Predictive Modeling of Multi-Peril Homeowners Insurance

Edward W. (Jed) Frees,

Glenn Meyers and Dave Cummings

University of Wisconsin – Madison and ISO Innovative Analytics

March, 2011

Outline

Homeowners Insurance

- Homeowners represents a large segment of the personal property and casualty (general) insurance business
- In the US, premiums are over $57 billions of US dollars (*I.I.I. Insurance Fact Book 2010*)
  - This is 13.6% of all property and casualty insurance premiums
  - This is 26.8% of personal lines insurance.
- It is difficult to think about buying a house without purchasing homeowners insurance
- Homeowners is typically sold as an all-risk policy, which covers all causes of loss except those specifically excluded.

Perils of Homeowners Insurance

- Many actuaries interested in pricing homeowners insurance are now decomposing the risk by *peril*, or cause of loss (e.g., Modlin, 2005).

Perils of Homeowners Insurance

- Many actuaries interested in pricing homeowners insurance are now decomposing the risk by *peril*, or cause of loss (e.g., Modlin, 2005).
- Decomposing risks by peril is not unique to personal lines insurance nor is it new.
  - Customary in population projections to study mortality by cause of death (e.g. Board of Trustees, 2009).
  - Robert Hurley (Hurley, 1958) discussed statistical considerations of multiple peril rating in the context of homeowner 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.*

Perils of Homeowners Insurance

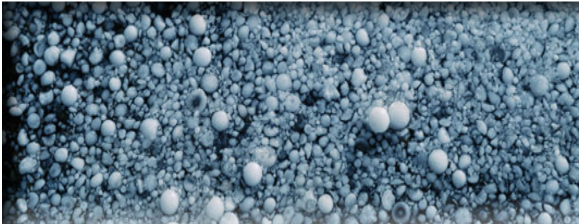
- Many actuaries interested in pricing homeowners insurance are now decomposing the risk by *peril*, or cause of loss (e.g., Modlin, 2005).
- Decomposing risks by peril is not unique to personal lines insurance nor is it new.
  - Customary in population projections to study mortality by cause of death (e.g. Board of Trustees, 2009).
  - Robert Hurley (Hurley, 1958) discussed statistical considerations of multiple peril rating in the context of homeowner 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.*
- Rollins (2005) - multi-peril rating is critical for maintaining economic efficiency and actuarial equity.

Perils of Homeowners Insurance

- Many actuaries interested in pricing homeowners insurance are now decomposing the risk by *peril*, or cause of loss (e.g., Modlin, 2005).
- Decomposing risks by peril is not unique to personal lines insurance nor is it new.
  - Customary in population projections to study mortality by cause of death (e.g. Board of Trustees, 2009).
  - Robert Hurley (Hurley, 1958) discussed statistical considerations of multiple peril rating in the context of homeowner 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.*
- Rollins (2005) - 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 others.
  - Example - "dwelling in an urban area" may be an excellent predictor for the theft peril but provide little useful information for the hail peril.

Some Perils - Hail

- What Is Hail?
  - a large frozen raindrop produced by intense thunderstorms
    - As the snowflakes fall, liquid water freezes onto them, forming ice pellets that will continue to grow as more and more droplets accumulate.
    - Upon reaching the bottom of the cloud, some of the ice pellets are carried by the updraft back up to the top of the storm.
    - As the ice pellets once again fall through the cloud, another layer of ice is added and the hail stone grows even larger.
  - The Largest Hailstone
    - Recorded fell in Coffeyville, Kansas, on September 3, 1970.
    - It measured about 17.5 inches in circumference (over 5.6 inches in diameter) and weighed more than 26 ounces (almost 2 pounds)!
    - Most hail is small – usually less than two inches in diameter.



Some Perils - Lightning

- What is Lightning?
  - Lightning is caused by the attraction between positive and negative charges in the atmosphere, resulting in the buildup and discharge of electrical energy.
    - Twenty percent of lightning strike victims die and 70% of survivors suffer serious long-term after-effects.

Some Perils - Fire

Some Perils - Wind

Source: Federal Alliance for Safe Homes (http://www.flash.org/)

Sample Selection

- We drew a random sample of size $n = 404,664$ from a homeowners database maintained by the ISO Innovative Analytics.
  - This database contains over 4.2 million policyholder years.
  - Based on the policies issued by several major insurance companies in the US, thought to be representative of most geographic areas.
- For covariates, there 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.
- See the web site http://www.iso.com/Products/ISO-Risk-Analyzer/ISO-Risk-Analyzer-
for more info.

9 Perils in Homeowners Insurance

Table: Summarizing 404,664 Policy-Years

Peril (j)	Frequency (in percent)	Number of Claims	Median Claims
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
WaterNonWeather	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

Types of Models

- Single Cause of Loss (Single-Peril)
  - Frequency-Severity
  - Pure Premium
- Multiple Causes of Loss (Multi-Peril)
  - Independent Perils
    - Frequency-Severity
    - Pure Premium
  - Models of Dependence
    - Instrumental Variables
    - Alternative Approaches

Single-Peril Models

- Some notation
  - y_i - describes the amount of the loss.
  - $\mathbf{x}_i$ - the complete set of explanatory variables.
  - r_i - a binary variable indicating whether or not the ith subject has a loss.

Single-Peril Models

- Some notation
  - y_i - describes the amount of the loss.
  - $\mathbf{x}_i$ - the complete set of explanatory variables.
  - r_i - a binary variable indicating whether or not the ith subject has a loss.
- Pure Premium (Tweedie) Modeling Strategy:
  - y_i is the dependent variable, $\mathbf{x}_i$ is the set of explanatory variables.
  - Loss distribution contains many zeros (corresponding to no claims) and positive amounts
  - Tweedie distribution - motivated as a Poisson mixture of gamma random variables.
  - Readily estimated using generalized linear model (GLM) techniques
  - Logarithmic link function - the mean parameter may be written as $\mu_i = \exp(\mathbf{x}_i'\beta)$.

Single-Peril Models

- Some notation
  - y_i - describes the amount of the loss.
  - $\mathbf{x}_i$ - the complete set of explanatory variables.
  - r_i - a binary variable indicating whether or not the ith subject has a loss.
- Pure Premium (Tweedie) Modeling Strategy:
  - y_i is the dependent variable, $\mathbf{x}_i$ is the set of explanatory variables.
  - Loss distribution contains many zeros (corresponding to no claims) and positive amounts
  - Tweedie distribution - motivated as a Poisson mixture of gamma random variables.
  - Readily estimated using generalized linear model (GLM) techniques
  - Logarithmic link function - the mean parameter may be written as $\mu_i = \exp(\mathbf{x}_i'\beta)$.
- Frequency-Severity (Two-Part Models) Modeling Strategy:
  - Use a binary regression model with r_i as the dependent variable and $\mathbf{x}_{1i}$ as the set of explanatory variables. (Typical models: logit, probit).
  - Conditional on $r_i = 1$, specify a regression model with y_i as the dependent variable and $\mathbf{x}_{2i}$ as the set of explanatory variables. (Typical models: lognormal, gamma).

Multi-Peril Independence Frequency Severity

- Decompose the risk into one of 9 types.
  - r_{ij} - binary variable to indicate a claim due to the jth type, $j = 1, \ldots, c$.
  - y_{ij} - the amount of the claim due to the jth type.
- Explanatory variables selected by peril j for the frequency, $\mathbf{x}_{F,i,j}$, and severity, $\mathbf{x}_{S,i,j}$, portions, $j = 1, \ldots, 9$.
  - For example, these variables range in number from eight for the Other peril to nineteen for the Water Weather peril.

Multi-Peril Independence Frequency Severity

- Decompose the risk into one of 9 types.
  - r_{ij} - binary variable to indicate a claim due to the jth type, $j = 1, \ldots, c$.
  - y_{ij} - the amount of the claim due to the jth type.
- Explanatory variables selected by peril j for the frequency, $\mathbf{x}_{F,i,j}$, and severity, $\mathbf{x}_{S,i,j}$, portions, $j = 1, \ldots, 9$.
  - For example, these variables range in number from eight for the Other peril to nineteen for the Water Weather peril.
- Modeling Strategy
  - Frequency - a logistic regression model with $r_{i,j}$ as the dependent variable and $\mathbf{x}_{F,i,j}$ as the set of explanatory variables, with corresponding set of regression coefficients $\beta_{F,j}$.
  - Severity - gamma regression model with $y_{i,j}$ as the dependent variable and $\mathbf{x}_{S,i,j}$ as the set of explanatory variables, with corresponding set of regression coefficients $\beta_{S,j}$.
  - We do this for each peril, $j = 1, \ldots, 9$.

Multi-Peril Independence Frequency Severity

- Decompose the risk into one of 9 types.
  - r_{ij} - binary variable to indicate a claim due to the jth type, $j = 1, \ldots, c$.
  - y_{ij} - the amount of the claim due to the jth type.
- Explanatory variables selected by peril j for the frequency, $\mathbf{x}_{F,i,j}$, and severity, $\mathbf{x}_{S,i,j}$, portions, $j = 1, \ldots, 9$.
  - For example, these variables range in number from eight for the Other peril to nineteen for the Water Weather peril.
- Modeling Strategy
  - Frequency - a logistic regression model with $r_{i,j}$ as the dependent variable and $\mathbf{x}_{F,i,j}$ as the set of explanatory variables, with corresponding set of regression coefficients $\beta_{F,j}$.
  - Severity - gamma regression model with $y_{i,j}$ as the dependent variable and $\mathbf{x}_{S,i,j}$ as the set of explanatory variables, with corresponding set of regression coefficients $\beta_{S,j}$.
  - We do this for each peril, $j = 1, \ldots, 9$.
- Modeling - equivalent to assuming that
  - perils are independent of one another and
  - that sets of parameters from each peril are unrelated to one another.
- We call these the *"independence" frequency-severity models*

Multi-Peril Independence Pure Premium Model

- For each peril, $j = 1, \ldots, 9$, we:
  - y_{ij} is the dependent variable
  - Define the union of the frequency $\mathbf{x}_{F,i,j}$ and severity $\mathbf{x}_{S,i,j}$ variables to be our set of explanatory variables for the jth peril, $\mathbf{x}_{i,j}$
  - Fit the model using generalized linear model (GLM) techniques with
  - Logarithmic link function - the mean parameter may be written as $\mu_{i,j} = \exp(\mathbf{x}'_{i,j}\beta_j)$.
- We call these the *"independence" pure premium models*.

Dependencies Among Perils

- Current actuarial practice involves modeling each peril in isolation of the others.
  - Use a set of variables $\mathbf{x}_{1,j}$ to predict the frequency and
  - another a set $\mathbf{x}_{2,j}$ to predict the severity for each peril, $j = 1, \ldots, c$.
- This amounts to assuming that perils are independent of one another
- We anticipate dependence among perils
  - Event classification can be ambiguous (e.g., fires triggered by lightning)
  - Unobserved latent characteristics of policyholders (cautious homeowners who are sensitive to potential losses due to theft-vandalism and liability) may induce dependencies among perils

Dependencies - Empirical Evidence

- We found substantial evidence of dependencies among frequencies - less evidence among severities

Dependencies - Empirical Evidence

- We found substantial evidence of dependencies among frequencies - less evidence among severities
- To see this, for $j = 1, \ldots, 9$,
  - Run a logistic regression model for each peril.
  - Calculate fitted probabilities $\hat{q}_{ij}$ - estimates of the probability of a claim for policyholder i, peril j
  - Number of *joint* claims (jth and kth perils) = $\sum_{i=1}^n \ r_{ij} \times r_{ik}$.
    - 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

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

  - To assess dependencies, use a t-statistic

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

- This t-statistic is a standard two-sample t-statistic *except* that we allow the probability of a claim to vary by policy i.

Dependencies - Empirical Evidence

Table: Test Statistics From Logistic Regression Fits

	Fire	Lightning	Wind	Hail	Water Weather	Water Non Weather	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

Strong statistical evidence of dependencies!!

Instrumental Variables

Instrumental variable (IV) estimation is a classic econometric technique.

Here is a quick overview of the basic idea.

- Suppose that theory suggests a linear model :

$$y_1 = \mathbf{x}'\beta_1 + \beta_2 y_2 + \varepsilon$$

- Ordinary least squares is not available because y_2 is related to ε

Instrumental Variables

Instrumental variable (IV) estimation is a classic econometric technique.

Here is a quick overview of the basic idea.

- Suppose that theory suggests a linear model :

$$y_1 = \mathbf{x}'\beta_1 + \beta_2 y_2 + \varepsilon$$

- Ordinary least squares is not available because y_2 is related to ε
- The instrumental variable strategy
  - assumes that you have available "instruments" **w** to approximate y_2
  - First stage: Run a regression of **w** on y_2 to get fitted values for y_2 of the form $\mathbf{w}'\mathbf{g}$
  - Second stage: Run a regression of **x** and $\mathbf{w}'\mathbf{g}$ on y_1

Instrumental Variables

Instrumental variable (IV) estimation is a classic econometric technique.

Here is a quick overview of the basic idea.

- Suppose that theory suggests a linear model :

$$y_1 = \mathbf{x}'\beta_1 + \beta_2 y_2 + \varepsilon$$

- Ordinary least squares is not available because y_2 is related to ε
- The instrumental variable strategy
  - assumes that you have available "instruments" **w** to approximate y_2
  - First stage: Run a regression of **w** on y_2 to get fitted values for y_2 of the form $\mathbf{w}'\mathbf{g}$
  - Second stage: Run a regression of **x** and $\mathbf{w}'\mathbf{g}$ on y_1
- There are conditions on the instruments. Typically, they may include a subset of **x** but must also include additional variables.
- Instrumental variables are employed when there are (1) systems of equations, (2) errors in variables and (3) omitted variables.

Instrumental Variables Approach to Dependence Modeling

- First consider the distribution of r_1
  - We believe that $r_2, \ldots, r_9$ may affect the distribution of r_1
  - The variables $r_2, \ldots, r_9$ are not sensible explanatory variables but we can use *estimates* of them.

Instrumental Variables Approach to Dependence Modeling

- First consider the distribution of r_1
  - We believe that $r_2, \ldots, r_9$ may affect the distribution of r_1
  - The variables $r_2, \ldots, r_9$ are not sensible explanatory variables but we can use *estimates* of them.
- Here is an outline of our proposed procedure:
  - For each of the nine perils
    - Fit a logistic regression model using an initial set of explanatory variables. These explanatory variables differ by peril.
    - Calculate fitted values to get predicted probabilities (by peril).
  - For each of the nine perils, fit a logistic regression model using
    - the initial set of explanatory variables and
    - the logarithmic predicted probabilities developed above.
- The paper contains extensions to incorporate severities

IV Pure Premium Model Coefficients

Table: 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

- The additional variables are statistically significant for each peril.
- This is just 3 of the 9 perils. Others are in the appendix.

Homeowners Data

- The “gold standard” in predictive modeling is model validation through examining performance of an independent held-out sample of data (e.g., Hastie, Tibshirani and Friedman, 2001)

Homeowners Data

- The "gold standard" in predictive modeling is model validation through examining performance of an independent held-out sample of data (e.g., Hastie, Tibshirani and Friedman, 2001)
- We drew two random samples from a homeowners database maintained by the Insurance Services Office.
- Our in-sample, or "training," dataset consists of a representative sample of 404,664 records taken from this database.
  - We estimated several competing models from this dataset
- We use a held-out, or "validation" subsample of 359,454 records, whose claims we wish to predict.
  - We present 8 scores that were calculated using the estimated models from the in-sample data and the explanatory variables from the held-out sample
  - The paper includes additional scoring methods

Scores from the Homeowners Example

Score	*Description*
	Basic, Single-peril
BasicFS	Frequency and Severity model
BasicTweedie	Pure premium (Tweedie) model
INDFreqSev	Multi-peril Frequency and Severity model
	Assumes independence among perils
	Instrumental Variable Multi-peril Frequency and Severity models
IVFreqSevA	Uses instruments for frequency component
IVFreqSevB	Uses instruments for severity component
IVFreqSevC	Uses instruments for frequency and severity components
	Multi-peril pure premium (Tweedie) models
INDTweedie	Assumes independence among perils
IVTweedie	Instrumental Variable version

Out-of-Sample Results

- Figure 1 emphasizes that there are important differences among scoring methods
- The paper documents several methods for comparing scores to held-out losses
  - This presentation focuses on the “Gini” index

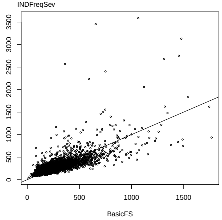


Figure: Single versus Multi-Peril Frequency-Severity Scores. This graph is based on a 1 in 100 random sample of size 3,594. The correlation coefficient is only 79.4%.

Gini Results from the Homeowners Example

Base Premium	Comparison Score: Basic FS	Basic TW	IND FreqSev	IVFreqSev A	IVFreqSev B	IVFreqSev C	IND Tweedie	IV Tweedie	*Maxima*
ConsPrem	28.81	28.11	28.00	29.42	28.18	29.44	28.46	28.42	29.44
BasicFS	-	4.41	7.15	9.15	7.32	9.09	9.25	9.49	9.49
BasicTW	9.13	-	8.55	10.31	8.79	10.53	9.68	9.54	10.53
INDFreqSev	11.28	8.99	-	10.47	4.42	10.26	9.55	11.09	11.28
IVFreqSevA	7.15	3.98	-2.27	-	-2.15	1.93	4.48	5.07	7.15
IVFreqSevB	11.03	8.52	-1.62	10.13	-	9.92	8.87	10.32	11.03
IVFreqSevC	7.43	3.89	-0.91	0.82	-1.68	-	4.50	4.55	7.43
INDTweedie	8.57	6.82	4.20	7.40	4.25	7.30	-	3.66	8.57
IVTweedie	8.38	6.58	5.40	7.21	5.55	7.50	4.11	-	8.38

- Standard errors are about 1.4 for each Gini coefficient
- When constant exposure is the base, all of the comparison scores do so well it is difficult to distinguish among them

Gini Results from the Homeowners Example

Base Premium	Comparison Score: Basic FS	Basic TW	IND FreqSev	IVFreqSev A	IVFreqSev B	IVFreqSev C	IND Tweedie	IV Tweedie	*Maxima*
ConsPrem	28.81	28.11	28.00	29.42	28.18	29.44	28.46	28.42	29.44
BasicFS	-	4.41	7.15	9.15	7.32	9.09	9.25	9.49	9.49
BasicTW	9.13	-	8.55	10.31	8.79	10.53	9.68	9.54	10.53
INDFreqSev	11.28	8.99	-	10.47	4.42	10.26	9.55	11.09	11.28
IVFreqSevA	7.15	3.98	-2.27	-	-2.15	1.93	4.48	5.07	7.15
IVFreqSevB	11.03	8.52	-1.62	10.13	-	9.92	8.87	10.32	11.03
IVFreqSevC	7.43	3.89	-0.91	0.82	-1.68	-	4.50	4.55	7.43
INDTweedie	8.57	6.82	4.20	7.40	4.25	7.30	-	3.66	8.57
IVTweedie	8.38	6.58	5.40	7.21	5.55	7.50	4.11	-	8.38

- The relativities are based on ratios of scores
  - The two-sample test shows that relativities based on differences of scores are statistically indistinguishable - we need not consider both
- The two-sample test shows that the IVFreqSevB performs more poorly than "A" and "C" on a number of tests - not a viable candidate
- A “mini-max” strategy for selecting a score suggests that IVFreqSevA is our top performer.

Concluding Remarks

- We examined other types of multivariate frequency models, including alternating logistic regressions and dependence ratio models. See Frees, Meyers and Cummings (2010, *Astin Bulletin*). These did not fare as well.

Concluding Remarks

- We examined other types of multivariate frequency models, including alternating logistic regressions and dependence ratio models. See Frees, Meyers and Cummings (2010, *Astin Bulletin*). These did not fare as well.
- The instrumental variable estimation technique is motivated by systems of equations, where the presence and amount of one peril may affect another.

Concluding Remarks

- We examined other types of multivariate frequency models, including alternating logistic regressions and dependence ratio models. See Frees, Meyers and Cummings (2010, *Astin Bulletin*). These did not fare as well.
- The instrumental variable estimation technique is motivated by systems of equations, where the presence and amount of one peril may affect another.
- 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.
  - We conjecture that instrumental variable estimators will be even more helpful for companies that track accident event level data.
  - This is also true for other lines of business, e.g., personal auto.

Concluding Remarks

- Incorporating dependencies into pricing structure can provide substantial additional predictive abilities.

Concluding Remarks

- Incorporating dependencies into pricing structure can provide substantial additional predictive abilities.
- One could also use this strategy to model homeowners and automobile policies jointly or umbrella policies, that consider several coverages simultaneously.

Some References

Papers are available at

`http://research3.bus.wisc.edu/jfrees`

- Dependent Multi-Peril Ratemaking Models, by EW Frees, G. Meyers and D. Cummings, 2010. To appear in *Astin Bulletin: Journal of the International Actuarial Association*
- Summarizing Insurance Scores Using a Gini Index, by EW Frees, G. Meyers and D. Cummings, 2010. To appear in *Journal of the American Statistical Association*.
- Predictive Modeling of Multi-Peril Homeowners Insurance, by EW Frees, G. Meyers and D. Cummings, 2011. Approved by the Casualty Actuarial Society's Ratemaking Committee. Submitted to *Variance*.
- *Regression Modeling with Actuarial and Financial Applications*, Cambridge University Press (2010), by EW Frees. Support materials available at `http://research.bus.wisc.edu/RegActuaries`.

Instrumental Variable Pure Premium Model Coefficients

Table: 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		

Model

1. Use a multivariate binary regression model with $\mathbf{r}_i = (r_{i,1}, \ldots, r_{i,c})'$ as the dependent variable.
2. Conditional on the frequency $\mathbf{r}_i$, for the severity we specify a multivariate regression with $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,c})'$ as the dependent variable.

Multivariate Severity Models

- Marginal distributions
  - For all perils j, gamma regressions with a logarithmic link
  - Differing for each peril j, explanatory variables $\mathbf{x}_{2i,j}$, regression parameters β_{2j} and scale parameters $scale_j$.
- Association, use a gaussian (normal) copula

$$\mathrm{cop}_N(u_1,\ldots,u_c) = \phi_N\left(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_c)\right)\prod_{j=1}^{c}\frac{1}{\phi(\Phi^{-1}(u_j))}.$$

  - Φ and ϕ are the standard normal distribution and density functions.
  - The multivariate normal density is

$$\phi_N(\mathbf{z}) = \frac{1}{(2\pi)^{c/2}\sqrt{\det\Sigma}}\exp\left(-\frac{1}{2}\mathbf{z}'\Sigma^{-1}\mathbf{z}\right).$$

  - The matrix Σ is a correlation matrix, with ones on the diagonal.
- For a single association parameter, the maximum likelihood estimator turned out to be 0.0746 with a t-statistic = 3.256, positively statistically significant.
- For other specifications, there are not enough joint claims to model the association among severities in a significant fashion.

IV Approach in Severity

- Here is a way to incorporate pure premiums, say $PREM_j$, that may vary by peril
  - In our data work, we will use base cost loss costs to approximate $PREM_j$.
- The IV approach provides motivation for using frequency to predict severity:
  - 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 from a stage 1 regression as, say, $\widehat{\pi}_j$
  - Because we use a log-link function, this suggests including $\ln(PREM_j/\widehat{\pi}_j)$. Often, logarithmic base cost loss cost are already in the regression, so
- Include $\ln \widehat{\pi}_j$ as a predictor of severity.
- Now, reverse the roles of frequency and severity – include $\ln \widehat{\mathrm{E}\ y_j}$ as a predictor of frequency.

Summary of IV Approach

- 1. Stage 1 - For each of the nine perils:
  - 1a. Fit a logistic regression model using an initial set of explanatory variables. These explanatory variables differ by peril. Calculate fitted values to get predicted probabilities (by peril).
  - 1b. Fit a gamma regression model using an initial set of explanatory variables with a logarithmic link function. These explanatory variables differ by peril and differ from those used in the frequency model. Calculate fitted values to get predicted severities (by peril).
- 2. Stage 2 - For each of the nine perils:
  - 2a. Fit a logistic regression model using
    - (i) an initial set of explanatory variables ,
    - (ii) the logarithm of the predicted probabilities developed in step 1(a) and
    - (iii) the logarithm of the fitted values in step 1(b).
  - 2b. Fit a gamma regression model using
    - (i) an initial set of explanatory variables and
    - (ii) the logarithm of the fitted values in step 1(a).