Predictive Modeling of Multi-Peril Homeowners Insurance

by Edward W. Frees, Glenn Meyers, and A. David Cummings

ABSTRACT

Predictive models are used by insurers for underwriting and ratemaking in personal lines insurance. Focusing on homeowners insurance, this paper examines many predictive generalized linear models, including those for pure premium (Tweedie), frequency (logistic) and severity (gamma). We compare predictions from models based on a single peril, or cause of loss, to those based on multiple perils. For multi-peril models, we introduce an instrumental variable approach to account for dependencies among perils. We calibrate these models using a database of detailed individual policyholder experience. To evaluate these many alternatives, we emphasize out-of-sample model comparisons. We utilize Gini indices for global comparisons of models and, for local comparisons, introduce nonparametric regression techniques. We find that using several different comparison approaches can help the actuary critically evaluate the effectiveness of alternative prediction procedures.

KEYWORDS

Instrumental variables, Tweedie distribution,
Gini index, insurance pricing

1. Introduction

This paper explores the use of predictive models that can be used for underwriting and ratemaking in homeowners insurance. Homeowners represents a large segment of the personal property and casualty insurance business; for example, in the United States, homeowners accounted for 13.6% of all property and casualty insurance premiums and 26.8% of personal lines insurance, for a total of over $57 billion (III 2010). Many actuaries interested in pricing homeowners insurance are now decomposing the set of dependent variables (r_i, y_i) by *peril,* or cause of loss (e.g., Modlin 2005). Homeowners is typically sold as an all-risk policy, which covers all causes of loss except those specifically excluded.

Decomposing risks by peril is not unique to personal lines insurance, nor is it new. For example, it is customary in population projections to study mortality by cause of death (e.g., Board of Trustees 2009). Further, in 1958, Robert Hurley discussed statistical considerations of multiple peril rating in the context of homeowners insurance. Referring to "multiple peril rating," Hurley stated: "The very name, whatever its inadequacies semantically, can stir up such partialities that the rational approach is overwhelmed in an arena of turbulent emotions."

Rating by multiple perils does not cause nearly as much excitement in today's world. Nonetheless, Rollins (2005) argues that multi-peril rating is critical for maintaining economic efficiency and actuarial equity. Decomposing risks by peril is intuitively appealing because some predictors do well in predicting certain perils but not in others. For example, "dwelling in an urban area" may be an excellent predictor for the theft peril but provide little useful information for the hail peril.

Current multi-peril rating practice is based on modeling each peril in isolation from the others. From a modeling point of view, this amounts to assuming that

- perils are independent of one another, and
- sets of parameters from each peril are unrelated to one another.

Although allowing sets of parameters to be unrelated to one another (sometimes called *functionally independent*) is plausible, it seems unlikely that perils are independent. Event classification can be ambiguous (e.g., fires triggered by lightning) and unobserved latent characteristics of policyholders (e.g., cautious homeowners who are sensitive to potential losses due to theft-vandalism as well as liability) may induce dependencies among perils. Prior empirical investigations reported in Frees, Meyers, and Cummings (2010) demonstrated statistically significant dependence among perils.

To accommodate potential dependencies, we introduce an *instrumental variables* approach. Instrumental variables is an estimation technique that is commonly used in econometrics to handle dependencies that arise among systems of equations. In this paper, we hypothesize that multiple peril models are jointly determined and that a methodology such as instrumental variables can be used to quantify these dependencies.

Although examining the multiple peril nature of homeowners insurance is intuitively plausible, not all insurers will wish to consider this complex model. In homeowners, consumers are charged a single price, meaning that the decomposition by peril may not be necessary for financial transactions. Moreover, from statistical learning it is well known (e.g., Hastie, Tibshirani, and Friedman 2001) that there is a price to be paid for complexity; other things equal, more complex models fare poorly compared to simpler alternatives for prediction purposes.

Thus, in this paper we compare our many alternative models using out-of-sample validation techniques. Section 1 introduces our data and Section 2 presents several baseline models. We consider both pure premium and frequency-severity approaches, as well as both a single- and multi-peril modeling framework in this work. Section 3 introduces the instrumental variable approach. We then show how these competing approaches fare in the context of a held-out validation sample in Section 4.

Loss distributions are not even approximately symmetric nor normally distributed; to illustrate, in our data 94% of the losses are zeros (corresponding to no claims) and when losses are positive, the distribution tends to be right-skewed and thick-tailed. Thus, the usual mean square metrics, such as variance and R^2, are not informative for capturing differences between predictions and held-out data. Thus, we use recent developments (Frees, Meyers, and Cummings 2011) on a statistical measure called a Gini index to compare predictors in Section 2. Section 5 explores nonparametric regression, an alternative validation measure. Both approaches allow us to compare, among other things, a single peril pure premium model with one dependent variable to a multiple peril model with many dependent variables. Section 6 closes with a summary and a few additional remarks.

2. Data and preliminary models

2.1. Data

To calibrate our models, we drew two random samples from a homeowners database maintained by the Insurance Services Office. This database contains over 4.2 million policyholder years. It is based on the policies issued by several major insurance companies in the United States, thought to be representative of most geographic areas. These policies are almost all for one year and so we will use a constant exposure (one) for our models.

Our in-sample, or "training," dataset consists of a representative sample of 404,664 records taken from this database. The summary measures in this section are based on this training sample. In Section 4, we will test our calibrated models on a second held-out, or "validation," sample that was also randomly selected from this database.

For each record, we have information on whether there are any claims due to a peril and the amount associated with that peril. Table 1 displays summary statistics for nine perils from our sample of 404,664 records. This table shows that WaterNonWeather is the most frequently occurring peril, whereas Liability is the least frequent. (WaterNonWeather is water damage from causes other than weather, e.g., the bursting of a water pipe in a house.) When a claim occurs, Hail is the most severe peril (based on the median claim amount), whereas the Other category is the least severe. In Table 1, we note that neither the Frequency nor the Number of claims sum to the totals due to jointly occurring perils within a policy.

Table 1. Homeowners summary statistics

Peril	Frequency (in percent)	Number of Claims	Median Claim Amount
Fire	0.310	1,254	4,152
Lightning	0.527	2,134	899
Wind	1.226	4,960	1,315
Hail	0.491	1,985	4,484
WaterWeather	0.776	3,142	1,481
Water NonWeather	1.332	5,391	2,167
Liability	0.187	757	1,000
Other	0.464	1,877	875
Theft-Vandalism	0.812	3,287	1,119
Total	5.889	23,834	1,661

In this work, we consider two sets of explanatory variables. The goal is to show how the predictive modeling techniques work over a range of information available to the analyst. The first set of variables is a base set that consists of the amount of insurance dwelling coverage, a building adjustment, the construction age of the building, policy deductibles, the homeowners policy form, and base cost loss costs. Here, the "base cost loss costs" are the ISO base class loss costs at the home address, a very reasonable proxy for territory.

The second set of variables is an "extended" list of variables that consists of many (over 100) explanatory variables to predict homeowners claims. These are a variety of geographic-based plus several standard industry variables that account for

- weather and elevation,
- vicinity,
- commercial and geographic features,
- experience and trend, and
- rating variables.

The Web site http://www.iso.com/Products/ISO-Risk-Analyzer/ISO-Risk-Analyzer-Homeowners.html provides more information on these explanatory variables.

2.2. Baseline models

Like most analysts, we do not wish to advocate one model as superior to others for every dataset. Rather, we view the collection of models as tools that the analyst has at his or her disposal—the job of the analyst is to pick the right tool for a dataset under consideration. Below is the collection of baseline models that we consider here. Note that in the empirical data section, we calibrate each model with the base and the extended list of predictor variables described in Section 1.

- Single-peril models—the dependent variable is the outcome for the total policy, not disaggregated by peril.
    - Pure premium models—there is a single dependent variable for this model that represents the total loss for a policy.
    - Frequency-severity models—there are two dependent variables in this type of model, one for the frequency and one for the severity.
- Multi-peril models—there are $c = 9$ separate outcomes for each policy, one for each peril.
    - Independence pure premium models—there is a dependent variable for each peril, resulting in $c = 9$ dependent variables. Under the "Independence" framework, one assumes independence among dependent variables.
    - Independence frequency-severity models—for each peril, there are two dependent variables, one for the frequency and one for the severity. This means that there are 2 × 9 = 18 dependent variables. Under the "Independence" framework, one assumes independence among dependent variables.
    - Dependence ratio models—these have the same set of dependent variables as the Independence Frequency-Severity Models. However, a dependence structure is introduced in the frequencies to accommodate potential dependencies.

A more detailed description of these models may be found in Appendix B.

A piece of advice, sometimes attributed to Albert Einstein, is to "Use the simplest model possible, but no simpler." Many analysts prefer the simpler single-peril models because of their ease of implementation and interpretability. Moreover, based on the discussion in the introductory Section 1, industry analysts also find a need for multi-peril models. Previous work in Frees, Meyers, and Cummings (2010) established statistically significant dependence among perils. (Appendix Section A gives readers a feel for the type of dependencies discussed in that work.) Thus, based on empirical evidence and intuition, a goal of this paper is to improve upon the assumption of independence in the multi-peril models. One approach is the "Dependence Ratio Model" that we introduced in Frees, Meyers, and Cummings (2010). A drawback of this approach is that it is based on a maximum likelihood estimation routine that is cumbersome to interpret and calibrate. Thus, in this paper, we introduce an instrumental variable approach that can be implemented without specialized software to accommodate the dependence among perils.

2.3. Overview of the instrumental variables approach

An instrumental variable approach to estimation can be used to improve upon the predictions under the independence models. To illustrate, suppose that we are interested in predicting fire claims and believe that there exists an association between fire and theft/vandalism claims. One would like to use the information in theft/vandalism claims to predict fire claims; however, the number and severity of theft/vandalism claims are unknown when making the predictions. We can, however, use *estimates* of theft/vandalism claims as predictors of fire claims. This is the essence of the instrumental variable estimation method where one substitutes proxies for variables that are not available a priori.

To provide motivation for someone to adopt this approach, consider a classic economic demand and supply problem that is summarized by two equations:

$$\begin{aligned} y_{1i} &= \beta_1 y_{2i} + \gamma_{10} + \gamma_{11} x_{1i} + \varepsilon_{1i} \left(\text{price}\right) \\ y_{2i} &= \beta_2 y_{1i} + \gamma_{20} + \gamma_{21} x_{2i} + \varepsilon_{2i} \left(\text{quantity}\right) \end{aligned} \tag{2.1}$$

Here, we assume that quantity (y_2) linearly affects price (y_1), and vice versa. Further, let x_1 be the purchasers' income and x_2 be the suppliers' wage rate. The other explanatory variables (x's) are assumed to be exogenous for the demand and supply equations.

For simplicity, assume that we have $i = 1, \ldots, n$ independent observations that follow display (1). One estimation strategy is to use ordinary linear regression. This strategy yields biased regression coefficient estimates because the right-hand side of display (2.1), the "conditioning" or explanatory variables, contains a y variable that is also a dependent variable. The difficulty with ordinary least squares estimation of the model in display (2.1) is that the right-hand side variables are correlated with the disturbance term. For example, looking at the price equation, one can see that quantity (y_2) is correlated with ε_1. This is because y_2 depends on y_1 (from the supply equation), which in turn depends on ε_1 (from the demand equation). This circular dependency structure induces the correlation that leads to biased regression coefficient estimation.

The instrumental variable approach is to use ordinary least squares with approximate values for the right-hand side dependent variables. To see how this works, we focus on the price equation and assume that we have available "instruments" $\mathbf{w}$ to approximate y_2. Then, we employ a two-stage strategy:

1. Run a regression of $\mathbf{w}$ on y_2 to get fitted values of the form $\widehat{y_2}$.
2. Run a regression of x_1 and $\widehat{y_2}$ on y_1.

As one would expect, the key difficulties are coming up with suitable instruments $\mathbf{w}$ that provide the basis for creating reasonable proxies for y_2 that do not have endogeneity problems. In our example, we might use x_2, the suppliers' wage rate, as our instrument in the stage 1 estimate of y_2, the quantity demanded. This variable is exogenous and not perfectly related to x_1, purchaser's income. Not surprisingly, there are conditions on the instruments. Typically, they may include a subset of the model predictor variables but must also include additional variables. Appendix C provides additional details.

3. Multi-peril models with instrumental variables

As discussed above, when modeling systems of $c = 9$ perils, it seems reasonable to posit that there may be associations among perils and, if so, attempt to use these associations to provide better predictors. For example, in prior work (see Appendix A), statistically significant associations between claims from fire and theft/vandalism were established. Sections 1 and 2 describe the estimation procedures in the pure premium and frequency/severity contexts, respectively.

3.1. Pure premium modeling

Under our independence pure premium model framework, we assume that the claim amount follows a Tweedie (1984) distribution. The shape and dispersion parameters vary by peril and the mean parameter is a function of explanatory variables available for that peril. Using notation, we assume that

$$y_{ij} \sim Tweedie\left(\mu_{i,j}, \phi_j, p_j\right),$$
$$i = 1, \ldots, n = 404{,}664, \; j = 1, \ldots, c = 9. \quad (3.1)$$

Here, ϕ_j is the dispersion parameter, p_j is the shape parameter, and $\mu_{i,j} = \exp(\boldsymbol{x}'_{i,j}\boldsymbol{\beta}_j)$ is the mean parameter using a logarithmic link function. There are many procedures for estimating the parameters in Equation (3.1); we use maximum likelihood. See, for example, Frees (2010) for an introduction to the Tweedie distribution in the context of regression modeling.

Estimating independence pure premium models with Equation (3.1) allows us to determine regression coefficient estimates $\boldsymbol{b}_{IND,j}$. These coefficients allow us to compute (independence model) pure premium estimates of the form $\hat{\mu}_{IND,i,j} = \exp(\boldsymbol{x}'_{i,j}\boldsymbol{b}_{IND,j})$.

For instrumental variable predictors, we use logarithmic fitted values from *other* perils as additional explanatory variables. For example, suppose we wish to estimate a pure premium model for the first peril. For the $j = 1^{st}$ peril, we already have predictors $\boldsymbol{x}_{i,1}$. We augment $\boldsymbol{x}_{i,1}$ with the additional predictor variables

$$\ln \hat{\mu}_{IND,i,j},\ j = 2, \ldots, c = 9.$$

We then estimate the pure premium model in Equation (3.1) using both sets of explanatory variables.

We summarize the procedure as follows.

- Stage 1. For each of the nine perils, fit a pure premium model in accordance with Equation (3.1). These explanatory variables differ by peril. Calculate fitted values, denoted as $\hat{\mu}_{IND,i,j}$. Because these fits are unrelated to one another, these are called the "independence" pure premium model fits.
- Stage 2. For each of the nine perils, fit a pure premium model using the Stage 1 explanatory variables as well as logarithmic fitted values from the other eight perils. Denote the predictions resulting from this model as $\hat{\mu}_{IV,i,j}$.

Table 2 summarizes the regression coefficient estimates for the fit of the instrumental variable pure premium model. This table shows results only for the additional instruments, the logarithmic fitted values. This is because our interest is in the extent that these additional variables improve the model fit when compared to the independence models. Table 2 shows that the additional variables are statistically significant, at least when one examines individual *t*-statistics. Although we do not include the calculations here, this is also true when examining collections of variables (using a likelihood ratio test). However, this is not surprising because we are working with a relatively large sample size, $n = 404{,}664$. We defer our more critical assessment of model comparisons to Section 4 where we compare models on an out-of-sample basis. There, we will label the resulting insurance scores as "IV_PurePrem."

We use logarithmic fitted values because of the logarithmic link function; in this way the additional predictors are on the same scale as the fitted values. Moreover, by using a natural logarithm, they can be interpreted as elasticities, or percentage changes. For example, to interpret the lightning coefficient of the fire fitted value, we have

$$0.220 = \frac{\partial \ln \hat{\mu}_{IV,FIRE}}{\partial \ln \hat{\mu}_{IND,LIGHT}} = \frac{\left(\dfrac{\partial \hat{\mu}_{IV,FIRE}}{\hat{\mu}_{IV,FIRE}}\right)}{\left(\dfrac{\partial \hat{\mu}_{IND,LIGHT}}{\hat{\mu}_{IND,LIGHT}}\right)}.$$

That is, holding other variables fixed, a 1% change in the fitted value for lightning is associated with a 0.22% change in the fitted value for fire.

3.2. Frequency and severity modeling

The approach to instrumental variable estimation for frequency and severity modeling is similar to the pure premium case but more complex. At the first stage, we calculate independence, frequency, and severity fits; we now have many instruments that can be used as predictor variables for second stage instrumental variable estimation. That is, in principle it is possible to use both fitted probabilities and severities in our instrumental variable frequency and severity models.

Based on our empirical work, we have found that the fitted probabilities provide better predictions than using both fitted probabilities and severities as instruments. Intuitively, coefficients for fitted severities are based on smaller sample sizes (when there is claim) and may contain less information in some sense than fitted probabilities. Thus, for our main model we feature fitted probabilities and include fitted severities for a robustness check (Appendix D).

The algorithm is similar to the pure premium modeling in Section 1. We summarize the procedure as follows.

- Stage 1. Compute independence frequency and severity model fitted values. Specifically, for each of the $j = 1, \ldots, 9$ perils:
  1a. Fit a logistic regression model using the explanatory variables $\boldsymbol{x}_{F,i,j}$. These explanatory variables differ by peril j. Calculate fitted values to get predicted probabilities, denoted as $\hat{\pi}_{IND,i,j}$.
  1b. Fit a gamma regression model using the explanatory variables $\boldsymbol{x}_{S,i,j}$ with a logarithmic

Table 2. Instrumental variable pure premium model coefficients
Shown are coefficients associated with the instruments, logarithmic fitted values

	Dependent Variables					
	Fire		Lightning		Wind	
Explanatory Variables	Estimate	*t*-statistic	Estimate	*t*-statistic	Estimate	*t*-statistic
Log Fitted Fire			0.3313	25.10	–0.0184	–1.52
Log Fitted Lightning	0.2200	15.49			0.4120	28.81
Log Fitted Wind	–0.0468	–3.16	0.2238	15.43		
Log Fitted Hail	–0.0196	–4.08	0.0702	14.04	–0.1021	–23.74
Log Fitted WaterWeather	0.2167	14.16	–0.2120	–11.98	–0.0706	–4.20
Log Fitted WaterNonWeat	–0.0568	–4.66	0.2822	12.54	0.3442	18.51
Log Fitted Liability	–0.0696	–6.05	–0.1667	–12.82	–0.0330	–2.82
Log Fitted Other	–0.0147	–1.34	0.0081	0.80	–0.2229	–20.45
Log Fitted Theft	0.7854	37.76	–0.1107	–4.77	–0.1815	–10.20

	Dependent Variables					
	Hail		Water Weather		Water NonWeather	
Explanatory Variables	Estimate	*t*-statistic	Estimate	*t*-statistic	Estimate	*t*-statistic
Log Fitted Fire	–0.0786	–7.08	0.1162	7.13	0.3789	33.24
Log Fitted Lightning	0.1291	9.36	0.0062	0.51	–0.0555	–3.58
Log Fitted Wind	0.1194	5.43	0.0504	3.76	0.0329	2.49
Log Fitted Hail			–0.0437	-8.74	0.0007	0.14
Log Fitted WaterWeather	0.2794	12.64			–0.2504	–16.37
Log Fitted WaterNonWeat	–0.1302	–7.48	0.2833	18.16		
Log Fitted Liability	–0.4527	–35.37	–0.1764	–14.95	–0.1297	–11.58
Log Fitted Other	–0.2411	–21.72	0.2419	20.33	0.0449	4.49
Log Fitted Theft	0.4334	27.43	0.2642	14.36	0.0827	5.10

	Dependent Variables					
	Liability		Other		Theft	
Explanatory Variables	Estimate	*t*-statistic	Estimate	*t*-statistic	Estimate	*t*-statistic
Log Fitted Fire	0.6046	50.38	–0.2285	–19.20	0.2881	25.72
Log Fitted Lightning	0.3883	31.83	0.1874	19.73	0.1567	11.36
Log Fitted Wind	–0.6248	–46.63	–0.1297	–11.09	–0.0907	–7.75
Log Fitted Hail	0.0822	16.12	–0.2128	–56.00	–0.0258	–6.00
Log Fitted WaterWeather	–0.4337	–22.71	0.2708	27.92	0.2515	18.22
Log Fitted WaterNonWeat	–0.2227	–12.80	0.5306	28.99	–0.2138	–15.06
Log Fitted Liability			–0.0341	–3.88	–0.1174	–11.40
Log Fitted Other	0.1258	12.21			0.1555	16.37
Log Fitted Theft	0.1447	7.13	–0.0658	–3.45		

link function. These explanatory variables may differ by peril and from those used in the frequency model. Calculate fitted values to get predicted severities (by peril), denoted as $\widehat{\mathrm{E}y}_{IND,i,j}$.

- Stage 2. Incorporate additional instruments into the frequency model estimation. Specifically, for each of the $j = 1, \ldots, 9$ perils:
  2. Fit a logistic regression model using the explanatory variables $\boldsymbol{x}_{F,i,j}$ and the logarithm of the predicted probabilities developed in step 1(a), $\ln \hat{\pi}_{IND,i,k}$, $k = 1, \ldots, 9$, $k \neq j$.

In Section 4 we will label the resulting insurance scores as "IV_FreqSevA." We remark that this procedure could easily be adapted to distributions other than the gamma, as well as link functions other than logarithmic. These choices simply worked well for our data.

As with the Section 1 pure premium instrumental variable model, we found many instruments to be statistically significant when this model was estimated with our in-sample data. This is not surprising because it is common to find effects that are "statistically significant" using large samples. Thus, we defer discussions of model selection to our out-of-sample validation beginning in Section 4. In this section, we examine alternative instrumental variable models. In particular, using additional instruments in the severity model (instead of the frequency model) will result in insurance scores labeled as "IV_FreqSevB." Use of additional instruments in frequency and severity, described in detail in Appendix D, will result in insurance scores labeled as "IV_FreqSevC."

4. Out-of-sample analysis

Qualitative model characteristics will drive some modelers to choose one approach over another. However, others will seek to understand how these competing approaches fare in the context of empirical evidence. As noted earlier, in-sample summary statistics are not very helpful for model comparisons. Measures of (in-sample) statistical significance provide little guidance because we are working with a large sample size (404,664 records); with large sample sizes, coefficient estimates tend to be statistically significant using traditional measures. Moreover, goodness-of-fit measures are also not very helpful. In the basic frequency-severity model, there are two dependent variables and in the multi-peril version, there are 18 dependent variables. Goodness-of-fit measures typically focus on a single dependent variable.

We rely instead on out-of-sample comparisons of models. In predictive modeling, the "gold standard" is model validation through examining performance of an independent held-out sample of data (e.g., Hastie, Tibshirani, and Friedman 2001). Specifically, we use our in-sample data of 404,664 records to compute parameter estimates. We then use the estimated parameters from the in-sample model fit as well as predictor variables from a held-out, or validation, sample of 359,454 records, whose claims we wish to predict. For us, the important advantage of this approach is that we are able to compare models with different dependent variables by aggregating predictions into a single score for a record.

To illustrate, consider the independence frequency severity model with 18 dependent variables. We can use estimators from this model to compute an overall predicted amount as

$$\begin{aligned}
\mathrm{IND\_FreqSev}_i &= \sum_{j=1}^{c} \widehat{\mathrm{Prob}}_{i,j} \times \widehat{\mathrm{Fit}}_{i,j} \\
&= \sum_{j=1}^{c} \frac{\exp\left(\boldsymbol{x}'_{F,i,j}\, \boldsymbol{b}_{F,j}\right)}{1+\exp\left(\boldsymbol{x}'_{F,i,j}\, \boldsymbol{b}_{F,j}\right)} \\
&\quad \times \exp\left(\boldsymbol{x}'_{S,i,j}\, \boldsymbol{b}_{S,j}\right). \qquad (4.1)
\end{aligned}$$

Here, $\widehat{\mathrm{Prob}}_{i,j}$ is the predicted probability using logistic regression model parameter estimates, $\boldsymbol{b}_{F,j}$, and frequency covariates $\boldsymbol{x}_{F,ij}$, for the jth peril. Further, $\widehat{\mathrm{Fit}}_{i,j}$ is the predicted amount based on a logarithmic link using gamma regression model parameter estimates, $\boldsymbol{b}_{S,j}$, and severity covariates $\boldsymbol{x}_{S,i,j}$, for the jth peril. This predicted amount, or "score," provides a basic input for ratemaking. We focus on this measure in this section.

In the following, Section 4.1 provides global comparisons of scores to actual claims. Section 4.2 provides cumulative comparisons using a Gini index. Section 5 provides local comparisons using nonparametric regression.

4.1. Comparison of scores

We examine the 14 scores that are listed in the legend of Table 3. This table summarizes the distribution of each score on the held-out data. Not surprisingly, each distribution is right-skewed.

Table 3 also shows that the single-peril frequency severity model using the extended set of variables (SP_FreqSev) provides the lowest score, both for the mean and at each percentile (below the 75th percentile). Except for this, no model seems to give a score that is consistently high or low for all percentiles. All scores have a lower average than the average held-out actual claims (TotClaims).

Table 3 shows that the distributions for the 14 scores appear to be similar. For an individual policy, to what extent do the scores differ? As one response to this

Table 3. Summary statistics of 14 scores and total claims

			Percentiles							
Score	Mean	Minimum	1st	5th	25th	50th	75th	95th	99th	Maximum
SP_FreqSev_Basic	291.10	20.48	85.00	120.25	182.74	240.37	334.62	618.37	1,025.88	8,856.79
SP_PurePrem_Basic	289.91	33.01	89.48	127.80	189.87	246.44	329.79	586.33	1,050.15	5,467.41
IND_PurePrem_Basic	290.91	37.49	92.08	124.04	182.68	240.30	328.87	612.47	1,087.06	13,577.91
IV_PurePrem_Basic	293.55	36.61	93.91	128.21	187.57	241.29	327.75	616.05	1,122.84	15,472.82
SP_FreqSev	287.79	8.78	71.55	105.39	171.55	237.95	339.40	631.98	1,039.19	6,864.46
SP_PurePrem	290.00	10.23	72.17	107.90	175.83	242.17	338.64	616.64	1,113.73	7,993.52
IND_FreqSev	294.93	33.05	97.14	126.61	185.07	244.99	333.68	606.03	1,106.17	22,402.49
IND_PurePrem	292.18	28.04	86.53	119.74	181.22	240.52	326.60	592.07	1,078.25	49,912.59
IV_PurePrem	294.06	12.42	78.41	113.14	178.62	240.38	330.21	614.22	1,095.70	107,158.09
IV_FreqSevA	290.91	23.99	88.70	121.70	182.29	241.42	327.81	606.23	1,096.86	18,102.93
IV_FreqSevB	295.32	28.52	94.58	124.77	184.29	245.26	335.38	606.63	1,100.61	24,394.06
IV_FreqSevC	291.17	20.88	84.78	118.21	180.63	241.57	329.92	608.28	1,098.40	20,046.03
DepRatio1	301.12	33.38	98.80	128.95	188.73	249.97	340.64	619.79	1,129.96	23,255.94
DepRatio36	302.39	33.48	99.27	129.65	189.87	251.41	342.30	620.38	1,132.36	23,092.35
TotClaims	332.89	0.00	0.00	0.00	0.00	0.00	0.00	660.00	5,916.33	350,000.00

Legend:

Score	**Interpretation**
Scores using the basic set of explanatory variables	
SP_FreqSev_Basic	Single-peril, frequency and severity model
SP_PurePrem_Basic	Single-peril, pure premium model
IND_PurePrem_Basic	Multi-peril independence, pure premium model
IV_PurePrem_Basic	Instrumental variable multi-peril pure premium model
Scores using the extended set of explanatory variables	
SP_FreqSev	Single-peril, frequency and severity model
SP_PurePrem	Single-peril, pure premium model
IND_FreqSev	Multi-peril frequency and severity model assuming independence among perils
IND_PurePrem	Multi-peril pure premium model assuming independence among perils
IV_PurePrem	Instrumental variable multi-peril pure premium model
Instrumental variable multi-peril frequency and severity models, using the extended set of explanatory variables	
IV_FreqSevA	Uses instruments in frequency model
IV_FreqSevB	Uses instruments in severity model
IV_FreqSevC	Uses instruments in frequency and severity models
Dependence ratio multi-peril frequency and severity models, using the extended set of explanatory variables	
DepRatio1	Uses a single parameter for frequency dependencies
DepRatio36	Uses 36 parameters for frequency dependencies

Table 4. Spearman correlations of 14 scores and total claims

	Basic Explanatory Variables				Extended Explanatory Variables									
	Single Peril		IND_	IV_	Single Peril		IND_		IV_	IV_FreqSev			DepRatio	
	Freq Sev	Pure Prem	Pure Prem	Pure Prem	Freq Sev	Pure Prem	Freq Sev	Pure Prem	Pure Prem	A	B	C	1	36
SP_FreqSev_Basic	1.000													
SP_PurePrem_Basic	0.949	1.000												
IND_PurePrem_Basic	0.922	0.948	1.000											
IV_PurePrem_Basic	0.924	0.941	0.965	1.000										
SP_FreqSev	0.880	0.842	0.817	0.811	1.000									
SP_PurePrem	0.818	0.855	0.809	0.801	0.892	1.000								
IND_FreqSev	0.808	0.834	0.875	0.850	0.798	0.802	1.000							
IND_PurePrem	0.850	0.875	0.899	0.888	0.849	0.872	0.905	1.000						
IV_PurePrem	0.830	0.842	0.852	0.862	0.850	0.879	0.858	0.962	1.000					
IV_FreqSevA	0.850	0.878	0.885	0.883	0.853	0.881	0.943	0.939	0.936	1.000				
IV_FreqSevB	0.797	0.826	0.858	0.843	0.800	0.806	0.994	0.916	0.871	0.948	1.000			
IV_FreqSevC	0.831	0.860	0.858	0.868	0.846	0.876	0.927	0.941	0.938	0.994	0.944	1.000		
DepRatio1	0.808	0.834	0.875	0.850	0.798	0.802	1.000	0.905	0.859	0.943	0.994	0.928	1.000	
DepRatio36	0.808	0.835	0.876	0.850	0.798	0.803	1.000	0.905	0.859	0.943	0.994	0.928	0.999	1.000
TotClaims	0.043	0.043	0.040	0.041	0.052	0.053	0.032	0.048	0.051	0.048	0.033	0.049	0.032	0.032

question, Table 4 provides correlations among the 14 scores and total claims. This table shows strong positive correlations among the scores, and a positive correlation between claims and each score. Because the distributions are markedly skewed, we use a nonparametric Spearman correlation to assess these relationships. Recall that a Spearman correlation is a regular (Pearson) correlation based on ranks, so that skewness does not affect this measure of association. See, for example, Miller and Wichern (1977) for an introduction to the Spearman correlation coefficient.

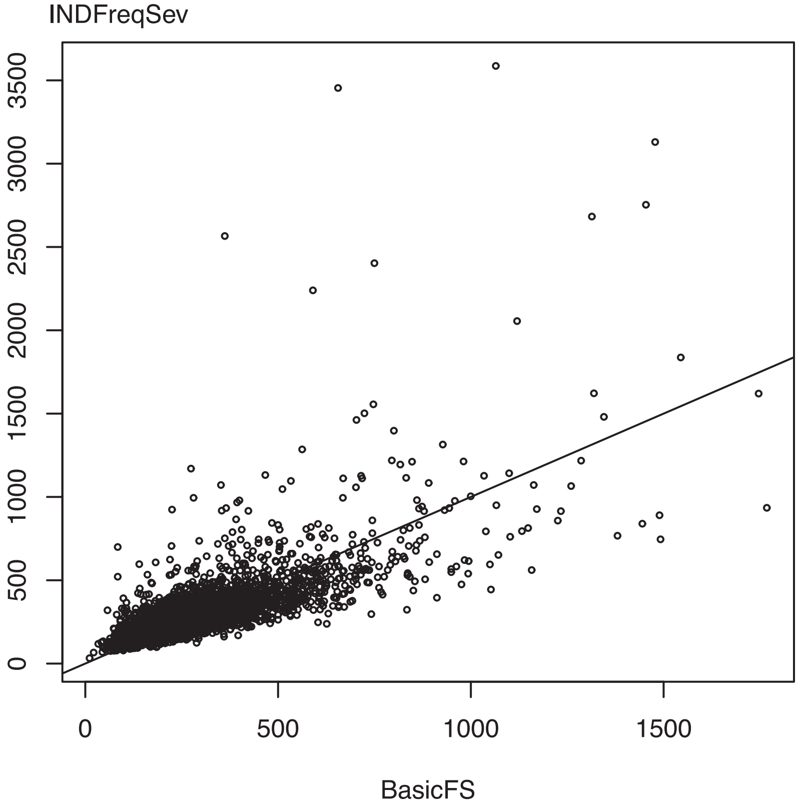
Table 4 shows strong associations within scores based on the basic explanatory variables (SP_FreqSev_Basic, SP_PurePrem_Basic, IND_PurePrem_Basic, and IV_PurePrem_Basic). In contrast, associations are weaker between scores based on basic explanatory variables and those based on the extended set of explanatory variables. For scores based on the extended set of explanatory variables, there is a strong association between the single peril scores (0.892, for SP_FreqSev and SP_PurePrem). It also shows strong associations within the multi-peril measures, particularly those of the same type (either frequency-severity or pure premium). The weakest associations are between the single- and multi-peril measures. For example, the smallest correlation, 0.798, is between SP_FreqSev and IND_FreqSev.

Although strongly associated, do the different scoring methods provide economically important differences in predictions? To answer this, Figure 1 shows the relationship between SP_FreqSev and IND_FreqSev. So that patterns are not obscured, only a 1% sample is plotted. This figure shows substantial variation between the two sets of scores. Particularly for larger scores, we see percentage differences that are 20% and higher.

4.2. Out-of-sample analysis using a Gini index

In insurance claims modeling, standard out-of-sample validation measures are not the most informative due to the high proportions of zeros (corresponding to no claim) and the skewed fat-tailed distribution of the positive values. We use an alternative validation measure, the Gini index, that is motivated by the economics of insurance. Properties of the insurance scoring ver-

Figure 1. Single versus multi-peril frequency-severity scores. This graph is based on a 1 in 100 random sample of size 3,594. The correlation coefficient is only 79.4%; the figure shows substantial variation between the two sets of scores.



sion of the Gini index have been recently established in Frees, Meyers, and Cummings (2011). Intuitively, the Gini index measures the negative covariance between a policy's "profit" ($P - y$, premium minus loss) and the rank of the relativity (R, score divided by premium). That is, the close approximation

$$\widehat{Gini} \approx -\frac{2}{n}\widehat{Cov}\big((P-y),\ rank(R)\big)$$

was established in Frees, Meyers, and Cummings (forthcoming).

4.2.1. Comparing scoring methods to a selected base premium

Assume that the insurer has adopted a base premium for rating purposes; to illustrate, we use the "SP_FreqSev_Basic" for this premium. Recall from Section 1 that this method uses only a basic set of rating variables to determine insurance scores from a single-peril, frequency and severity model. Assume that the insurer wishes to investigate alternative scoring methods to understand the potential vulnerabilities of this premium base; Table 5 summarizes several comparisons using the Gini index. This table includes the comparison with the alternative score IND_FreqSev as well as twelve other scores.

The standard errors were derived in Frees, Meyers, and Cummings (2011) where the asymptotic normality of the Gini index was proved. Thus, to interpret Table 5, one may use the usual rules of thumb and reference to the standard normal distribution to assess statistical significance. For the three scores that use the basic set of variables, SP_PurePrem_Basic, IND_PurePrem_Basic, and IV_PurePrem_Basic, all have Gini indices less than two standard errors, indicating a lack of statistical significance. In contrast, the other Gini indices all are more than three standard errors above zero, indicating that the ordering used by each

Table 5. Gini indices and standard errors

Alternative Score	Gini	Standard Error	Alternative Score	Gini	Standard Error
SP_PurePrem_Basic	4.89	2.74	IV_FreqSevA	12.59	2.50
IND_PurePrem_Basic	4.01	2.77	IV_FreqSevB	10.61	2.54
IV_PurePrem_Basic	4.33	2.75	IV_FreqSevC	12.80	2.49
SP_FreqSev	11.15	2.54	DepRatio1	10.09	2.56
SP_PurePrem	9.97	2.59	DepRatio36	10.06	2.56
IND_FreqSev	10.03	2.56			
IND_PurePrem	10.96	2.57			
IV_PurePrem	11.29	2.55			

Note: Base Premium is SP_FreqSev_Basic.

score helps detect important differences between losses and premiums.

The paper of Frees, Meyers, and Cummings (2011) also derived distribution theory to assess statistical differences between Gini indices. Although we do not review that theory here, we did perform these calculations for our data. It turns out that there are no statistically significant differences among the ten Gini indices that are based on the extended set of explanatory variables.

In summary, Table 5 suggests that there are important advantages to using extended sets of variables compared to the basic variables, regardless of the scoring techniques used. Moreover, this table suggests that the instrumental variable scores provide improved "lift" when compared to the scores generated by the independence model.

5. Out-of-sample analysis using local comparisons of claims to scores

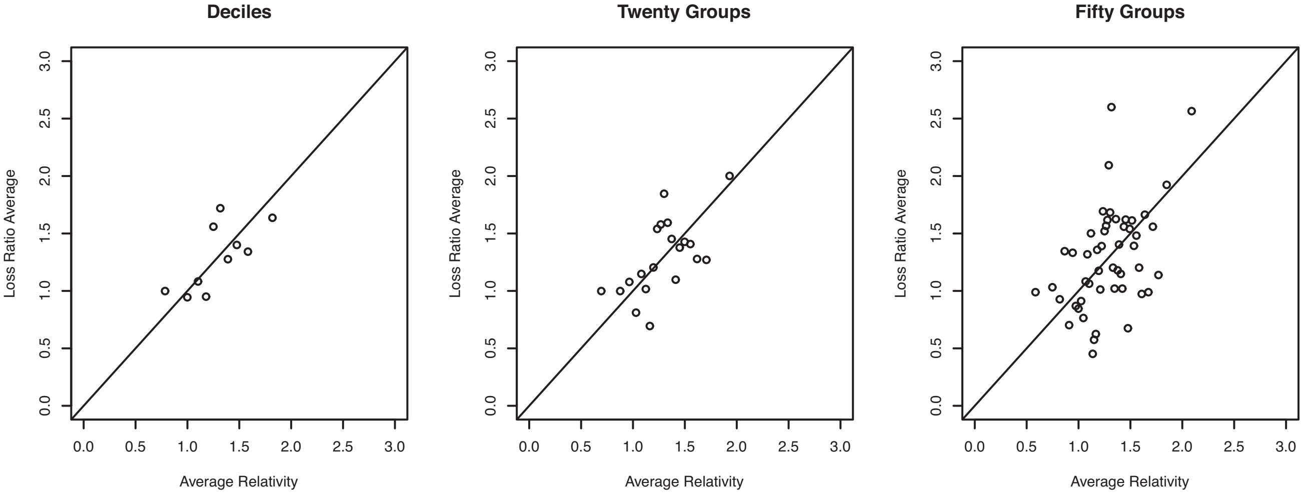
As noted in Section 2, one interpretation of the Gini index is as the covariance between $y - P$ (loss minus premium) and the rank of relativities. Another interpretation is as an area between *cumulative* distributions of premiums and losses. Through the accumulation process, models may be locally inadequate and such deficiencies may not be detected by a Gini index. Thus, this section describes an alternative graphical approach that can help us assess the performance of scores locally.

One method of making local comparisons used in practice involves comparing averages of relativities and loss ratios for homogenous subgroups. Intuitively, if a score S is a good predictor of loss y, then a graph of scores versus losses should be approximately a straight line with slope one. This is also true if we rescale by a premium P. To illustrate, let (S_i, y_i) represent the score and loss for the ith policy and, when rescaled by premium P_i, let $R_i = S_i/P_i$ and $LR_i = y_i/P_i$ be the corresponding relativity and loss ratio. To make homogenous subgroups, we could group the policies by relativity deciles and compare average loss ratios for each decile.

The left-hand panel of Figure 2 shows this comparison for the premium P = "SP_FreqSev_Basic" and score S = "SP_FreqSev". A more primitive comparison of relativities and loss ratios would involve a plot of R_i versus LR_i; however, personal lines insurance typically has many zero losses, rendering such a graph ineffective. For our application, each decile is the average over 35,945 policies, making this comparison reliable. This panel shows a linear relation between the average loss ratio and relativity, indicating that the score SP_FreqSev is a desirable predictor of the loss.

An overall summary of the plot of relativities to loss ratios is analogous to the Gini index calculation. In the former, the relationship of interest is $LR = \frac{y}{P}$ versus R; in the latter, it is $y - P$ versus $rank(R)$. The differences are (a) the rescaling of losses by premiums and (b) the use of rank relativities versus relativities. The Gini index summarizes the entire curve whereas the graphs in this section will allow us to examine relationships "locally," as described below.

Figure 2. Average relativities and loss ratios by groups of scores. Each panel displays a linear relationship. The variability about the relationship increases as the number of bins increases.



Of course, extensive aggregation such as at the decile level may hide important patterns. The middle and right-hand panels of Figure 2 show comparisons for 20 and 50 bins, respectively. In the right-hand panel, each of the 50 bins represents an average of 2% of our hold-out data (= 7,189 records per bin). This panel shows substantial variability between the average relativity and loss ratio, so we consider alternative comparison methods.

Specifically, we use nonparametric regression to assess score performance. Although nonparametric regression is well known in the predictive modeling community (e.g., Hastie, Tibshirani, and Friedman 2001), it is less widely used in actuarial applications. The ideas are straightforward. Consider a set of relativities and loss ratios of the form (R_i, LR_i), $i = 1, \ldots, n$. Suppose that we are interested in a prediction at relativity x. Then, for some neighborhood about x, say, $[x - b, x + b]$, one takes the average loss ratio over all sets whose score falls in that neighborhood. Using notation, we can express this average as

$$\hat{\mathrm{m}}(x) = \frac{\sum_{i=1}^{n} w(x, R_i) LR_i}{\sum_{i=1}^{n} w(x, R_i)}, \qquad (5.1)$$

where the weight function w(x, R_i) is 1 if R_i falls in $[x - b, x + b]$ and 0 otherwise. By taking an average of all those observations with scores that are "close" to $R = x$, we get a good idea as to what one can expect LR to be—that is, $\mathrm{E}(LR|R = x)$, the regression function. It is called "nonparametric" because there is no assumption about a functional form such as linearity.

To see how this works, Figure 3 provides a plot for the basic frequency severity score, SP_FreqSev, and its multi-peril version assuming independence, IND_FreqSev. To calculate the nonparametric fits, this figure is based on $b = 0.1$. For our data, this choice of b (known as a "bandwidth") means that the averages were calculated using at least 13,000 records. For example, at $x = 0.6$, there were 27,492 policies with relativities that fell in the interval [0.5,0.7]. These policies had an average loss ratio of 0.7085, resulting in a deviation of 0.1085. We plot the fits in increments of 0.05 for the value of x, meaning that there is some overlap in adjacent neighborhoods. This overlap is not a concern for estimating average fits, as we are doing here. We plot only relativities in the interval [0.6, 1.6] because the data become sparse outside of that interval. Figure 3 shows that the deviations from IND_FreqSev and SP_FreqSev are comparable; it is difficult to say which score is uniformly better.

Figure 4 provides additional comparisons. The left panel compares the error in IND_FreqSev to one of the instrumental variable alternatives, IV_FreqSevA. Here, the IV_FreqSevA error is smaller for low relativities (0.6 through 0.8) and medium size

Figure 3. Comparison of single and multi-peril frequency-severity loss ratios. The deviations from IND_FreqSev and SP_FreqSev are comparable and it is difficult to say which score is uniformly better.

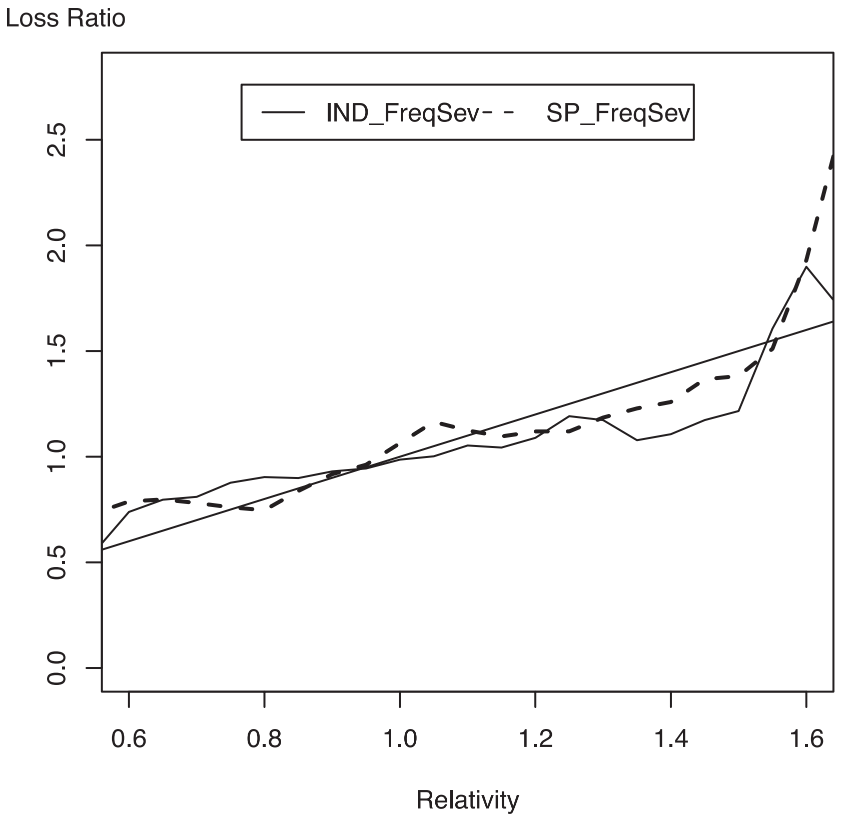
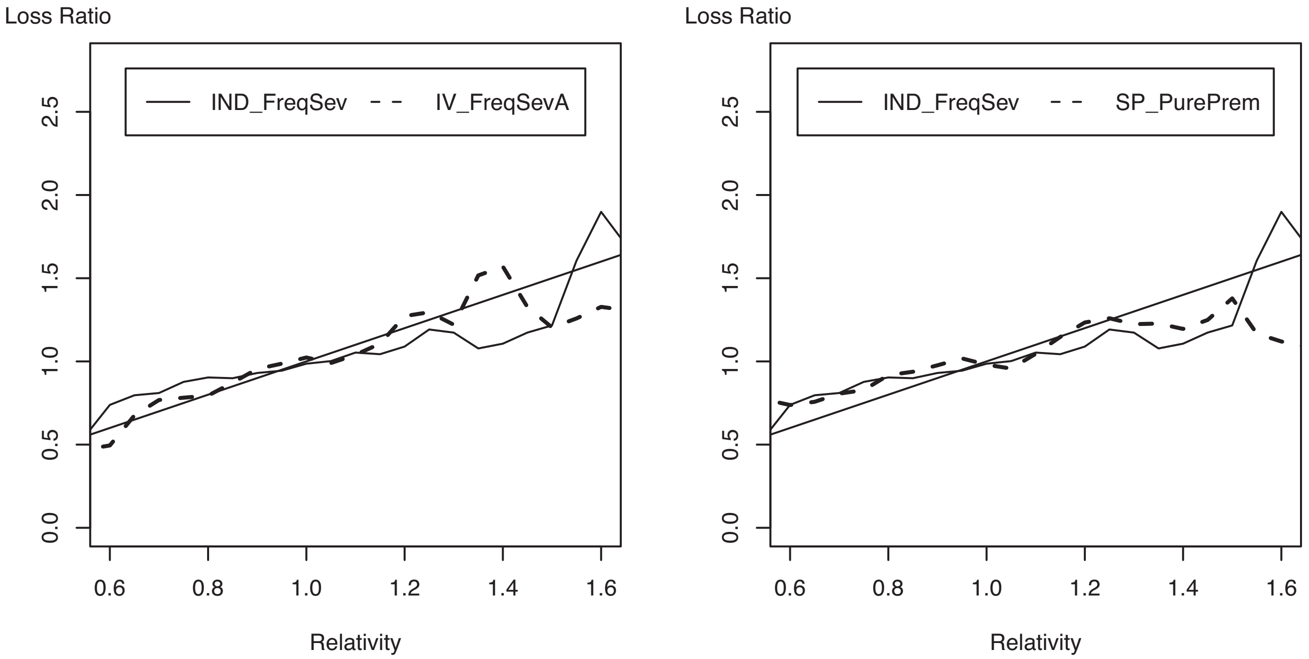


Figure 4. Comparison of loss ratios from several scoring methods. The left panel compares the independence to an instrumental variable frequency-severity approach; the latter is clearly preferred to the former. The right panel compares the independence frequency-severity approach to the single peril pure premium (Tweedie) method. These two measures perform about the same for most of the data.



relativities (1.2 through 1.4) and approximately similar elsewhere (0.8 through 1.2). Based on this comparison, the score IV_FreqSevA is clearly preferred to the score IND_FreqSev.

The right panel of Figure 4 compares the error in IND_FreqSev to the basic pure premium score, SP_PurePrem. This panel shows that these two measures perform about the same for most of the data, suggesting that neither is uniformly superior to the other.

For each illustration, we seek a score that is close to the 45 degree line. For our applications, we interpret $\widehat{m}(x) - x$ to be the deviation when using the relativity R to predict loss ratios LR. Compared to Gini indices, this measure allows us to see the differences between relativities and loss ratios locally over regions of x.

6. Summary and concluding remarks

In this paper, we considered several models for predicting losses for homeowners insurance. The models considered include

- single versus multiple perils, and
- pure premium versus frequency-severity approaches.

Moreover, in the case of multiple perils, we also compared

- independence to instrumental variable models.

The instrumental variable estimation technique is motivated by systems of equations, where the presence and amount of one peril may affect another. We show in Section 3 that instrumental variable estimators accommodate statistically significant relationships that we attribute to associations among perils.

For our data, each accident event was assigned to a single peril. For other databases where an event may give rise to losses for multiple perils, we expect greater association among perils. Intuitively, more severe accidents give rise to greater losses and this severity tendency will be shared among losses from an event. Thus, we conjecture that instrumental variable estimators will be even more helpful for companies that track accident event level data.

This paper applies the instrumental variable estimation strategy to homeowners insurance, where a claim type may be due to fire, liability, and so forth. One could also use this strategy to model homeowners and automobile policies jointly or umbrella policies that consider several coverages simultaneously. As another example, in health care, expenditures are often broken down by diagnostic-related groups.

Although an important contribution of our work is the introduction of instrumental variable techniques to handle dependencies among perils, we do not wish to advocate one technique or approach as optimal in all situations. Sections 2 and 3, as well as Appendix B, introduce many models, each of which has advantages compared to alternatives. For example, models that do not decompose claims by peril have the advantage of relative simplicity and hence interpretability. The "independence" multi-peril models allow analysts to separate claims by peril, thus permitting greater focus in the choice of explanatory variables. The instrumental variable models allow analysts to accommodate associations among perils. When comparing the pure premium to the frequency-severity approaches, the pure premium has the advantage of relative simplicity. In contrast, the frequency-severity has the advantage of permitting greater focus, and hence interpretability, on the choice of explanatory variables.

This paper supplements these qualitative considerations through quantitative comparisons of predictors based on a held-out, validation sample. For our data, we found substantial differences among scoring methods, suggesting that the choice of methods could have an important impact on an insurer's pricing structure. We found that the instrumental variable alternatives provided genuine "lift" compared to baseline multi-peril rating methods that implicitly assume independence, for both the pure premium and frequency-severity approaches. We used nonparametric regression techniques to explore local differences in the scores. Although we did not develop this point extensively, we conjecture that insurers could use the nonparametric techniques to identify regions where one scoring method is superior to an alternative (using covariate information) and possibly develop a next stage "hybrid" score.

References

Arellano, M., *Panel Data Econometrics.* Oxford, U.K.: Oxford University Press, 2003.

Board of Trustees, Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds, *2009 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds,* Washington, D.C.: Government Printing Office, 2009. Available at: http://www.socialsecurity.gov/OACT/TR/2009/tr09.pdf.

Frees, E. W., *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences,* New York: Cambridge University Press, 2004.

Frees, E. W., *Regression Modeling with Actuarial and Financial Applications,* New York: Cambridge University Press, 2010.

Frees, E. W., G. Meyers, and A. D. Cummings, "Dependent Multi-Peril Ratemaking Models," *ASTIN Bulletin* 40(2), 2010, pp. 699–726.

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

Frees, E. W., G. Meyers, and A. D. Cummings, "Insurance Ratemaking and a Gini Index," *Journal of Risk and Insurance* (forthcoming).

Hastie, T., R. Tibshirani, and J. Friedman, *Elements of Statistical Learning: Data Mining, Inference and Prediction.* New York: Springer, 2001.

Hurley, R. L., "Multiple Peril Rating Problems—Some Statistical Considerations," *Proceedings of the Casualty Actuarial Society* 46, 1958, pp. 196–213.

III (Insurance Information Institute), *III Insurance Fact Book,* New York: III, 2010.

Miller, R. B. and D. W. Wichern, *Intermediate Business Statistics: Analysis of Variance, Regression and Time Series,* New York: Holt, Rinehart and Winston, 1977.

Modlin, C., "Homeowners' Modeling." Presentation at the 2005 *Casualty Actuarial Society Seminar on Predictive Modeling,* available at http://www.casact.org/education/specsem/f2005/handouts/modlin.pdf.

Rollins, J. W., "A Modern Architecture for Residential Property Insurance Ratemaking," *Proceedings of the Casualty Actuarial Society* 92, 2005, pp. 486–578.

Tweedie, M. C. K., "An Index which Distinguishes between Some Important Exponential Families" in *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Golden Jubilee International Conference,* eds. J. K. Ghosh and J. Roy, 1984, Calcutta: Indian Statistical Institute, pp. 579–604.

Wooldridge, J. M., *Econometric Analysis of Cross Section and Panel Data,* Cambridge, MA: MIT Press, 2002.

Appendices

A. Summary statistics of the homeowners data

This section displays summary statistics of the frequency portion of the homeowners data, the purpose being to illustrate the dependence among perils. There were relatively few joint claims and so it is difficult to intuitively argue for a severity dependency. Many of these statistics appeared in Frees, Meyers, and Cummings (2010). To keep this paper self-contained, these summary measures are provided here to familiarize readers with our data.

Table 6 gives the number of joint claims among perils. For example, we see that there were only three records that had a Lightning and a Liability claim within the year.

To measure association among perils, Table 7 provides the dependence ratios among perils. A dependence ratio is the ratio of the joint probability to the product of the marginal probabilities. For example, for perils 1 and 2, the dependence ratio is

$$\text{dependence ratio} = \frac{Pr(r_1 = 1,\ r_2 = 1)}{Pr(r_1 = 1)\ Pr(r_2 = 1)}.$$

For example, from Table 6, we would calculate this as

$$\frac{\frac{11}{404664}}{\frac{1254}{404664} \times \frac{2134}{404664}} = 1.663.$$

A dependence ratio equal to one indicates independence among perils.

Table 7 suggests dependence among perils. However, these statistics do not control for the effects of explanatory variables. For example, combinations of explanatory variables that mean a high probability of one peril may also induce a high probability of another peril, thus leading to seeming positive association.

For assessing frequency dependencies in the presence of explanatory variables, recall that r denotes the binary variable that indicates a claim ($y = 1$). Let q_{ij} be the corresponding probability of a claim. The number of claims that is joint between the jth and kth perils is $\Sigma_{i=1}^{n} r_{ij} \times r_{ik}$. Assuming independence among perils, this has mean and variance

$$\mathrm{E}\left(\sum_{i=1}^{n} r_{ij} \times r_{ik}\right) = \sum_{i=1}^{n} q_{ij} \times q_{ik}$$

and

$$\mathrm{Var}\left(\sum_{i=1}^{n} r_{ij} \times r_{ik}\right) = \sum_{i=1}^{n} q_{ij} q_{ik} - \left(q_{ij} q_{ik}\right)^2.$$

Table 6. Joint claim counts among perils

	Fire	Lightning	Wind	Hail	Water Weather	Water NonWeather	Liability	Other	Theft Vand
Lightning	11								
Wind	23	17							
Hail	7	11	23						
WaterWeather	23	12	62	13					
WaterNWeath	27	32	92	43	93				
Liability	4	3	17	3	7	16			
Other	16	18	45	2	18	48	13		
TheftVand	20	25	55	16	38	71	9	31	
Totals	1254	2134	4960	1985	3142	5391	757	1877	3287

Note: Totals refer to all claims from a peril, not just those occurring jointly with another peril.

Table 7. Dependence ratios among perils

	Fire	Lightning	Wind	Hail	Water Weather	Water NonWeather	Liability	Other
Lightning	1.663							
Wind	1.496	1.338						
Hail	1.138	1.051	0.945					
WaterWeath	2.362	0.724	1.610	0.843				
WaterNWeath	1.616	1.126	1.392	1.626	2.222			
Liability	1.705	0.751	1.832	0.808	1.191	1.587		
Other	2.751	1.818	1.956	0.217	1.235	1.920	3.702	
TheftVand	1.963	1.442	1.365	0.992	1.489	1.621	1.464	2.033

To assess dependencies among the claim frequencies, we employ the *t*-statistic

$$t_{jk} = \frac{\sum_{i=1}^{n} r_{ij} \times r_{ik} - \sum_{i=1}^{n} q_{ij} \times q_{ik}}{\sqrt{\sum_{i=1}^{n} q_{ij} q_{ik} - \left(q_{ij} q_{ik}\right)^2}}. \qquad \text{(A.1)}$$

The *t*-statistic in Equation (A.1) would be a standard two-sample *t*-statistic, except that we allow the probability of a claim to vary by policy *i*. To estimate these probabilities, we fit a logistic regression model for each peril *j*, where the explanatory variables are peril-specific. Each model was fit in isolation of the others, thus implicitly using the null hypothesis of independence among perils.

Table 8 summarizes the test statistics for assessing independence among the frequencies. Not surprisingly, the strongest relationship was between water damage due to weather and water damage from causes other than weather. The largest dependence ratio in Table 7, between fire and the "Other" category, was the second largest *t*-statistic—this indicates strong dependence even after covariates are introduced. Interestingly, the only significant negative relationship was between hail and the "Other" category.

For the degrees of freedom of the *t*-statistic, we have followed the usual rule of the number of observations minus the number of parameters. Because our sample size is large ($n = 404{,}664$) relative to the number of parameters, the reference distribution is essentially normal.

B. Notation and baseline models

In a multi-peril model, one decomposes the risk into one of c types ($c = 9$ in Table 1). To set notation, define $r_{i,j}$ to be a binary variable indicating whether or not the ith record has an insurance claim due to

Table 8. Test statistics from logistic regression fits

	Fire	Lightning	Wind	Hail	Water Weather	Water NonWeather	Liability	Other
Lightning	1.472							
Wind	1.662	1.530						
Hail	0.754	0.247	–1.240					
WaterWeath	3.955	–1.166	3.185	–0.100				
WaterNWeath	2.732	0.837	3.369	1.697	7.429			
Liability	1.023	–0.485	2.436	–0.303	0.333	1.825		
Other	4.048	2.229	3.919	–2.616	0.478	4.004	4.929	
TheftVand	3.085	1.816	2.270	–0.235	2.227	3.503	1.147	3.766

the jth type, $j = 1, \ldots, c$. Similarly, $y_{i,j}$ denotes the amount of the claim due to the jth type. To relate the multi- to the single-peril variables, we have

$$r_i = 1 - \left(1 - r_{i,1}\right) \times \cdots \times \left(1 - r_{i,c}\right) \tag{B.1}$$

and

$$y_i = \sum_{j=1}^{c} r_{i,j} \times y_{i,j}. \tag{B.2}$$

We interpret r_i to be a binary variable indicating whether or not the ith policyholder has an insurance claim and y_i describes the amount of the claim, if positive.

Single-Peril Frequency-Severity Model

In homeowners, insurers typically have available many home and a few policyholder characteristics upon which rates are based. For notation, let $\boldsymbol{x}_i$ be a complete set of explanatory variables that is available to the analyst. In the frequency-severity approach, models are specified for both the frequency and severity components. For example, for the frequency component we might fit a logistic regression model with r_i as the dependent variable and $\boldsymbol{x}_{Fi}$ as the set of explanatory variables. Denote the corresponding set of regression coefficients as $\boldsymbol{\beta}_F$. For the severity component, we condition on the occurrence of a claim ($r_i = 1$), and might use a gamma regression model with y_i as the dependent variable and $\boldsymbol{x}_{Si}$ as the set of explanatory variables. Denote the corresponding set of regression coefficients as $\boldsymbol{\beta}_S$. In this paper, we call this the *single-peril frequency-severity model.* Beginning in Section 4, we label the resulting insurance scores as "SP_FreqSev."

Single-Peril Pure Premium Model

An alternative approach is to model the claim amount y_i directly using the entire dataset. Because the distribution of $\{y_i\}_{i=1}^{n}$ contains many zeros (corresponding to no claims) and positive amounts, it is common to use a distribution attributed to Tweedie (1984). This distribution is motivated as a Poisson mixture of gamma random variables. Moreover, because it is a member of the linear exponential family, it may be readily estimated using generalized linear model techniques. In our empirical work, we use a logarithmic link function so that the mean parameter may be written as $\mu_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, thus incorporating all of the explanatory variables. We call this the *single-peril pure premium model.* For readers wishing a review of the Tweedie distribution, see Frees (2010, Chapter 13). We will label the resulting insurance scores as "SP_PurePrem."

Multi-Peril Independence Models

In both the frequency-severity and pure premium approaches, dependent variables can be readily decomposed by peril. From our database, explanatory variables have been selected by peril j for the frequency, $\boldsymbol{x}_{F,i,j}$, and severity, $\boldsymbol{x}_{S,i,j}$, portions, $j = 1, \ldots, 9$. For example, these variables range in number from eight for the Other peril to 19 for the Water Weather peril. A multi-peril frequency-severity approach is:

- For frequency, we fit a logistic regression model with $r_{i,j}$ as the dependent variable and $\boldsymbol{x}_{F,i,j}$ as the set of explanatory variables, with a corresponding set of regression coefficients $\boldsymbol{\beta}_{F,j}$.
- For severity, after conditioning on the occurrence of a claim ($r_{i,j} = 1$), we use a gamma regression model with $y_{i,j}$ as the dependent variable and $\boldsymbol{x}_{S,i,j}$ as the set of explanatory variables, with corresponding set of regression coefficients $\boldsymbol{\beta}_{S,j}$.
- We do this for each peril, $j = 1, \ldots, 9$.

From a modeling point of view, this amounts to assuming that perils are independent of one another and that sets of parameters from each peril are unrelated to one another. Thus, we call these the *"independence" frequency-severity models.* We will label the resulting insurance scores as "IND_FreqSev."

Following a similar set of reasoning, for pure premium modeling we define the union of the frequency $\boldsymbol{x}_{F,i,j}$ and severity $\boldsymbol{x}_{S,i,j}$ variables to be our set of explanatory variables for the jth peril, $\boldsymbol{x}_{i,j}$. With these, one can estimate a pure premium model for each peril, $j = 1, \ldots, 9$. We call these the *"independence" pure* *premium models.* We will label the resulting insurance scores as “IND_PurePrem.”

To compare the basic (single-peril) and independence (multi-peril) models, we look to out-of-sample results beginning in Section 4. In Frees, Meyers, and Cummings (2010), we introduced a multivariate binary model that accounts for dependencies among the peril frequencies. This work established statistical significance among the perils. Thus, for completeness, in Section 4 we include these scores labeled as “DepRatio1” and “DepRatio36”, for 1 and 36 dependency parameters, respectively. Additional details on this method are in Frees, Meyers, and Cummings (2010).

C. The instrumental variables approach

This section provides a brief introduction of the instrumental variable method of estimation that is widely used in econometrics. Our treatment follows that in Frees (2004).

To motivate this approach, consider a classical economic demand and supply problem introduced in Section 3. One estimation approach is to organize all of the dependent variables on the left-hand side and estimate the model using likelihood inference. Specifically, with some algebra, we could rewrite display (2.1) as

$$\boldsymbol{y}_i = \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} = \boldsymbol{B} \begin{pmatrix} \gamma_{10} & \gamma_{11} & 0 \\ \gamma_{20} & 0 & \gamma_{21} \end{pmatrix} \begin{pmatrix} 1 \\ x_{1i} \\ x_{2i} \end{pmatrix} + \boldsymbol{B} \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix}$$

where $\boldsymbol{B} = \begin{pmatrix} 1 & -\beta_1 \\ -\beta_2 & 1 \end{pmatrix}^{-1}$. Then, parameters of this equation could be estimated using, for example, maximum likelihood.

Using instrumental variables is an alternative approach. That is, suppose that theory suggests a linear equation of the form

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i .$$

We may consider an explanatory variable to be endogenous if it is correlated with the disturbance term. Because zero covariance implies zero correlation and because disturbance terms are mean zero, we require only that E $\varepsilon_i \boldsymbol{x}_i = \boldsymbol{0}$ for exogeneity. When not all of the regressors are exogenous, the instrumental variable technique employs a set of variables, $\boldsymbol{w}_i$, that are correlated with the regressors specified in the structural model. Specifically, we assume

$$\mathrm{E}\varepsilon_i \boldsymbol{w}_i = \mathrm{E}\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)\boldsymbol{w}_i = \boldsymbol{0}$$

for the instruments to be exogenous. With these additional variables, an instrumental variable estimator of $\boldsymbol{\beta}$ is of the form $\boldsymbol{b}_{IV} = (\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{W}}\boldsymbol{y}$. Here, $\boldsymbol{P}_{\boldsymbol{W}} = \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'$ is a projection matrix and $\boldsymbol{W} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$ is the matrix of instrumental variables.

In many situations, instrumental variable estimators can be easily computed using two-stage least squares. In the first stage, one regresses each endogenous regressor on the set of exogenous explanatory variables and calculates fitted values of the form $\widehat{\boldsymbol{X}} = \boldsymbol{P}_{\boldsymbol{W}}\boldsymbol{X}$. In the second stage, one regresses the dependent variable on the fitted values using ordinary least squares to get the instrumental variable estimator, that is, $\boldsymbol{b}_{IV} = (\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1}\widehat{\boldsymbol{X}}'\boldsymbol{y}$.

As noted in Section 3, there are conditions on the instruments. Typically, they may include a subset of **x** but must also include additional variables. For example, if they did not include additional variables, then linear combinations of instruments yield perfect linear combinations of **x**, resulting in perfect collinearity and non-identifiability of the coefficients. Further, the new explanatory variables in **w** must also be exogenous (unrelated to ε), otherwise we have done nothing to solve our initial problem.

Instrumental variables are employed when there are (1) systems of equations, (2) errors in variables, and (3) omitted variables. For the applications in this paper, we will use this concept for both systems of equations and for omitted variables. Extensions to non-linear systems are readily available in standard econometric texts, including Arellano (2003) and Wooldridge (2002).

D. Frequency and severity instrumental variable estimation

Section 2 describes the instrumental variable approach focusing on the frequency portion of the model. We also found that fitted probabilities of a peril help to predict the severity *from that peril* (and, vice versa, fitted severities help to predict probabilities). To provide intuition, we focus on the severity model to begin and, as we will see, we will be able to easily reverse the roles of frequency and severity. In our database, we have a variable "base cost loss costs" that we use to approximate $PREM_j$, pure premium, in our empirical work.

- Pure premium is expected frequency times severity, that is, $PREM_j = \pi_j \times \mathrm{E}\, y_j$.
- This suggests that a good explanatory variable for the severity portion is $PREM_j/\pi_j$.
- Of course, we do not know π_j but can estimate it from a stage 1 regression as, say, $\hat{\pi}_j$.
- Because we use a log-link function, this suggests including $\ln(PREM_j/\hat{\pi}_j)$. Often, logarithmic base cost loss costs are already in the regression, so we include $\ln \hat{\pi}_j$ as a predictor of severity.

An interesting aspect of this logic is that the instrumental variable approach provides motivation for using frequency to predict severity.

Now, reverse the roles of frequency and severity—include $\ln \widehat{\mathrm{E}\, y_j}$ as a predictor of frequency. We remark that when one does this, it is not quite as clean an argument because we typically use the logit link with logistic model. However, for small probabilities, these two are quite close and so a log-fitted severity works well at this stage.

We summarize the procedure as follows.

- Stage 1. Compute independence frequency and severity model fitted values. Specifically, for each of the $j = 1, \ldots, 9$ perils:
  - 1a. Fit a logistic regression model using the explanatory variables $\boldsymbol{x}_{F,i,j}$. These explanatory variables differ by peril j. Calculate fitted values to get predicted probabilities, denoted as $\hat{\pi}_{IND,i,j}$.
  - 1b. Fit a gamma regression model using the explanatory variables $\boldsymbol{x}_{S,i,j}$ with a logarithmic link function. These explanatory variables may differ by peril and from those used in the frequency model. Calculate fitted values to get predicted severities (by peril), denoted as $\widehat{\mathrm{E}\, y}_{IND,i,j}$.
- Stage 2. Incorporate additional instruments into the model estimation. Specifically, for each of the $j = 1, \ldots, 9$ perils:
  - 2a. Fit a logistic regression model using
    - the explanatory variables $\boldsymbol{x}_{F,i,j}$,
    - the logarithm of the predicted probabilities developed in step 1(a), $\ln \hat{\pi}_{IND,i,j}$, $k = 1, \ldots, 9$, $k \neq j$ and
    - the logarithm of the fitted values in step 1(b), $\ln \hat{\pi}_{IND,i,j}$.
  - 2b. Fit a gamma regression model using
    - the explanatory variables xS,i,j and
    - the logarithm of the fitted predicted probabilities in step 1(a), $\ln \hat{\pi}_{IND,i,j}$.