPIF is similar to small commercial property insurance, making our work of interest to a broad readership.

Second, we have given a detailed analysis in the manner of a case study so that other analysts may replicate parts of our approach. Specifically, through our use of GLM techniques, we provide relativities not only for our primary rating variables but also for endorsements. We provided an approach for handling these optional coverages when it is not known whether or not a claim is due to an endorsement.

Third, we have explored the use of shrinkage estimation in ratemaking. Although applications can be general, we find them particularly appealing in the case of endorsements. For our data set, we found that little predictive ability was lost by using shrinkage methods and they gave much more intuitively appealing relativities. Particularly in a political environment such as that enjoyed by government insurance, it is helpful to have relativities that can be calibrated in a disciplined manner and are consistent with sound economic, risk management, and actuarial practice.

Acknowledgments

References

Brockett, P. L., S-L Chuang, and U. Pitaktong, "Generalized Additive Models and Non-parametric Regression," in E. W. Frees, G. Meyers, and R. A. Derrig, eds., *Predictive Modeling Applications in Actuarial Science,* Cambridge, MA: Cambridge University Press, 2014.

Dean, G. C., "Generalized Linear Models," in E. W. Frees, G. Meyers, and R. A. Derrig, eds., *Predictive Modeling Applications in Actuarial Science,* Cambridge, MA: Cambridge University Press, 2014.

Fahrmeier, L., and J. Klinger, "Estimating and Testing Generalized Linear Models under Inequality Restrictions," *Statistical Papers* 35(1), 1994, pp. 211–229.

Frees, E. W., "Frequency and Severity Models," in E. W. Frees, G. Meyers, and R. A. Derrig, eds., *Predictive Modeling Applications in Actuarial Science,* Cambridge, MA: Cambridge University Press, 2014.

Frees, E. W., G. Meyers, and D. Cummings, "Summarizing Insurance Scores Using a Gini Index," *Journal of the American Statistical Association* 106, 2011, pp. 1085–1098.

Hastie, T., R. Tibshirani, and J. Friedman, *Elements of Statistical Learning,* vol. 2, New York: Springer, 2009.

Nyquist, H., "Restricted Estimation of Generalized Linear Models," *Journal of the Royal Statistical Society,* Series C, 40(1), 1991, pp. 133–141.

Werner, G., and C. Modlin, *Basic Ratemaking,* Arlington, VA: Casualty Actuarial Society, 2010.

7. Appendices

7.1. Appendix: Estimation details

In this appendix, we briefly introduce recent developments in shrinkage estimation so that the reader can have an intuitive understanding of the various methods available. In the following section, we show results from an alternative model specification using the pure premium approach. We will use this specification to explain why we have chosen the two-part model as our suggested model, and equip the reader with an idea of the pros and cons of each method.

GLM estimation

As described in Dean (2014), estimation of generalized linear models is based on a likelihood function of the form

$$l = \sum_{i=1}^{n} \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right].$$

Table 14. Exponential family distributions summary

Family	$a(\phi)$	$b(\theta)$	$c(y, \phi)$
Poisson	1	e^{θ}	$-\ln(y!)$
Gamma	ϕ	$-\ln(-\theta)$	$\dfrac{1}{\phi}\ln\dfrac{y}{\phi}-\ln y-\Gamma\left(\dfrac{1}{\phi}\right)$

This form of the likelihood function allows one to write a generalized routine for fitting a family of distributions by specifying a family object within statistical software.

Table 14 shows the distributions used in this paper. In a typical application, each response is recorded with a set of explanatory variables $\mathbf{X}_{it}$, so that the linear predictor based on a logarithmic link is given by the exponential $\mu_{it} = \exp(\mathbf{x}'_{it}\boldsymbol{\beta})$. By parameterizing the mean and the dispersion in the log likelihood function, the problem of estimating the coefficient $\boldsymbol{\beta}$ boils down to solving a nonlinear optimization problem. The typical method to solve this problem in the GLM context is by Fisher scoring, which is essentially a type of Newton-Raphson iteration

$$\boldsymbol{\beta}_{(\tau+1)} = \boldsymbol{\beta}_{(\tau)} + \left[\nabla^2 l\right]^{-1} \nabla l.$$

Standard statistical packages for generalized linear models implement this iteration scheme.

For the equality restricted problem, we wish to impose a constraint of the form

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}. \tag{7.1}$$

For estimation under this constraint, one may use the approach in Nyquist (1991), where the equality restricted problem is solved by a modified version of Fisher scoring iterations. For covariate matrix $\mathbf{X}$ and weight matrix $\mathbf{W}$, this formula is

$$\boldsymbol{\beta}_{(\tau+1)} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{z}$$

$$+ (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{R}'\left[\mathbf{R}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}$$

$$\times [\mathbf{r} - \mathbf{R}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{z}],$$

where

$$z_i = \mathbf{x}'_i\boldsymbol{\beta}_{(\tau)} + (y_i - \mu_i)\left.\frac{\partial\eta_i}{\partial\mu_i}\right|_{\boldsymbol{\beta} = \boldsymbol{\beta}_{(\tau)}}.$$

Here, $\eta_i = \mathbf{x}'_i\boldsymbol{\beta}$ is the usual systematic component.

Inequality restricted optimization

In contrast to the traditional equality restrictions in equation (7.1), one may impose inequality restrictions. In this subsection, we provide an overview of how this alternative optimization problem is implemented in statistical packages. This provides an intuition behind the shrinkage estimation used in Section 3. One approach for the inequality restricted problem

$$\mathbf{R}\boldsymbol{\beta} \leq \mathbf{r}$$

has been published by Fahrmeir and Klinger (1994) in the GLM context. With high speed computers available nowadays, the inequality restricted problem may be simpler to solve by directly optimizing the likelihood function.

We may formulate the inequality restricted problem as an optimization problem of the form

$$\text{maximize} \quad l(\boldsymbol{\beta},\phi)$$

$$\text{subject to} \quad f_j(\boldsymbol{\beta},\phi) \leq 0, \quad j = 1,\ldots,m$$

for coefficients $\boldsymbol{\beta}$ and dispersion ϕ, and m constraints. Nowadays, high speed computers provide routines for solving this problem easily. These packages would usually transform the constrained problem into an unconstrained problem, so that we have
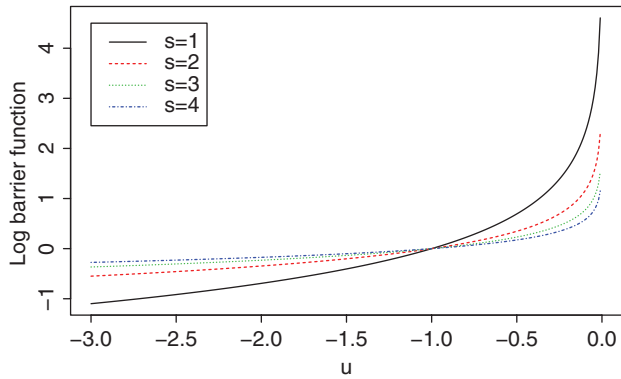
$$\text{maximize} \quad l(\boldsymbol{\beta},\phi) - \sum_{j=1}^{m} I\left(f_j(\boldsymbol{\beta},\phi)\right)$$

with the indicator function

$$I(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0 \end{cases}$$

which blows to infinity when the argument is positive. Because the indicator function is difficult to optimize,

Figure 3. Depiction of the log barrier function

statistical packages usually approximate it with a smooth log barrier function of the following form

$$I(u) \approx -\frac{1}{s}\log(-u)$$

which tends to infinity as $s \to \infty$, cf. Hastie, Tibshirani, and Friedman (2009). The log barrier function is depicted in Figure 3. The method is often called the interior point method, as the iteration begins from an initial value within the constraint region, and continues iterating in a direction optimizing the objective function, without exiting from the interior of the constraint region, as s is increased in each step of the iteration.

For the LGPIF application, the number of constraints m is the number of endorsements. Hence, the constraint functions f_j are

$$\mathbf{R} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} \quad \mathbf{r} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

$$f_j(\boldsymbol{\beta},\phi) = -\left(\mathbf{I}_{jj}[\boldsymbol{\beta}_2]_j - 0\right)^2 = -[\boldsymbol{\beta}_2]_j^2$$

$$j = 1,\ldots,m \quad \text{where} \quad m = \dim(\boldsymbol{\beta}_2)$$

Maximizing the penalized likelihood is hence identical to maximizing the likelihood, and the log barrier function. Maximization of $I(f_j)$ is accomplished by f_j approaching zero, and thus the coefficients $[\boldsymbol{\beta}_2]_j$ approaching zero. The second panel of Figure 4 in the following section illustrates this for $m = 2$, in relation with the alternative more general *elastic net* penalty.

Elastic nets

The *elastic net* has become a popular method among actuarial analysts recently. This method adds a penalty term to the likelihood function, which corresponds to a weighted average of two constraint regions, as shown in Figure 4.

The elastic net maximizes the following objective function

$$l(\boldsymbol{\beta}) - \lambda\left[(1-\alpha)\sum_{i=1}^m \beta_i^2 + \alpha\sum_{i=1}^m |\beta_i|\right] \qquad (7.2)$$

Figure 4. Illustration of the effect of LASSO and Ridge Penalties, for smaller and smaller constraint regions around the origin
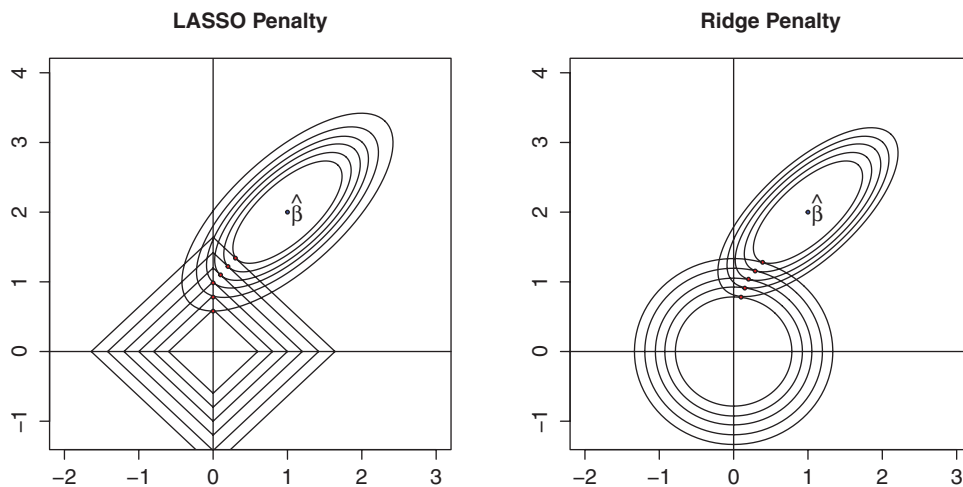
The special case of $\alpha = 0$ corresponds to the classical ridge regression and is the method used in this paper. The closed form equation in Section 3.1 is simply the Lagrangian form of the constrained optimization problem, where λ is the Lagrangian multiplier. As described in Hastie, Tibshirani, and Friedman (2009), there is a one-to one correspondence between the arc length of the constraint region, and the tuning parameter λ.

The analyst may alternatively use the LASSO penalty, which is the special case when $\alpha = 1$. The LASSO penalty is motivated by creating an angular constraint region around the origin, as in the left panel of Figure 4. The advantage of the LASSO penalty is that it induces sparsity in the coefficients, which makes it a favorable choice for variable selection problems. Figure 5 shows the estimation result

Figure 5. Coefficient estimates for the Poisson frequency model, for various tuning parameters using Elastic Net penalty. The top left panel shows the ridge penalty, and the bottom panel shows the LASSO penalty
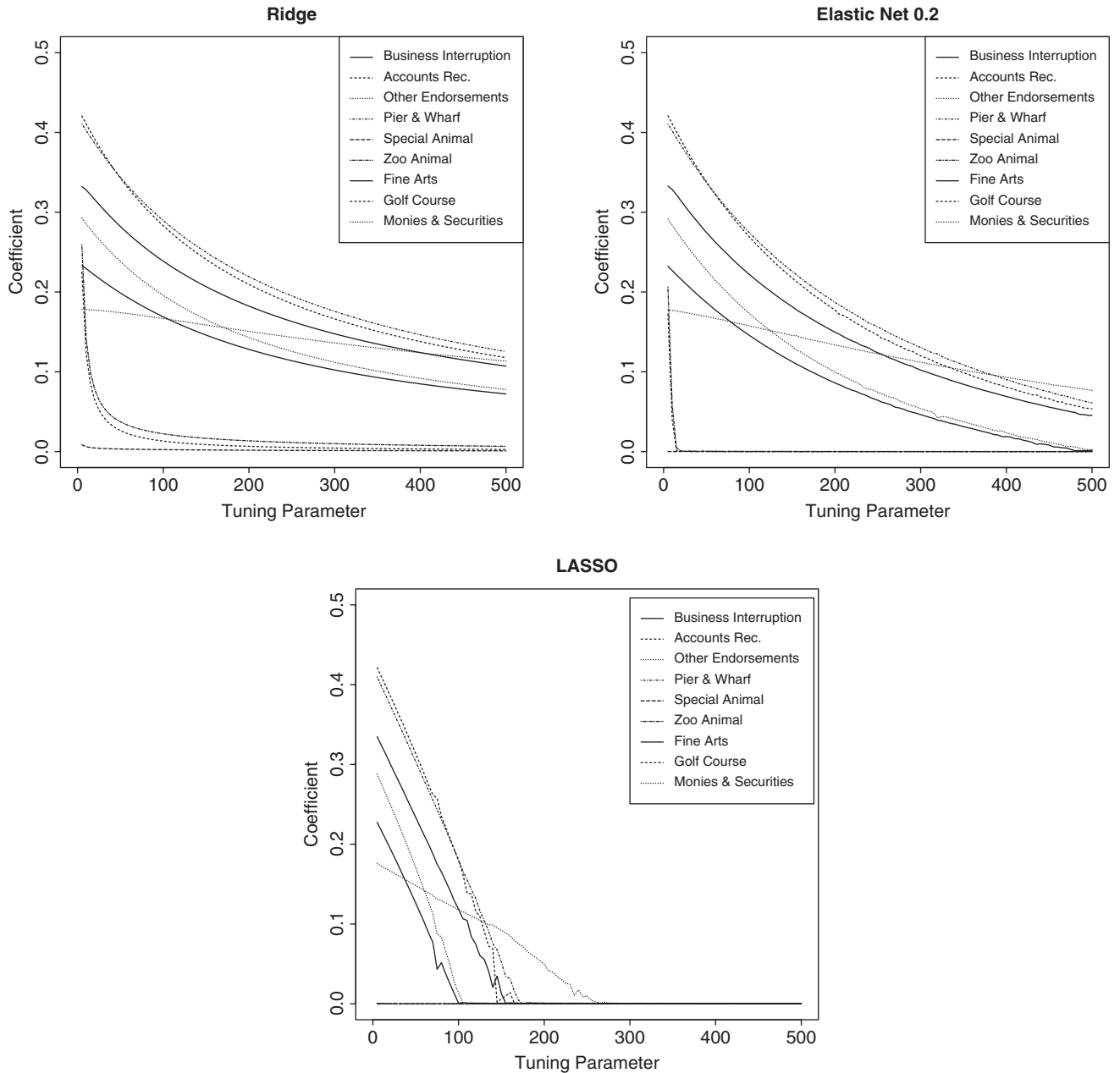
of the endorsements for the LGPIF using different values of α, including the special case when α = 1, where the reader may observe that all of the coefficients except for MoneySec have shrunk to zero as λ reaches 200, and MoneySec also shrunk to zero by the time λ reaches 300. For the LGPIF, we recommended the ridge penalty, as our goal is to provide an interpretable rate for all endorsements, without inducing sparsity.

The elastic net allows the analyst to shrink the coefficients toward zero, by constraining the coefficient to smaller and smaller constraint regions, as the tuning parameter λ is increased. The method allows the analyst to select a weight α to control the shrinkage behavior of coefficients. Figure 5 shows the different convergence behavior, resulting from different weighting parameters.

7.2. Appendix: Pure premium method

In this section, we provide coefficient estimates for the pure premium method. In the pure premium method, the likelihood function written for all claims uses the Tweedie distribution, which has mass at zero. This likelihood is optimized for the coefficients β and dispersion φ. Because the endorsements are allowed to influence the severity of the claims as well as the frequency, we observe substantially different results.

Table 15 shows the estimates and standard errors. When fitting this model, we include the coefficient estimates using the same offset variables as those used in the frequency-severity model. This permits a fair comparison of models for the out of sample comparison described in Section 5. In this section, we see that the claim scores for the pure premium model were generated using the coefficient estimates in Table 15 and these scores show a strong correlation with externally available premiums, using fewer parameters than the frequency-severity approach. This feature is the main strength of the pure premium approach.

The main limitation of the pure premium approach is the interpretation of coefficients. Table 15 shows

that the coefficient estimates for LnSpecialAnimalCovRat, LnOtherCovRat, AccRec and PierWharf are negative without shrinkage, and are therefore difficult to interpret. When shrinkage is applied some of the coefficients turn positive, however LnOtherCovRat and MoneySec remain negative. We experimented with various combinations of covariates, and could not achieve interpretable coefficients. Because of this limitation, in the main body of the report, we focus on the frequency-severity approach. The additional flexibility of this model allowed us to achieve interpretable coefficients, and resulting relativities, that are critical for our application.

7.3. Appendix: Endorsements by entity type

The endorsements in the dataset exhibit interactions with the entity type. These interactions are worth documenting, hence we provide summary tables here.

Table 16 shows the claim frequencies and mean claim for the Monies and Securities categorical variable. Tables 17, 18, 19 show interaction of various other endorsements with the entity type. Table 20 shows the claim frequencies and severity means for entities receiving different alarm credit amounts.

Tables 16, 17, 18, and 19 give an idea of the varying effects of the endorsements over entity types. These tables reveal interesting interactions. For example, Table 16 describes, by entity type, frequency and severity for four categories of monies and securities endorsements: (i) an "A, B" type, (ii) a "limited term" type, (iii) both types, and (iv) neither types. In Table 6, we show experience for the 2,137 policyholder-year observations that have the "A, B" type (with or without limited term) and the 556 observations that have limited term (with or without the "A, B" type). The endorsement amount of each type is used in the shrinkage analysis, to assess proper relativities. Table 17 shows that the high frequency and severity of Golf Course Grounds are mainly due to counties.

Table 15. Tweedie model with and without shrinkage

	λ = 0		λ = 1,000	
	Estimate	Standard Error	Estimate	Standard Error
	Base Rating Variables			
(Intercept)	5.945	0.317	5.781	0.315
LnCoverage	0.867	0.099	0.888	0.098
LnDeduct	0.089	0.038	0.106	0.038
TypeCity	1.070	0.380	1.053	0.382
TypeCounty	0.937	0.701	−0.251	0.507
TypeMisc	−0.376	0.286	−0.324	0.283
TypeSchool	−0.968	0.336	−0.843	0.333
TypeTown	0.250	0.258	0.473	0.252
Fire5=1	0.263	0.093	0.207	0.092
NoClaimCredit=1	0.374	0.180	0.562	0.176
	Interaction Terms			
LnCoverage* TypeCity	−0.281	0.128	−0.254	0.128
LnCoverage* TypeCounty	−0.242	0.182	0.057	0.138
LnCoverage* TypeMisc	−0.090	0.117	−0.076	0.115
LnCoverage* TypeSchool	0.265	0.117	0.217	0.116
LnCoverage* TypeTown	−0.406	0.165	−0.451	0.164
LnCoverage* NoClaimCredit	−0.259	0.053	−0.318	0.052
	Endorsements			
LnBusInterCovRat	0.110	0.156	0.003	0.022
LnSpecialAnimalCovRat	−0.621	1.927	0.000	0.022
LnZooAnimalCovRat	8.568	6.896	0.000	0.022
LnFineArtsCovRat	0.472	0.317	0.004	0.022
LnGolfCourseCovRat	1.295	0.842	0.001	0.022
LnOtherCovRat	−0.643	0.238	−0.005	0.022
	Endorsement Indicators			
AccRec	−0.042	0.225	0.000	0.022
PierWharf	0.809	0.153	0.018	0.022
MoneySec	−0.136	0.085	−0.009	0.022
φ (dispersion)	165.022	3.438	166.727	3.472
−2 Log L	44,196		44,244	

7.4. Appendix: Cross-sectional out-of-sample validation

Figure 6 shows a robustness check of our model for seasonal variations, using a randomly selected training sample, consisting of 80% of the policy-holders, and the remaining 20% of the policyholders used as the holdout sample. Here, the training sample consists of 988 policyholders (5377 observations), and the holdout sample 248 policyholders (1356 observations).

Table 16. Claims summary by monies and securities category

| Entity Type | No MS nor MS Limited | | | MS Only | | |
	Claim Frequency	Avg. Severity	Num. Policies	Claim Frequency	Avg. Severity	Num. Policies
Village	0.391	7,780	844	0.546	14,540	366
City	1.652	16,770	423	2.145	10,657	207
County	3.646	10,248	79	7.258	16,870	155
Misc	0.187	43,999	514	0.179	38,060	95
School	0.529	67,245	891	2.685	71,798	568
Town	0.102	23,769	706	0.089	13,027	235
Total	0.566	32,754	3,457	2.049	34,478	1,626
Entity Type	MS Limited Only			Both MS and MS Limited		
	Claim Frequency	Avg. Severity	Num. Policies	Claim Frequency	Avg. Severity	Num. Policies
Village	0.263	2,195	19	0.634	15,141	112
City	0.800	7,658	10	2.536	23,844	153
County	2.200	3,953	5	2.056	18,416	89
Misc	—	—	—	—	—	—
School	0.200	8,366	5	2.203	23,329	133
Town	—	—	6	0.292	2,794	24
Total	0.556	4,699	45	1.843	20,628	511

Table 17. Summary by golf course, pier & wharf, vacancy permit

| Entity Type | Category No/Yes | Golf Course | | | Pier & Wharf | | | Vacancy Permit | | |
		Claim Frequency	Avg. Severity	Num. Policies	Claim Frequency	Avg. Severity	Num. Policies	Claim Frequency	Avg. Severity	Num. Policies
Village	0	0.452	10,668	1,339	0.438	10,929	1,265	0.444	10,979	1,282
Village	1	0.500	2,285	2	0.684	6,976	76	0.627	5,480	59
City	0	1.938	16,740	785	1.647	18,488	654	1.799	15,330	703
City	1	2.250	27,464	8	3.324	11,067	139	3.044	28,336	90
County	0	3.568	15,292	315	4.985	16,323	272	2.867	16,001	293
County	1	37.154	18,906	13	4.482	11,461	56	21.914	11,511	35
Misc	0	0.184	43,423	604	0.184	24,307	602	0.186	43,036	609
Misc	1	0.400	14,827	5	0.286	717,267	7	—	—	—
School	0	1.434	64,346	1,597	1.434	64,346	1,597	1.453	64,860	1,559
School	1	—	—	—	—	—	—	0.632	44,485	38
Town	0	0.103	19,831	971	0.090	6,300	937	0.102	20,080	968
Town	1	—	—	—	0.471	116,053	34	0.333	1,899	3
Total	0	1.025	31,331	5,611	1.027	31,930	5,327	0.952	31,946	5,414
Total	1	18.000	20,866	28	2.510	24,649	312	4.902	21,232	225

Table 18. Summary by fine arts, business interruption, other endorsements

Entity Type	Category No/Yes	Fine Arts			Business Interruption			Other Endorsements		
		Claim Frequency	Avg. Severity	Num. Policies	Claim Frequency	Avg. Severity	Num. Policies	Claim Frequency	Avg. Severity	Num. Policies
Village	0	0.451	10,694	1,336	0.440	10,473	1,294	0.451	10,666	1,338
Village	1	0.800	1,816	5	0.787	14,131	47	1.000	3,249	3
City	0	1.860	15,752	763	1.555	16,533	741	1.953	16,978	786
City	1	4.000	38,526	30	7.442	22,233	52	0.571	9,573	7
County	0	2.699	15,414	309	2.289	16,179	270	4.942	14,867	325
County	1	40.684	15,980	19	17.052	12,382	58	0.333	159,109	3
Misc	0	0.184	39,912	602	0.180	19,450	562	0.186	43,036	607
Misc	1	0.286	155,513	7	0.255	237,624	47	—	—	2
School	0	1.436	64,067	1,591	1.439	61,637	1,577	1.306	65,086	1,561
School	1	0.833	136,028	6	1.050	217,002	20	7.000	22,705	36
Town	0	0.103	19,831	971	0.103	19,831	970	0.103	19,831	969
Town	1	—	—	—	—	—	1	—	—	2
Total	0	0.960	31,026	5,572	0.888	29,998	5,414	1.073	31,231	5,586
Total	1	13.493	37,896	67	6.427	48,612	225	4.906	28,245	53

Table 19. Summary by Fire5, and no claim credit

Entity Type	Category No/Yes	Fire Class < 5			No Claim Credit		
		Claim Frequency	Avg. Severity	Num. Policies	Claim Frequency	Avg. Severity	Num. Policies
Village	0	0.387	10,436	809	0.541	10,564	901
Village	1	0.551	10,928	532	0.270	10,901	440
City	0	1.342	8,001	295	2.023	15,792	644
City	1	2.295	21,636	498	1.584	24,191	149
County	0	2.300	13,931	110	5.123	15,076	310
County	1	6.211	16,187	218	1.056	23,507	18
Misc	0	0.167	6,954	228	0.253	44,676	336
Misc	1	0.197	61,445	381	0.103	39,619	273
School	0	0.420	20,678	514	1.956	68,512	1,103
School	1	1.915	80,131	1,083	0.267	46,802	494
Town	0	0.076	38,653	552	0.122	4,288	492
Town	1	0.138	4,303	419	0.084	39,745	479
Total	0	0.502	13,935	2,508	1.501	31,365	3,786
Total	1	1.596	41,421	3,131	0.310	30,499	1,853

Table 20. Claims summary by entity type and alarm credit category

Entity Type	No Alarm Credit			Alarm Credit 5%		
	Claim Frequency	Avg. Severity	Num. Policies	Claim Frequency	Avg. Severity	Num. Policies
Village	0.326	11,078	829	0.278	8,086	54
City	0.893	7,576	244	2.077	4,150	13
County	2.140	16,013	50	—	—	1
Misc	0.117	15,122	386	0.278	13,064	18
School	0.422	25,523	294	0.410	14,575	122
Town	0.083	25,257	808	0.194	3,937	31
Total	0.318	15,118	2,611	0.431	10,762	239

Entity Type	Alarm Credit 10%			Alarm Credit 15%		
	Claim Frequency	Avg. Severity	Num. Policies	Claim Frequency	Avg. Severity	Num. Policies
Village	0.500	8,792	50	0.725	10,544	408
City	1.258	8,625	31	2.485	20,470	505
County	2.125	11,688	8	5.513	15,476	269
Misc	0.077	3,923	26	0.341	87,021	179
School	0.488	11,597	168	2.008	85,140	1,013
Town	0.091	2,338	44	0.261	9,490	88
Total	0.517	10,194	327	2.093	41,458	2,462

Figure 6. For a randomly selected training sample, the frequency-severity model showed 95.59% with the premiums, and 45.62% Spearman correlation with the holdout sample claims. For the Tweedie model, the Spearman correlation coefficient is 94.10% with the premiums, and 43.86% with the holdout sample claims