Modeling Loss Data: Endorsements and Portfolio Management

Travelers

Edward W. (Jed) Frees

University of Wisconsin – Madison

11 November 2016
Outline

1. Resources for Introductions to Insurance Analytics
   - Book Series
   - Insurance Company Analytics

2. Rating Endorsements

3. Portfolio Management
   - Unbundling of Coverages
   - Gini Statistics
   - Portfolio Risk Retention

4. Concluding Remarks
Research Team
In this first part of the talk, I intend to

- Introduce a resource for actuaries wishing to learn more about analytics
  - A two volume series, published by *Cambridge University Press*.
- Introduce a resource for statisticians/machine learners/financial engineers wishing to learn more about insurance company operations
Welcome!

This is the new website for Predictive Modeling Applications in Actuarial Science, a two volume series that we are creating.

This website currently focuses on Volume 1. We provide content preview, data (txt or csv format) and R code (R format) for each chapter here. Contact authors for further information about data and code.
I'm an actuary. Look it up.
# Table of Contents

1. Predictive Modeling in Actuarial Science

## Predictive Modeling Foundations

2. Overview of Linear Models
3. Regression with Categorical Dependent Variables
4. Regression with Count Dependent Variables
5. Generalized Linear Models
6. Frequency and Severity Models

## Predictive Modeling Methods

7. Longitudinal and Panel Data Models
8. Linear Mixed Models
9. Credibility and Regression Modeling
10. Fat-Tail Regression Models
11. Spatial Statistics
12. Unsupervised Learning

## Bayesian and Mixed Modeling

13. Bayesian Computational Methods
14. Bayesian Regression Models
15. Generalized Additive Models and Nonparametric Regression
16. Non-Linear Mixed Models

## Longitudinal Modeling

17. Time Series Analysis
18. Claims Triangles/Loss Reserves
19. Survival Models
20. Transition Modeling
Volume II: Case Studies in Insurance

Table of Contents

1. Pure Premium Modeling Using Generalized Linear Models
2. Applying Generalized Linear Models to Insurance Data: Frequency-Severity Versus Pure Premium Modeling
3. GLMs as Predictive Claim Models
4. Framework for General Insurance Ratemaking: Beyond the Generalized Linear Model
5. Using Multilevel Modeling for Group Health Insurance Ratemaking: A Case Study from the Egyptian Market
6. Clustering in General Insurance Pricing
8. The Predictive Distribution of Loss Reserve Estimates Over a Finite Time Horizon
9. Finite Mixture Model and Workers Compensation Large Loss Regression Analysis
10. A Framework for Managing Claim Escalation Using Predictive Modeling
11. Predictive Modeling for Usage-Based Auto Insurance
We coordinated and co-authored this two volume set, published by Cambridge University Press, that provides the foundations of statistical modeling for actuaries interested in learning about predictive analytics.

- I am a co-Editor, along with Glenn Meyers and Richard Derrig.
- Authors from 7 different countries.

**Book URL** [http://research.bus.wisc.edu/PredModelActuaries](http://research.bus.wisc.edu/PredModelActuaries)

- Volume 2, on case studies, appeared this Fall.
What is Analytics?


- Insurance is a data-driven industry – analytics is a key to deriving information from data.
- But what is analytics?

- Insurance is a data-driven industry – analytics is a key to deriving information from data.
- But what is analytics? Some alternative descriptors:
  - “business intelligence” may focus on processes of collecting data, often through databases and data warehouses
  - “business analytics” utilizes tools and methods for statistical analyses of data
  - “data science” can encompass broader applications in many scientific domains

- Insurance is a data-driven industry – analytics is a key to deriving information from data.
- But what is analytics? Some alternative descriptors:
  - “business intelligence” may focus on processes of collecting data, often through databases and data warehouses
  - “business analytics” utilizes tools and methods for statistical analyses of data
  - “data science” can encompass broader applications in many scientific domains
- **Analytics** – the process of using data to make decisions.
  - This process involves gathering data, understanding models of uncertainty, making general inferences, and communicating results.
Led by statistician W. Edwards Deming, an earlier generation sought to utilize quality improvement techniques to improve business processes, resulting in the field now known as “total quality management.”
What is Analytics?

- Led by statistician W. Edwards Deming, an earlier generation sought to utilize quality improvement techniques to improve business processes, resulting in the field now known as “total quality management.”
- Analytics continues to enjoy increasing popularity among businesses.
Why “Predictive”? 
Why “Predictive”?

- Statisticians think about the traditional triad of inference: hypothesis testing, parameter estimation, and prediction.
- In insurance, predictions are useful for existing risks in future periods as well as not yet observed risks in a current period.

Figure: Predictive Features of Insurance Analytics, Norberg (1979).
Analytics can feed into an insurance company at three levels. These are:

- Individual Insurance Processes
- Insurance Company Operations
- Insurance Company Enterprise

Naturally, the three are highly inter-connected.
One way to describe the operations of a company that sells insurance products is to adopt a granular approach, that is, a “micro” oriented view, thinking specifically about what happens to a contract at various stages of its existence within a company.

**Figure:** Timeline of a Typical Insurance Policy. Arrows mark the occurrences of random events.
Insurance Processes

- Claims process

Figure: Development of a Typical Claim.
Another way is to aggregate detailed insurance processes into larger “operational” units that many companies use as functional areas to segregate employee activities and areas of responsibilities. Consider the following:

- **Initial Underwriting and Ratemaking**
  - Offer right price for the right risk
  - Avoid adverse selection
- **Renewal Underwriting and Ratemaking**
  - Retain profitable customers longer
  - Update prices using experience
- **Claims and Product Management**
- **Reserving**
- **Capital Allocation and Solvency**
Insurance Company Operations

- Initial Underwriting and Ratemaking
- Renewal Underwriting and Ratemaking
- Claims and Product Management
  - Detect and manage claims fraud
  - Manage claims costs
  - Understand excess layers for reinsurance and retention.
- Reserving
  - Predict future obligations
  - Quantify the uncertainty of the estimates
  - Match projections of obligations to income streams
- Capital Allocation and Solvency
  - Decide appropriate level of necessary capital
  - Manage external stakeholders’ expectations; regulators, rating agencies, reputation
Insurance is big business – insurance activities comprised about 2.5% of the US gross domestic product in 2012.

Because of the size, it is not surprising that these firms employ analytics in the same manner as other large corporations.

These areas include (i) sales and marketing, (ii) compensation analysis, (iii) productivity analysis, and (iv) financial forecasting. For example, in sales and marketing:

- Predict customer behavior/needs (target appropriate customers)
- Anticipate customer reactions to promotions/rate changes
- Manage acquisition costs (online sales, agent compensation)
<table>
<thead>
<tr>
<th>FUNCTION</th>
<th>DESCRIPTION</th>
<th>EXEMPLARYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply chain</td>
<td>Simulate and optimize supply chain flows; reduce inventory and stock-outs.</td>
<td>Dell, Wal-Mart, Amazon</td>
</tr>
<tr>
<td>Customer selection,</td>
<td>Identify customers with the greatest profit potential; increase likelihood that</td>
<td>Harrah's, Capital One, Barclays</td>
</tr>
<tr>
<td>loyalty, and service</td>
<td>they will want the product or service offering; retain their loyalty.</td>
<td></td>
</tr>
<tr>
<td>Pricing</td>
<td>Identify the price that will maximize yield, or profit.</td>
<td>Progressive, Marriott</td>
</tr>
<tr>
<td>Human capital</td>
<td>Select the best employees for particular tasks or jobs, at particular</td>
<td>New England Patriots,</td>
</tr>
<tr>
<td></td>
<td>compensation levels.</td>
<td>Oakland A's, Boston Red Sox</td>
</tr>
<tr>
<td>Product and service</td>
<td>Detect quality problems early and minimize them.</td>
<td>Honda, Intel</td>
</tr>
<tr>
<td>quality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial performance</td>
<td>Better understand the drivers of financial performance and the effects of</td>
<td>MCI, Verizon</td>
</tr>
<tr>
<td></td>
<td>nonfinancial factors.</td>
<td></td>
</tr>
<tr>
<td>Research and development</td>
<td>Improve quality, efficacy, and, where applicable, safety of products and</td>
<td>Novartis, Amazon, Yahoo</td>
</tr>
<tr>
<td></td>
<td>services.</td>
<td></td>
</tr>
</tbody>
</table>

HARVARD BUSINESS REVIEW • JANUARY 2006
One can also use analytics to model the entire company. This feeds into enterprise risk management.

Figure: Flowchart of a Typical Dynamic Risk Model (DRM). Adapted from IAA (2010).
Paper entitled “Rating Endorsements using Generalized Linear Models”
- by myself and doctoral student Gee Lee.
- Accepted for publication in *Variance*, flagship publication of the Casualty Actuarial Society.
An endorsement, or a rider,

- provides optional insurance coverage
- may include alternative deductibles and coverage limits
- also provides extensions to the type of peril (e.g., stolen jewelry in homeowners insurance) covered

If there were no charge, it is not optional
How do we charge for an endorsement in a generalized linear model setting?
How do we charge for an endorsement in a generalized linear model setting?

1. Endorsements form a relatively small fraction of the premium base and so only informal, ad hoc, approaches are needed.
2. Use information from an external agency for this set of relativities
3. Treat endorsements as another type of coverage and use GLM techniques to determine this set of prices.
   - Requires a substantial amount of data
   - Requires claims that are identified by type of endorsement.

Property coverage has been available since 1911.

The fund insures property such as government buildings, schools, libraries, and motor vehicles.

“Local government” entities include counties, cities, towns, villages, school districts, and library boards

- The fund has over 1,000 such entities.


Property coverage has been available since 1911.

The fund insures property such as government buildings, schools, libraries, and motor vehicles.

“Local government” entities include counties, cities, towns, villages, school districts, and library boards

- The fund has over 1,000 such entities.
Determining Effective Relativities

- Because of the size of the fund, there are few difficulties using GLM to determine relativities/rates for the basic variables.
- Endorsements are more difficult:
  1. Fund is undergoing a major rate restructuring, politically sensitive.
  2. Information from external agencies is expensive and not particularly relevant.
  3. LGPIF data for optional coverages is limited.
Determining Effective Relativities

- Because of the size of the fund, there are few difficulties using GLM to determine relativities/rates for the basic variables.
- Endorsements are more difficult
  1. Fund is undergoing a major rate restructuring, politically sensitive
  2. Information from external agencies is expensive and not particularly relevant
  3. LGPIF data for optional coverages is limited
- We employed GLM techniques with restrictions on the coefficients through shrinkage using well-known penalized likelihood methods. Advantages:
  1. We provide relativities for endorsements in a disciplined manner, mitigating ad hoc adjustments
  2. Because we use GLM techniques, our approach is naturally multivariate and the introduction of endorsements accounts for the presence of other rating variables.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EntityType</td>
<td>Categorical variable that is one of six types: (Village, City, County, Misc, School, or Town)</td>
</tr>
<tr>
<td>LnCoverage</td>
<td>Total building and content coverage, in logarithmic millions of dollars</td>
</tr>
<tr>
<td>LnDeduct</td>
<td>Deductible, in logarithmic dollars</td>
</tr>
<tr>
<td>NoClaimCredit</td>
<td>Binary variable to indicate no claims in the past two years</td>
</tr>
<tr>
<td>Fire5</td>
<td>Binary variable to indicate the fire class is below 5 (The range of fire class is 0 ~ 10)</td>
</tr>
</tbody>
</table>
## Description of Endorsements

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

<table>
<thead>
<tr>
<th>Endorsements</th>
<th>Num of Obs</th>
<th>Average Frequency</th>
<th>Average Claim</th>
<th>Average Endorsement Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Interruption</td>
<td>225</td>
<td>6.493</td>
<td>228,393</td>
<td>2,679,595</td>
</tr>
<tr>
<td>Accounts Receivable</td>
<td>172</td>
<td>5.360</td>
<td>283,634</td>
<td>853,966 7</td>
</tr>
<tr>
<td>Pier and Wharf</td>
<td>312</td>
<td>2.599</td>
<td>41,262</td>
<td>245,445</td>
</tr>
<tr>
<td>Fine Arts</td>
<td>67</td>
<td>13.537</td>
<td>419,083</td>
<td>12,160,956</td>
</tr>
<tr>
<td>Golf Course Grounds</td>
<td>28</td>
<td>18.036</td>
<td>469,986</td>
<td>237,500</td>
</tr>
<tr>
<td>Zoo Animals</td>
<td>10</td>
<td>73.900</td>
<td>1,615,405</td>
<td>1,102,790</td>
</tr>
<tr>
<td>Special Use Animal</td>
<td>256</td>
<td>5.617</td>
<td>95,790</td>
<td>21,903</td>
</tr>
<tr>
<td>Vacancy Permit</td>
<td>225</td>
<td>4.902</td>
<td>158,402</td>
<td>1,779,212</td>
</tr>
<tr>
<td>Monies and Securities</td>
<td>2,137</td>
<td>2.071</td>
<td>60,868</td>
<td>58,928</td>
</tr>
<tr>
<td>Other Endorsements</td>
<td>53</td>
<td>5.000</td>
<td>40,819</td>
<td>4,763,019</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5,639</strong></td>
<td><strong>1.169</strong></td>
<td><strong>49,359</strong></td>
<td></td>
</tr>
</tbody>
</table>

*We observe only whether an entity has an endorsement (and the amount of the additional coverage), **not** whether a claim is due to an endorsement.*
Endorsements

- Theory/practice suggests that *endorsement coverage amount* may influence claims outcomes
- To capture this, using GLMs
  - \( y_B \) represents claims from a base coverage, mean \( \mu_B = \exp(x'\beta) \)
  - Let \( y_E \) be the claims from an endorsement, mean \( \mu_E \).
  
  \[
  \mu = E(y) = \begin{cases} 
  \mu_B = \exp(x'\beta) & \text{ endorsement not present} \\
  \mu_B + \mu_E = \exp(x'\beta + \beta_E x_E) & \text{ endorsement present}
  \end{cases}
  \]
  
- Let \( \text{Coverage}_E \) and \( \text{Coverage}_B \) represent amount of coverage for the endorsement and base (building and contents)
Endorsements

- Base mean $\mu_B = \exp(x'\beta)$, Endorsement mean $\mu_E$.

$$\mu = E (y = \begin{cases} 
\mu_B = \exp(x'\beta) & \text{endorsement not present} \\
\mu_B + \mu_E = \exp(x'\beta + \beta Ex_E) & \text{endorsement present}
\end{cases}) .$$

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

- With this specification, we have

$$\mu_E = \exp(x'\beta + \beta Ex_E) - \mu_B$$

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

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

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

using the approximation $(1 + z)^b \approx 1 + bz$.

- Endorsement Price $\mu_E$ is a factor times the endorsement coverage, rescaled by the overall cost per unit coverage.
  - The factor, $\beta_E$, is estimated from the data.
### Relativities for Base Rating Variables and Endorsements

<table>
<thead>
<tr>
<th></th>
<th>$\lambda = 0$</th>
<th>$\lambda = 5$</th>
<th>$\lambda = 1,000$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basic Rating Variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LnCoverage</td>
<td>3.024</td>
<td>3.029</td>
<td>3.065</td>
</tr>
<tr>
<td>LnDeduct</td>
<td>0.941</td>
<td>0.938</td>
<td>0.958</td>
</tr>
<tr>
<td>TypeCity</td>
<td>0.885</td>
<td>0.891</td>
<td>0.874</td>
</tr>
<tr>
<td>TypeCounty</td>
<td>0.414</td>
<td>0.348</td>
<td>0.166</td>
</tr>
<tr>
<td>TypeMisc</td>
<td>1.129</td>
<td>1.169</td>
<td>1.229</td>
</tr>
<tr>
<td>TypeSchool</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
</tr>
<tr>
<td>TypeTown</td>
<td>3.804</td>
<td>3.797</td>
<td>3.814</td>
</tr>
<tr>
<td>Fire5</td>
<td>0.857</td>
<td>0.853</td>
<td>0.839</td>
</tr>
<tr>
<td>NoClaimCredit</td>
<td>0.815</td>
<td>0.818</td>
<td>0.827</td>
</tr>
<tr>
<td><strong>Endorsement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LnBusInterCovRat</td>
<td>1.207</td>
<td>1.208</td>
<td>1.036</td>
</tr>
<tr>
<td>LnAccRecCovRat</td>
<td>1.141</td>
<td>1.018</td>
<td>1.008</td>
</tr>
<tr>
<td>LnAddInsCovRat</td>
<td>1.343</td>
<td>1.329</td>
<td>1.045</td>
</tr>
<tr>
<td>LnPierWarfCovRat</td>
<td>1.040</td>
<td>1.046</td>
<td>1.003</td>
</tr>
<tr>
<td>LnSpecialAnimalCovRat</td>
<td>1.416</td>
<td>1.069</td>
<td>1.001</td>
</tr>
<tr>
<td>LnZooAnimalCovRat</td>
<td>32.816</td>
<td>1.538</td>
<td>1.004</td>
</tr>
<tr>
<td>LnFineArtsCovRat</td>
<td>1.379</td>
<td>1.481</td>
<td>1.071</td>
</tr>
<tr>
<td>LnGolfCourseCovRat</td>
<td>2.771</td>
<td>1.293</td>
<td>1.002</td>
</tr>
<tr>
<td>LnMoneySecCovRat</td>
<td>1.165</td>
<td>1.165</td>
<td>1.165</td>
</tr>
<tr>
<td>LnMoneySecLimCovRat</td>
<td>1.276</td>
<td>1.276</td>
<td>1.276</td>
</tr>
</tbody>
</table>
Begin with classic linear model shrinkage estimation, minimize

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

Values of \( \lambda \) control the complexity of the model; smaller values mean less shrinkage.

Can write this in terms of classical “ridge regression”

\[
\hat{\beta}_{\text{shrink}} = (X'X + \lambda I)^{-1} X'y
\]

appealing in instances of collinearity.

For (nonlinear) GLMs, we use a penalized likelihood of the form

\[
l(\beta) = \sum_{i=1}^{n} \log f(y_i) - \lambda \| R\beta - r \|^2,
\]
Traditionally, insurers use information reported by policyholders on application forms, combined with selected external sources.

- E.g., police reports for automobile insurance or medical exam results for life insurance.
- Many variables are categorical, making even limited info complex.

Many analysts are now exploring the use of shrinkage and related “regularization” techniques for use in handling “big data”

- Our contribution reminds analysts of another classical purpose of such techniques, to smooth erratic estimates in a disciplined way.
Contributions of this Work

1 Detailed analysis of the Wisconsin Local Government Property Insurance Fund
   - There is little in the literature on government property and casualty actuarial applications.
   - The LGPIF is similar to small commercial property insurance, making our work of interest to a broad readership.
Contributions of this Work

1. Detailed analysis of the Wisconsin Local Government Property Insurance Fund
   - There is little in the literature on government property and casualty actuarial applications.
   - The LGPIF is similar to small commercial property insurance, making our work of interest to a broad readership.

2. Detailed analysis in the manner of a case study so that other analysts may replicate parts of our approach.
   - We provide relativities not only for our primary rating variables but also for endorsements.
   - Introduce an approach for handling these optional coverages when it is not known whether or not a claim is due to an endorsement.
Contributions of this Work

2. Case study
3. Explored the use of shrinkage estimation in ratemaking
   - Shrinkage is particularly appealing in the case of endorsements.
   - Little predictive ability was lost by using shrinkage methods and they gave much more intuitively appealing relativities.
   - Helpful to have relativities that can be calibrated in a disciplined manner and are consistent with sound economic, risk management, and actuarial practice.
In this part of the talk, I summarize my work on developing tools to manage insurance portfolios

- Unbundling of coverages was the first investigation
- The work on Gini statistics is motivated by the ratemaking problem
- As another example, think about an analogy to investment managers
  - By allocating investments among risks, they seek to optimize on risk versus reward trade-offs
- Insurers also maintain portfolios (of insurance policies) whose risks must be managed
  - Management tools include policy renewals, changing policy limits and deductibles, facultative reinsurance and the like.
In two papers, *Journal of the American Statistical Association*, Frees and Valdez (2008), *Astin Bulletin: Journal of the International Actuarial Association*, Frees, Shi, and Valdez (2009), we showed how to incorporate several claim types:

- $y_{ij,1}$ - claim for injury to a party other than the insured - “injury”;  
- $y_{ij,2}$ - claim for damages to the insured, including injury, property damage, fire and theft - “own damage”; and  
- $y_{ij,3}$ - claim for property damage to a party other than the insured - “third party property”.

Distribution for each claim is typically medium to long-tail. The full multivariate claim may not be observed. For example:

<table>
<thead>
<tr>
<th>Claim Combination</th>
<th>$(y_1)$</th>
<th>$(y_2)$</th>
<th>$(y_3)$</th>
<th>$(y_1, y_2)$</th>
<th>$(y_1, y_3)$</th>
<th>$(y_2, y_3)$</th>
<th>$(y_1, y_2, y_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>0.4</td>
<td>73.2</td>
<td>12.3</td>
<td>0.3</td>
<td>0.1</td>
<td>13.5</td>
<td>0.2</td>
</tr>
</tbody>
</table>

We introduced a copula model to incorporate dependencies. I am now thinking of this as a first set of papers on “portfolio management.”
Unbundling of Coverages

From Frees, Shi, Valdez (2009)

- Decompose the comprehensive coverage into more “primitive” coverages: third party injury, own damage and third party property.
- Calculate a risk measure for each unbundled coverage, as if separate financial institutions owned each coverage.
- Compare to the bundled coverage that the insurance company is responsible for.
- Despite positive dependence, there are still size advantages to bundling.

<table>
<thead>
<tr>
<th>Unbundled Coverages</th>
<th>VaR 95%</th>
<th>VaR 99%</th>
<th>CTE 95%</th>
<th>CTE 99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third party injury</td>
<td>309,881</td>
<td>1,163,855</td>
<td>964,394</td>
<td>2,657,911</td>
</tr>
<tr>
<td>Own damage</td>
<td>59,898</td>
<td>86,421</td>
<td>76,951</td>
<td>104,576</td>
</tr>
<tr>
<td>Third party property</td>
<td>209,509</td>
<td>264,898</td>
<td>248,793</td>
<td>324,262</td>
</tr>
<tr>
<td>Sum of Unbundled Coverages</td>
<td>579,288</td>
<td>1,515,174</td>
<td>1,290,137</td>
<td>3,086,749</td>
</tr>
<tr>
<td>Bundled (Comprehensive) Coverage</td>
<td>324,611</td>
<td>763,042</td>
<td>652,821</td>
<td>1,537,692</td>
</tr>
</tbody>
</table>
How Important is the Copula?

Very!!

<table>
<thead>
<tr>
<th>Table. VaR and CTE for Bundled Coverage by Copula</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Copula</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Independence</td>
</tr>
<tr>
<td>Normal</td>
</tr>
<tr>
<td>$t$</td>
</tr>
</tbody>
</table>

- **VaR** (Value at Risk) and **CTE** (Conditional Tail Expectation) are two measures that suggest how much capital needed by the insurer.
- The choice of the copula significantly affects the required capital.
With large sample sizes, how do we tell if a new variable is useful for rating?

Predictive validation is the key. But traditional measures (e.g., root mean square error) are inadequate because

- They lack economic context
- They perform poorly for non-normal data, e.g., mass at zero and skewed, long-tailed, claims distributions

We introduced a “Gini statistic” that is based on forming insurance portfolios and see how a rating scheme (with new variables) would perform on a held-out validation sample.
The Gini Index

- The 45 degree line is known as the “line of equality”
  - In welfare economics, this represents the situation where each person has an equal share of income (or wealth)

- To read the Lorenz Curve
  - Pick a point on the horizontal axis, say 60% of households
  - The corresponding vertical axis is about 40% of income
  - This represents income inequality
  - The farther the Lorenz curve from the line of equality, the greater is the amount of income inequality

- The Gini index is defined to be (twice) the area between the Lorenz curve and the line of equality

![Lorenz Curve Graph]

41 / 58
**Lorenz curves and Gini indexes for Brazil and Hungary**

To measure income inequality in a country and compare this phenomenon among countries more accurately, economists use Lorenz curves and Gini index. A Lorenz curve plots the cumulative percentages of total income received against the cumulative percentages of recipients, starting with the poorest individual or household (Figure 5.2). How is it constructed?

First, economists rank all the individuals or households in a country by their income level, from the poorest to the richest. Then all of these individuals or households are divided into 5 groups (20% in each) or 10 groups (10% in each) and the income of each group is calculated and expressed as a percentage of GDP (see Figure 5.1). Next, economists plot the shares of GDP received by these groups cumulatively— that is, plotting the income share of the poorest quintile against 20% of population, the income share of the poorest quintile and the next (fourth) quintile against 40% of population, and so on, until they plot the aggregate share of all five quintiles (which equals 100%) against 100% of the population. After connecting all the points on the chart— starting with the 0% share of income received by 0% of the population— they get the Lorenz curve for this country.

The deeper a country's Lorenz curve, the less equal its income distribution. For Figure 5.2:

- **Hungary** (Gini index = 27.0%)
- **Brazil** (Gini index = 63.4%)

<table>
<thead>
<tr>
<th>Percentage of total population</th>
<th>Percentage of total income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorest</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>40</td>
<td>80</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>80</td>
<td>40</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

- Line of absolute equality
- Line of absolute inequality
We consider an “ordered” Lorenz curve, that varies from the usual Lorenz curve in two ways

- Instead of counting people, think of each person as an insurance policyholder and look at the amount of insurance premium paid
- Order losses and premiums by a third variable that we call a relativity

Policies are profitable when expected claims are less than premiums

Expected claims are unknown but we will consider one or more candidate insurance scores, $S(x)$, that are approximations of the expectation

- We are most interested in polices where $S(x_i) < P(x_i)$

One measure (that we focus on) is the relative score

$$R(x_i) = \frac{S(x_i)}{P(x_i)},$$

that we call a relativity.
Ordered Lorenz Curve

- **Notation**
  - $x_i$ - explanatory variables, $P(x_i)$ - premium, $y_i$ - loss, $R_i = R(x_i)$, $I(\cdot)$ - indicator function, and $E(\cdot)$ - mathematical expectation

- **The Ordered Lorenz Curve**
  - **Vertical axis**
    \[
    F_L(s) = \frac{E[y I(R \leq s)]}{E y} = \frac{\sum_{i=1}^{n} y_i I(R_i \leq s)}{\sum_{i=1}^{n} y_i}
    \]
  
    that we interpret to be the *market share of losses*.

  - **Horizontal axis**
    \[
    F_P(s) = \frac{E[P(x) I(R \leq s)]}{E P(x)} = \frac{\sum_{i=1}^{n} P(x_i) I(R_i \leq s)}{\sum_{i=1}^{n} P(x_i)}
    \]

    that we interpret to be the *market share of premiums*.

- **The distributions are unchanged when we**
  - rescale either (or both) losses ($y$) or premiums ($P(x_i)$) by a positive constant
  - transform relativities by any (strictly) increasing function
Homeowners Example

- To read the ordered Lorenz Curve
  - Pick a point on the horizontal axis, say 60% of premiums
  - The corresponding vertical axis is about 53.8% of losses
  - This represents a profitable situation for the insurer
    - Uses “SP_FreqSev_Basic” = base premium, relativity uses score “IND_FreqSev”
    - The “line of equality” represents a break-even situation
  - An Ordered Lorenz Curve. For this curve, the corresponding Gini index is 10.03% with a standard error of 1.45%.

![Ordered Lorenz Curve diagram](image)
When regression functions are used for scoring, the Gini index can be viewed as a goodness-of-fit measure.

We have introduced measures to quantify the statistical significance of empirical Gini coefficients.

- The theory allows us to compare different Gini indices.
- It is also useful in determining sample sizes.

We provided a few alternative ways to think about our new Gini index, e.g., as an area, profit measure, etc.

In particular, interpret this index as proportional to the correlation between a policy’s “profit” \((P - y)\) and the rank of the relative premium \(\text{rank}(S/P)\). Very nice intuition.

The Gini index is a little like a hypothesis test in that one identifies a “null hypothesis” - this is the base score in the relativity.

- There is an asymmetry in the treatment of scores.

It gives an economically meaningful way to assess out-of-sample fit.
We (Frees, Meyers and Cummings) wrote a series of papers that appeared in the top statistical and actuarial journals.

- Dependent Multi-Peril Ratemaking Models

- Summarizing Insurance Scores Using a Gini Index

- Predictive Modeling of Multi-Peril Homeowners Insurance

- Insurance Ratemaking and a Gini Index

Our work was recognized in the 2015 ARIA Prize given by the Casualty Actuarial Society.
The ordered Lorenz curve allows us to capture the separation between losses and premiums in an order that is most relevant to potential vulnerabilities of an insurer’s portfolio. The corresponding Gini index summarizes this potential vulnerability. It provides a tool for “portfolio management” - identification of good and bad risks in a portfolio.
For motivation, let us recall the now classic Markowitz investment portfolio problem.

The goal is to find the allocation of assets that provides the least risky portfolio return for a given expected portfolio return. Specifically,

- $p$ risks with returns: $R_1, \ldots, R_p$
- The investor allocates assets $A_i$ to each risk for a sum of $A = A_1 + \cdots + A_p$ invested in total
- The expected portfolio return is $A_1 E R_1 + \cdots + A_p E R_p$.
- Mathematically, we seek to find allocation of assets $A_1, \ldots, A_p$ to minimize

$$\text{Var} \left( \sum_{i=1}^{p} A_i R_i \right)$$

subject to a desired minimum expected return

$$\mathbb{E} \left( \sum_{i=1}^{p} A_i R_i \right) \geq \mu_{\text{min}}$$

and a budget constraint $A = A_1 + \cdots + A_p$

This is now recognized to be a quadratic programming problem which can be readily solved.
More generally, we can replace the variance function with any convex function.
This is now a convex optimization problem.
- Many algorithms exist for the solution of these types of problems.
First example, and historically most prominent, is to use the quantile, or “value at risk”
Another is to use the conditional tail expectation.
Insurance Portfolio Problem

- Let me think of $A_iR_i = Y_i$ as an insurance loss
- This has expectation $P_i = E(Y_i)$, the price
- The amount retained by the company is $c_iY_i$ for premium $c_iP_i$
  - Interpret $c_i$ to be a coinsurance parameter (selected in conjunction with the policyholder).
  - Alternatively, $c_i$ may be a parameter in a “quota share reinsurance” agreement, a type of proportional reinsurance
- Mathematically, we seek to find risk retention parameters $c_1, \ldots, c_p$ to minimize the value at risk
  \[
  \text{VaR} \left( \sum_{i=1}^{p} c_iY_i \right)
  \]
  subject to a revenue constraint $c_1P_1 + \cdots + c_pP_p \geq P_{\min}$
Both the finance/investment and insurance portfolio problems consider risks $Y_i$ in a portfolio

- In investments, the risk is an asset
- In insurance, the risk is an obligation of the insurer

Like the finance portfolio optimization problem, in insurance

- One makes a decision about each risk $Y_i$, in the context of a portfolio
- Risks $Y_i$ are non-identical
- Risks may be dependent
  - may have different coverages for the same policyholder
  - different policyholders share latent characteristics (e.g., economy, geography)
In insurance, risks are illiquid, e.g., contracts renew at different times.

There are also non-linear mechanisms that can be used to control the amount of risk retained:

- An upper policy limit, say $u$, limits the liability of the insurer to $\min(Y_i, u)$. (Could also cede excess to a reinsurer).
- A deductible, say $d$, provides a floor for risk taking, e.g., the insurer’s liability is $\max(Y_i - d, 0)$.

This nonlinearity means the problem is non-convex; hence no guarantee of global solutions:

- Optimize over each risk retention parameter
- My work focuses on local changes
This is summarized in a working paper entitled “Insurance Portfolio Risk Retention.”

The insurer’s portfolio consists of:

- \( p \) risks: \( Y_1, \ldots, Y_p \)
- For each risk, there is a set of insurance risk retention parameters \( \theta_i \); may correspond to a deductible, coinsurance, or upper policy limit
- \( i \)th retained risk: \( g_{\theta_i}(Y_i) \)
- Decision variables: portfolio retention parameters \( \theta = (\theta_1, \ldots, \theta_p) \)
- Portfolio: \( S_\theta = \sum_{i=1}^{p} g_{\theta_i}(Y_i) \)

There is a trade-off between:

- premium, \( P(S_\theta) = e.g. \ E S_\theta \)
- risk measure \( R(S_\theta) = e.g. \xi_\theta \) (quantile)
  in the sense that they move together with each element of \( \theta \)
Motivation - Portfolio Problem

- Could treat as constrained optimization problem
  - Minimize: $R(S_\theta)$
  - Subject to: $P(S_\theta) \geq P_{\text{min}}$
  - over a choice of $\theta$
- The Lagrangian is $R(S_\theta) - \lambda (P(S_\theta) - P_{\text{min}})$
- For a single $\theta$, the Lagrange multiplier is

$$\lambda = \frac{\partial \theta R(S_\theta)}{\partial \theta P(S_\theta)} = RM^2$$

This uses the short-hand notation $\partial \theta = \frac{\partial}{\partial \theta}$.
- I show how to compute this statistic for various risk retention parameters under sensible probabilistic distributions, e.g., Tweedie regressions (with rating covariates) and (copula-based) dependencies among risks
For a portfolio

- The tool allows the insurer to select among deductible, coinsurance, and upper policy limits strategies.
- The tool can be used to identify the “best” and “worst” risks in terms of opportunities for risk management.
- The tool can also be used by a manager of a policy with multiple coverages (e.g., building and contents, motor vehicle, and equipment).
- The tool recognizes the impact of dependence modeling.

This tool has the advantage of being in a form similar to an investment portfolio problem. This familiar form should be comforting to insurance managers.
We are pushing the frontiers on modeling insurance losses
We are also developing ideas on how to handle dependencies among risks (using copulas and related functions). Current projects include:
- Incorporating dependencies for discrete data (e.g., claims counts)
- Dependencies for frequencies and severities
- Incorporating spatial and temporal relationships
- Many others...
Overheads are available at:
https://sites.google.com/a/wisc.edu/jed-frees/

Thank you for your kind attention.