

**Casualty Actuarial Society
E-Forum, Winter 2011
Volume 2**



The CAS *E-Forum*, Winter 2011-Volume 2

The Winter 2011-Volume 2 edition of the CAS *E-Forum* is a cooperative effort between the Committee for the CAS *E-Forum* and various other CAS committees.

The CAS Committee on Ratemaking presents for discussion nine papers prepared in response to the 2011 call for ratemaking papers. This *E-Forum* also includes one additional paper.

Some of the Ratemaking Call Papers will be discussed by the authors at the 2011 CAS Ratemaking and Product Management Seminar on March 20-22, 2011, in New Orleans, LA.

CAS Committee on Ratemaking

Todd W. Lehmann, *Chairperson*

John L. Baldan
Angelo E. Bastianpillai
LeRoy A. Boison
James M. Boland
Lee M. Bowron
William M. Carpenter
Donald L. Closter
Christopher L. Cooksey
Kiera Elizabeth Doster

John S. Ewert
Dennis L. Lange
Pierre Lepage
Taylan Matkap
Robert W. Matthews
Dennis T. McNeese
Benjamin R. Newton
Baohui Ning
Joseph M. Palmer

Jane C. Taylor
Jonathan White
Richard P. Yocius
Ronald Joseph Zaleski
Yi Zhang
Cheri Widowski, *CAS Staff*
Liaison

CAS *E-Forum*, Winter 2011-Volume 2

Table of Contents

2011 CAS Ratemaking Call Papers

GAP Insurance—Techniques and Challenges

Lee Bowron, ACAS, MAAA, and John Kerper, FSA, MAAA 1-13

Credibility for a Tower of Excess Layers

David R. Clark, FCAS, MAAA 1-20

Predictive Modeling of Multi-Peril Homewoners Insurance

Edward W. (Jed) Frees, Glenn Meyers, FCAS, and A. David Cummings, FCAS 1-33

Deductibles, Policy Limits, and Reinsurance: A Case Study in Malaysia

Noriszura Ismail, Ph.D., and Ansar Asnawi Ahmad Anuar 1-47

Generalized Linear Mixed Models for Ratemaking: A Means of Introducing Credibility into a Generalized Linear Model Setting

Fred Klinker, FCAS, MAAA 1-25

Reserving in the Age of Obesity

Chris Laws and Frank Schmid 1-29

Towards Multivariate Ratemaking: Claim Frequency Analysis Examples

Hernán L. Medina, CPCU, API, AU, AIM, ARC 1-49

Multi-Year Policy Pricing

Benjamin R. Newton, FCAS, RPLU 1-15

Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

Frank Schmid 1-35

Additional Paper

Mortality Trend Models

Gary G. Venter 1-30

***E-Forum* Committee**

Windrie Wong, *Chairperson*

Mark A. Florenz

Karl Goring

Dennis L. Lange

Elizabeth A. Smith, *Staff Liaison*

John Sopkowicz

Zongli Sun

Yingjie Zhang

For information on submitting a paper to the *E-Forum*, visit
<http://www.casact.org/about/policiesProc/index.cfm?fa=forum>.

GAP Insurance—Techniques and Challenges

Lee Bowron, ACAS, MAAA, and John Kerper, FSA, MAAA

Abstract: GAP (Guaranteed Asset Protection) insurance is an insurance product that insures the difference (if any) between the loan balance and the actual value of the underlying asset. Typically, this insurance is sold in conjunction with a traditional insurance product and guarantees that an insurable event will be sufficient to satisfy any lien upon the asset. While this type of insurance is used to cover a variety of exposures, the largest asset class is private passenger vehicles.

1. WHAT IS GAP INSURANCE?

The origins of GAP insurance are a little murky—the product has existed for about 25 years and originally may have been underwritten by car dealers as a sort of “quasi-insurance” product.

GAP is similar to credit life and credit A&H because it pays the vehicle loan in the event of certain contingencies, namely the car being deemed a “total loss” by the physical damage insurer. GAP will cover the shortfall between the loan payoff and the insurance recovery (typically book value less the deductible.)

While the term for gap coverage matches the term of the loan, the possibility of a claim is zero once the book value (less the deductible) of the vehicle exceeds the loan payoff. Also, there can be only one claim on a GAP policy—once a claim has been made the policy is expired and any remaining unearned premium is fully earned.

The regulatory framework for GAP differs from state to state. In some states it is not technically considered insurance. Other states may require that an insurance policy ultimately back the liabilities of a program (contractual liability), while others may consider the full premium insurance. Regardless of the regulatory framework, the techniques developed in this paper would be applicable since the consideration for pricing applications should be the ultimate projection of the underlying losses.

GAP products are structurally different from most other property/casualty products and an understanding of the structure and terminology may be helpful for the actuary who is unfamiliar with the business.

GAP is generally sold for a single payment for the entire term of the underlying loan and the sale is made at the time that the covered vehicle is purchased. A GAP policy can be cancelled and a refund processed. This will happen if the vehicle is sold or the policyholder requests a cancellation.

The refund method varies by state, with most using a Rule of 78s amortization due to the declining value of the coverage, but some, notably Texas, requiring pro rata for the return of premium. In addition, some lienholders may specifically require pro rata refunds in order to finance the GAP policy with the auto loan.

In addition to GAP, the consumer may encounter several other ancillary products during the inevitable visit to the dealer's finance and insurance department. These products include pre-paid maintenance, a vehicle service contract, VIN etch, etc. All of these products are almost always financed with the vehicle.

In states where GAP is not regulated as an insurance product, the price charged by the dealer is made up of three components: (1) GAP reserve, (2) administrative fees, and (3) dealer markup. It's also important to note that component (1) is the only portion that is paid to the insurer. Components (2) and (3) are not paid to the insurer, nor are they included in premium for purposes of calculating premium tax or risk-based capital. The portion of the price remitted to the insurance company may be the entire GAP reserve (1) or the GAP reserve may be placed in trust and a contractual liability policy can be issued to guarantee the performance of the trust.

An administrator typically will perform all the processing and servicing of the GAP contract. An agent will represent the administrator to the dealer clients. The GAP reserve may be remitted to an insurance company, or it may not be considered insurance in a regulatory sense. For the actuary, there are two items of note:

The terminology of reserve is misleading because "reserve" in GAP typically refers to all funds used to pay claims, not just the outstanding portion, and is more analogous to written premium. For our purposes, we will use the term *premium*.

Since the majority of expenses are paid prior to the remittance of funds to the insurance company, the expected loss ratio on net of expense premium on a book is higher than other property/casualty products. Often, a book will be priced at an expected loss ratio of 80 to 90 percent.

One should also note that GAP insurance represents a "moral hazard" since, after a loss, the insured will be in a better financial position than before since the negative equity on the vehicle has been removed. Of course, this is not different than replacement cost on homeowners insurance. Based on reviews of proprietary data, there is evidence of this hazard by the noticeable rise in frequency of GAP claims during the recession of 2008/2009.

2. CONSIDERATIONS WHEN PRICING GAP

GAP claims are dependent on two criteria: the occurrence of a total loss by the contract holder and the loan balance at the time of loss exceeding the book value of the vehicle.

Since the underlying product is private passenger insurance, we would expect that the same rating variables that are prevalent in private passenger pricing would also be predictive for the GAP pricing.

However, one must remember that a claim is only generated by a total loss, which would indicate that the frequency of more expensive or higher symbol vehicles might be lower than budget-priced vehicles, since they may be less likely to be declared a total loss by the insurance company. Even a state insurance department may not fully understand this difference. Texas, for example, mandates GAP rates by the amount of the loan, which has little correlation with loss.

GAP losses on vehicles will be driven more by the depreciation of the vehicle, which historically has been faster for American-made sedans and slower for some of the European and Japanese makes. Depreciation rates can fluctuate and are often a function of consumer preference. So the current depreciation rate for a vehicle may be subject to change in the future.

The severity of the loss will, of course, depend on the loan balance and the book value at the time of loss. Since the loan will amortize more slowly on a longer-term loan (and provide a greater length of coverage), the length of the loan is a factor in the severity. It also may affect the frequency because, once the value of the vehicle exceeds the loan, the claim would not be compensated.

Another major factor is the book value of the vehicle at the time of the purchase. While one might assume that the price of the vehicle would be equal to its underlying value, this is not necessarily true.

In many sales, the purchasers will owe more on their existing vehicles than the trade-in values. In the industry, this is known as “negative equity” or being “upside down.” These customers are typically offered more for their trade than the vehicle is worth and the difference is reflected in the retail price. This inflated purchase price creates an immediate GAP exposure at the inception of the policy.

For example, suppose that a customer’s current vehicle is worth \$8,000 but the customer owes \$12,000 on the vehicle. The new vehicle can be purchased for \$25,000, which is the book value of the vehicle. In this case, a dealer may increase the price of the new vehicle to \$29,000 and the value of the trade to \$12,000. This will allow the existing note to be settled and a new loan for \$29,000 will be originated. Therefore, the negative equity is “rolled” into the new loan. If a total loss occurred immediately on this new vehicle, the purchaser would face a shortfall of \$4,000.

While these types of transactions may not be the majority of overall vehicle purchases, they will be a substantial part of a GAP portfolio, because these purchasers recognize their negative equity situations and will seek to insure the exposures.

In general, used vehicles will show fewer propensities for initial negative equity than new vehicles, and the frequency and the severity will be lower.

Unfortunately, many GAP insurance writers do not capture both the loan amount and the vehicle value at the time of purchase, which makes the analysis difficult. If loan amount and terms are captured, one can model the potential GAP severities on the book by examining the difference between the amortized value of the loan and the book value of the vehicle less the deductible.

While the typical automobile liability rating variables (such as age, credit score, marital status, driving record, garaging zip code, etc.) would likely be predictive for GAP coverage, in reality the rating plans for GAP coverage are currently very simple, with the vast majority only varying on the term of the loan.

Finally, the actuary must consider the catastrophe exposure for this line, which would include the typical catastrophe perils that affect the automobile physical damage coverage. Hail, wind and flood would be typical causation factors—the largest exposure is likely flood as most other catastrophes would not result in total losses to the vehicle and because most states require a vehicle that has been flooded to be declared a total loss. There was a significant amount of GAP catastrophe loss associated with Hurricane Katrina, although this was mitigated by insureds driving their vehicles out of the flood zones prior to the hurricane. Since vehicles with significant GAP exposure are also likely newer vehicles, they may be more prone to be removed from a potential catastrophe exposure.

3. THE LEVERAGED IMPACT OF USED VEHICLE PRICING

The biggest uncertainty with the analysis of a GAP program is the future direction of used vehicle pricing. Used vehicle prices are subject to volatile shifts. This occurs because of economic shifts that can impact the market value of used vehicles.

It is important to note that for GAP pricing, late-model used vehicle prices are more important, as the sale prices of older vehicles (more than three years from the current model year) will not be subject to significant GAP claims.

Since GAP will cover the difference between the book value and the loan balance, the difference between book value and loan amount acts like a very high deductible in a traditional automobile physical damage insurance policy. As the Table 1 below shows, changes in used vehicle values are significantly leveraged up into changes in the GAP severity.

Table 1

	(1)	(2)	(3)	(4)	(5)	(6)
	Loan Amount	Book Value	GAP Deductible	GAP Coverage	Change in Book Value	Change in GAP Coverage
Base	16,000	13,000	500	3,500		
Increase in Book Value	16,000	14,300	500	2,200	10%	-37%
Decrease in Book Value	16,000	11,700	500	4,800	-10%	37%

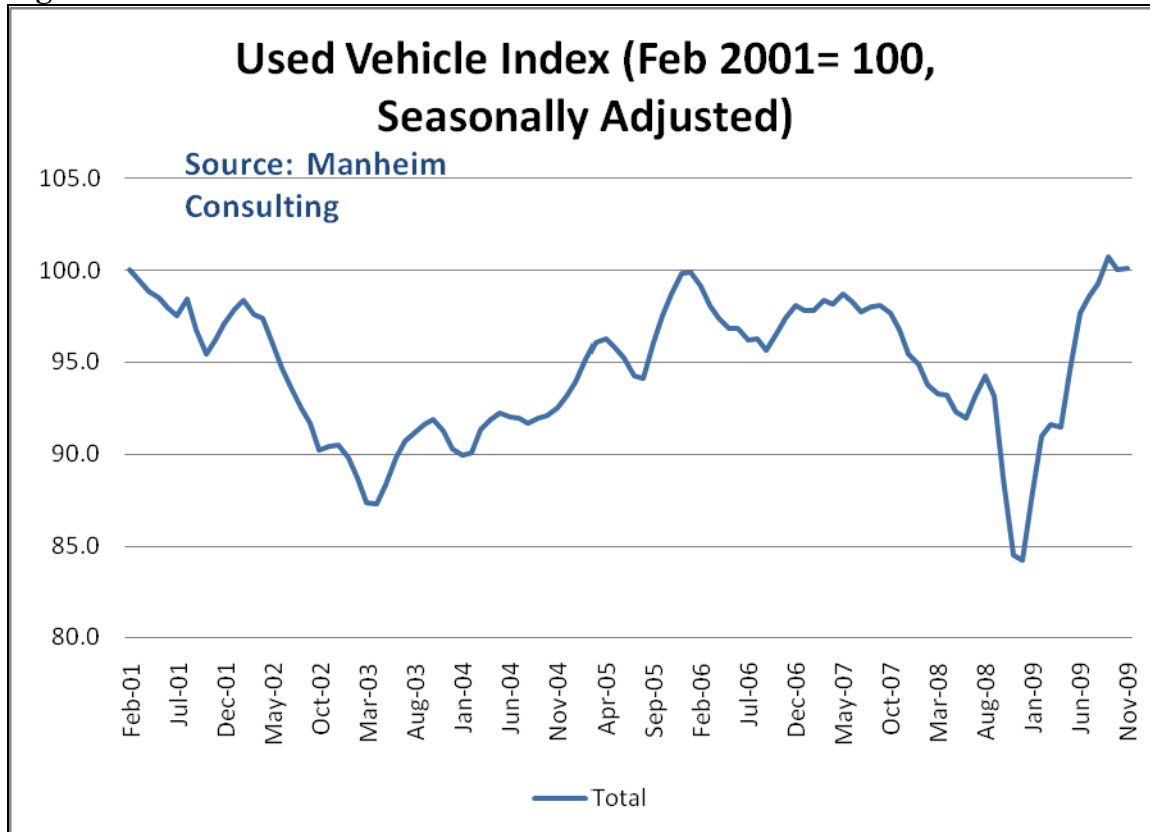
(4) (1) - (2) + (3)

(5) (2) / Base (2)

(6) (4) / Base (4)

Another issue is shifting vehicle preference among types of vehicles, such as the definite relationship between small vehicle prices and gasoline prices. Alternatively, there is an inverse relationship between large trucks prices and gasoline prices. Dramatic shifts in consumer preferences will cause GAP claims to increase, even if overall prices remain stable. This is because increases in underlying asset prices are capped by the amount of the loan while decreases remain uncapped.

Figure 1



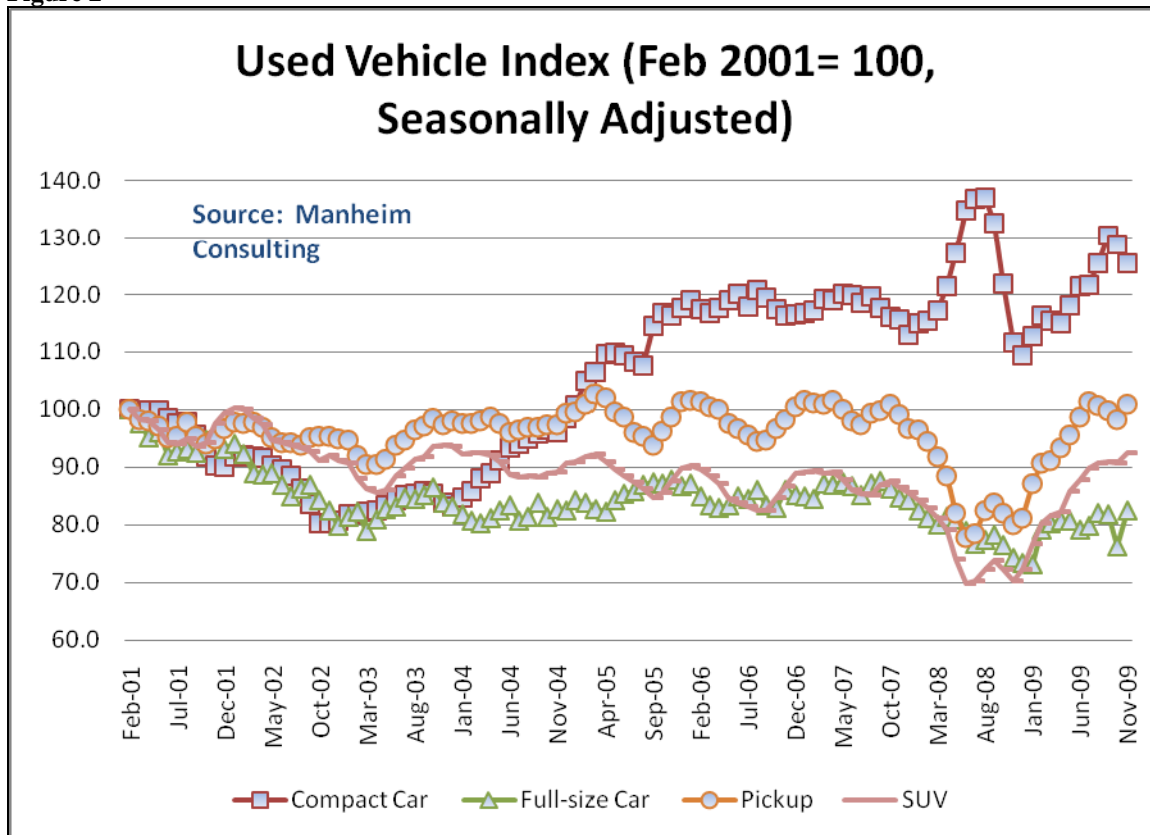
As Figure 1 above shows, the index of used vehicle prices is subject to significant variation. Vehicles in the index are compiled from auction sales which focus mostly on late model used vehicles.

For example, in the aftermath of 9/11, vehicle manufacturers began to heavily incentivize the purchase of new vehicles through “zero percent financing” and other enticements. The result was a strong decline in the value of used vehicles, as these prices adjusted to the corresponding new vehicle price.

The “great recession” officially began in December 2007, which is the beginning of a decrease in the price of used vehicles.

In late 2008, used vehicle prices showed a dramatic improvement. There is some evidence that this increase may be more due to a lower supply of late model used vehicles in the marketplace rather than an increase in demand. This supply constraint may be due to decreased new vehicle sales in the prior years, as well as less leasing of vehicles (which generates a sale when the lease terminates). In addition, rental car fleets (who are a major source of late-model used vehicles) purchased fewer vehicles.

Figure 2



The chart in Figure 2 shows the same data, but is broken out by popular vehicle segments. For example, compact cars have shown significant price appreciation since 2001.

In conclusion, GAP is a leverage product in which small changes in the underlying book value of the asset will cause large swings in projected results. In addition, private passenger vehicles are subject to dramatic and somewhat unpredictable changes in price due to economic forces, petroleum prices, and consumer preferences. Furthermore, future regulatory requirements such as increased mileage standards may affect asset prices. Forecasting future GAP claims is subject to significant variation.

4. EARNINGS PATTERNS

GAP is a multi-year policy for which premium is earned through the use of earnings factors or earnings curves. These earnings are subject to actuarial review during evaluation before issuing a loss reserve opinion.

Typically, earnings are done on a “Rule-of-78s”-basis, which implies a quicker earnings pattern than a pro rata or even earnings typical for most property casualty products. The Rule-of-78s will earn premium as a function of the sum of the digits of the remaining term with the sum of digits of all term values.

Earnings for Rule-of-78s

N = Term in months.

M = Evaluated month.

Earnings factor to apply to written premium for this contract:

$$\text{Earnings Factor} = \sum_{M=1}^N \frac{2(N+1-M)}{(N)(N+1)}.$$

As the Table 2 below shows, this pattern closely resembles the balance of a loan. Research indicates that this pattern is slower than the actual emerging experience.¹ A more accurate earnings pattern can be obtained by assuming a reduction in the term of the loan by 25% (Term Elimination Factor) and using the Rule-of-78s pattern on these numbers).

Abbreviating the term for GAP insurance also makes sense when determining earned premium. GAP insurance does not cover the loan balance; it covers the difference between the loan balance and the book value. Once the loan amortizes to the point which the loan balance plus the

¹ The 25% reduction is suggested by industry data in the 2009 CCIA GAP Study which was based on approximately 489,000 exposures and 5,800 claims. Several methods were employed, with the 25% reduction having the best fit if one refunds based on Rule-of-78s. Mature books may also be analyzed directly for the underlying claim distribution pattern.

deductible is less than the book value of the vehicle, there is no severity associated with a claim. Therefore, we would expect that the severity of a GAP policy to reach zero more quickly than the loan balance.

Earnings for Abbreviated Rule-of-78s

N = Term in months.

M = Evaluated month.

A = Term Elimination Factor ($0 < A \leq 1$).

Earnings factor to apply to written premium for this contract:

$$Z = (N \times (1 - A)).$$

For $M < Z$.

$$\text{Earnings Factor} = \sum_{M=1}^Z \frac{2(Z+1-M)}{(Z)(Z+1)}.$$

Else 1.

Table 2 illustrates these calculations for a 60-month loan.

Table 2

Month	(1) Balance	(2) Balance Earnings	(3) Rule of 78s	(4) Abbreviat ed Rule of 78s	(5) Example Book Value	(6) Deduct	(7) GAP Severity	(8) Example Book Value Earnings
1	10,000	3.1%	3.3%	4.3%	8,000	500	2,500	3.0%
2	9,860	3.1%	3.2%	4.3%	7,840	500	2,520	3.0%
3	9,720	3.0%	3.2%	4.2%	7,683	500	2,537	3.0%
4	9,579	3.0%	3.1%	4.1%	7,530	500	2,549	3.1%
5	9,436	2.9%	3.1%	4.0%	7,379	500	2,557	3.1%
6	9,293	2.9%	3.0%	3.9%	7,231	500	2,562	3.1%
7	9,150	2.8%	3.0%	3.8%	7,087	500	2,563	3.1%
8	9,005	2.8%	2.9%	3.7%	6,945	500	2,560	3.1%
9	8,859	2.7%	2.8%	3.6%	6,806	500	2,553	3.1%
10	8,713	2.7%	2.8%	3.5%	6,670	500	2,543	3.1%
11	8,566	2.7%	2.7%	3.4%	6,537	500	2,529	3.0%
12	8,418	2.6%	2.7%	3.3%	6,406	500	2,512	3.0%

GAP Insurance—Techniques and Challenges

Month	(1) Balance	(2) Balance Earnings	(3) Rule of 78s	(4) Abbreviat ed Rule of 78s	(5) Example Book Value	(6) Deduct	(7) GAP Severity	(8) Example Book Value Earnings
13	8,269	2.6%	2.6%	3.2%	6,278	500	2,491	3.0%
14	8,119	2.5%	2.6%	3.1%	6,152	500	2,467	3.0%
15	7,969	2.5%	2.5%	3.0%	6,029	500	2,439	2.9%
16	7,817	2.4%	2.5%	2.9%	5,909	500	2,409	2.9%
17	7,665	2.4%	2.4%	2.8%	5,790	500	2,374	2.9%
18	7,511	2.3%	2.3%	2.7%	5,675	500	2,337	2.8%
19	7,357	2.3%	2.3%	2.6%	5,561	500	2,296	2.8%
20	7,202	2.2%	2.2%	2.5%	5,450	500	2,252	2.7%
21	7,046	2.2%	2.2%	2.4%	5,341	500	2,205	2.7%
22	6,889	2.1%	2.1%	2.3%	5,234	500	2,155	2.6%
23	6,731	2.1%	2.1%	2.2%	5,129	500	2,102	2.5%
24	6,573	2.0%	2.0%	2.1%	5,027	500	2,046	2.5%
25	6,413	2.0%	2.0%	2.0%	4,926	500	1,987	2.4%
26	6,252	1.9%	1.9%	1.9%	4,828	500	1,925	2.3%
27	6,091	1.9%	1.9%	1.8%	4,731	500	1,860	2.2%
28	5,928	1.8%	1.8%	1.7%	4,637	500	1,792	2.2%
29	5,765	1.8%	1.7%	1.6%	4,544	500	1,721	2.1%
30	5,600	1.7%	1.7%	1.5%	4,453	500	1,648	2.0%
31	5,435	1.7%	1.6%	1.4%	4,364	500	1,571	1.9%
32	5,269	1.6%	1.6%	1.4%	4,277	500	1,492	1.8%
33	5,102	1.6%	1.5%	1.3%	4,191	500	1,410	1.7%
34	4,933	1.5%	1.5%	1.2%	4,107	500	1,326	1.6%
35	4,764	1.5%	1.4%	1.1%	4,025	500	1,239	1.5%
36	4,594	1.4%	1.4%	1.0%	3,945	500	1,149	1.4%
37	4,423	1.4%	1.3%	0.9%	3,866	500	1,057	1.3%
38	4,250	1.3%	1.3%	0.8%	3,788	500	962	1.2%
39	4,077	1.3%	1.2%	0.7%	3,713	500	865	1.0%

Month	(1) Balance	(2) Balance Earnings	(3) Rule of 78s	(4) Abbreviat ed Rule of 78s	(5) Example Book Value	(6) Deduct	(7) GAP Severity	(8) Example Book Value Earnings
40	3,903	1.2%	1.1%	0.6%	3,638	500	765	0.9%
41	3,728	1.2%	1.1%	0.5%	3,566	500	662	0.8%
42	3,551	1.1%	1.0%	0.4%	3,494	500	557	0.7%
43	3,374	1.0%	1.0%	0.3%	3,424	500	450	0.5%
44	3,196	1.0%	0.9%	0.2%	3,356	500	340	0.4%
45	3,016	0.9%	0.9%	0.1%	3,289	500	228	0.3%
46	2,836	0.9%	0.8%	0.0%	3,223	500	113	0.1%
47	2,655	0.8%	0.8%	0.0%	3,159	500	–	0.0%
48	2,472	0.8%	0.7%	0.0%	3,095	500	–	0.0%
49	2,288	0.7%	0.7%	0.0%	3,033	500	–	0.0%
50	2,104	0.7%	0.6%	0.0%	2,973	500	–	0.0%
51	1,918	0.6%	0.5%	0.0%	2,913	500	–	0.0%
52	1,731	0.5%	0.5%	0.0%	2,855	500	–	0.0%
53	1,543	0.5%	0.4%	0.0%	2,798	500	–	0.0%
54	1,354	0.4%	0.4%	0.0%	2,742	500	–	0.0%
55	1,164	0.4%	0.3%	0.0%	2,687	500	–	0.0%
56	973	0.3%	0.3%	0.0%	2,633	500	–	0.0%
57	781	0.2%	0.2%	0.0%	2,581	500	–	0.0%
58	587	0.2%	0.2%	0.0%	2,529	500	–	0.0%
59	393	0.1%	0.1%	0.0%	2,479	500	–	0.0%
60	197	0.1%	0.1%	0.0%	2,429	500	–	0.0%

- (1) Loan Balance for 60-month loan, 7% interest rate
- (2) Earnings based on (1)
- (3) Traditional Rule of 78s
- (4) Rule of 78s reducing term to 45 months from 60 months

- (5) Book value assuming monthly 2% depreciation
- (7) Maximum of (1) - (5) + (6) and 0
- (8) Using (7) to form earnings curve.

A mature book of business can be analyzed to see the indicated underlying earnings patterns without relying on formulaic earnings patterns, but results should be reasonably close to an abbreviated Rule of 78s.

5. PROJECTING THE RESULTS

Once an appropriate earned premium has been calculated using either the Rule of 78s, the Abbreviated Rule of 78s, or actual historical data, the actuary should project the losses. It would be incorrect to merely assume that the current loss ratio will continue into the future; explicit assumptions should be made about future frequencies and severities. In addition, the book should be analyzed separately by credible class, such as term, initial GAP (if available), vehicle type and other relevant private passenger rating variables.

Of course, one should be careful because GAP is subject to volatile results due to fluctuations in the financing and used-vehicle market. GAP pure premiums will increase when financing standards become more lax because lenders will allow more negative equity (as described above) to be rolled into new loans. In addition, changes in used vehicle pricing will affect future loss rates.

Ideally, the book would contain loan amounts, terms, and book values (both historical and current). With this information, every current GAP could be modeled and the future GAP could be forecasted using different economic scenarios. On a more practical level, it is easier to calculate the earned contracts (using the same earnings factors described above) for different policy years and compare the results under different economic conditions.

A generalized linear model can be utilized to forecast the frequency and severity by class. For frequency, a logistic regression model is appropriate (since there can be only either zero or one claim.)

A logistic model would be specified by the claims divided by the earned contracts—with the earned factor calculated above applied to the contract. Contracts that have incurred a claim could be considered fully earned.

Logistic regression uses the natural logs of the frequencies of claims.

$$\text{logit}(p_i) = \ln \left(\frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i}.$$

Where p = observed frequency and B = parameters and x = observed values for significant rating variables.

For severity, a gamma model provides a decent fit since there is not a significant tail on the severity amounts.

Unless the data set contains a long time period with periods of inflation and deflation in used-vehicle pricing, it is not possible through statistical models to capture all the variability associated with GAP claims.

Therefore, one could explicitly model various used vehicle pricing scenarios to better understand the potential variability of the results.

Profit provisions for GAP should recognize the variability of results due to economic conditions. This may imply a larger profit and contingencies provision than for more stable lines of insurance. Analyzing the proper profit and contingencies provision for GAP is beyond the scope of this paper, but would be a good topic for further research.

6. ENHANCEMENTS TO GAP

Recently, companies have introduced enhancements or added features to GAP products that offer down payment assistance or additional consideration when an insured files a claim. For example, a product might offer an additional \$1,000 for any total loss during the policy period. In this case, the earnings curve would substantially be reversed, since it is more likely that a vehicle would be declared a total loss at the end of the contract.

These cases must be analyzed independently. Ideally, one could forecast the expected pure premium at each month in the contract for the additional coverage. This could be combined with the expected pure premium for the traditional GAP coverage and a new earnings curve would be formed.

7. SIMILAR PRODUCTS

Other asset classes that may have GAP policies are boats, recreational vehicles, and commercial equipment. In addition, financial institutions such as banks may purchase a GAP policy to cover their entire portfolio of loans when the borrower is unable to satisfy the lien after a total loss.

Another type of insurance similar to GAP is called Residual Value Insurance (RVI). RVI is often purchased in situations where an owner is leasing some type of property to another party. Examples of property that may be covered by RVI are rental real estate, leased automobiles, rolling stock of railroads (i.e., train cars), and leased airplanes. RVI will cover the difference between actual book value of the property at the end of the lease and the residual value specified in the lease. The specified residual value will be a forecast of the book value at the time of the purchase of the property.

In the past, RVI products have incurred significant underwriting losses and may not currently be available for all types of property. This is likely due to the difficulty of forecasting future residual values.

These products may be analyzed in similar fashion, though the terms may differ by product since

they can be customized. In addition, less data will typically be available for asset classes other than private passenger vehicles.

REFERENCES

- [1] Hause, Christopher, “2009 CCIA GAP Study,” Presentation at the 2009 Consumer Credit Insurers Association Conference for Actuaries and Credit Insurance Practitioners, Chicago, IL, July 28, 2009.
- [2] Kerper, John, “Reserving for Automobile Warranty and Other Long Duration Contracts—GAP & Residual Value,” Presentation at the 2002 Casualty Loss Reserve Seminar, Arlington, VA, September 23-24, 2002.
- [3] Kerper, John, “Managing Non-Life Insured Products Sold Through Auto Dealers, Part II—Gap” *Society of Actuaries—NewsDirect*, No. 38, October 2001, pp. 10-11.

Credibility for a Tower of Excess Layers

David R. Clark, FCAS, MAAA

Abstract:

In pricing excess of loss reinsurance, the traditional method for applying credibility is as a weighted average of two estimates of expected loss: one from experience rating and a second from exposure rating. This paper will show how this method can be improved by incorporating loss estimates from lower layers, thus producing a multi-factor, credibility-weighted estimate of expected loss.

The method described is based on minimum variance criteria, whereby the resulting credibility-weighted estimator has a lower variance than any other combination of the individual estimators. It is shown that the multi-factor credibility model can be presented as a simple recursive procedure for practical application.

Keywords. Excess of loss reinsurance, exposure rating, credibility.

1. INTRODUCTION

This paper will address a particular problem in pricing excess reinsurance that can benefit from an application of credibility theory.

In reinsurance, an actuary or underwriter is required to estimate losses in a per-occurrence excess layer. For example, a treaty may cover loss occurrences that exceed a retention of \$1,000,000 up to a limit of an additional \$1,000,000; this would be referred to as a \$1,000,000 xs \$1,000,000 layer.

In order to estimate the expected losses in the excess layer, there may be several tools available. The first is a pure experience rating, sometimes called a “burn cost” because of its use in rating fire policies. An experience rating looks at the actual historical losses for the ceding company that have penetrated the excess layer – including adjustments for trend and development – relative to the historical exposures.

In addition to the experience rating, there is usually an industry-based, size-of-loss distribution available. This size-of-loss distribution gives the probability of a loss penetrating into the excess layer and the expected severity in the layer. It is the basis for an “exposure rating” estimate. More precisely, it is the basis for multiple exposure rating estimates, because there are a variety of ways that the size-of-loss distribution can be used.

The exposure rating curve can be used to divide an overall (primary or ground-up) expected loss into the losses expected in various layers. The overall expected loss can be a permissible loss ratio (for example, 100% minus expenses) applied to manual premium. More often, it is calculated from the ceding company’s experience. In such case, the exposure rating is clearly not independent from the experience rating.

Alternatively, the size-of-loss curve can be applied to an estimate of total claim counts for the ceding company. It could also be applied to a lower excess layer; for example, we use the size-of-loss distribution to estimate the \$1,000,000 xs \$1,000,000 layer relative to the \$500,000 xs \$500,000 layer.

We see, therefore, that the analyst has a collection of estimators available. These estimators are not independent from one another but instead are related in many ways. Our goal is to select among these estimators, or combine them, in an optimal way.

Credibility theory can help us accomplish this goal.

1.1 Research Context

This paper builds upon existing credibility theory. However, much of the past literature has been concerned with primary ratemaking, comparing loss experience in one class of business with others. For the reinsurance context, our concern will be “vertical” rather than “horizontal,” as we look at a tower of contiguous excess layers.

The excess reinsurance problem was taken up by Mashitz and Patrik (1990), who limited their discussion to the problem of layer counts. More recently, papers by Cockroft (2004), Goulet, Forgues and Lu (2006), Parodi and Bonche (2008), and Marcus (2010) have included analysis that addresses severity as well as frequency. In general, these papers do not include methods that capture all the ways that exposure and experience ratings are interrelated.¹

The present paper will examine expected losses to excess layers including some of the interrelationships between how exposure and experience rating are applied in practice. The focus will be on showing how the credibility procedure actually reduces the variance in the estimate of expected loss.

1.2 Objective

The goal is to outline a procedure that will produce an optimal or best estimate of expected loss for the excess layer being priced. “Best” will mean a minimum variance unbiased estimator.

Informally stated, the minimum variance criterion says that an estimate that incorporates all available information is more reliable than one that ignores some information (such as losses in lower excess layers).

¹ See for example the “Practical considerations” (section 6) in Cockroft (2004).

1.3 Outline

The remainder of the paper proceeds as follows.

Section 2 will describe the basics of credibility theory supporting the proposed method.

Section 3 will show how this theory can be applied in practice as a recursive credibility method. In order to illustrate the technique, a simple example using a Pareto distribution will be traced throughout the paper.

The final result of this paper will be a very practical method for applying credibility that works recursively. It starts with a simple weighting of experience and exposure rates for a low layer and then uses a layer relativity from the exposure curve to provide an estimate for the next layer up. This practical implementation can be used even without direct reference to the theoretical model that is demonstrated.

2. BACKGROUND AND METHODS

Our goal in setting up a credibility procedure is to find the best possible estimate of future expected losses, making use of all available relevant information. A best estimate will generally have two main properties:

- The estimate will be unbiased; meaning that its expected value will be equal to the true expected value.
- The estimate will have minimum variance; meaning informally that it will tend to be closer to the true expected value than other possible estimates.

We will assume that all of the estimators used in our discussion are unbiased. If some are biased, then they need to be adjusted to an unbiased basis before they are included in a credibility-weighted average.²

The focus of this paper will be on the minimum variance criterion. The best combination of estimators will be one that minimizes the overall variance. For this reason, this approach is also known as “least squares” or “greatest accuracy” credibility (cf. Boor 1992, Venter 2003, Marcus 2010).

If we have two or more estimates available that make use of different information, the best estimate may be some combination of those estimates. Credibility theory allows us to properly

² See Section 3 of Marcus (2010) for a good discussion on testing the validity of the unbiasedness assumption.

combine these different estimates, so that we have a single final estimate that makes the best use of all of the available information.

2.1 The Two-Factor Model

We can begin with the familiar case in which credibility is applied as the weighted average of two estimators, $\widehat{\mu}_1$ and $\widehat{\mu}_2$, which are assumed to be unbiased estimators of a true value μ .

$$\widehat{\mu}_{cw} = w \cdot \widehat{\mu}_1 + (1 - w) \cdot \widehat{\mu}_2. \quad (2.1)$$

The assumption that these estimates are unbiased is expressed as follows.

$$E(\widehat{\mu}_1) = E(\widehat{\mu}_2) = \mu. \quad (2.2)$$

The variance of the credibility-weighted (cw) average of the two estimators is a linear combination of the variances and covariances.

$$Var(\widehat{\mu}_{cw}) = w^2 \cdot Var(\widehat{\mu}_1) + 2w \cdot (1 - w) \cdot Cov(\widehat{\mu}_1, \widehat{\mu}_2) + (1 - w)^2 \cdot Var(\widehat{\mu}_2). \quad (2.3)$$

The optimal value of the credibility weight can be found by least squares by setting:

$\frac{\partial Var(\widehat{\mu}_{cw})}{\partial w} = 0$. This produces the following weight.

$$\widehat{w} = \frac{Var(\widehat{\mu}_2) - Cov(\widehat{\mu}_1, \widehat{\mu}_2)}{Var(\widehat{\mu}_1) + Var(\widehat{\mu}_2) - 2 \cdot Cov(\widehat{\mu}_1, \widehat{\mu}_2)}. \quad (2.4)$$

The calculated weights can be substituted back into the formula for the variance of the credibility-weighted estimator (formula 2.5). This form is instructive because it shows that the variance of the credibility-weighted estimator is less than either of the individual estimators' variances. We can therefore see the value in a rigorous credibility formula as improving our ability to estimate an expected loss more accurately.

$$Var(\widehat{\mu}_{cw}) = \frac{Var(\widehat{\mu}_1) \cdot Var(\widehat{\mu}_2) \cdot (1 - \rho^2)}{Var(\widehat{\mu}_1) + Var(\widehat{\mu}_2) - 2 \cdot Cov(\widehat{\mu}_1, \widehat{\mu}_2)}. \quad (2.5)$$

In this expression, the correlation coefficient is defined as follows.

$$\rho = \frac{Cov(\widehat{\mu}_1, \widehat{\mu}_2)}{\sqrt{Var(\widehat{\mu}_1) \cdot Var(\widehat{\mu}_2)}}. \quad (2.6)$$

2.2 Multi-factor Model

The multi-factor theory can be expanded to include multiple estimators. In this case, we need to define a covariance matrix, Σ , which includes the variances and covariances between each pair of estimators.

$$\Sigma = \begin{bmatrix} \text{Var}(\widehat{\mu}_1) & \text{Cov}(\widehat{\mu}_1, \widehat{\mu}_2) & \cdots & \text{Cov}(\widehat{\mu}_1, \widehat{\mu}_n) \\ \text{Cov}(\widehat{\mu}_2, \widehat{\mu}_1) & \text{Var}(\widehat{\mu}_2) & \cdots & \text{Cov}(\widehat{\mu}_2, \widehat{\mu}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\widehat{\mu}_n, \widehat{\mu}_1) & \text{Cov}(\widehat{\mu}_n, \widehat{\mu}_2) & \cdots & \text{Var}(\widehat{\mu}_n) \end{bmatrix}. \quad (2.7)$$

The credibility-weighted average of the n unbiased estimators is again a linear function of the individual estimators.

$$\widehat{\mu}_{cw} = w_1 \cdot \widehat{\mu}_1 + w_2 \cdot \widehat{\mu}_2 + \cdots + w_n \cdot \widehat{\mu}_n. \quad (2.8)$$

The set of these weights is defined as a vector of parameters.

$$\overrightarrow{W}^T = \langle w_1, w_2, \dots, w_n \rangle. \quad (2.9)$$

The constraint that these weights must add up to 1.00 (or 100%) can be written as $\mathbf{1} = \overrightarrow{W}^T \cdot \mathbf{1}_n$ where $\mathbf{1}_n$ is a column vector of ones.

The variance of the credibility-weighted estimator is then a weighted average of the variance and covariance terms in Σ .

$$\text{Var}(\widehat{\mu}_{cw}) = \overrightarrow{W}^T \cdot \Sigma \cdot \overrightarrow{W}. \quad (2.10)$$

The least-squares estimate for these weights can be found by solving the equation above. The result is that the weights are proportional to the row (or column) totals of the inverse of the covariance matrix.³

$$\overrightarrow{W} = (\mathbf{1}_n^T \cdot \Sigma^{-1} \cdot \mathbf{1}_n)^{-1} \cdot \Sigma^{-1} \cdot \mathbf{1}_n. \quad (2.11)$$

For the special case in which all of the estimators are independent, this reduces to having the weights proportional to the inverse of each estimator's variance.⁴

$$w_i = \frac{\text{Var}(\widehat{\mu}_i)^{-1}}{\sum_{k=1}^n \text{Var}(\widehat{\mu}_k)^{-1}}. \quad (2.12)$$

One final observation before showing how this applies to excess pricing is that the multivariate

³ This result is well known in other branches of finance and represents the solution to the minimum variance or efficient portfolio weights. See for example Theorem 17.1 in Hardle and Hlavka (2007).

⁴ This result is given as Theorem A.3 in Bühlmann and Gisler (2005), p.280. It is also a standard feature of weighted regression theory.

case can alternatively be written in a recursive form. For example, a three-variable case can be viewed as a weighted average between one variable and the weighted average of the other two variables.

$$\begin{aligned}\widehat{\mu}_{cw} &= w_1 \cdot \widehat{\mu}_1 + w_2 \cdot \widehat{\mu}_2 + w_3 \cdot \widehat{\mu}_3 \\ &= z_1 \cdot \widehat{\mu}_1 + (1 - z_1) \cdot \{z_2 \cdot \widehat{\mu}_2 + (1 - z_2) \cdot \widehat{\mu}_3\}.\end{aligned}\tag{2.13}$$

3. CREDIBILITY APPLIED TO EXCESS-OF-LOSS REINSURANCE

The specific problem that we are examining is to find the best estimate of expected loss in an excess layer.

In order to make this discussion more practical, we will make an assumption that the true severity distribution is a single parameter Pareto, defined as in Section 3.1. In Section 3.2, we will then show first how exposure and experience rating estimates are combined. Finally, in Section 3.3, we will show how lower excess layers can also be incorporated in the method using a recursive form of the multi-factor credibility formula.

3.1 Defining the Reinsurance Problem

In order to describe the expected loss in the reinsurance application, we need to start with some definitions:

X	random variable representing a single loss event
$F(x)$	Cumulative Distribution Function; probability that a loss is x or less
R	Retention taken by the ceding company
L	Limit above the Retention covered by the reinsurer
$Layer$	Function representing loss taken by the reinsurer
	Defined as: $Layer = MIN\{MAX(x - R, 0), L\}$
N	Random variable representing the number of losses in the historical period

In order to make this discussion more realistic, we will define a simple curve form to use in the calculation of the credibility factors. For our example, we will use the single parameter Pareto

distribution,⁵ defined as follows:

$$F(x) = 1 - \left(\frac{\theta}{x}\right)^\alpha \quad \text{for } x \geq \theta. \quad (3.1)$$

The value theta, θ , is known as the loss threshold, and represents the smallest loss amount that is part of the analysis. For example, in a reinsurance submission, we might ask for all losses of \$500,000 and greater.

The expected loss in an excess layer is calculated as follows:

$$E(\text{Layer}) = \begin{cases} \left(\frac{\theta}{\alpha - 1}\right) \cdot \left[\left(\frac{\theta}{R}\right)^{\alpha-1} - \left(\frac{\theta}{R+L}\right)^{\alpha-1}\right] & \alpha \neq 1 \\ \theta \cdot \ln\left(1 + \frac{L}{R}\right) & \alpha = 1 \end{cases} \quad (3.2)$$

Similarly, the second moment of an excess layer is calculated as follows:

$$E(\text{Layer}^2) = \begin{cases} \left(\frac{2\theta^2}{(\alpha - 1) \cdot (\alpha - 2)}\right) \cdot \left[\left(\frac{\theta}{R}\right)^{\alpha-2} - \left(\frac{R + (\alpha - 1)L}{R + L}\right) \cdot \left(\frac{\theta}{R + L}\right)^{\alpha-2}\right] & \alpha \neq 1, 2 \\ 2\theta \cdot \left\{L - R \cdot \ln\left(1 + \frac{L}{R}\right)\right\} & \alpha = 1 \\ 2\theta^2 \cdot \left\{\ln\left(1 + \frac{L}{R}\right) - \left(\frac{L}{R+L}\right)\right\} & \alpha = 2 \end{cases} \quad (3.3)$$

For our example, we will have the following information available:

- All losses above a threshold $\theta = \$500,000$.
- Experience rating for the \$500,000 xs \$500,000 (or \$500xs\$500) layer (*Layer*₁).
- Experience rating for the \$1,000,000 xs \$1,000,000 (or 1Mxs1M) layer (*Layer*₂).
- An insurance industry-based, Pareto distribution with parameter α_0 .
- An estimate of the expected number of losses above θ , denoted n_0 .

⁵ This distribution, along with the formulas for capped moments related to (3.2) and (3.3), can be found in Appendix A.4.1.4 of Klugman et al. (2004).

(This estimate n_0 comes from manual rating, not from account experience.)

3.2 Combining Exposure and Experience Rating Estimates

We now proceed to define exposure and experience rating models and how they can be combined.

3.2.1 Exposure Rating

An exposure rate is an estimate of expected losses in an excess layer based on external insurance data. It is sometimes called the “prior estimate” because it can be calculated prior to seeing the actual loss experience for the ceding company.

The exposure rate requires two pieces of information: a severity (size-of-loss) curve from industry-wide data, and an expected number of total losses. Because we are assuming that the severity follows a Pareto distribution, we only need a single parameter α_0 to describe it. For the expected number of losses in the prospective period, we likewise have a prior estimate \hat{n}_0 .

In addition to our prior estimates α_0 and \hat{n}_0 , we also need to have estimates of the variances around these estimates $Var(\alpha_0)$ and $Var(\hat{n}_0)$. The coefficient of variation (CV) related to the frequency is given below.

$$CV_{n_0} = \frac{\sqrt{Var(\hat{n}_0)}}{\hat{n}_0}. \quad (3.4)$$

We can also approximate the variance of the severity using the “delta method”⁶ relative to the variance of the parameter α_0 :

$$Var(E(Layer|\alpha_0)) \approx Var(\alpha_0) \cdot \left[\frac{\partial E(Layer|\alpha_0)}{\partial \alpha_0} \right]^2. \quad (3.5)$$

The derivative with respect to the Pareto alpha is easily calculated.

⁶ A multivariate version of the delta method is described in Loss Models (Klugman et al). For the single parameter Pareto, the variance approximation is much simpler.

The univariate delta method is based on approximating a function using the first two terms of the Taylor series expansion, $g(x) \approx g(a) + g'(a) \cdot (x - a)$, which results in $Var(g(x)) \approx (g'(a))^2 \cdot Var(x)$. This method would provide an exact result if the *Layer* formula were a linear function of α ; because it is not, our results are only approximate.

$$\begin{aligned} \frac{\partial E(Layer|\alpha_0)}{\partial \alpha_0} & \qquad \qquad \qquad (3.6) \\ &= \frac{\theta}{(\alpha_0 - 1)} \cdot \left\{ \ln\left(\frac{\theta}{R}\right) \cdot \left(\frac{\theta}{R}\right)^{\alpha_0-1} - \ln\left(\frac{\theta}{R+L}\right) \cdot \left(\frac{\theta}{R+L}\right)^{\alpha_0-1} \right\} \\ &\quad - \frac{\theta}{(\alpha_0 - 1)^2} \cdot \left\{ \left(\frac{\theta}{R}\right)^{\alpha_0-1} - \left(\frac{\theta}{R+L}\right)^{\alpha_0-1} \right\}. \end{aligned}$$

The exposure rate and the variance⁷ around the exposure rate are therefore estimated as follows:

$$\widehat{\mu}_{expos} = \hat{n}_0 \cdot E(Layer|\alpha_0). \qquad (3.7)$$

$$Var(\widehat{\mu}_{expos}) = \hat{n}_0^2 \cdot CV_{n_0}^2 \cdot E(Layer|\alpha_0)^2 + \hat{n}_0^2 \cdot (CV_{n_0}^2 + 1) \cdot Var(E(Layer|\alpha_0)). \qquad (3.8)$$

From these expressions for the exposure rate, we may observe that both the mean and standard deviation are proportional to the expected number of losses above the threshold θ . This allows us to scale the exposure rate for any change in subject premium.

Having defined the components of exposure rating, it is useful to show representative values⁸ for these calculations.

Following our earlier introduction, we will assume that the severity is a single parameter Pareto with a threshold θ of \$500,000. For the parameter α , we will select a value of 1.500. The variance around this Pareto parameter can be roughly estimated by first selecting a range of possible values. For our example, we will assume that the variance is .05, with this amount selected by the user.

For expected counts \hat{n}_0 above the threshold for the future period, we will select an average value of five losses. The variance around this number is more difficult to estimate, as it may be dependent on how much variance there is for risks within a manual rating classification. If the frequency is more judgmentally selected, then there may be even more uncertainty. To illustrate the calculations, we will assume a coefficient of variation (CV) of .300 or 30%.

From these selected values, we can estimate the severity for both excess layers, the exposure rate (frequency times severity), and the parameter variance around our estimated exposure rate.

⁷ This formula assumes that the estimates for frequency and severity are independent, and then makes use of the relationship: $Var(X \cdot Y) = E(X)^2 \cdot Var(Y) + Var(X) \cdot E(Y)^2 + Var(X) \cdot Var(Y)$.

⁸ For all of these numerical examples, the numbers are purely for illustration purposes and should not be taken as recommendations for pricing.

Table 1 - Variance around Exposure Rate		
<u>Description</u>	<u>Notation</u>	<u>Value</u>
Pareto Threshold	θ	500,000
Pareto Alpha	α_0	1.500
Variance around Alpha	$Var(\alpha_0)$.05
Expected Severity 500xs500	$E(Layer_1 \alpha_0)$	292,893
Expected Severity 1Mxs1M	$E(Layer_2 \alpha_0)$	207,107
Variance around Layer 2 Severity	$Var(E(Layer_2 \alpha_0))$	2.230E+09
Expected Counts at Threshold	\hat{n}_0	5.0
Coefficient of Variation of Counts	CV_{n_0}	.300
Exposure Rate for 1Mxs1M	$\widehat{\mu}_{expos}$	1,035,535
Variance around Exposure Rate	$Var(\widehat{\mu}_{expos})$	1.573E+11

3.2.2 Experience Rating

An experience rate is an estimate of expected losses in an excess layer based on the actual loss experience for the ceding company. For our notation, this will be denoted a “burn cost” with the subscript “bc.”

In our estimate of the experience rate, we need to adjust the sum of historical losses in the layer to the prospective period based on the relative exposure volumes (V).

$$\widehat{\mu}_{bc} = \frac{V_{prospective}}{V_{historical}} \cdot \sum_{k=1}^N Layer_{2,k}. \tag{3.9}$$

This expression is therefore simply the sum of the historical losses that penetrate into the second layer (\$1,000,000 xs \$1,000,000) adjusted to the volume of premium in the prospective period. It is assumed that these losses are trended to the future level and that the historical premium is likewise adjusted (“onleveled”) to the future level.

The excess development can be built into this calculation by using as the historical exposure volume the onlevel premium divided by excess development:

$$V_{historical} = \sum_{Years=t} \frac{V_t}{LDF_t}. \quad (3.10)$$

If we assume that the frequency distribution is Poisson, then we can estimate an expected variance around the experience rate.

$$Var(\widehat{\mu}_{bc}) = \left(\frac{V_{prospective}}{V_{historical}} \right)^2 \cdot E(N) \cdot E(Layer_2^2 | \alpha_0). \quad (3.11)$$

This variance is based on the expected process variance⁹ of the severity from the exposure rating model. The relationship between the prospective expected counts and the expected counts for the historical period is based on the assumption that the claim frequency relative to the onlevel premium is unchanged.

$$\frac{E(\widehat{n}_0)}{V_{prospective}} = \frac{E(N)}{V_{historical}}. \quad (3.12)$$

We can estimate the expected losses in the historical period, $E(N)$, by making use of the prospected expected losses from exposure rating, $E(\widehat{n}_0)$, and formula (3.12).

The table below shows the results of these calculations.

Table 2 - Variance around Experience Rate		
<u>Description</u>	<u>Notation</u>	<u>Value</u>
Expected Severity 1Mxs1M	$E(Layer_2 \alpha_0)$	207,107
Second Moment	$E(Layer_2^2 \alpha_0)$	1.716E+11
Expected Prospective Counts	$E(\widehat{n}_0)$	5.0
Prospective Premium	$V_{prospective}$	2,000,000
Historical Onlevel Premium	$V_{historical}$	10,000,000
Expected Historical Counts	$E(N)$	25.0
Experience Rate for 1Mxs1M	$E(\widehat{\mu}_{bc})$	1,035,534
Variance around Experience Rate	$Var(\widehat{\mu}_{bc})$	1.716E+11

⁹ We are making an approximation in this paper that the expected process variance $E_\alpha[Var(Layer|\alpha)]$ can be approximated as $Var(Layer|E(\alpha))$. Without this approximation, we would need to specify a complete prior distribution for the α instead of just the variance. Alternatively, the process variance could be estimated from the empirical experience rating.

3.2.3 Credibility weighting these two estimates

The experience and exposure rating models produce estimates of the future expected loss to an excess layer. Because they are working with very different information, they can be considered independent.

$$\widehat{\mu}_{cw} = w \cdot \widehat{\mu}_{bc} + (1 - w) \cdot \widehat{\mu}_{expos}. \quad (3.13)$$

The credibility weight for the experience rate is then written in a familiar form, based on the expected number of claims in the historical period (substituting in formulas 3.8 and 3.11).

$$w = \frac{Var(\widehat{\mu}_{expos})}{Var(\widehat{\mu}_{bc}) + Var(\widehat{\mu}_{expos})} = \frac{E(N)}{E(N) + k} \quad (3.14)$$

where

$$k = \frac{E(Layer_2^2 | \alpha_0)}{CV_{n_0}^2 \cdot E(Layer | \alpha_0)^2 + (CV_{n_0}^2 + 1) \cdot Var(E(Layer | \alpha_0))}. \quad (3.15)$$

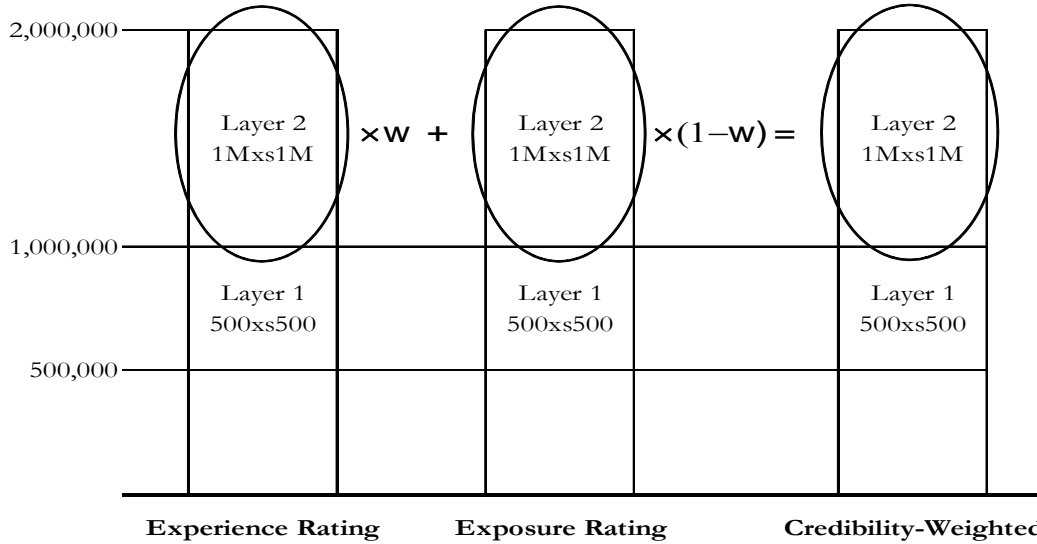
All of the elements of this credibility weight can be evaluated prior to actually estimating the experience rating. To illustrate, we continue with the numerical example.

Table 3 - Variance around Credibility Rate		
<u>Description</u>	<u>Notation</u>	<u>Value</u>
Variance Around Experience Rate	$Var(\widehat{\mu}_{bc})$	1.716E+11
Variance Around Exposure Rate	$Var(\widehat{\mu}_{expos})$	1.573E+11
Expected Historical Counts	$E(N)$	25.0
Credibility “k”	k	27.3
Credibility Weight to Experience	w	47.8%
Variance around Credibility Rate	$Var(\widehat{\mu}_{cw})$	8.206E+10

As expected, the variance around the credibility-weighted rate is less than the variance of either of the individual estimates from exposure or experience rating. This is consistent with our goal of finding an estimator with minimum variance.

We can illustrate the concept of the credibility weighting of experience and exposure rates by the graphic below.

Example of Standard Credibility Procedure



This illustrates the concept that the credibility weighting is based solely on the rates in the \$1,000,000 xs \$1,000,000 layer and makes no use of the information in the lower layer. We now proceed to show how the information in this lower layer can be used.

3.3 Including Losses from a Lower Layer

As noted above, the experience and exposure rating models make use of different sources of information. However, they do not make use of all the information that is available to the analyst. We are also able to price layers of insurance below the layer being quoted.

An additional estimate of expected loss is made by applying relativities from the exposure rating model to the expected loss in the first layer \$500,000 xs \$500,000. This is our “relativity” (rel) method.

$$\widehat{\mu}_{rel} = \left\{ \frac{V_{prospective}}{V_{historical}} \cdot \sum_{k=1}^N Layer_{1,k} \right\} \cdot \frac{E(Layer_2|\alpha_0)}{E(Layer_1|\alpha_0)} \tag{3.16}$$

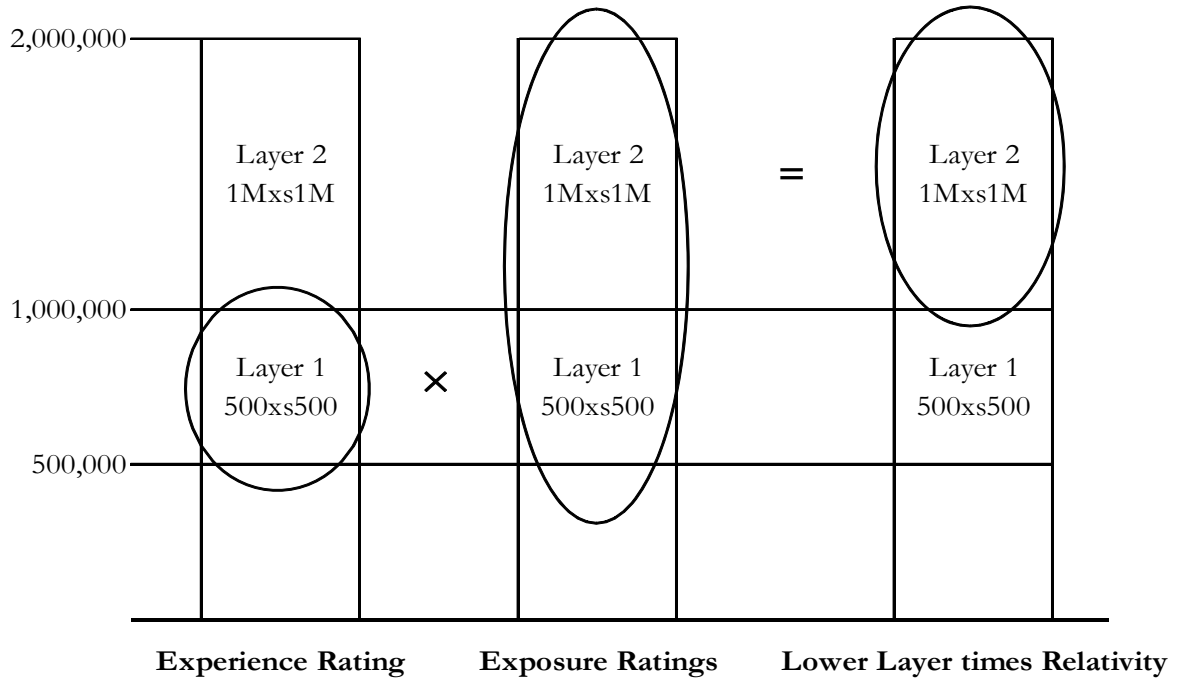
In this formula, we continue to use the shorthand notation:

$$Layer_{1,k} = MIN\{MAX(x_k - 500,000; 0); 500,000\} \text{ for loss } k$$

$$\text{and } Layer_{2,k} = MIN\{MAX(x_k - 1,000,000; 0); 1,000,000\} \text{ for loss } k.$$

The graphic below illustrates how the relativity method makes use of the experience in the lower layer.

Using Exposure-Rating Relativities



This relativity-based estimate is not independent from either the pure exposure rate $\widehat{\mu}_{expos}$ or from the pure experience burn cost $\widehat{\mu}_{bc}$. It shares dependence on the industry size-of-loss distribution with the exposure rate.¹⁰ The experience rates for first and second layers are also clearly related (for example, there can be no losses in the second layer without at least one loss in the first layer with limit L_1).

The remainder of this section will provide detailed formulas using the Pareto severity and Poisson frequency model. These formulas allow us to create a tractable numerical example that can be reproduced by the ambitious reader and may be helpful for gaining intuition about the sensitivity of the credibility weights to the variance assumptions.

However, the key result is not the Pareto/Poisson model itself but the recursive form of credibility that results. The more practical-minded reader can skip the detailed formulas and not miss this key result.

¹⁰ Here we deviate from Marcus (2010), who assumes independence of the exposure rating and the severity curve underlying the ILF. While that assumption avoids the need to calculate this additional covariance term, it does not lead to the practical implementation in a recursive form.

As a starting point, we may observe the covariance between the experience rates in the two layers.¹¹

$$\text{Cov}\left(\sum_{k=1}^N \text{Layer}_1, \sum_{k=1}^N \text{Layer}_2\right) = E(N) \cdot L_1 \cdot E(\text{Layer}_2). \quad (3.17)$$

The covariance of the relativity-based estimate and the experience rate for the second layer is given as follows:

$$\text{Cov}(\widehat{\mu}_{rel}, \widehat{\mu}_{bc}) = \left(\frac{V_{prospective}}{V_{historical}}\right)^2 \cdot E(N) \cdot L_1 \cdot E(\text{Layer}_2) \cdot \frac{E(\text{Layer}_2|\alpha_0)}{E(\text{Layer}_1|\alpha_0)}. \quad (3.18)$$

The severity used in the exposure rate and the layer relativity are closely dependent and may be treated as perfectly correlated.

$$\text{Cov}\left(E(\text{Layer}_2|\alpha_0), \frac{E(\text{Layer}_2|\alpha_0)}{E(\text{Layer}_1|\alpha_0)}\right) = \sqrt{\text{Var}(E(\text{Layer}_2|\alpha_0)) \cdot \text{Var}\left(\frac{E(\text{Layer}_2|\alpha_0)}{E(\text{Layer}_1|\alpha_0)}\right)}. \quad (3.19)$$

The bases to which these exposure rating factors apply are independent, so that the covariance between the exposure rate and the relativity-based estimate is as follows:

$$\begin{aligned} \text{Cov}(\widehat{\mu}_{expos}, \widehat{\mu}_{rel}) & \quad (3.20) \\ &= \hat{n}_0^2 \cdot E(\text{Layer}_1|\alpha_0) \cdot \sqrt{\text{Var}(E(\text{Layer}_2|\alpha_0)) \cdot \text{Var}\left(\frac{E(\text{Layer}_2|\alpha_0)}{E(\text{Layer}_1|\alpha_0)}\right)}. \end{aligned}$$

From the formulas given above, it is interesting to note that we can calculate all of the needed covariances without introducing any additional correlation assumptions into the model. All of the correlation is implied directly by the structure of the layers themselves.

The variance around the relativity-based estimate can also be estimated.

$$\begin{aligned} \text{Var}(\widehat{\mu}_{rel}) & \quad (3.21) \\ &= \left(\frac{V_{prospective}}{V_{historical}}\right)^2 \cdot \left\{ E(N) \cdot E(\text{Layer}_1^2|\alpha_0) \cdot \left(\frac{E(\text{Layer}_2|\alpha_0)}{E(\text{Layer}_1|\alpha_0)}\right)^2 \right. \\ & \quad \left. + (E(N)^2 \cdot E(\text{Layer}_1|\alpha_0)^2 + E(N) \cdot E(\text{Layer}_1^2|\alpha_0)) \right. \\ & \quad \left. \cdot \text{Var}\left(\frac{E(\text{Layer}_2|\alpha_0)}{E(\text{Layer}_1|\alpha_0)}\right) \right\}. \end{aligned}$$

¹¹ This formula is valid if the two layers are not overlapping – that is, the retention on the second layer is higher than the retention plus limit on the first layer.

As with the exposure rate, the variance of the relativity factor can be approximated via the “delta method.”

$$Var\left(\frac{E(Layer_2|\alpha_0)}{E(Layer_1|\alpha_0)}\right) \approx Var(\alpha_0) \cdot \left[\frac{\partial \frac{E(Layer_2|\alpha_0)}{E(Layer_1|\alpha_0)}}{\partial \alpha_0} \right]^2 \quad (3.22)$$

For the Pareto distribution, the relativity ratio is calculated as follows:

$$\frac{E(Layer_2|\alpha_0)}{E(Layer_1|\alpha_0)} = \left\{ \frac{R_2^{1-\alpha_0} - (R_2 + L_2)^{1-\alpha_0}}{R_1^{1-\alpha_0} - (R_1 + L_1)^{1-\alpha_0}} \right\} \quad (3.23)$$

The derivative with respect to the Pareto alpha is easily calculated.

$$\begin{aligned} \frac{\partial \frac{E(Layer_2|\alpha_0)}{E(Layer_1|\alpha_0)}}{\partial \alpha_0} & \quad (3.24) \\ &= \frac{E(Layer_2|\alpha_0)}{E(Layer_1|\alpha_0)} \\ & \cdot \left[\left\{ \frac{\ln(R_1) \cdot R_1^{1-\alpha_0} - \ln(R_1 + L_1) \cdot (R_1 + L_1)^{1-\alpha_0}}{R_1^{1-\alpha_0} - (R_1 + L_1)^{1-\alpha_0}} \right\} \right. \\ & \left. - \left\{ \frac{\ln(R_2) \cdot R_2^{1-\alpha_0} - \ln(R_2 + L_2) \cdot (R_2 + L_2)^{1-\alpha_0}}{R_2^{1-\alpha_0} - (R_2 + L_2)^{1-\alpha_0}} \right\} \right] \end{aligned}$$

The variance around the layer relativity is calculated below.

Table 4 - Variance around Relativity-based Rate		
Description	Notation	Value
Expected Severity 500xs500	$E(Layer_1 \alpha_0)$	292,893
Expected Severity 1Mxs1M	$E(Layer_2 \alpha_0)$	207,107
Layer Relativity	$\frac{E(Layer_2 \alpha_0)}{E(Layer_1 \alpha_0)}$.7071
Variance of Layer Relativity	$Var\left(\frac{E(Layer_2 \alpha_0)}{E(Layer_1 \alpha_0)}\right)$.0120
Variance of Relativity-based Rate	$Var(\widehat{\mu}_{rel})$	8.788E+10

The covariance matrix for the three estimators is:

Credibility for a Tower of Excess Layers

$$\Sigma = \begin{bmatrix} \text{Var}(\widehat{\mu}_{expos}) & 0 & \text{Cov}(\widehat{\mu}_{expos}, \widehat{\mu}_{rel}) \\ 0 & \text{Var}(\widehat{\mu}_{bc}) & \text{Cov}(\widehat{\mu}_{bc}, \widehat{\mu}_{rel}) \\ \text{Cov}(\widehat{\mu}_{rel}, \widehat{\mu}_{expos}) & \text{Cov}(\widehat{\mu}_{rel}, \widehat{\mu}_{bc}) & \text{Var}(\widehat{\mu}_{rel}) \end{bmatrix} \quad (3.25)$$

The inverse of this covariance matrix provides the credibility weights for the three estimates of expected loss. We calculate the inverse of the covariance matrix and then assign the credibility weights proportional to the row (or column) totals.

As the example below shows, this final three-factor credibility estimator has a smaller variance than any of the three individual variances. The resulting variance is also less than the variance from the two-factor credibility calculation (Table 3).

	Expos	bc	Relativity
Covariance Matrix:	1.573E+11	0	3.790E+10
	0	1.716E+11	7.322E+10
	3.790E+10	7.322E+10	8.788E+10
Inverse:	7.580E-12	2.165E-12	-5.073E-12
	2.165E-12	9.663E-12	-8.986E-12
	-5.073E-12	-8.986E-12	2.105E-11
Column Total:	4.672E-12	2.843E-12	6.996E-12
Weights:	32.2%	19.6%	48.2%
Total Variance:	6.891E+10		

The credibility-weighted estimate is a weighted average of the three separate estimates.

$$\begin{aligned} \widehat{\mu}_{cw} = & w_1 \cdot \widehat{n}_0 \cdot E(Layer_2 | \alpha_0) \\ & + w_2 \cdot \frac{V_{prospective}}{V_{historical}} \cdot \sum_{k=1}^N Layer_{2,k} \\ & + w_3 \cdot \left\{ \frac{V_{prospective}}{V_{historical}} \cdot \sum_{k=1}^N Layer_{1,k} \right\} \cdot \frac{E(Layer_2 | \alpha_0)}{E(Layer_1 | \alpha_0)} \end{aligned} \quad (3.26)$$

This can be rearranged in the recursive form discussed earlier. In this form, we see that a credibility weighting is performed between the exposure and experience rates for the first (\$500,000 xs \$500,000) layer. This credibility-weighted estimate for the first layer is then adjusted to the level of the second (\$1,000,000 xs \$1,000,000) layer using relativities, and that amount is weighted with

the experience rate for the second layer.

$$\begin{aligned} \widehat{\mu}_{cw} = & (1 - z_2) \cdot \left\{ (1 - z_1) \cdot \widehat{n}_0 \cdot E(\text{Layer}_1 | \alpha_0) \right. \\ & \left. + z_1 \cdot \frac{V_{prospective}}{V_{historical}} \cdot \sum_{k=1}^N \text{Layer}_{1,k} w_2 \right\} \cdot \frac{E(\text{Layer}_2 | \alpha_0)}{E(\text{Layer}_1 | \alpha_0)} \\ & + z_2 \cdot \frac{V_{prospective}}{V_{historical}} \cdot \sum_{k=1}^N \text{Layer}_{2,k} . \end{aligned} \tag{3.27}$$

If we had additional layers above the second layer, then this recursive procedure could be repeated.

As a practical matter, the variances needed for the rigorous multi-factor model are not known with certainty. Further, the pricing analyst may want to modify the weights based on other considerations such as data quality or potential changes in the underlying exposures that require expert judgment. The recursive form can still be used with judgmentally selected weights as a systematic way to incorporate all of the information from the lower layers.

4. RESULTS AND DISCUSSION

We have seen that a minimum variance or “best” estimate of expected losses in an excess layer is one that makes use of all the available information from both experience and exposure rating models. The combination of estimates from simple methods is conveniently performed in a linear credibility framework.

The final procedure derived from this credibility framework starts with a lower excess layer and credibility weights it with a complement from industry sources. The exposure distribution produces an expected layer relativity that can be applied to this lower layer to produce the complement for a second layer. Higher layers are likewise estimated by climbing recursively up the tower of excess layers.

This recursive procedure is grounded in credibility theory, but it also allows for a high degree of judgment as the analyst can adjust the credibility percentages for each step.

Some outstanding questions left from this research are:

- How can we improve on the estimate of the uncertainty in the exposure rating distribution?

- How can we include other sources of uncertainty, such as variability in trend, development or onlevel factors?
- Is there an optimal way of dividing the layers so that the best of all possible credibility-weighted averages is created?

5. CONCLUSIONS

The credibility procedure outlined in this paper should be useful for excess-of-loss reinsurance or other applications in which expected losses in excess layers need to be estimated. While this procedure was not invented by the author, the grounding in linear credibility theory gives a sound theory for systematically estimating expected losses.

Acknowledgment

The author gratefully acknowledges the helpful discussion with Michael Fackler, Ira Robbin, Jim Sandor, and Marc Shamula. Any errors remaining in the paper are, of course, solely the responsibility of the author.

Supplementary Material

An Excel example of the formulas in this paper is available from the author.

6. REFERENCES

- Bühlmann, Hans and Alois Gisler, *A Course in Credibility Theory and its Applications*, New York: Springer, 2005.
- Boor, Joseph A., “Credibility Based on Accuracy,” *Proceedings of the Casualty Actuarial Society* LXXIX, 1992, pp. 166-185.
- Cockroft, Mark, “Bayesian Credibility for Excess-of-Loss Reinsurance,” Paper presented at the GIRO Conference in Killarney, Ireland, October 12, 2004.
- Goulet, Vincent, Antoni Forgues, and Jiatao Lu, “Credibility for Severity Revisited,” *North American Actuarial Journal* 10 (1), 2006, pp. 49-62.
- Hardle, Wolfgang and Zdenek Hlavka, *Multivariate Statistics: Exercises and Solutions*, New York: Springer, 2007.
- Klugman, Stuart A., Harry H. Panjer, and Gordon E. Wilmot, *Loss Models: From Data to Decisions*, 2nd ed., New York: Wiley, 2004.
- Marcus, Lawrence F., “Credibility for Experience Rating, A Minimum Variance Approach,” *CAS E-Forum*, Summer 2010.
- Mashitz, Isaac and Gary Patrik, “Credibility for Treaty Reinsurance Excess Pricing,” *Casualty Actuarial Society Discussion Paper Program*, May 1990, Vol. 1, pp. 317-368.
- Parodi, Pietro and Stephane Bonche, “Uncertainty-Based Credibility and its Application to Excess-of-Loss Reinsurance,” *CAS E-Forum*, Winter 2008.
- Venter, Gary, “Credibility Theory for Dummies,” *CAS E-Forum*, Winter 2003, pp. 621-627.

Abbreviations and notations:

R	a retention or “attachment point” representing the amount retained by a ceding company before the reinsurance layer responds
L	a limit, representing the maximum amount payable in a treaty on a single loss
$Layer_i$	Formula for a loss in a layer; $Layer = MIN(MAX(X - R, 0), L)$
\hat{n}_0	a priori estimate of expected number of losses in the prospective period
$E(N)$	expected number of losses in the historical period
X	random variable representing a single loss amount
$\hat{\mu}$	an estimator of a true expected loss value μ ; the estimator is itself a random variable

Biography of the Author

Dave Clark is a senior actuary with Munich Reinsurance America. He is a Fellow of the Casualty Actuarial Society and a Member of the American Academy of Actuaries.

Deductibles, Policy Limits, and Reinsurance: A Case Study in Malaysia

Noriszura Ismail, Ph.D., and Ansar Asnawi Ahmad Anuar

Abstract

In developing countries such as Malaysia, the availability of reinsurance arrangements provides several advantages to primary insurers, such as keeping their risk exposures at prudent levels by having large risk exposures reinsured by another company, meeting client requests for larger insurance coverage by having their limited financial sources supported by another company, and acquiring another company's underwriting skills, experience and complex claim handling ability. These are essential considerations for primary insurers that wish to expand their insurance business and reduce the size of their loss exposure, especially in countries like Malaysia, where the number of primary insurers is large and the size of their resources is small. This paper aims to model the amount of insurance loss, to provide a range of deductibles and policy limits based on Loss Elimination Ratios (LER), to compute insolvency probabilities via linear loading and PH-Transform assumptions, to calculate Increased Limit Factors (ILF), to apply a frequency and severity approach to pricing excess-of-loss layers, and to assess the insolvency probability of a reinsurance treaty. In particular, the PH-Transform assumption is applied throughout as a means of incorporating a risk load, thus lowering the insolvency probability of a single excess-of-loss layer as well as multiple layers of a reinsurance treaty.

Keywords: Loss elimination ratio; insolvency probability; reinsurance; general insurance, PH-Transform.

1. INTRODUCTION

Reinsurance premiums in the Malaysian non-life insurance industry may be categorized into those ceded abroad and those ceded within Malaysia. In 1965 and 1975, for instance, reinsurance premiums ceded abroad were RM12 million and RM60 million, equivalent to 17% and 21% of written premiums respectively. These amounts increased to RM296 million and RM1223 million in 1985 and 1995, equivalent to 24% and 27% of written premiums respectively, but decreased to RM957 million in 2005, equivalent to 10% of written premiums (Lee [9], Bank Negara Malaysia [1], Bank Negara Malaysia [2]). Figures 1-2 show the reinsurance premiums ceded abroad (1965-2005) in terms of volume and proportion of written premium. It should be noted that the currency of Ringgit Malaysia (RM) was pegged at RM3.80=USD1 on 2 September 1998 and shifted to a managed float against a basket of currencies as of 21 July 2005.

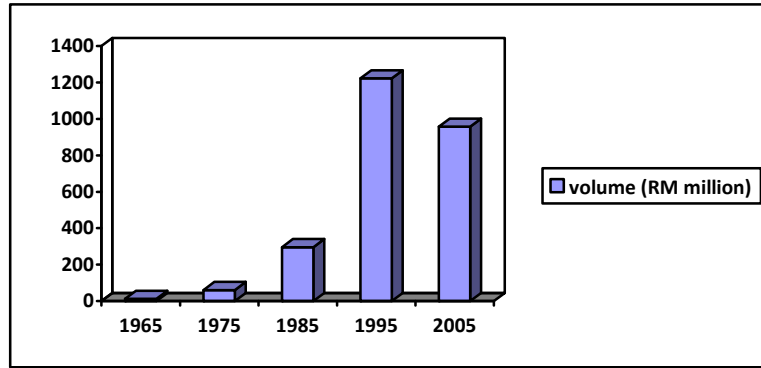


Figure 1: Volume of reinsurance premium ceded abroad (RM million)

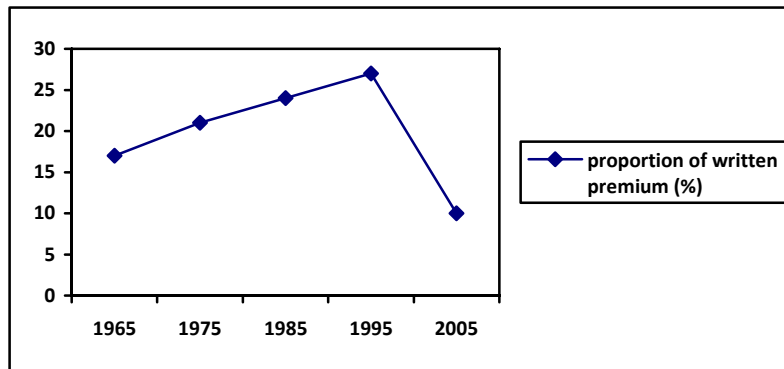


Figure 2: Proportion of reinsurance premium ceded abroad (% of written premium)

Based on the proportion of written premiums, there was a marked deterioration in 1985 and 1995 in terms of domestic retention compared to 1965 and 1975, due to the fact that Malaysia never imposed restrictions on foreign exchange outflows for reinsurance purposes. For most companies, their limited financial resources and expertise in underwriting and handling complex claims increased their dependence upon outside reinsurers, leading to the issue of unsatisfactory domestic retention of premium (Lee [9]). The level of retention improved in 2005, however, largely due to the continuous efforts taken by regulatory bodies and industry players, especially in encouraging domestic insurers and reinsurers to absorb higher proportions of large risks.

Over the past decade, there were many discussions on trade liberalization not only in Malaysia but also in the rest of the world, involving the removal of trade barriers or easing of regulations that inhibit the workings of the free market (Lau [8]). In March 2001, the central bank of

Malaysia, Bank Negara Malaysia (BNM), launched the Financial Sector Masterplan (FSMP). This fairly extensive ten-year road map for the banking and insurance sectors includes specific recommendations that are to be implemented in phases over a ten-year period to deregulate and liberalize the country's financial industry (Bank Negara Malaysia [3]). Even though the local tariff on motor and fire insurance has served its purpose well since its implementation, it is now considered outdated and not reflective of market realities (Lau [8]). The tariff mechanism specified floor rates for various risk classes, but sometimes resulted in cross-subsidization among risk classes, and also within risk classes, whereby better risks subsidized the worse ones (Cummins [9]). In addition, limitations on deductibles and limits have not been appropriately revised to reflect inflation and other economic changes (Rao [10]).

This study aims to model the amount of insurance loss, to provide a range of deductibles and policy limits based on Loss Elimination Ratios (LER), to compute insolvency probabilities via linear loading and PH-Transform assumptions, to calculate Increased Limit Factors (ILF), to apply a frequency and severity approach to pricing excess-of-loss layers, and to assess the insolvency probability of a reinsurance treaty. In particular, the PH-Transform assumption is applied throughout as a means of incorporating a risk load, thus lowering the insolvency probability of a single excess-of-loss layer as well as multiple layers of a reinsurance treaty.

Several studies focusing on reinsurance, deductibles and policy limits have been carried out in the insurance and actuarial literature. Zhuang [14] established orderings of optimal allocations of policy limits and deductibles with respect to the distortion of risk measures; Hua and Cheung [9] applied the equivalent utility premium principle and studied the worst allocations of policy limits and deductibles; Dimitriyadis and Oney [5] modeled loss distributions using the Allianz tool pack, derived premiums at different levels of deductibles, and computed ruin probabilities; and Wang [12] introduced the Proportional Hazard (PH) Transform and applied this method to price ambiguous risks, excess-of-loss coverage, increased limits, risk portfolios and reinsurance treaties.

In this study, the modeling of loss amount, the computation of insolvency probability and the pricing of excess-of-loss layers are based on loss data obtained from one of the leading insurers in Malaysia. The approach suggested in this study can be considered to be fair, as it serves to lower insolvency probability. The suggested approach can also be considered to be efficient, since it can be computed in a straightforward manner using R programming.

2. LOSS MODEL

2.1 Maximum Likelihood Method

Claims data on health insurance's critical illnesses was obtained from one of the leading insurers in Malaysia, providing information on gender (male and female) and age of policyholders (below 25, 25-50 and above 50) in year 2008. In particular, the loss data of sample size $n=192$ for female aged 25-50 is fitted using a maximum likelihood method. Preliminary analysis has been conducted prior to the fitting procedure to ensure that the sample data is trended and does not contain any anomalies or outliers.

The likelihood function for complete individual data is

$$L(\underline{\theta}) = \prod_{i=1}^n f(x_i | \underline{\theta}), \quad (1)$$

where $f(x_i | \underline{\theta})$ denotes the probability density function (p.d.f.) with parameters $\underline{\theta} = \theta_1, \theta_2, \dots, \theta_k$.

The maximum likelihood estimators are obtained by maximizing the log likelihood function:

$$\ln L(\underline{\theta}) = \sum_{i=1}^n \ln f(x_i | \underline{\theta}). \quad (2)$$

Table 1 shows the estimated parameters and the log likelihood of several parametric distributions fitted on the amount of loss, sorted by decreasing values of log-likelihood within the number of parameters. The best models for one-parameter, two-parameter and three-parameter distributions are selected by choosing the largest value of the log likelihood function, $\ln L(\underline{\theta})$.

2.2 Model Selection

The next step to select the best model is to perform the Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) tests. The K-S statistical test is defined as (Klugman et al. [7])

$$D = \max_{t \leq x \leq u} |F_n(x_i) - F^*(x_i)|, \quad i = 1, 2, \dots, n \quad (3)$$

where $F^*(x_i)$ denotes the parametric cumulative distribution function (c.d.f.), and $F_n(x_i)$ the empirical c.d.f. evaluated at x_i respectively. The best model is chosen by selecting the lowest D .

Table 1: Estimated parameters

Parametric distribution	Number of parameters	Estimated parameters	$\ln L(\underline{\theta})$
Exponential	1	$\lambda = 0.000025$	-2,207
Inverse exponential	1	$\theta = 8582.61$	-2,349
Gamma	2	$\alpha = 1.4637$ $\theta = 26,279.57$	-2199.9
Weibull	2	$\theta = 41,256.46$ $\tau = 1.2401$	-2200.4
Loglogistic	2	$\theta = 29,628.99$ $\gamma = 1.9801$	-2,205
Pareto	2	$\theta = 350,026.3$ $\alpha = 9.8929$	-2,211
Inverse Paralogistic	2	$\theta = 20,728.54$ $\tau = 1.4871$	-2,219
Lognormal	2	$\mu = 10.1786$ $\sigma = 1.0639$	-2,227
Inverse Pareto	2	$\theta = 13,487.44$ $\tau = 1.8890$	-2,243
Inverse Weibull	2	$\theta = 14,301.71$ $\tau = 0.6626$	-2,291
Inverse Gamma	2	$\alpha = 0.5573$ $\theta = 4,782.70$	-2,321
Inverse Gaussian	2	$\theta = 8,607.16$ $\mu = 6,000,000$	-2,322
Burr	3	$\theta = 86,426.43$ $\gamma = 1.5169$ $\alpha = 3.7783$	-2,197
Generalized Pareto	3	$\theta = 731,790.4$ $\tau = 1.5305$ $\alpha = 30.1434$	-2,200
Transformed Gamma	3	$\theta = 30,270.96$ $\tau = 1.0664$ $\alpha = 1.3183$	-2,200
Inverse Transformed Gamma	3	$\theta = 8 \times 10^{-12}$ $\tau = 0.1684$ $\alpha = 27.3012$	-2,238

The A-D statistical test, defined as the weighted average of the squared differences of the empirical and parametric c.d.f.s, emphasizes the goodness of fit of the tail over the middle of distribution (Klugman et al. [7]),

$$A^2 = -nF^*(u) + n \sum_{j=0}^k (1 - F_n(y_j))^2 [\ln(1 - F^*(y_j)) - \ln(1 - F^*(y_{j+1}))] + n \sum_{j=1}^k F_n(y_j)^2 [\ln(F^*(y_{j+1})) - \ln(F^*(y_j))], \tag{4}$$

where $y_0 < y_1 < \dots < y_k < y_{k+1} = u$ denote the unique non-censored data, $F^*(y_j)$ the parametric c.d.f. and $F_n(y_j)$ the empirical c.d.f. The best model is chosen by selecting the lowest A^2 .

Finally, the Schwarz Bayesian Criterion (SBC) penalizes models having a greater number of parameters. The SBC is defined as (Klugman et al. [7])

$$SBC = \ln L - \frac{r}{2} \ln n, \tag{5}$$

where r denotes the number of parameters and n the sample size. The best model is chosen by selecting the highest SBC. Table 2 shows the results of the K-S, A-D and SBC tests carried out on loss data. The best-fitting distribution for the loss amount is Burr with parameters $\theta = 86,426.43$, $\gamma = 1.5169$ and $\alpha = 3.7783$ and thus, the following discussion will use this distribution.

Table 2: Results of K-S, A-D and SBC tests

Parametric distribution	Numbers of parameters	K-S test	A-D test	SBC
Exponential	1	0.18655	389.31	-2209.63
Gamma	2	0.11098	384.68	-2205.16
Burr	3	0.09454	383.87	-2204.40

3. LOSS ELIMINATION RATIO (LER)

The Loss Elimination Ratio (LER) is the ratio of the decrease in expected loss for an insurer writing a policy with a deductible and/or policy limit to the expected loss for an insurer writing a full-coverage policy.

3.1 Deductible Policy

When an insurer introduces a deductible to a policy, say at the value of d , the loss retained by the insured may be represented by the random variable Y , where

$$Y = \begin{cases} X, & X < d \\ d, & X \geq d \end{cases}, \quad (6)$$

whereas the loss covered by the insurer and paid as claim may be represented by the random variable W , where

$$W = \begin{cases} 0, & X < d \\ X - d, & X \geq d \end{cases}, \quad (7)$$

so that $X = Y + W$.

Therefore, in terms of an insurer's perspective, the Loss Elimination Ratio (LER) is equal to

$$LER = \frac{E(X; d)}{E(X)}, \quad (8)$$

where

$$E(X; d) = \int_0^d xf(x)dx + d \int_d^{\infty} f(x)dx,$$

and

$$E(X) = \int_0^{\infty} xf(x)dx = \int_0^{\infty} S(x)dx,$$

where $S(x)$ denotes the survival function, which is equal to $1 - F(x)$.

Table 3 shows the LER, written in the currency of Ringgit Malaysia (RM), for several deductible values, assuming individual losses follow a Burr distribution with parameters $\theta = 86,426.43$, $\gamma = 1.5169$ and $\alpha = 3.7783$. As an example, the LER at $d = \text{RM}10,000$ is 0.25, implying that 25% of insurer's losses is eliminated by introducing a deductible of RM10,000. Appendix 1 shows the calculation of LER using R programming with the assistance of the *actuar* package.

Table 3: Values of d and LER

Burr distribution, $E(X) = RM\ 38,131$			
d (RM)	$E(X;d)$ (RM)	LER	Δ LER
0	0	0.000	-
1000	998.27	0.026	0.026
2000	1990.13	0.052	0.026
3000	2972.73	0.078	0.026
4000	3944.03	0.103	0.025
5000	4902.40	0.129	0.026
6000	5846.51	0.153	0.024
7000	6775.27	0.178	0.025
8000	7687.74	0.202	0.024
9000	8583.16	0.225	0.023
10000	9460.91	0.248	0.023
11000	10320.45	0.271	0.023
12000	11161.40	0.293	0.022
13000	11983.42	0.314	0.021
14000	12786.30	0.335	0.021
15000	13569.87	0.356	0.021
16000	14334.05	0.376	0.020
17000	15078.82	0.395	0.019
18000	15804.21	0.414	0.019
19000	16510.29	0.433	0.019
20000	17197.19	0.451	0.018

The graph of LER vs. d is shown in Figure 3, indicating that the ratio of eliminated loss is directly proportional to the deductible. However, after a certain point, a higher deductible can no longer provide a significant proportion of eliminated loss to an insurer.

In practice, the criteria for deductible may differ depending on the requirements and preferences of each insured. Nevertheless, an insurer may use the values shown in Table 3 and the graph shown in Figure 3 to indicate whether the deductible proposed by the insured provides a significant proportion of eliminated losses to the insurer. The insurer should also recognize that a high deductible is not attractive to policyholders since they have to retain a large portion of losses on their own.

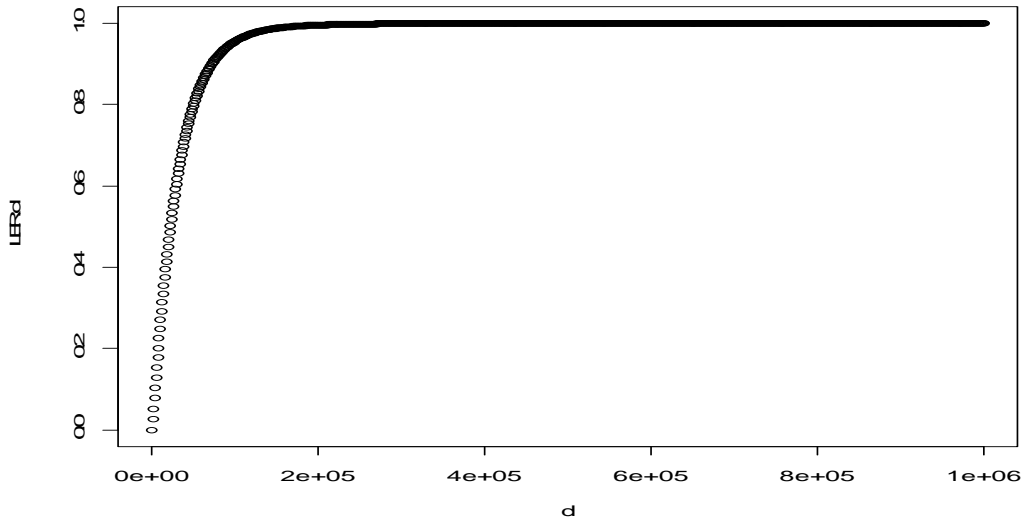


Figure 3: Graph of LER vs. deductible

3.2 Policy Limit

When an insurer introduces a policy limit in its coverage, say at the value of u , the loss covered by the insurer and paid as claim may be represented by the random variable K , where

$$K = \begin{cases} X, & X < u \\ u, & X \geq u \end{cases}, \quad (9)$$

whereas the loss covered by a reinsurer may be represented by the random variable L , where

$$L = \begin{cases} 0, & X < u \\ X - u, & X \geq u \end{cases}, \quad (10)$$

so that $X = K + L$.

Therefore, in terms of an insurer's perspective, the Loss Elimination Ratio (LER) is

$$LER = \frac{E(X) - E(X;u)}{E(X)}, \quad (11)$$

where

$$E(X;u) = \int_0^u xf(x)dx + u \int_u^\infty f(x)dx,$$

and

$$E(X) = \int_0^{\infty} xf(x)dx = \int_0^{\infty} S(x)dx.$$

Table 4 shows the LER for several policy limit values, assuming individual losses follow a Burr distribution with parameters $\theta = 86,426.43$, $\gamma = 1.5169$ and $\alpha = 3.7783$.

Table 4: Values of u and LER

Burr distribution, $E(X) = RM\ 38,131$			
u (RM)	$E(X;u)$ (RM)	LER	Δ LER
40000	27332.77	0.283	-0.010
41000	27686.41	0.274	-0.009
42000	28028.2	0.265	-0.009
43000	28358.49	0.256	-0.009
44000	28677.63	0.248	-0.008
60000	32528.78	0.147	-0.005
61000	32705.46	0.142	-0.005
62000	32876.09	0.138	-0.004
63000	33040.89	0.133	-0.005
64000	33200.05	0.129	-0.004
80000	35123.51	0.079	-0.002
81000	35212.36	0.077	-0.002
82000	35298.28	0.074	-0.003
83000	35381.36	0.072	-0.002
84000	35461.7	0.07	-0.002

As an example, the LER at $u = RM60,000$ is 0.15, implying that 15% of losses can be eliminated by introducing a policy limit of RM60,000. The graph of LER vs. u is shown in Figure 4, indicating that the ratio of eliminated loss is inversely proportional to the limit. However, after a certain point, a higher limit can no longer provide a significant proportion of eliminated loss to an insurer.

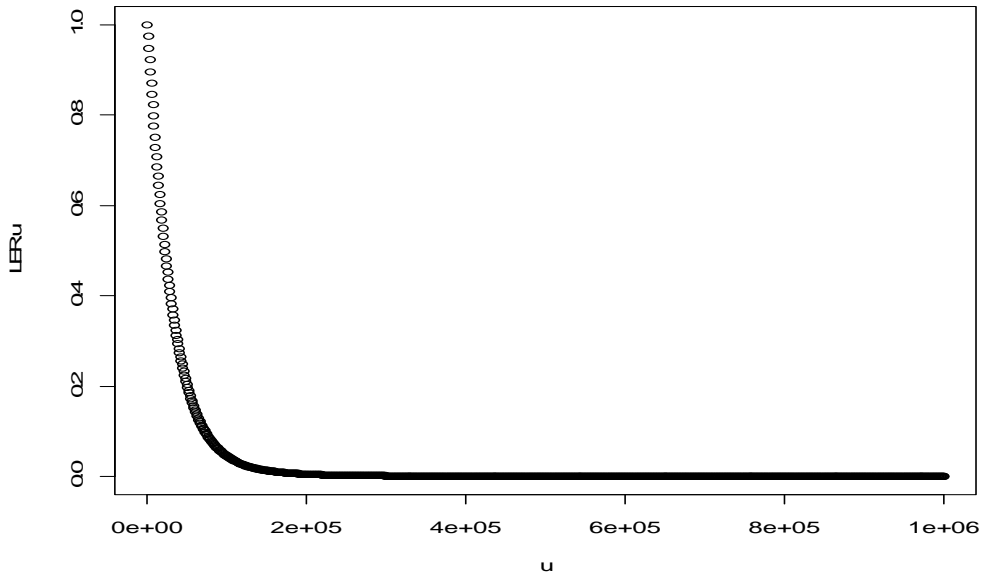


Figure 4: Graph of LER vs. policy limit

In practice, the criteria for policy limit may also differ depending on the requirements and preferences of both insurers and reinsurers. Nevertheless, an insurer may use the values shown in Table 4 and the graph illustrated in Figure 4 to indicate whether the proposed limit provides a significant proportion of eliminated losses.

4. LINEAR LOADING ASSUMPTION

4.1 Insolvency Probability of Deductible Policy

When an insurer introduces a policy with a deductible, at the value of d , the loss covered by insurer and paid as a claim may be represented by the random variable W as shown in equation (7). For an individual risk model, the aggregate claims of a deductible policy, with a deductible of d , may be defined as

$$S = W_1 + W_2 + \dots + W_n, \quad (12)$$

where W_1, W_2, \dots, W_n denote independent and identically distributed (i.i.d.) random variables.

The conditional mean and variance of W_i , respectively, are

$$E(W_i | X > 0) = \mu_w \quad (13)$$

and

$$Var(W_i | X > 0) = \sigma_w^2, \quad (14)$$

where the probability of loss greater than zero or equivalently the probability of incurring a claim is equal to

$$\Pr(X > 0) = q. \quad (15)$$

Therefore, for a deductible policy, $E(W | X > 0)$ and $E(W^2 | X > 0)$ can be written as

$$\mu_w = E(W | X > 0) = \int_d^\infty (x - d) f(x) dx = E(X) - E(X; d) \quad (16)$$

and

$$E(W^2 | X > 0) = \int_d^\infty (x - d)^2 f(x) dx = E(X^2) - E((X; d)^2) - 2dE(X) + 2dE(X; d), \quad (17)$$

so that

$$\sigma_w^2 = E(W^2 | X > 0) - (E(W | X > 0))^2. \quad (18)$$

Finally, the distribution of aggregate claims, S , for a single portfolio of risk in an individual risk model may be estimated by applying Central Limit Theorem (CLT). In particular, if the number of policies, n , is large, the distribution of S may be estimated by a normal distribution with mean,

$$\mu_{s,w} = E(S) = n\mu_w q, \quad (19)$$

and variance,

$$\sigma_{s,w}^2 = Var(S) = n(\sigma_w^2 q + \mu_w^2 q(1 - q)). \quad (20)$$

The same approach can also be applied to multiple portfolios of risks, whereby equation (19) is rewritten as $\mu_{s,w} = E(S) = \sum_i n_i \mu_{w,i} q_i$ where i denotes the i^{th} portfolio. Equivalently, equation (20) can be rewritten as $\sigma_{s,w}^2 = Var(S) = \sum_i n_i (\sigma_{w,i}^2 q_i + \mu_{w,i}^2 q_i (1 - q_i))$.

If the premium is calculated using a linear loading assumption, i.e., premium = $\mu_{s,w} (1 + \xi)$, where ξ denotes the relative loading, a simple definition of the probability of insolvency for a single portfolio of risk may be expressed as the probability of having aggregate claims larger than aggregate premiums, or, equivalently,

$$\Pr(S > (1 + \xi)\mu_{s,w}) = \Pr\left(\frac{S - \mu_{s,w}}{\sigma_{s,w}} > \frac{(1 + \xi)\mu_{s,w} - \mu_{s,w}}{\sigma_{s,w}}\right) = 1 - \Pr\left(Z < \frac{\mu_{s,w}}{\sigma_{s,w}} \xi\right). \quad (21)$$

It should be noted that when $\xi = 0$, the premium is equivalent to the expected aggregate claims of policies with a deductible at d . The linear loading assumption indicates that the relative loading, ξ , is fixed as a constant proportion of $\mu_{s,w}$ regardless of any values of d .

Tables 5-7 show the values of the insolvency probability for several values of ξ , n and q , assuming the amount of loss follows Burr with parameters $\theta = 86,426.43$, $\gamma = 1.5169$, and $\alpha = 3.7783$.

The graphs of insolvency probability vs. deductible for several values of d , ξ , n , and q are shown in Figures 5-7, indicating that under the assumption of linear loading, the insolvency probability increases as the deductible increases. One possible justification for this increase in the insolvency probability can be explained by observing the values of $\mu_{s,w}$ and $\sigma_{s,w}$ displayed in Table 5. Even though both $\mu_{s,w}$ and $\sigma_{s,w}$ decrease when the deductible increases, $\mu_{s,w}$ decreases faster than $\sigma_{s,w}$, causing the quantity $\mu_{s,w}(\sigma_{s,w})^{-1}$ to decrease. Based on equation (21), the probability of insolvency is therefore expected to increase.

In addition, the graphs in Figures 5-7 also show that the insolvency probability:

- decreases as the relative loading, ξ , increases
- decreases as the probability of incurring claim, q , increases
- decreases as the number of policies, n , increases

When the probability of incurring a claim or the number of policies increases, $\mu_{s,w}$ increases faster than $\sigma_{s,w}$, causing the quantity $\mu_{s,w}(\sigma_{s,w})^{-1}$ to increase. Therefore, based on equation (21), the probability of insolvency is expected to decrease.

Appendix 2 shows the calculation of the insolvency probability for a deductible policy using R programming with the assistance of the *actuar* package, assuming the amount of loss follows a Burr distribution.

Table 5: Values of d and insolvency probability ($n = 3000$, $q = 0.2$)

d (RM)	$\mu_{s,w}$ (RM)	$\sigma_{s,w}$ (RM)	$\xi = 0.25$ Insolvency probability	$\xi = 0.20$ Insolvency probability	$\xi = 0.15$ Insolvency probability	$\xi = 0.10$ Insolvency probability	$\xi = 0.05$ Insolvency probability	$\xi = 0.00$ Insolvency probability
5,000	19,937,056	1,071,492	0.000002	0.000099	0.002627	0.031395	0.176097	0.50
10,000	17,201,950	998,232	0.000008	0.000284	0.004871	0.042422	0.194448	0.50
15,000	14,736,570	929,117	0.000037	0.000757	0.008677	0.056360	0.213877	0.50
20,000	12,560,178	864,187	0.000140	0.001826	0.014624	0.073055	0.233703	0.50

Table 6: Values of d and insolvency probability ($n = 3000$, $\xi = 0.15$)

d	$q = 0.40$			$q = 0.30$			$q = 0.20$		
	$\mu_{s,w}$	$\sigma_{s,w}$	Insolvency probability	$\mu_{s,w}$	$\sigma_{s,w}$	Insolvency probability	$\mu_{s,w}$	$\sigma_{s,w}$	Insolvency probability
5,000	39,874,112	1,425,201	0.000014	29,905,584	1,273,880	0.000215	19,937,056	1,071,492	0.002627
10,000	34,403,900	1,340,023	0.000059	25,802,925	1,191,941	0.000583	17,201,950	998,232	0.004871
15,000	29,473,141	1,257,672	0.000220	22,104,856	1,113,820	0.001460	14,736,570	929,117	0.008677
20,000	25,120,356	1,178,332	0.000692	18,840,267	1,039,610	0.003280	12,560,178	864,187	0.014624

Table 7: Values of d and insolvency probability ($\xi = 0.15$, $q = 0.2$)

d	$n = 3000$			$n = 2000$			$n = 1000$		
	$\mu_{s,w}$	$\sigma_{s,w}$	Insolvency probability	$\mu_{s,w}$	$\sigma_{s,w}$	Insolvency probability	$\mu_{s,w}$	$\sigma_{s,w}$	Insolvency probability
5,000	19,937,056	1,071,492	0.002627	13,291,371	874,869	0.011338	6,645,685	618,626	0.053546
10,000	17,201,950	998,232	0.004871	11,467,967	815,053	0.017406	5,733,983	576,329	0.067801
15,000	14,736,570	929,117	0.008677	9,824,380	758,621	0.026035	4,912,190	536,426	0.084785
20,000	12,560,178	864,187	0.014624	8,373,452	705,606	0.037533	4,186,726	498,939	0.104071

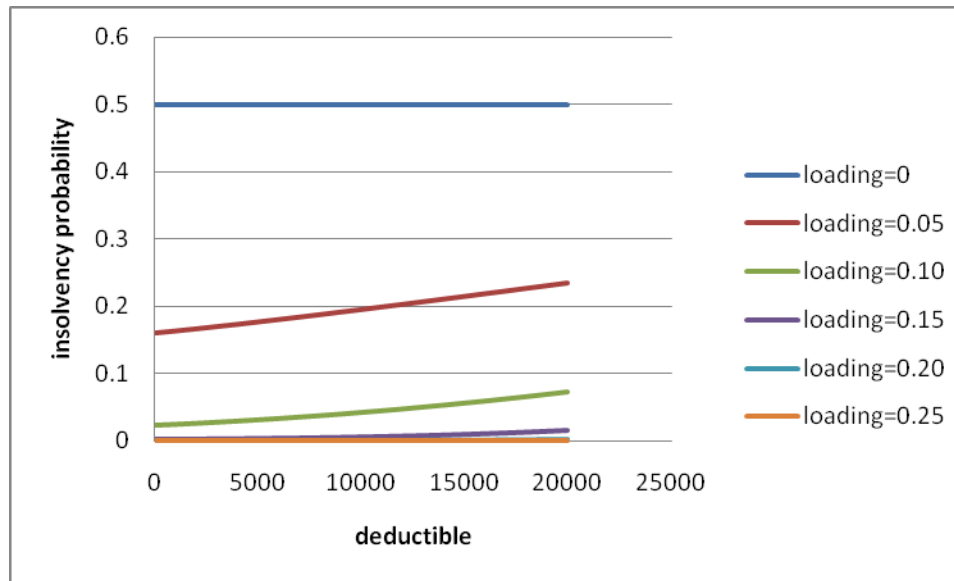


Figure 5: Graph of insolvency probability vs. deductible ($n = 3000$, $q = 0.2$)

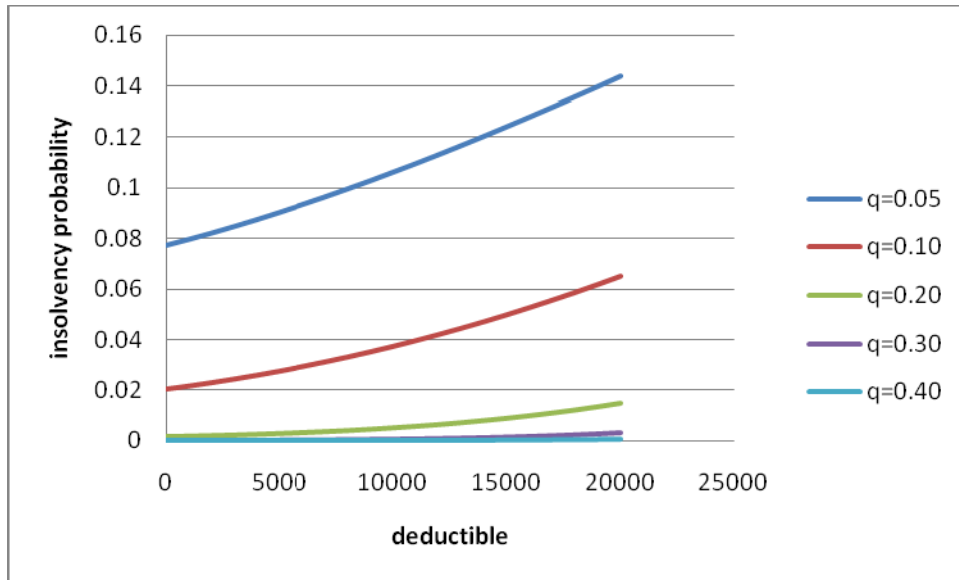


Figure 6: Graph of insolvency probability vs. deductible ($n = 3000$, $\xi = 0.15$)

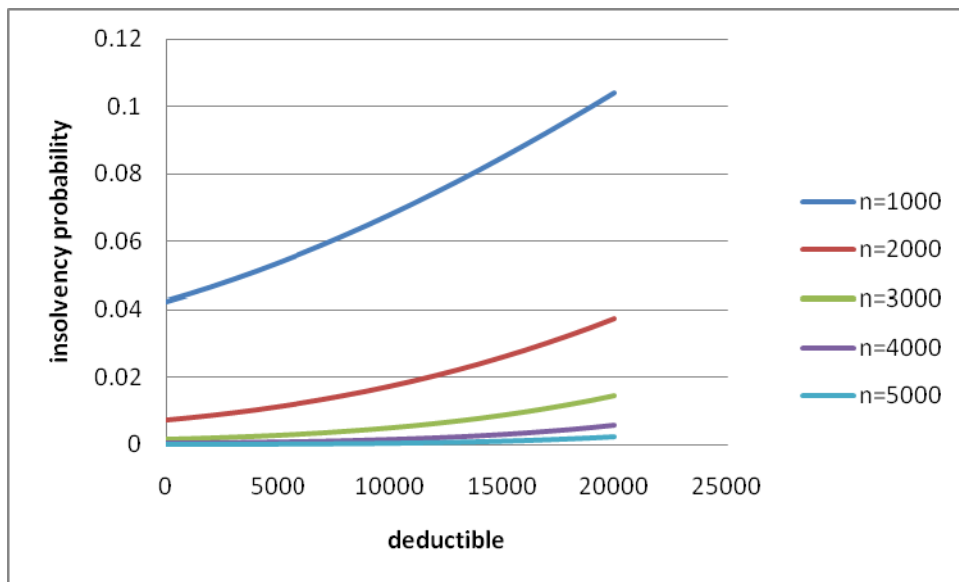


Figure 7: Graph of insolvency probability vs. deductible ($\xi = 0.15$, $q = 0.2$)

4.2 Insolvency Probability of Policy Limit

When an insurer introduces a policy limit, say at the value of u , the loss covered by insurer and paid as a claim may be represented by the random variable K as shown in equation (9). For an individual risk model, the aggregate claims of a policy with limit u may be defined as

$$S = K_1 + K_2 + \dots + K_n, \quad (22)$$

where K_1, K_2, \dots, K_n denote independent and identically distributed (i.i.d.) random variables.

The conditional mean and variance of K_i respectively are

$$E(K_i | X > 0) = \mu_K, \quad (23)$$

and

$$\text{Var}(K_i | X > 0) = \sigma_K^2. \quad (24)$$

Therefore, for a policy limit, $E(K | X > 0)$ and $E(K^2 | X > 0)$ can be written as

$$\mu_K = E(K | X > 0) = \int_0^{\infty} xf(x)dx - \int_u^{\infty} (x-u)f(x)dx = E(X; u) \quad (25)$$

and

$$E(K^2 | X > 0) = \int_0^{\infty} x^2 f(x)dx - \int_u^{\infty} (x-u)^2 f(x)dx = E((X; u)^2), \quad (26)$$

so that

$$\sigma_K^2 = E(K^2 | X > 0) - (E(K | X > 0))^2. \quad (27)$$

The distribution of S , by applying Central Limit Theorem (CLT), may be estimated by normal distribution with mean,

$$\mu_{S,K} = E(S) = n\mu_K q, \quad (28)$$

and variance,

$$\sigma_{S,K}^2 = \text{Var}(S) = n(\sigma_K^2 q + \mu_K^2 q(1-q)). \quad (29)$$

If the premium is calculated using a linear loading assumption, i.e. premium = $\mu_{S,K}(1 + \xi)$, the probability of insolvency for a single portfolio of risk may be equated as the probability of having aggregate claims larger than aggregate premiums, or, equivalently,

$$\Pr(S > \mu_{S,K}(1 + \xi)) = \Pr\left(\frac{S - \mu_{S,K}}{\sigma_{S,K}} > \frac{\mu_{S,K}(1 + \xi) - \mu_{S,K}}{\sigma_{S,K}}\right) = 1 - \Pr\left(Z < \frac{\mu_{S,K}}{\sigma_{S,K}} \xi\right). \quad (30)$$

It should be noted that when $\xi = 0$, the premium is equivalent to the expected aggregate claims of policies with a policy limit at u . The linear loading assumption indicates that the relative loading, ξ , is fixed as a constant proportion of $\mu_{s,K}$ regardless of any values of u .

Tables 8-10 show the values of the insolvency probability for several values of u , ξ , n and q , assuming the amount of loss follows Burr with parameters $\theta = 86,426.43$, $\gamma = 1.5169$ and $\alpha = 3.7783$.

The graphs of insolvency probability vs. policy limit for several values of ξ , n and q are shown in Figures 8-10, indicating that under the assumption of linear loading, the insolvency probability increases as the policy limit increases. Based on values of $\mu_{s,K}$ and $\sigma_{s,K}$ displayed in Table 8, even though both $\mu_{s,K}$ and $\sigma_{s,K}$ increase when the limit increases, $\sigma_{s,K}$ increases faster than $\mu_{s,K}$ causing the quantity $\mu_{s,K}(\sigma_{s,K})^{-1}$ to decrease. Based on equation (30), the probability of insolvency is expected to increase.

In addition, the graphs in Figures 8-10 also show that insolvency probability

- decreases as the relative loading, ξ , increases;
- decreases as the probability of incurring claim, q , increases; and
- decreases as the number of policies, n , increases.

When the probability of incurring a claim or the number of policies increases, $\mu_{s,K}$ increases faster than $\sigma_{s,K}$ causing the quantity $\mu_{s,K}(\sigma_{s,K})^{-1}$ to increase. Therefore, based on equation (30), the probability of insolvency is expected to decrease.

Table 8: Values of u and insolvency probability ($n = 3000$, $q = 0.2$)

u (RM)	$\mu_{S,K}$ (RM)	$\sigma_{S,K}$ (RM)	$\xi = 0.25$	$\xi = 0.24$	$\xi = 0.23$	$\xi = 0.22$	$\xi = 0.21$
			Insolvency probability	Insolvency probability	Insolvency probability	Insolvency probability	Insolvency probability
40,000	16,399,665	674,696	6.13E-10	2.71E-09	1.13E-08	4.46E-08	1.66E-07
60,000	19,517,266	849,996	4.72E-09	1.79E-08	6.42E-08	2.19E-07	7.11E-07
80,000	21,074,104	956,995	1.84E-08	6.28E-08	2.04E-07	6.34E-07	1.88E-06
100,000	21,866,758	1,022,471	4.48E-08	1.43E-07	4.35E-07	1.27E-06	3.54E-06

Table 9: Values of u and insolvency probability ($n = 3000$, $\xi = 0.15$)

u	$q = 0.40$			$q = 0.30$			$q = 0.20$		
	$\mu_{S,K}$	$\sigma_{S,K}$	Insolvency probability	$\mu_{S,K}$	$\sigma_{S,K}$	Insolvency probability	$\mu_{S,K}$	$\sigma_{S,K}$	Insolvency probability
40,000	32,799,329	855,061	0.0000000	24,599,497	784,592	0.0000013	16,399,665	674,696	0.0001332
60,000	39,034,532	1,091,347	0.0000000	29,275,899	994,238	0.0000050	19,517,266	849,996	0.0002863
80,000	42,148,208	1,239,193	0.0000002	31,611,156	1,123,712	0.0000122	21,074,104	956,995	0.0004780
100,000	43,733,516	1,331,212	0.0000004	32,800,137	1,203,592	0.0000218	21,866,758	1,022,471	0.0006685

Table 10: Values of u and insolvency probability ($\xi = 0.15$, $q = 0.2$)

u (RM)	$n = 3000$			$n = 2000$			$n = 1000$		
	$\mu_{s,K}$ (RM)	$\sigma_{s,K}$ (RM)	Insolvency probability	$\mu_{s,K}$	$\sigma_{s,K}$	Insolvency probability	$\mu_{s,K}$	$\sigma_{s,K}$	Insolvency probability
40,000	16,399,665	674,696	0.000133	10,933,110	550,887	0.001456	5,466,555	389,536	0.017644
60,000	19,517,266	849,996	0.000286	13,011,511	694,019	0.002460	6,505,755	490,746	0.023376
80,000	21,074,104	956,995	0.000478	14,049,403	781,383	0.003498	7,024,701	552,521	0.028255
100,000	21,866,758	1,022,471	0.000668	14,577,839	834,844	0.004406	7,288,919	590,324	0.032006

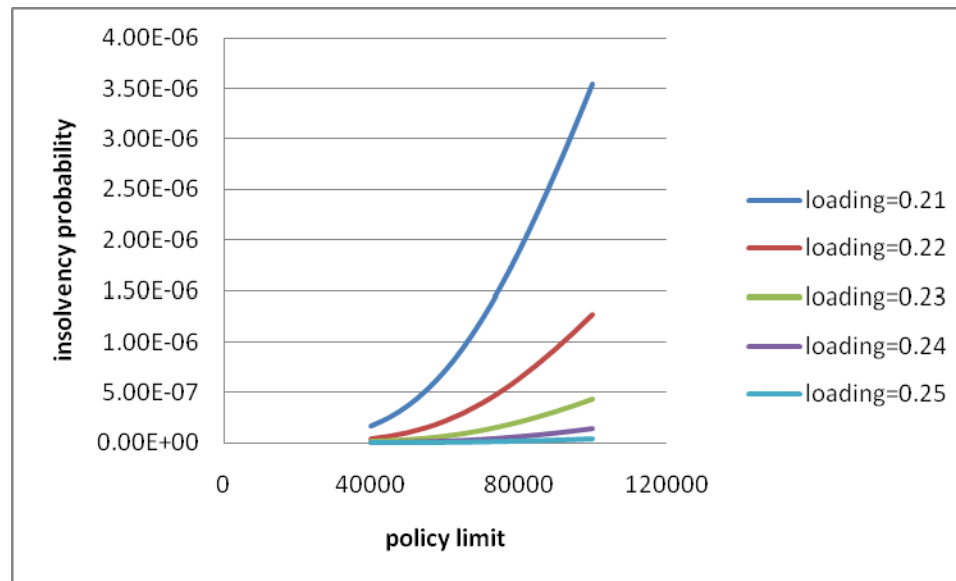


Figure 8: Graph of insolvency probability vs. policy limit ($n = 3000$, $q = 0.2$)

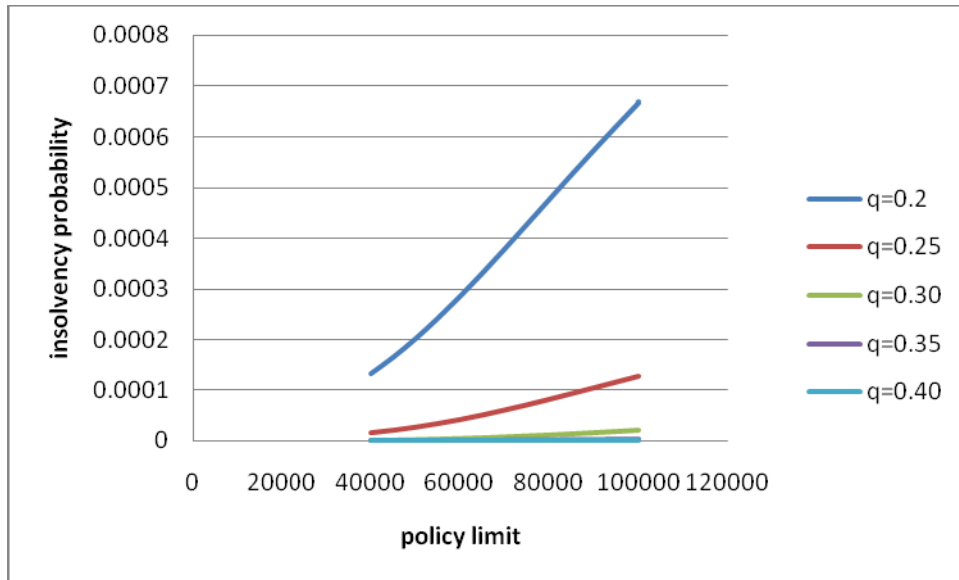


Figure 9: Graph of insolvency probability vs. policy limit ($n = 3000$, $\xi = 0.15$)

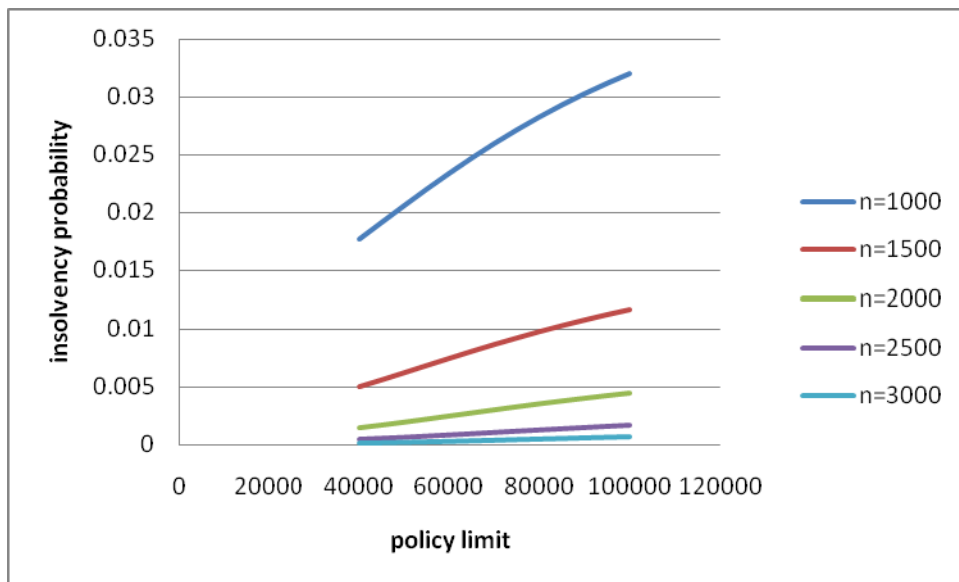


Figure 10: Graph of insolvency probability vs. policy limit ($\xi = 0.15$, $q = 0.2$)

5. PH-TRANSFORM ASSUMPTION

The determination of expected loss or mean severity based on the Proportional Hazard Transform (PH-Transform) assumption introduced by Wang [12] may be used as an alternative to reduce the probability of insolvency at a higher deductible or policy limit. In particular, the PH-Transform assumption incorporates an “appropriate” risk load in the severity distribution at a higher deductible or policy limit, and thus allows the probability of insolvency to be lower.

The mean severity under the PH-Transform assumption can be calculated as (Wang [12]-[13])

$$H(X) = \int_0^{\infty} (S(x))^r dx, \quad 0 < r \leq 1, \quad (31)$$

where r denotes the index of ambiguity degree. The PH-mean shown in equation (31) represents a risk-adjusted premium and is quite sensitive to the choice of r . Index r can be assigned to the level of confidence in the estimation of loss, where a lower value of r implies a more ambiguous situation. For example, a non-ambiguous scenario for the best estimate could occur when there is little ambiguity regarding the best estimate of the severity distribution, such as when all experts agree with confidence in the estimate, whereas an ambiguous scenario could occur when there is considerable ambiguity regarding the best estimate of the severity distribution, such as when experts disagree and have little confidence in such estimate. From a broader perspective, examples of conditions contributing to greater ambiguity include uncertainty of the underlying loss distribution, incomplete information, insufficient data, changes in claim generating mechanisms, extra expenses associated with risk-sharing transactions, and difference in local market climates due to differences in geographic areas and/or lines of insurance (Wang [11]).

The PH-Transform can also be applied using subjective guidelines for the error of estimation; an actuary may construct his own table for index r to reflect different levels of ambiguity. One such example is given by Wang [11]:

Table 11: Ambiguity level and index r

Ambiguity level	Index r
Slightly ambiguous	0.96 – 1.00
Moderately ambiguous	0.90 – 0.95
Highly ambiguous	0.80 – 0.89
Extremely ambiguous	0.50 - 0.79

Source: Wang [11]

In addition to the severity distribution, the PH-Transform assumption can be applied on the frequency distribution where appropriate. As an example, in pricing a reinsurance contract, the PH-Transform can be applied separately on the severity and frequency distributions. The choice of r depends on the level of confidence in the estimate of claim severity and frequency. If the actuary has higher confidence in the estimate of claim frequency distribution but lower confidence in the estimate of claim severity distribution, he should chose a higher r for claim frequency, say 0.95, and a lower r for claim severity, say 0.85. For example, higher confidence for the frequency distribution and lower confidence for the severity distribution should be applied on types of insurance risks that provide considerable past data on the probability of occurrence but much uncertainty on the size of loss due to arbitrary court awards.

5.1 Insolvency Probability of Deductible Policy

The same approach may be used to find the expected loss of a deductible policy,

$$H(W) = \int_d^{\infty} (S(x))^r dx, \quad 0 < r \leq 1. \quad (32)$$

where W is defined as equation (7).

For example, assume that the amount of loss follows a Burr distribution with parameters (α, θ, γ) . The survival function is equal to

$$S(x) = \left(\frac{\theta^\gamma}{\theta^\gamma + x^\gamma} \right)^\alpha, \quad (33)$$

and if the PH-Transform assumption is applied, the survival function also follows a Burr distribution, but with parameters $(r\alpha, \theta, \gamma)$,

$$S(x) = \left(\frac{\theta^\gamma}{\theta^\gamma + x^\gamma} \right)^{r\alpha} \tag{34}$$

Therefore, the equation of expected loss shown by equation (32) can also be rewritten as $H(W) = E(X) - E(X;d)$, this time assuming that the loss distribution follows a Burr distribution with parameters $(r\alpha, \theta, \gamma)$. In addition, $H(W)$ can be rewritten as a function of $E(W)$,

$$H(W) = (1 + \psi)E(W) \tag{35}$$

where $E(W) = \int_d^\infty S(x)dx$, and ψ denotes the equivalent relative loading of a policy with deductible valued at d .

Table 12 shows the expected loss, $H(W)$, and the equivalent relative loading, ψ , under the PH-Transform assumption for several values of r . For example, the expected loss with no loading, i.e. the expected loss at $r = 1$, for a deductible valued at RM5,000 is equivalent to RM33,228. If the PH-Transform assumption with $r = 0.9$ is applied, the expected loss is RM36,804 and the equivalent relative loading, ψ , is equal to 0.11.

Table 12: Expected loss and relative loading (deductible policy)

d (RM)	Expected loss $r = 1$ (RM)	Expected loss $r = 0.9$ (RM)	Relative loading	Expected loss $r = 0.7$ (RM)	Relative loading
5,000	33,228	36,804	0.11	47,426	0.43
10,000	28,670	32,203	0.12	42,740	0.49
15,000	24,561	28,013	0.14	38,382	0.56
20,000	20,934	24,267	0.16	34,389	0.64

Figure 11 shows the graph of expected loss vs. deductible for several values of r under the assumption of PH-Transform. It can be seen that the expected loss calculated under the PH-Transform ($r = 0.9$ and $r = 0.7$) is higher than the basic expected loss ($r = 1$), implying that the expected loss is higher when the estimation of loss amount becomes more ambiguous.

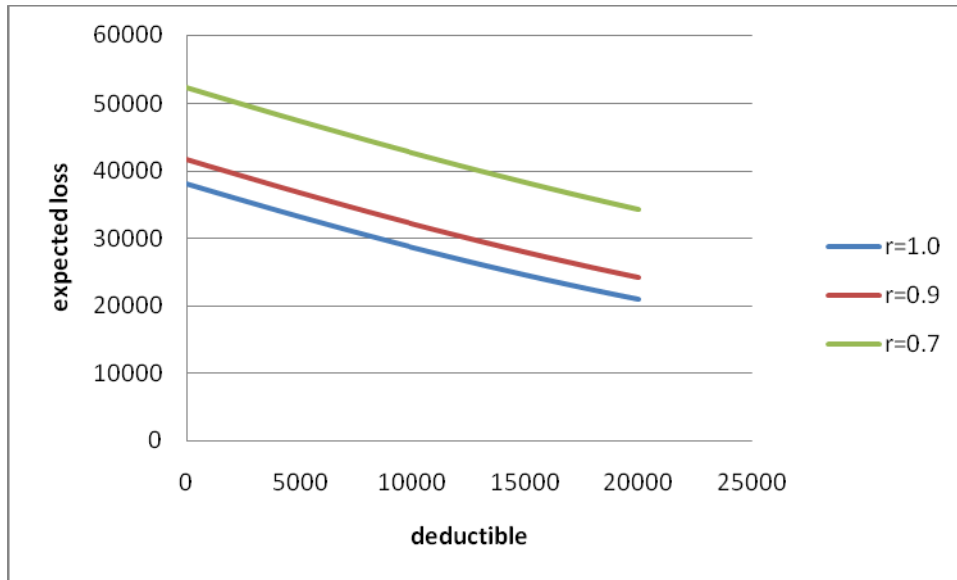


Figure 11: Graph of expected loss vs. deductible

If the probability of insolvency is calculated using equation (21), the linear loading assumption and PH-Transform assumption can be compared by using ξ as the relative loading for linear assumption and ψ as the relative loading for PH-Transform assumption. The main difference between the assumptions is that the relative loading for PH-Transform increases when d increases, whereas for linear loading, the relative loading remains fixed when d increases. Table 13 shows the values for insolvency probability for several values of ξ and ψ assuming $n = 3000$ and $q = 0.2$. Figure 12 shows the graph of insolvency probability vs. deductible under several linear loading and PH-Transform assumptions, also assuming $n = 3000$ and $q = 0.2$. It can be seen that the insolvency probability is lower for higher deductibles under the PH-Transform assumption. Thus, the PH-Transform can be used as an alternative to reduce the probability of insolvency at higher deductible values by incorporating an “appropriate” risk load in the severity distribution.

Appendix 3 uses R programming with the assistance of the *actuar* package to calculate the expected loss for a deductible policy under the PH-Transform assumption, assuming the amount of loss follows a Burr distribution.

Table 13: Insolvency probability for linear loading and PH-Transform

d (RM)	Linear loading				PH-Transform $r = 0.9$	
	ξ	Insolvency probability	ξ	Insolvency probability	ψ	Insolvency probability
5,000	0.15	0.003	0.10	0.031	0.11	0.023
10,000	0.15	0.005	0.10	0.042	0.12	0.017
15,000	0.15	0.009	0.10	0.056	0.14	0.013
20,000	0.15	0.015	0.10	0.073	0.16	0.010

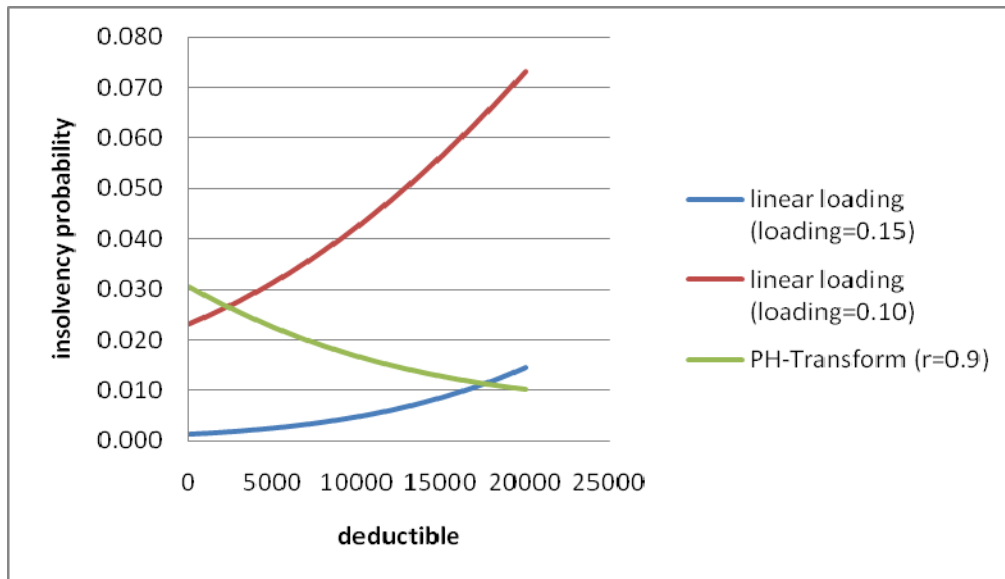


Figure 12: Graph of insolvency probability vs. deductible

5.2 Insolvency Probability of Policy Limit

Similar to a deductible policy, the expected loss of a policy limit under PH-Transform assumption can be calculated as

$$H(K) = \int_0^u (S(x))^r dx, \quad 0 < r \leq 1, \quad (36)$$

where K is defined as equation (9).

If the amount of loss follows a Burr distribution with parameters (α, θ, γ) , the equation of expected loss shown by equation (36) can be rewritten as $H(K) = E(X; u)$, this time assuming that the loss distribution follows a Burr distribution with parameters $(r\alpha, \theta, \gamma)$. In addition, $H(K)$ can be rewritten as a function of $E(K)$,

$$H(K) = (1 + \eta)E(K), \tag{37}$$

where $E(K) = \int_0^u S(x)dx$, and η denotes the equivalent relative loading of a policy with limit valued at u .

Table 14 provides the expected loss, $H(K)$, and the equivalent relative loading, η , under the PH-Transform assumption for several values of r assuming $n = 3000$ and $q = 0.2$. Figure 13 shows the graph of expected loss vs. policy limit for several values of r under the PH-Transform assumption, also assuming $n = 3000$ and $q = 0.2$. It can be seen that the expected loss calculated under the PH-Transform assumption ($r = 0.8$ and $r = 0.7$) is higher than the basic expected loss ($r = 1$), also implying that the expected loss is higher when the estimation of loss amount becomes more ambiguous.

Table 14: Expected loss and relative loading (policy limit)

u (RM)	Expected loss $r = 1$ (RM)	Expected loss $r = 0.8$ (RM)	Relative loading	Expected loss $r = 0.7$ (RM)	Relative loading
40,000	27,333	29,286	0.07	30,353	0.11
60,000	32,529	36,068	0.11	38,106	0.17
80,000	35,124	39,960	0.14	42,875	0.22
100,000	36,445	42,228	0.16	45,849	0.26

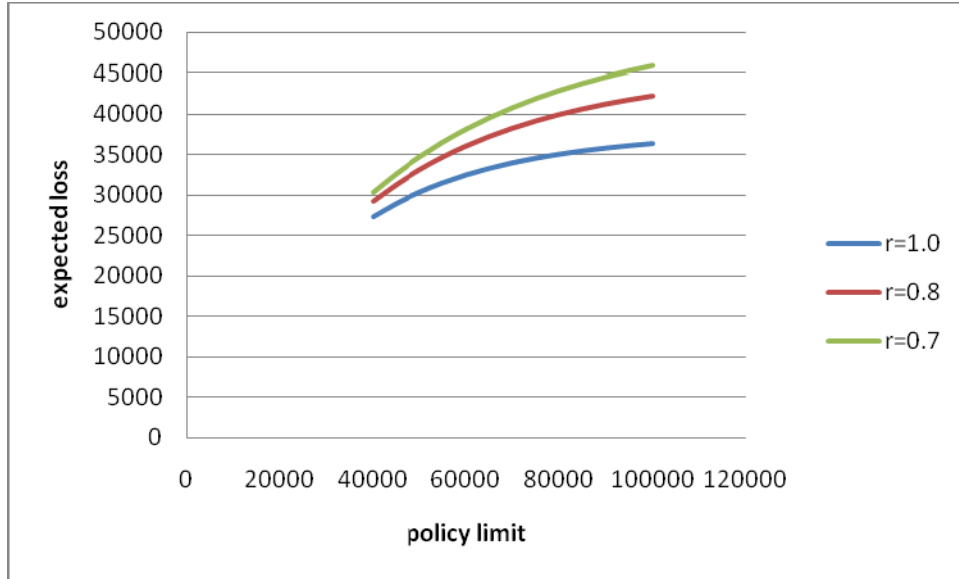


Figure 13: Graph of expected loss vs. policy limit

If the probability of insolvency is calculated using equation (30), the linear loading assumption and PH-Transform assumption can also be compared by using ξ as the relative loading for the linear assumption and η as the relative loading for PH-Transform assumption. The main difference between the assumptions is that the relative loading for PH-Transform increases when u increases, whereas for linear loading, the relative loading remains fixed when u increases. Table 15 shows the values for insolvency probability for several values of ξ and η .

Table 15: Insolvency probability for linear loading and PH-Transform

u (RM)	Linear loading				PH-Transform $r = 0.8$	
	ξ	Insolvency probability	ξ	Insolvency probability	η	Insolvency probability
40,000	0.15	0.0001	0.10	0.0075	0.07	0.0412
60,000	0.15	0.0003	0.10	0.0108	0.11	0.0062
80,000	0.15	0.0005	0.10	0.0138	0.14	0.0012
100,000	0.15	0.0007	0.10	0.0162	0.16	0.0003

Figure 14 shows the graph of insolvency probability vs. policy limit under several linear loading and PH-Transform assumptions. It can be seen that the insolvency probability is lower for higher limits under the PH-Transform assumption. Thus, the PH-Transform can be used as an alternative to reduce the probability of insolvency at higher limit values by incorporating an “appropriate” risk load in the severity distribution.

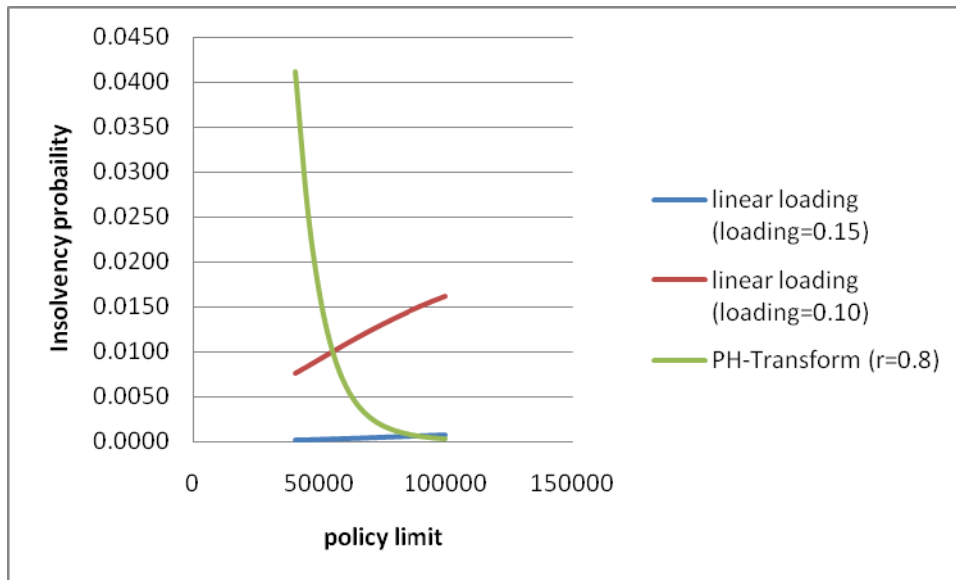


Figure 14: Graph of insolvency probability vs. policy limit

6. EXCESS LAYERS OF A SINGLE RISK

6.1 Pricing of Excess Layers

In an insurance contract containing both a deductible d and a policy limit u , the loss of a layer $(d, d + u]$ of a risk X can be defined by the random variable M , where

$$M = \begin{cases} 0, & X \leq d \\ X - d, & d < X < d + u \\ u, & X \geq d + u \end{cases} \quad (38)$$

Therefore, the average loss or mean severity of a layer $(d, d + u]$ may be written as

$$E(M) = \int_d^{d+u} S(x) dx, \quad (39)$$

whereas under the PH-Transform assumption, the average loss of the same layer is

$$H(M) = \int_d^{d+u} (S(x))^r dx. \quad (40)$$

If the amount of loss follows a Burr distribution with parameters (α, θ, γ) , the equation of expected loss or mean severity shown by equation (39) can also be rewritten as

$$E(M) = E(X; d + u) - E(X; d), \quad (41)$$

whereas under the PH-Transform assumption, equation (40) can also be rewritten as

$$H(M) = E(X; d + u) - E(X; d), \quad (42)$$

this time assuming the amount of loss follows a Burr distribution with parameters $(r\alpha, \theta, \gamma)$.

For a single risk, the expected aggregate claims shown by equations (19) and (28) can be simplified into

$$E(S) = E(M)q, \quad (43)$$

i.e., assuming $n = 1$.

Under the PH-Transform assumption, the expected aggregate claim amount can also be calculated, and it is equal to

$$E(S) = H(M)q. \quad (44)$$

$H(M)q$ can also be rewritten as a function of $E(M)q$,

$$H(M)q = (1 + \xi)E(M)q, \quad (45)$$

where ξ denotes the equivalent relative loading of a policy with deductible d and limit u .

Table 16 shows the expected aggregate claims and equivalent relative loading, ξ , for several values of d and u under the PH-Transform assumption, where $n = 1$, $q = 0.1$ and the individual loss amount follows a Burr distribution with parameters $\theta = 86,426.43$, $\gamma = 1.5169$, and $\alpha = 3.7783$. For example, the expected aggregate claim amount or the premium with no loading, i.e., $r = 1$, for layer $(0, 5000]$, is equivalent to RM490.24. If the PH-Transform assumption with $r = 0.92$ is applied, the premium is RM491.01 and the equivalent relative loading is $\xi = 0.002$. It can be observed from the table that the relative loading, ξ , under the PH-Transform assumption increases as the layer, $(d, d + u]$, increases.

Figure 15 shows the graph of expected aggregate claim amount vs. layer for several values of the ambiguity index, r , assuming $q = 0.1$ for the same loss distribution assumption. The graph shows that the expected aggregate claim amount decreases when the value of the layer, $(d, d + u]$, increases. Equations (39) and (40) imply that the expected aggregate claim amount depends on the integrals of $S(x)$ and $S(x)^r$. Since $S(x)$ is a decreasing function, the areas under the curves of $S(x)$ and $S(x)^r$ are smaller as the value of $(d, d + u]$ is higher, which causes the expected aggregate claim amount to decrease. In addition, the graph also shows that the expected aggregate claim amount increases when the ambiguity index, r , decreases, indicating that the relative loading, ξ , is higher when the estimation of loss is more ambiguous.

Table 16: Expected aggregate claim amount and relative loading (single risk, PH Transform)

d (RM)	$d + u$ (RM)	Aggregate claims (RM) ($r = 1$)	Aggregate claims (RM) ($r = 0.92$)	Relative loading	Aggregate claims (RM) ($r = 0.90$)	Relative loading
0	5,000	490.24	491.01	0.002	491.20	0.002
5,000	10,000	455.85	459.22	0.007	460.07	0.009
10,000	15,000	410.90	417.38	0.016	419.02	0.020
20,000	25,000	315.38	327.20	0.037	330.23	0.047
40,000	45,000	165.32	180.61	0.092	184.65	0.117
80,000	85,000	41.59	50.74	0.220	53.33	0.282
100,000	105,000	21.68	27.87	0.285	29.67	0.368
160,000	165,000	3.93	5.80	0.473	6.39	0.623

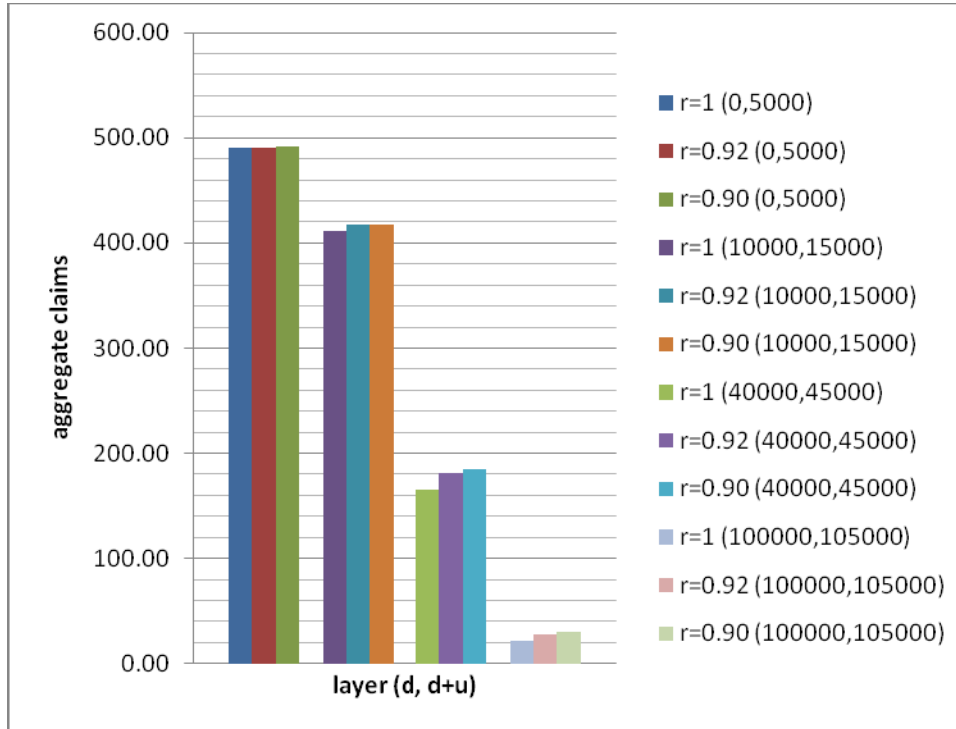


Figure 15: Graph of expected aggregate claim amount vs. layer (single risk, $n = 1$, $q = 0.1$)

Appendix 4 shows the calculation of the expected aggregate claim amount for a single risk and a single layer using R programming with the assistance of the *actuar* package, assuming that the severity follows a Burr distribution.

6.2 Increased Limit Factor (ILF)

In liability insurance, a policy generally provides coverage up to a specified maximum amount that will be paid on any individual loss. In the U.S., it is general practice to publish rates for some standard limit, the “basic limit” (for example, USD\$100,000), to which rates the increased limit factors (ILF) are applied to calculate increased limit rates (Wang [11]). In Malaysia, however, the practice has not been implemented; therefore, the ILF calculated in this study may be used as some indication or basis for possible basic and increased rates.

If the basic limit is valued at RM100,000, the ILF can be calculated as the expected loss at the increased limit divided by the expected loss at the basic limit,

$$ILF(a) = \frac{E(X;a)}{E(X;100000)}. \tag{46}$$

If a risk load is to be included, equation (46) can be rewritten as

$$ILF(a) = \frac{E(X;a) + RL(a)}{E(X;100000) + RL(100000)}, \tag{47}$$

where $RL(a)$ and $RL(100000)$ denote the risk load.

Under the PH-Transform assumption, equation (47) can be rewritten as

$$ILF(a) = \frac{H(X;a)}{H(X;100000)}, \tag{48}$$

where $H(X;a)$ and $H(X;100000)$ denote the mean severity calculated under the PH-Transform assumption. Since $H(X;a) > E(X;a)$ and $H(X;100000) > E(X;100000)$, the equivalent risk load for the PH-Transform assumption can be calculated. Table 17 shows the ILFs under the PH-Transform assumption assuming that the loss distribution follows a Burr distribution with parameters $\theta = 86,426.43$, $\gamma = 1.5169$ and $\alpha = 3.7783$. However, the ILFs calculated appear to be extremely flat, indicating that larger claims may be under-represented by fitting a Burr distribution. Additional treatment is needed in this situation, such as considering a mixed distribution which may produce a more appropriate result for fitting large claims.

Figure 16 shows the graph of ILF vs. a under the PH-Transform assumption for the same severity distribution. The graph shows that the ILFs increase when a increases but remain at a fixed value for large values of a . In addition, the graph shows that the ILFs increase when the ambiguity index, r , decreases, implying that the risk load is higher when loss estimation is more ambiguous.

Table 17: ILF

a (RM)	$E(X;a)$ (RM)	ILF without RL	Risk Load (RM) ($r = 0.9$)	ILF ($r = 0.9$)	Risk Load (RM) ($r = 0.85$)	ILF ($r = 0.85$)
100,000	36,444.60	1.000000	2,678.91	1.000000	4,172.73	1.000000
200,000	37,960.89	1.041605	3,412.12	1.057497	5,401.38	1.067581
300,000	38,097.00	1.045340	3,535.89	1.064140	5,624.74	1.076431
400,000	38,120.88	1.045995	3,566.56	1.065534	5,683.37	1.078462
500,000	38,127.10	1.046166	3,576.64	1.065951	5,703.53	1.079112

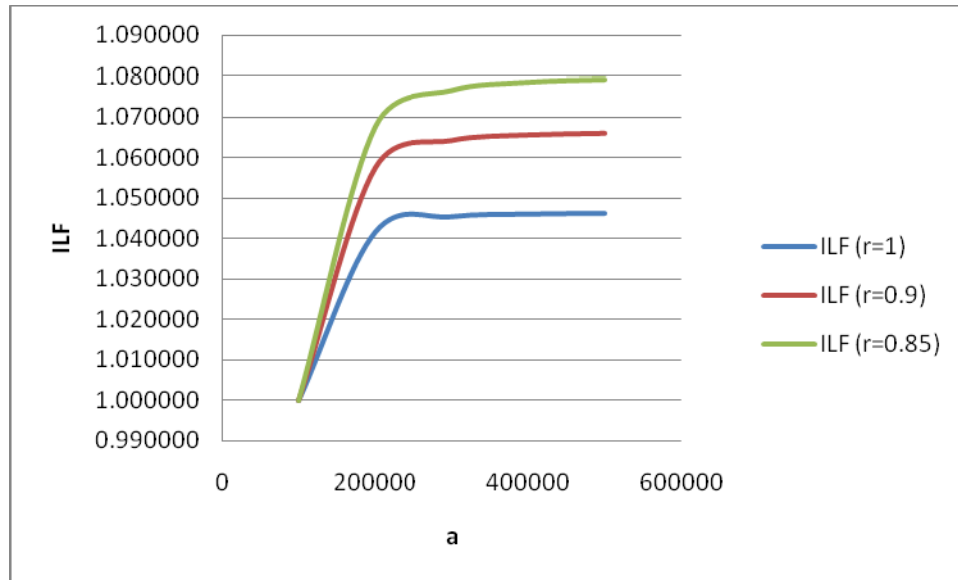


Figure 16: Graph of ILF vs. a

Appendix 5 shows the calculation of ILFs using R programming with the assistance of *actuar* package, assuming that the amount of loss follows a Burr distribution.

7. EXCESS-OF-LOSS FOR REINSURANCE TREATY

In a developing country such as Malaysia, we seldom have a single local insurer covering a single large risk, especially in non-life insurance businesses. In practice, a large risk is usually divided into several excess-of-loss layers shared and insured by several local or multinational insurers or reinsurers. The pricing of layers, therefore, is crucial, especially in the process of dividing risk and pricing risk fairly for each insurer. In this paper, we would like to introduce an approach which may be considered as fair and efficient for pricing excess-of-loss layers of a reinsurance treaty. The fairness in pricing may be achieved by implementing a PH-Transform assumption whereby the insolvency probability is lowered. In addition, the efficiency in pricing may be obtained by using R programming with the *actuar* package to allow the pricing by layer to be computed with less effort.

Let N denote the random variable for claim frequency. Hence, the expected frequency can be calculated as

$$E(N) = \sum_{k=0}^{\infty} S(k), \quad k = 0, 1, \dots, \quad (49)$$

whereas under a PH-Transform assumption, the expected frequency is equivalent to (Wang [11],

$$H(N) = \sum_{k=0}^{\infty} (S(k))^r. \tag{50}$$

Let X denote the random variable for loss severity. The expected severity is $E(X) = \int_0^{\infty} S(x)dx$, whereas under the PH-Transform assumption, the expected severity is equal to $H(X) = \int_0^{\infty} (S(x))^r dx$.

By implementing both frequency and severity approaches, the expected aggregate claims can be calculated as

$$E(S) = E(N)E(X), \tag{51}$$

whereas under the assumption of PH-Transform, the expected aggregate claims is equal to

$$H(N)H(X). \tag{52}$$

The same approach may also be implemented for calculating the price of several excess-of-loss layers. The mean severity for layer $(d, d + u]$ is the same as equation (41) whereas under a PH-Transform assumption, the mean severity for the same layer is the same as equation (42). Therefore, the expected aggregate claims is

$$E(S) = E(N)E(M), \tag{53}$$

whereas under a PH-Transform assumption, the expected aggregate claims is

$$H(N)H(M). \tag{54}$$

If the amount of loss follows a Burr distribution with parameters (α, θ, γ) , the calculation of $H(M)$ in equation (54) also follows a Burr distribution, this time with parameters $(r\alpha, \theta, \gamma)$.

If the claim frequency follows a Poisson distribution with parameter λ , the aggregate claims, S , follow a compound Poisson distribution whereby the variance of aggregate claims can be written as

$$Var(S) = \lambda E(M^2), \tag{55}$$

where $E(M^2) = E((X; d + u)^2) - E((X; d)^2) - 2dE(X; d + u) + 2dE(X; d)$.

Table 18 shows the mean severity, mean frequency, burning cost, loaded rate, and relative loading under a PH-Transform assumption for several excess-of-loss layers, assuming N is Poisson with parameter $\lambda = 100$, X is Burr with parameters $\theta = 86,426.43$, $\gamma = 1.5169$ and $\alpha = 3.7783$, and $r = 0.95$ for both frequency and severity distributions.

Table 18: Mean severity, mean frequency, burning cost, loaded rate and relative loading

Layer (RM)	$E(M)$ (RM)	$H(M)$ ($r = 0.95$) (RM)	$E(N)$	$H(N)$ ($r = 0.95$)	Burning Cost	Loaded Rate	Relative Loading
(100k,300k]	1,652.40	2,033.77	100	100.47	0.016524	0.020434	0.24
(300k,500k]	30.10	46.15	100	100.47	0.000301	0.000464	0.54
(500k,700k]	2.91	5.04	100	100.47	0.000029	0.000051	0.74
(700k,900k]	0.56	1.06	100	100.47	0.000006	0.000011	0.90
(100k,900k]	1,685.97	2,086.01	100	100.47	0.016860	0.020959	0.24

The burning cost is calculated as (Wang [11])

$$\frac{E(M)E(N)}{SEP}, \tag{56}$$

where SEP denotes the subject earned premium. In this study, the SEP is assumed to be RM10,000,000.

The loaded rate is calculated as (Wang [11])

$$\frac{H(M)H(N)}{SEP}, \tag{57}$$

whereby it can also be written as a function of the burning cost,

$$\frac{H(M)H(N)}{SEP} = (1 + \xi) \frac{E(M)E(N)}{SEP}, \tag{58}$$

where ξ denotes the equivalent relative loading. Based on Table 18, the relative loading, ξ , under a PH-Transform assumption increase as the excess-of-loss layer, $(d, d + u]$, increase. In addition, the values of $E(M)$ and $H(M)$ decrease when the layer, $(d, d + u]$, increases.

The distribution of aggregate claims, S , by applying Central Limit Theorem, may be estimated by the Normal distribution with mean $E(S) = \lambda E(M)$ and variance $Var(S) = \lambda E(M^2)$. The probability of insolvency, i.e. the probability of having aggregate claims larger than aggregate premiums, for a PH-Transform assumption can be calculated as

$$\Pr(S > H(N)H(M)) = \Pr(S > (1 + \xi)E(S)) = \Pr\left(Z > \frac{E(S)}{\sqrt{\text{Var}(S)}} \xi\right). \quad (59)$$

In terms of insolvency probability, the main difference between a linear loading assumption and a PH-Transform assumption is that the relative loading for a PH-Transform increases when the layer $(d, d + u]$ increases, whereas the relative loading remains fixed at ξ for all layers under the linear loading.

Table 19 provides the value of mean severity, mean frequency, mean aggregate claims, and variance aggregate claims. It should be noted that both $E(S)$ and $\text{Var}(S)$ decrease when excess-of-loss layer, $(d, d + u]$, increases.

Table 20 shows the values of premium and relative loading for several excess-of-loss layers under the PH-Transform assumptions ($r = 0.95$, $r = 0.90$ and $r = 0.85$). It should be noted that the lower the ambiguity index, r , the higher the premium layer, implying that the relative loading is higher when ambiguity increases. In addition, the premium is lower when the layer, $(d, d + u]$, increases. The relative loading is also higher when the layer, $(d, d + u]$, increases.

Table 21 shows the values of insolvency probability under a linear loading assumption for several values of relative loading ($\xi = 0.10$, $\xi = 0.15$ and $\xi = 0.20$), and a PH-Transform assumption for several values of ambiguity index ($r = 0.95$, $r = 0.90$ and $r = 0.85$). The table shows that the insolvency probability for the PH-Transform is lower than the linear loading for all layers, but the difference is lower when the layer of $(d, d + u]$ increases. Therefore, a PH-Transform assumption may be used as an alternative to reduce insolvency probability of excess-of-loss layers in reinsurance treaties by incorporating “appropriate” risk loads in the frequency and severity distributions of all layers.

Table 19: Mean severity, mean frequency, mean aggregate claims and variance aggregate claims

Layer	$E(M)$ (RM)	$E(N)$	$E(S) = \lambda E(M)$ (RM)	$Var(S) = \lambda E(M^2)$ (RM)
(100k,300k]	1,652.40	100	165,240	12,596,760,695
(300k,500k]	30.10	100	3,010	356,232,253
(500k,700k]	2.91	100	291	41,096,487
(700k,900k]	0.56	100	56	8,650,994
(100k,900k]	1,685.97	100	168597	14,506,333,740

Table 20: Premium and relative loading (PH-Transform)

Layer	$H(M)H(N)$ (RM) ($r = 0.95$)	Relative loading	$H(M)H(N)$ (RM) ($r = 0.9$)	Relative loading	$H(M)H(N)$ (RM) ($r = 0.85$)	Relative loading
(100k,300k]	204,337	0.24	253,397	0.53	315,181	0.91
(300k,500k]	4,637	0.54	7,154	1.38	11,055	2.67
(500k,700k]	507	0.74	884	2.04	1,543	4.31
(700k,900k]	106	0.90	201	2.60	383	5.84
(100k,900k]	209,587	0.24	261,635	0.55	328,162	0.95

Table 21: Insolvency probability

Layer (RM)	Linear loading			PH-Transform		
	$\Pr(S > E(S)(1 + \xi))$ $\xi = 0.1$	$\Pr(S > E(S)(1 + \xi))$ $\xi = 0.15$	$\Pr(S > E(S)(1 + \xi))$ $\xi = 0.2$	$\Pr(S > H(X)H(N))$ ($r = 0.95$)	$\Pr(S > H(X)H(N))$ ($r = 0.9$)	$\Pr(S > H(X)H(N))$ ($r = 0.85$)
(100k, 300k]	0.4415	0.4126	0.3842	0.3638	0.2161	0.0908
(300k, 500k]	0.4936	0.4905	0.4873	0.4657	0.4131	0.3350
(500k, 700k]	0.4982	0.4973	0.4964	0.4866	0.4632	0.4226
(700k, 900k]	0.4992	0.4989	0.4985	0.4932	0.4803	0.4558
(100k, 900k]	0.4443	0.4168	0.3898	0.3668	0.2199	0.0926

Figures 17-20 show the graphs of insolvency probability for several values of ξ (under a linear loading assumption) and r (under a PH-Transform assumption) for each layer of $(d, d + u]$. Figure 21 shows the graph of insolvency probability for all layers. The equivalent loading, ξ , for each r is also shown in the figures. As an example, when $r = 0.95$ under the PH-Transform, the equivalent ξ for layer $(100k, 300k]$ is $\xi = 0.24$, as shown in Figure 17.

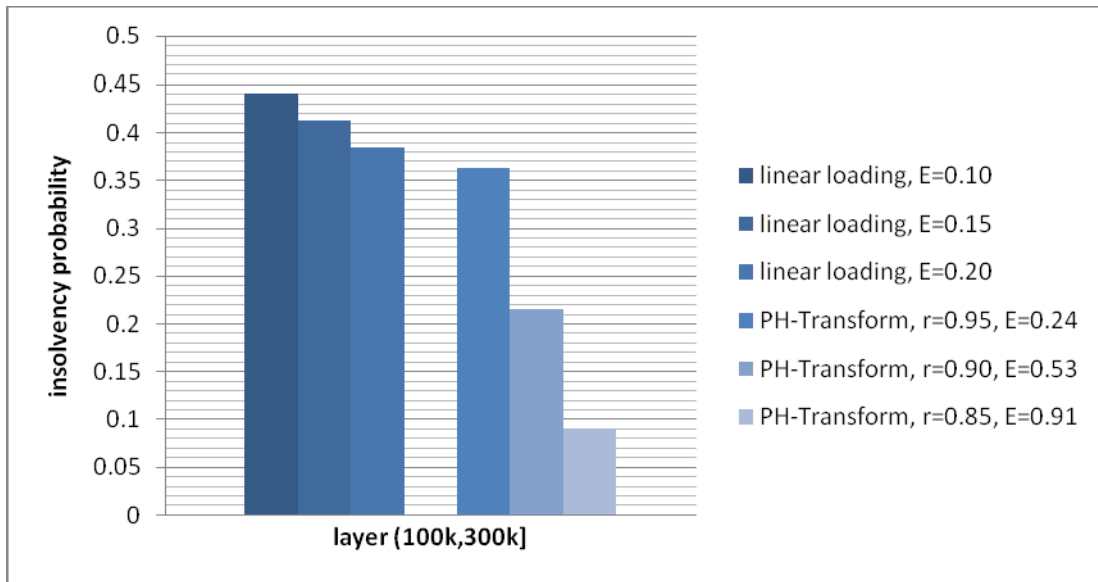


Figure 17: Graph of insolvency probability (layer (100k,300k])

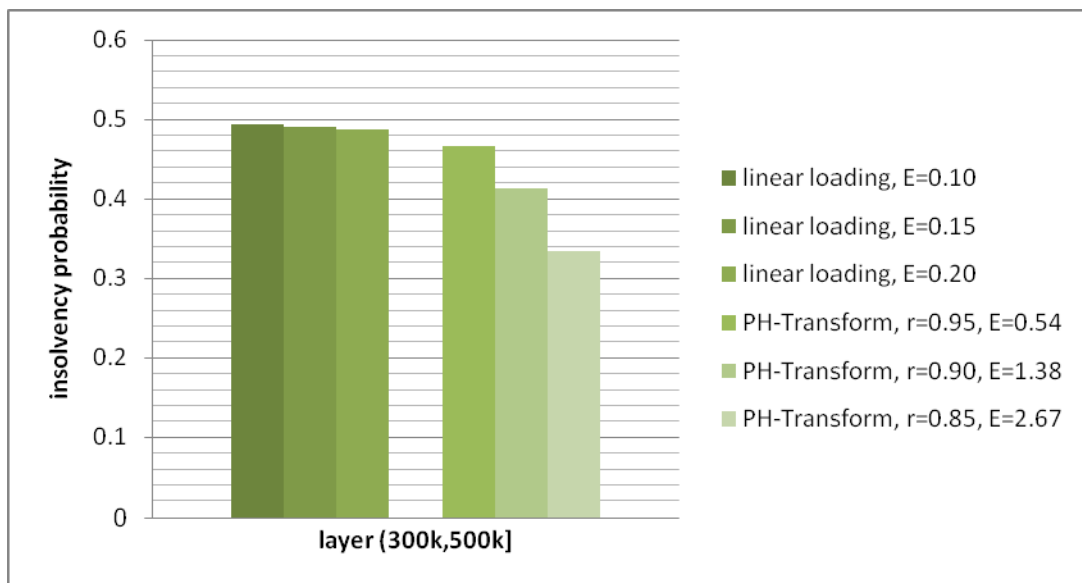


Figure 18: Graph of insolvency probability (layer (300k,500k])

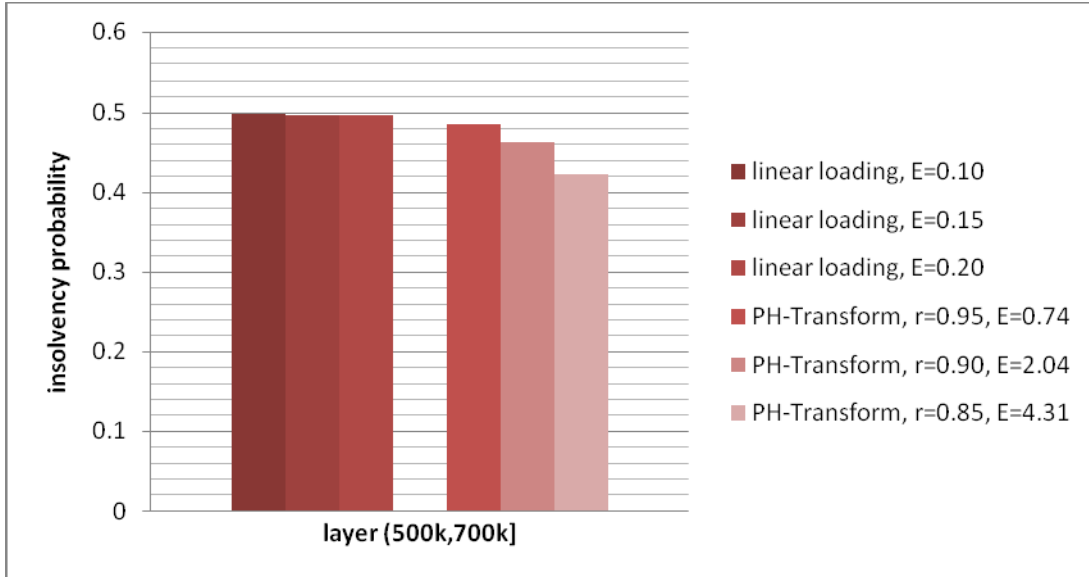


Figure 19: Graph of insolvency probability (layer (500k,700k])

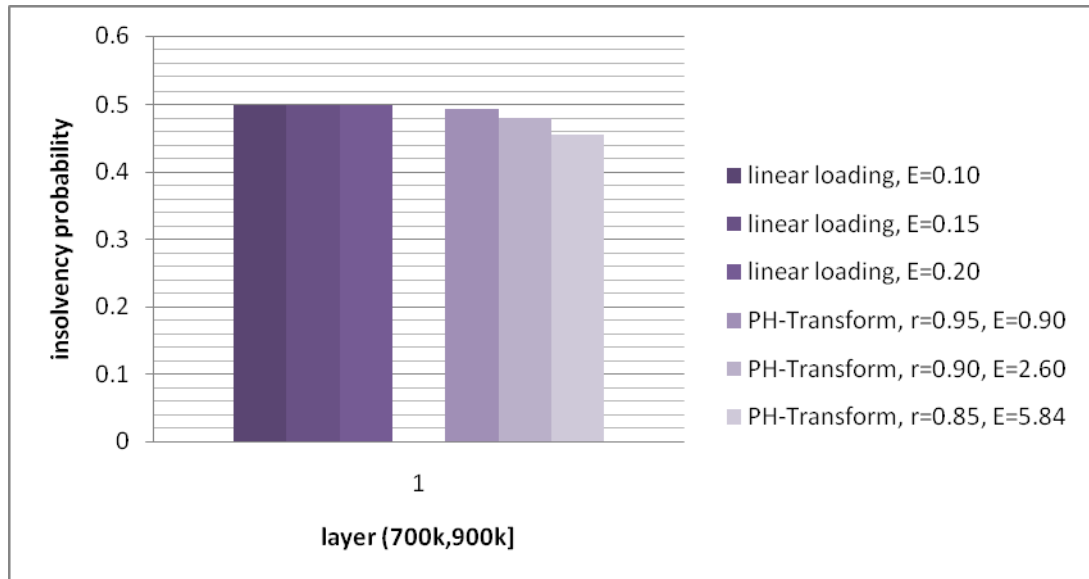


Figure 20: Graph of insolvency probability (layer (700k,900k])

The graphs show that under the linear loading assumption, insolvency probability decreases when relative loading increases. When the PH-Transform assumption is applied, the insolvency probability is reduced to a lower level compared to the linear loading assumption, and the reason for this is that the equivalent risk load is higher under the PH-Transform.

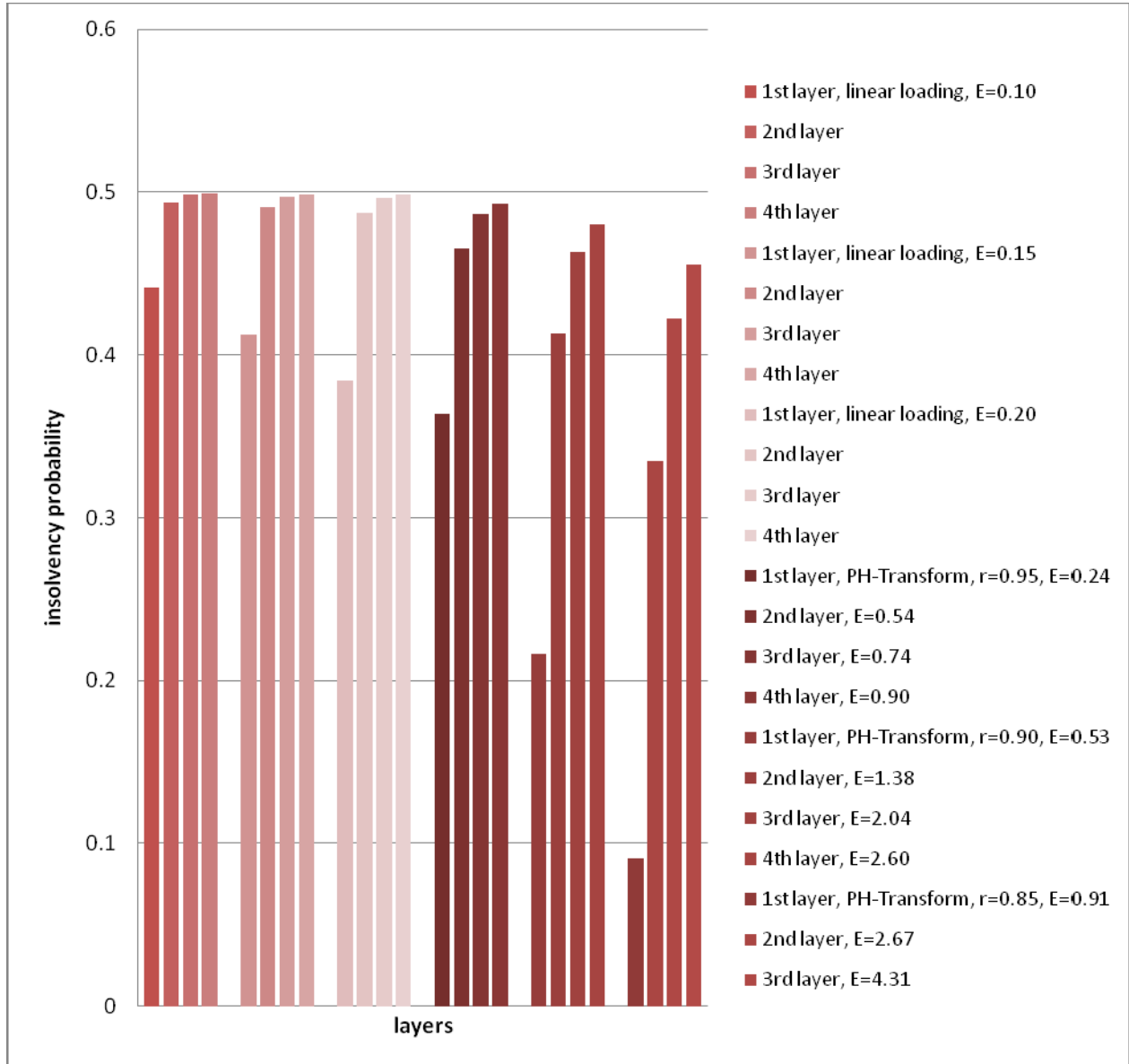


Figure 21: Graph of insolvency probability (all layers)

Appendix 6 shows the calculation of mean severity, mean frequency, mean aggregate claims, variance aggregate claims and insolvency probability under linear loading and PH-Transform assumptions, using R programming with the assistance of *actuar* package, assuming the severity distribution is Burr and the frequency distribution is Poisson.

8. CONCLUSION

In this paper, we have modeled individual loss amount, selected the best model using Kolmogorov-Smirnov, Anderson-Darling and Schwarz Bayesian Criterion, provided a range of deductible and policy limits based on Loss Elimination Ratio (LER), calculated insolvency probability under linear loading and PH-Transform assumptions, priced excess-of-loss of layer $(d, d + u]$ assuming a single risk, calculated increased limit factors (ILF), priced layers of a reinsurance treaty using a frequency and severity approach, and calculated the insolvency probability of a reinsurance treaty. Our proposed approach may be considered fair and efficient for two main reasons; the PH-Transform assumption may be implemented to lower the insolvency probability, and the R programming with the *actuar* package may be used for pricing excess-of-loss layers with less effort. In particular, the PH-Transform assumption is applied as a means of incorporating a risk load in the severity and/or frequency distributions and can be used to lower the insolvency probability of a single excess-of-loss layer as well as multiple layers of a reinsurance treaty. In addition, the ILF calculated in this study may be used as some indication or basis for possible basic and increased rates of the Malaysian insurance losses.

It is noteworthy that different distributions for loss severity and frequency can also be applied. Besides Burr distribution, Wang [12] showed that the PH-Transform assumption can be applied to several loss amount distributions such as exponential, uniform, Pareto and Weibull. The mean severity for a PH-Transform assumption, i.e., $H(M) = \int_d^{d+u} (S(x))^r dx$, can easily be computed using R programming with *actuar* package for such distributions. In addition, the computation of mean frequency for a PH-Transform assumption, i.e., $H(N) = \sum_{k=0}^{\infty} (S(k))^r$, for other frequency distributions such as binomial or negative binomial, can be also be implemented using R programming with the *actuar* package.

REFERENCES

- [1] Bank Negara Malaysia, "Annual report of the Director General of Insurance," 1995.
- [2] Bank Negara Malaysia, "Annual report of the Director General of Insurance," 2005.
- [3] Bank Negara Malaysia, "The Financial Sector Masterplan," 2001.
- [4] Cummins, J.D., "Property-Liability Insurance Price Regulation: The Last Bastion?", Chapter 1 in *Deregulating Property-Liability Insurance: Restoring Competition and Increasing Market Efficiency*, Washington, DC: AEI-Brookings Joint Center for Regulatory Studies, the American Enterprise Institute for Public Policy Research, 2002.
- [5] Dimitriyadis, I., and U.N. Oney, "Deductibles in Health Insurance," *Journal of Computational and Applied Mathematics* 233(1), 2009, pp. 51-60.

- [6] Hua, L. and K.C. Cheung, "Worst Allocation of Policy Limits and Deductibles," *Insurance: Mathematics and Economics* 40(1), 2008, pp. 93-98.
- [7] Klugman, S., H. Panjer, and G. Willmot, *Loss Models: From Data to Decisions*, 3rd ed., Hoboken, NJ: John Wiley and Sons, 2008.
- [8] Lau, S.W., "Malaysia Reinsurance Market Environment: Where Are We Going?," *Insurance*, July-September 2009, pp. 19-24.
- [9] Lee, H.L., *The Insurance Industry in Malaysia: A Study in Financial Development and Regulation*, Kuala Lumpur: Oxford University Press, 1997.
- [10] Rao, K., "The Impact of Detariffing," *Newstrack: Corporate Newsletter of Bajaj Allianz General Insurance*, December, 6, 2006.
- [11] Wang, S., "Implementation of Proportional Hazards Transforms in Ratemaking," *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society LXXXV*, 1998, pp. 940-979.
- [12] Wang, S., "Insurance Pricing and Increased Limits Ratemaking by Proportional Hazards Transforms," *Insurance: Mathematics and Economics* 17, 1995, pp. 43-54.
- [13] Wang, S., "Premium Calculation by Transforming the Layer Premium Density," *ASTIN Bulletin* 26, 1996, pp. 71-92.
- [14] Zhuang, W., Z. Chen, and T. Hu, "Optimal Allocation of Policy Limits and Deductibles Under Distortion Risk Measures," *Insurance: Mathematics and Economics* 44(3), 2008, pp. 409-414.

Appendix 1: R programming for LER (deductible policy, Burr distribution)

```
deduktibel <- function(alfa, gama, teta)
{
# to calculate E(X), d, E(X:d) and LER
EX <- mburr(1, alfa, gama, 1, teta)
d <- seq(0, 20000, by=1000)
EX.d <- levburr(d, alfa, gama, 1, teta, 1)
LER.d <- EX.d/EX
result.d <- cbind(d, EX.d, LER.d)
# to plot LER vs. d
plot.LERvsD <- plot(d, LER.d, type="p")
# to print result
list(EX=EX, result.d=result.d, plot.LERvsD)
}
deduktibel(alfa=3.778263226, gama=1.516886923, teta=86426.43339)
```

Appendix 2: R programming for insolvency probability (deductible policy, Burr distribution, linear loading)

```
insolvent.prob <- function(alfa, gama, teta, n, prob.claim, loading)
{
# to calculate d, E(W), E(W^2), Var(W), E(S), Var(S) and insolvency probability
d <- seq(0, 20000, by=1000)
```

```

EW <- mburr(1,alfa,gama,1,teta) - levburr(d,alfa,gama,1,teta,1)
EW2 <- mburr(2,alfa,gama,1,teta) - levburr(d,alfa,gama,1,teta,2) -
      2*d*mburr(1,alfa,gama,1,teta) + 2*d*levburr(d,alfa,gama,1,teta,1)
VW <- EW2 - (EW^2)
ES <- n*EW*prob.claim
VS <- n*(VW*prob.claim+(EW^2)*prob.claim*(1-prob.claim))
sigmaS <- VS^0.5
insolven.prob <- pnorm(ES*loading/sigmaS, 0, 1, FALSE, FALSE)
result <- cbind(d, ES, sigmaS, insolven.prob)
# to plot insolvency probability vs. deductible
plot.PROBvsD <- plot(d,insolven.prob,type="p")
# to print result
list(n=n, prob.claim=prob.claim, loading=loading, result=result, plot.PROBvsD)
}
insolvent.prob(alfa=3.778263226, gama=1.516886923, teta=86426.43339, n=3000, prob.claim=0.2,
loading=0.25)

```

Appendix 3: R programming for expected loss (deductible policy, Burr distribution, PH Transform)

```

explossPH <- function(alfa, gama, teta, r)
{
# to compute d, E(X) and loading
d <- seq(0, 20000, by=1000)
EX.basic <- mburr(1,alfa,gama,1,teta) - levburr(d,alfa,gama,1,teta,1)
EX.r <- mburr(1,r*alfa,gama,1,teta) - levburr(d,r*alfa,gama,1,teta,1)
loading <- (EX.r-EX.basic)/EX.basic
result <- cbind(d, EX.basic, EX.r, loading)
# to plot E(X) vs. deductible
plot.EXvsD <- plot(c(d,d), c(EX.basic, EX.r), type="p")
# to print result
list(r=r, result=result, plot.EXvsD)
}
explossPH(alfa=3.778263226, gama=1.516886923, teta=86426.43339, r=0.8)

```

Appendix 4: R programming for expected aggregate premium (single layer, single risk, Burr distribution, PH Transform)

```
layer <- function(alfa, gama, teta, d, u, prob.claim, r)
{
# to compute E(S) and loading
ES <- prob.claim*(levburr(u,alfa,gama,1,teta,1) - levburr(d,alfa,gama,1,teta,1))
ESr <- prob.claim*(levburr(u,r*alfa,gama,1,teta,1) - levburr(d,r*alfa,gama,1,teta,1))
loading <- (ESr-ES)/ES
result <- cbind(d, u, ES, ESr, loading)
# to print result
list(prob.claim=prob.claim, r=r, result=result)
}
d<-scan(n=8)
0 5000 10000 20000 40000 80000 100000 160000
u<-d+5000
layer(alfa=3.778263226, gama=1.516886923, teta=86426.43339, d, u, prob.claim=0.1, r=0.92)
```

Appendix 5: R programming for ILF (Burr distribution)

```
ILF <- function(alfa, gama, teta, r)
{
# to calculate a, E(X), risk load and ILF
a <- seq(100000,2000000,by=100000)
EX.a <- levburr(a,alfa,gama,1,teta,1)
EX.ar <- levburr(a,alfa*r,gama,1,teta,1)
EX.100k <- levburr(100000,alfa,gama,1,teta,1)
EX.100kr <- levburr(100000,alfa*r,gama,1,teta,1)
riskload <- EX.ar - EX.a
ILF <- EX.a/EX.100k
ILF.r <- EX.ar/EX.100kr
result <- cbind(a, EX.a, ILF, riskload, ILF.r)
# to print result
list(r=r, result=result)
}
ILF(alfa=3.778263226, gama=1.516886923, teta=86426.43339, r=0.9)
```

Appendix 6: R programming for mean severity, mean frequency, mean aggregate claims, variance aggregate claims and insolvency probability (excess-of-loss layers, Burr and Poisson distributions)

```
reinsurans <- function(alfa, gama, teta, lamda, d, u, r, SEP, loading)
{
# to compute E(M), H(M), E(N) and H(N)
EM <- levburr(d+u,alfa,gama,1,teta,1) - levburr(d,alfa,gama,1,teta,1)
HM <- levburr(d+u,r*alfa,gama,1,teta,1) - levburr(d,r*alfa,gama,1,teta,1)
data.diskret <- 0:10000
EN <- lamda
HN <- sum((1-ppois(data.diskret,lamda))^r)
# to compute E(S), Var(S) and insolvency probability
ES <- EM*EN
VS <- lamda*(levburr(d+u,alfa,gama,1,teta,2)-levburr(d,alfa,gama,1,teta,2)-
2*d*levburr(d+u,alfa,gama,1,teta,1)+2*d*levburr(d,alfa,gama,1,teta,1))
insolvency.prob <- pnorm(ES*loading/(VS^(0.5)),0,1,FALSE,FALSE)
insolvency.probr <- pnorm(((HM*HN)-ES)/(VS^(0.5)),0,1,FALSE,FALSE)
# to compute H(M)H(N), burning cost, loaded rate and relative loading
HMHN <- HM*HN
burning.cost <- (EM*EN)/SEP
loaded.rate <- (HM*HN)/SEP
relative.loading <- (loaded.rate-burning.cost)/burning.cost
result <- cbind(d, d+u, EM=EM, HM=HM, EN=EN, HN=HN, HMHN=HMHN,
burning.cost=burning.cost, loaded.rate=loaded.rate, relative.loading=relative.loading, ES=ES, VS=VS,
insolvency.prob=insolvency.prob, insolvency.probr=insolvency.probr)
# to print output
list(r=r, loading=loading, SEP=SEP, result=result)
}
d <- scan(n=5)
100000 300000 500000 700000 100000
u <- scan(n=5)
200000 200000 200000 200000 800000
reinsurans(alfa=3.778263226, gama=1.516886923, teta=86426.43339, lamda=6, d, u, r=0.9, SEP=10000000,
loading=0.1)
```

Predictive Modeling of Multi-Peril Homeowners Insurance

Edward W. (Jed) Frees^{*} Glenn Meyers[†] A. David Cummings^{‡§}

Abstract. Predictive models are used by insurers for underwriting and ratemaking in personal lines insurance. Focusing on homeowners insurance, this paper provides a systematic comparison of many predictive generalized linear models. We compare pure premium (Tweedie) and frequency/severity models based on single perils as well as multiple perils. With multiple perils, we also introduce instrumental variable models that 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 show how to use Gini indices for economic validation. We also consider a nonparametric regression that is used extensively by the statistical learning community. We find that different validation measures can help the actuary critically evaluate the effectiveness of alternative scoring procedures.

^{*}University of Wisconsin and ISO Innovative Analytics

[†]ISO Innovative Analytics

[‡]ISO Innovative Analytics

[§]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 US, 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 billions of US dollars (*I.I.I. Insurance Fact Book 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 (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.”

Rating by multiple perils does not cause nearly as much excitement in today’s world. 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 of the others. 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.

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. Our 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. To set the stage, Section 2.1 introduces our data and several baseline models. We consider both pure premium as well as frequency-severity approaches in this work. Thus, Section 2.1 introduces these approaches in both a single- and multi-peril

modeling framework. Section 3 introduces the instrumental variable approach, in both the pure premium and frequency-severity context. 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, for 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, 2010b) on a statistical measure called a Gini index to compare predictors in Section 5. Section 6 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 7 closes with a summary and a few additional remarks.

2 Data and Notations

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 were 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,” subsample that was also randomly selected from this database.

For each record, we have information on whether there was one or more 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 (according to the median severity) whereas the Other category is the least severe. In Table 1, we note that neither the frequency nor the number sum to the totals due to jointly occurring perils within a policy.

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 “basic” set consists of 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.

The second set 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.

Our 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.

2.2 Notations 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 i th record has an insurance claim due to the j th type, $j = 1, \dots, c$. Similarly, $y_{i,j}$ denotes the amount of the claim due to the j th type. To relate the multi- to the single-peril variables, we have

$$r_i = 1 - (1 - r_{i,1}) \times \dots \times (1 - r_{i,c}) \quad (1)$$

and

$$y_i = \sum_{j=1}^c r_{i,j} \times y_{i,j}. \quad (2)$$

We interpret r_i to be a binary variable indicating whether or not the i th 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 \mathbf{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 \mathbf{x}_{Fi} as the set of explanatory variables. Denote the corresponding set of

regression coefficients as β_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 \mathbf{x}_{S_i} as the set of explanatory variables. Denote the corresponding set of regression coefficients as β_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 $\mu_i = \exp(\mathbf{x}'_i\beta)$, 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 be decomposed by peril. From our database, explanatory variables have been selected by peril j for the frequency, $\mathbf{x}_{F,i,j}$, and severity, $\mathbf{x}_{S,i,j}$, portions, $j = 1, \dots, 9$. For example, these variables range in number from eight for the Other peril to nineteen 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 $\mathbf{x}_{F,i,j}$ as the set of explanatory variables, with corresponding set of regression coefficients $\beta_{F,j}$.

- 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 $\mathbf{x}_{S,i,j}$ as the set of explanatory variables, with corresponding set of regression coefficients $\beta_{S,j}$.
- We do this for each peril, $j = 1, \dots, 9$.

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 $\mathbf{x}_{F,i,j}$ and severity $\mathbf{x}_{S,i,j}$ variables to be our set of explanatory variables for the j th peril, $\mathbf{x}_{i,j}$. With these, one can estimate a pure premium model for each peril, $j = 1, \dots, 9$. 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 will look to out-of-sample results beginning in Section 4. In Frees et al. (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 will include these scores labeled as “DepRatio1” and “DepRatio36”, for 1 and 36 dependency parameters, respectively. Additional details on this method are in Frees et al. (2010).

3 Multi-Peril Models with Instrumental Variables

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 our prior work (see Appendix A), we established statistically significant associations between claims from fire and theft/vandalism.

In this paper, we introduce an instrumental variable method of estimation to improve upon the predictions under the independence models. For example, 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 keep this paper self-contained, Appendix Section B provides an introduction to instrumental variable estimation as it appears in a classical linear system of equations. Sections 3.1 and 3.2 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 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

$$y_{ij} \sim \text{Tweedie}(\mu_{i,j}, \phi_j, p_j), \quad i = 1, \dots, n = 404,664, \quad j = 1, \dots, c = 9. \quad (3)$$

Here, ϕ_j is the dispersion parameter, p_j is the shape parameter and $\mu_{i,j} = \exp(\mathbf{x}'_{i,j}\boldsymbol{\beta}_j)$ is the mean parameter using a logarithmic link function. There are many procedures for estimating the parameters in equation (3), we use maximum likelihood.

Estimating independence pure premium models with equation (3) allows us to determine regression coefficient estimates $\mathbf{b}_{IND,j}$. These coefficients allow us to compute (independence model) pure premium estimates of the form $\hat{\mu}_{IND,i,j} = \exp(\mathbf{x}'_{i,j} \mathbf{b}_{IND,j})$.

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 $\mathbf{x}_{i,1}$. We augment $\mathbf{x}_{i,1}$ with the additional predictor variables

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

We then estimate the pure premium model in equation (3) 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). 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 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

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

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 Section C).

The algorithm is similar to the pure premium modeling in Section 3.1. We summarize the procedure as follows.

- Stage 1 - Compute independence frequency and severity model fitted values. Specifically, for each of the $j = 1, \dots, 9$ perils:
 - 1a. Fit a logistic regression model using the explanatory variables $\mathbf{x}_{F,i,j}$. These explanatory variables differ by peril j . Calculate fitted values to get predicted probabilities, denoted as $\widehat{\pi}_{IND,i,j}$.
 - 1b. Fit a gamma regression model using the explanatory variables $\mathbf{x}_{S,i,j}$ 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 $\widehat{E} y_{IND,i,j}$.
- Stage 2. Incorporate additional instruments into the frequency model estimation. Specifically, for each of the $j = 1, \dots, 9$ perils:
 - 2. Fit a logistic regression model using the explanatory variables $\mathbf{x}_{F,i,j}$ and the logarithm of the predicted probabilities developed in step 1(a), $\ln \widehat{\pi}_{IND,i,k}$, $k = 1, \dots, 9, k \neq j$.

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 3.1 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 Section C, 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 subsample 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

$$\text{IND_FreqSev}_i = \sum_{j=1}^c \widehat{\text{Prob}}_{i,j} \times \widehat{\text{Fit}}_{i,j} = \sum_{j=1}^c \frac{\exp(\mathbf{x}'_{F,i,j} \mathbf{b}_{F,j})}{1 + \exp(\mathbf{x}'_{F,i,j} \mathbf{b}_{F,j})} \times \exp(\mathbf{x}'_{S,i,j} \mathbf{b}_{S,j}). \quad (4)$$

Here, $\widehat{\text{Prob}}_{i,j}$ is the predicted probability using logistic regression model parameter estimates, $\mathbf{b}_{F,j}$, and frequency covariates $\mathbf{x}_{F,i,j}$, for the j th peril. Further, $\widehat{\text{Fit}}_{i,j}$ is the predicted amount based on a logarithmic link using gamma regression model parameter estimates, $\mathbf{b}_{S,j}$, and severity covariates $\mathbf{x}_{S,i,j}$, for the j th 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 5 provides cumulative comparisons using a Gini index. Section 6 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 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.

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.

5 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 version 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 (score divided by premium).

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 2.1 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 et al. (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 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 is 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.

6 Out-of-Sample Analysis Using Local Comparisons of Claims to Scores

As described in Section 5, 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 i th 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 “SP_FreqSev_Basic” and score “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.

Summarizing 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.

Of course, extensive aggregation such as at the decile level may hide important patterns. The middle and right-hand panels of Figure 2 shows 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 et al., 2001), it is less widely used in actuarial applications. The ideas are straight-forward. Consider a set of relativities and loss ratios of the form $(R_i, LR_i), i = 1, \dots, n$. 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

$$\hat{m}(x) = \frac{\sum_{i=1}^n w(x, R_i) LR_i}{\sum_{i=1}^n w(x, R_i)}, \quad (5)$$

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 $E(LR|R = x)$, 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 relativities (1.2 through 1.4) and approximately similar elsewhere. The right panel compares the error in IND_FreqSev to the basic pure premium score, SP_PurePrem, showing that

these two measures perform about the same for most of the data.

For our application, we interpret $\hat{m}(x) - x$ 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 .

7 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 showed 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 healthcare, 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 2.2 and 3 introduced many models, each of which has advantages compared to alternatives. For example, the “basic” 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

1. Arellano, M. (2003). *Panel Data Econometrics*. Oxford University Press, Oxford, UK.
2. Board of Trustees, Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds (2009), *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. Available at:
<http://www.socialsecurity.gov/OACT/TR/2009/tr09.pdf>.
3. Frees, Edward W. (2004). *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. Cambridge University Press, New York.
4. Frees, Edward W. (2010). *Regression Modeling with Actuarial and Financial Applications*. Cambridge University Press, New York.
5. Frees, Edward W., Glenn Meyers, and A. David Cummings (2010). Dependent multi-peril ratemaking models. *Astin Bulletin* 40(2), 699-726.
6. Frees, Edward W., Glenn Meyers, and A. David Cummings (2011). Summarizing insurance scores using a Gini index. To appear in the *Journal of the American Statistical Association*.
7. Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2001). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York.
8. Hurley, Robert L. (1958). Multiple peril rating problems – Some statistical considerations. *Proceedings of the Casualty Actuarial Society* 46, 196-213.
9. Meyers, Glenn and A. David Cummings (2009). “Goodness of Fit” vs. “Goodness of Lift”. *The Actuarial Review: Newsletter of the Casualty Actuarial Society*, August, p. 16. Available at http://www.casact.org/newsletter/pdfUpload/ar/AR_Aug2009_1.pdf.
10. Modlin, Claudine (2005). 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>.
11. Rollins, John W. (2005). A modern architecture for residential property insurance ratemaking. *Proceedings of the Casualty Actuarial Society* 92, 486-578.
12. Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. In *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Golden Jubilee International Conference* (Editors J. K. Ghosh and J. Roy), pp. 579-604. Indian Statistical Institute, Calcutta.
13. Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Part I

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

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

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 j th and k th perils is $\sum_{i=1}^n r_{ij} \times r_{ik}$. Assuming independence among perils, this has mean and variance

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

and

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

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

$$t_{jk} = \frac{\sum_{i=1}^n r_{ij} \times r_{ik} - \sum_{i=1}^n q_{ij} \times q_{ik}}{\sqrt{\sum_{i=1}^n q_{ij}q_{ik} - (q_{ij}q_{ik})^2}}. \quad (6)$$

The t -statistic in equation (6) 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 Overview of the Instrumental Variables Approach

To keep this paper self-contained, this section provides a brief introduction of the instrumental variable methods 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 that is summarized by two equations:

$$\begin{aligned} y_{1i} &= \beta_1 y_{2i} + \gamma_{10} + \gamma_{11} x_{1i} + \varepsilon_{1i} && \text{(price)} \\ y_{2i} &= \beta_2 y_{1i} + \gamma_{20} + \gamma_{21} x_{2i} + \varepsilon_{2i} && \text{(quantity)}. \end{aligned} \tag{7}$$

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. These other explanatory variables (x ’s) are assumed to be exogenous for the demand and supply equations.

For simplicity, assume that we have $i = 1, \dots, n$ independent observations that follow display (7). One estimation strategy is to use ordinary linear regression. As we will see, this strategy yields biased regression coefficients estimates. This is because on the right-hand side of display (7), the “conditioning” or explanatory variables, contains a y variable that is also a dependent variable.

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 re-write display (7) as

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

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

A simpler approach, that has advantages in many situations, is to use ordinary least squares with approximate values for the right-hand side dependent variables. To see how this approach works, we focus on the price equation and assume that we have available “instruments” \mathbf{w} to approximate y_2 . Then, we employ the following two-stage strategy:

1. Run a regression of \mathbf{w} on y_2 to get fitted values of the form \hat{y}_2 .
2. Run a regression of x_1 and \hat{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 a 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.

The difficulty with ordinary least squares estimation of the model in display (7) is that the right-hand side variables are correlated with the disturbance term. For example, looking 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 turns depends on ε_1 (from the demand equation). This circular dependency structure induces the correlation that leads to biased regression coefficient estimation.

Extending this line of thought, suppose that theory suggests a linear equation of the form

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

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 $E \varepsilon_i \mathbf{x}_i = \mathbf{0}$ for exogeneity. When not all of the regressors are exogenous, the instrumental variable technique employs a set of variables, \mathbf{w}_i , that are correlated with the regressors specified in the structural model. Specifically, we assume

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

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

Although we will not go into the details, there are conditions on the instruments. Typically, they may include a subset of \mathbf{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 \mathbf{x} , resulting in perfect collinearity and non-identifiability of the coefficients. Further, the new explanatory variables in \mathbf{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).

C Frequency and Severity Instrumental Variable Estimation

Section 3.2 described 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 approximate $PREM_j$, pure premium, in our empirical work.

- Pure premium is expected frequency times severity, that is, $PREM_j = \pi_j \times E y_j$
- 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 $\ln \widehat{E} y_j$ 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 $j = 1, \dots, 9$ perils:
 - 1a. Fit a logistic regression model using the explanatory variables $\mathbf{x}_{F,i,j}$. 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 $\mathbf{x}_{S,i,j}$ 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 $\widehat{E} y_{IND,i,j}$.
- Stage 2. Incorporate additional instruments into the model estimation. Specifically, for each of the $j = 1, \dots, 9$ perils:
 - 2a. Fit a logistic regression model using
 - * the explanatory variables $\mathbf{x}_{F,i,j}$,
 - * the logarithm of the predicted probabilities developed in step 1(a), $\ln \hat{\pi}_{IND,i,k}$, $k = 1, \dots, 9, k \neq j$ and
 - * the logarithm of the fitted values in step 1(b), $\ln \widehat{E} y_{IND,i,j}$.
 - 2b. Fit a gamma regression model using
 - * the explanatory variables $\mathbf{x}_{S,i,j}$ and
 - * the logarithm of the fitted predicted probabilities in step 1(a), $\ln \hat{\pi}_{IND,i,j}$.

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
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

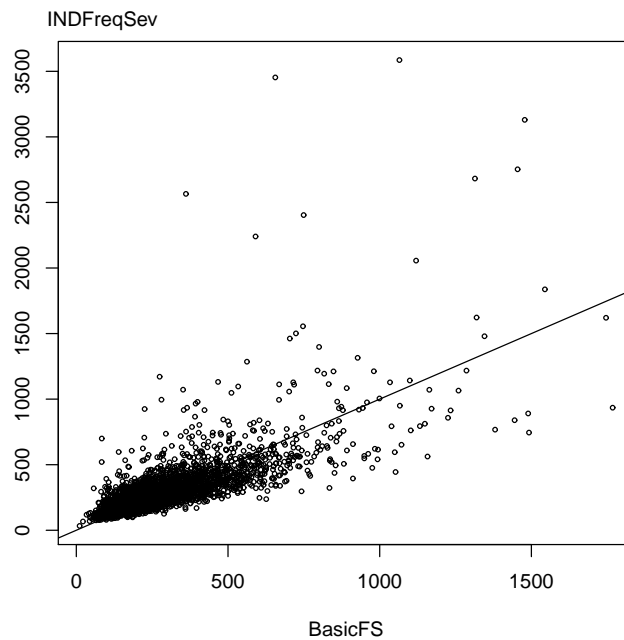


Figure 1: Single versus Multi-Peril Frequency-Severity Scores. This graph is based on a 1 in 100 random sample of size 3,594. The correlation coefficient is only 79.4%.

Table 2: Instrumental Variable Pure Premium Model Coefficients.
Shown are coefficients associated with the instruments, logarithmic fitted values.

Explanatory Variables	Dependent Variables					
	Fire		Lightning		Wind	
	Estimate	<i>t</i> -statistic	Estimate	<i>t</i> -statistic	Estimate	<i>t</i> -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

Explanatory Variables	Dependent Variables					
	Hail		Water Weather		Water NonWeather	
	Estimate	<i>t</i> -statistic	Estimate	<i>t</i> -statistic	Estimate	<i>t</i> -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

Explanatory Variables	Dependent Variables					
	Liability		Other		Theft	
	Estimate	<i>t</i> -statistic	Estimate	<i>t</i> -statistic	Estimate	<i>t</i> -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		

Table 3: Summary Statistics of Fourteen Scores and Total Claims

Score	Mean	Mini- mum	Percentiles							Maxi- mum
			1st	5th	25th	50th	75th	95th	99th	
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
<i>Scores using the basic set of explanatory variables</i>	
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
<i>Scores using the extended set of explanatory variables</i>	
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.
<i>Instrumental variable multi-peril frequency and severity models, using the extended set of explanatory variables</i>	
IV_FreqSevA	Uses instruments in frequency model
IV_FreqSevB	Uses instruments in severity model
IV_FreqSevC	Uses instruments in frequency and severity models
<i>Dependence ratio multi-peril frequency and severity models, using the extended set of explanatory variables</i>	
DepRatio1	Uses a single parameter for frequency dependencies
DepRatio36	Uses 36 parameters for frequency dependencies

Table 4: Spearman Correlations of Fourteen Scores and Total Claims

	Basic Explanatory Variables						Extended Explanatory Variables						DepRatio											
	Single Peril			IV_			Single Peril			IND_			IV_			DepRatio								
	Freq	Pure	Sev	Prem	Pure	Prem	Freq	Pure	Sev	Prem	Pure	Prem	Freq	Pure	Sev	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	0.048	0.051	0.048	0.051	0.048	0.033	0.049	0.032	0.032	0.032

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.

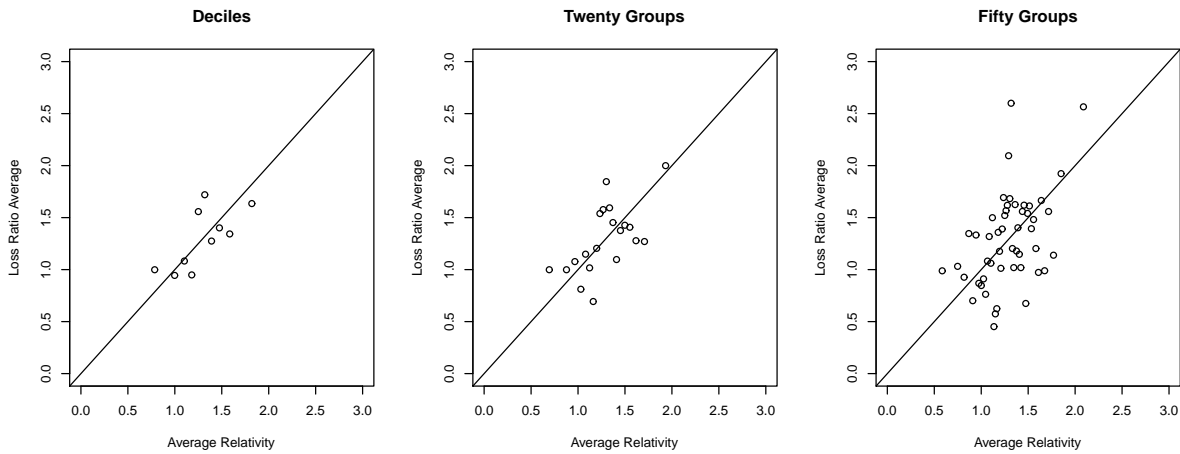


Figure 2: Average Relativities and Loss Ratios, by Groups of Scores.

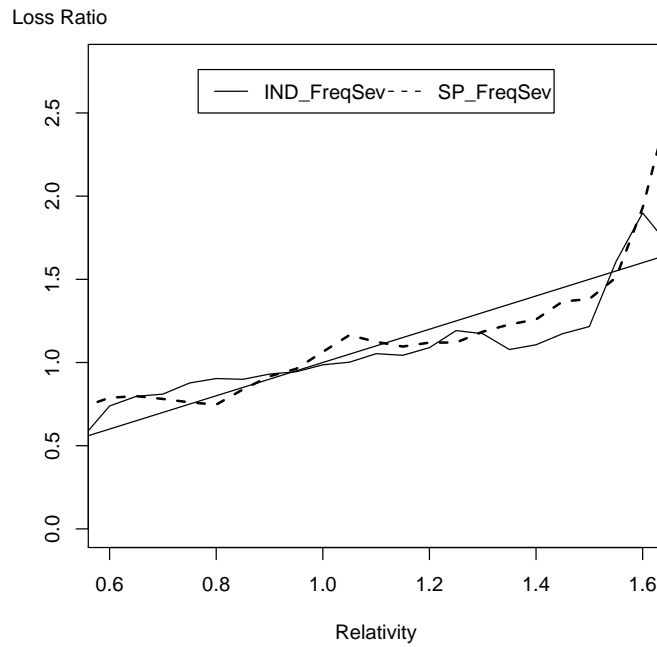


Figure 3: Comparison of Single and Multi-Peril Frequency-Severity Loss Ratios.

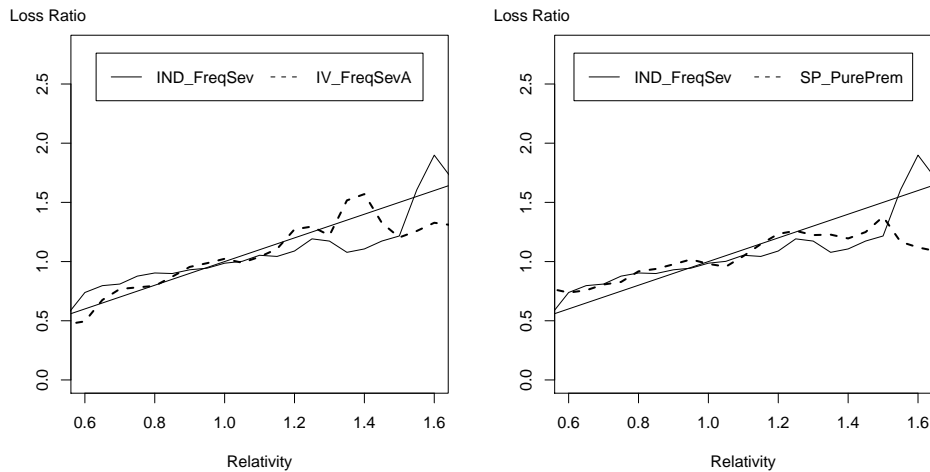


Figure 4: Comparison of Loss Ratios from several Scoring Methods. The left panel compares the independence to an instrumental variable frequency-severity approach. The right panel compares the independence frequency-severity approach to the single peril pure premium (Tweedie) method.

Table 6: Joint Claim Counts Among Perils

	Fire	Lightning	Wind	Hail	Water Weather	Water Non Weather	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 Non Weather	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

Table 8: 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

Generalized Linear Mixed Models for Ratemaking: A Means of Introducing Credibility into a Generalized Linear Model Setting

Fred Klinker, FCAS, MAAA

Abstract: GLMs that include explanatory classification variables with sparsely populated levels assign large standard errors to these levels but do not otherwise shrink estimates toward the mean in response to low credibility. Accordingly, actuaries have attempted to superimpose credibility on a GLM setting, but the resulting methods do not appear to have caught on. The Generalized Linear Mixed Model (GLMM) is yet another way of introducing credibility-like shrinkage toward the mean in a GLM setting. Recently available statistical software, such as SAS PROC GLIMMIX, renders these models more readily accessible to actuaries. This paper offers background on GLMMs and presents a case study displaying shrinkage towards the mean very similar to Buhlmann-Straub credibility.

Keywords: Credibility, Generalized Linear Models (GLMs), Linear Mixed Effects (LME) models, Generalized Linear Mixed Models (GLMMs).

1. INTRODUCTION

Generalized Linear Models (GLMs) are by now well accepted in the actuarial toolkit, but they have at least one glaring shortcoming--there is no statistically straightforward, consistent way of incorporating actuarial credibility into a GLM.

Explanatory variables in GLMs can be either continuous or classification. Classification variables are variables such as state, territory within state, class group, class within class group, vehicle use, etc. that take on a finite number of discrete values, commonly referred to in statistical terminology as "levels." The GLM determines a separate regression coefficient for each level of a classification variable. To the extent that some levels of some classification variables are only sparsely populated, there is not much data on which to base the estimate of the regression coefficient for that level. The GLM will still provide an estimated coefficient for that level but will assign it a large standard error of estimation. In effect, the GLM warns the user to exercise considerable care in interpreting that coefficient but doesn't otherwise adjust the estimated coefficient to take into account the low volume of data. When faced with this situation, the natural inclination of an actuary is to shrink low credibility levels towards the mean, but the GLM quotes a large standard error of estimation and leaves it at that.

There have been a number of responses to this unsatisfactory state of affairs. Some actuaries have been known to apply an ad hoc credibility adjustment to coefficients output by a GLM. In some cases this even produces results similar to those arrived at by more statistically rigorous

methods. If so, then what is so wrong with the ad hoc credibility adjustment of GLM output? First, we cannot guarantee the ad hoc results will always agree closely with results from those other methods. Second, the statisticians who designed our GLMs were unaware we intended to subject GLM estimates to the violence of a subsequent round of ad hoc credibility adjustments. If they had known, they might have suggested a better starting point than GLM estimates. This gets back to the old issue that a sequence of steps, each optimal individually, may not be optimal in the aggregate.

Turning to other, more statistically rigorous attempts to incorporate credibility in a GLM setting, it would be desirable to find a method that estimates both the GLM and the credibility adjustment in a single, statistically consistent step where each GLM estimation and credibility adjustment takes into account the fact that the other estimation process is also going on. A number of authors have indeed produced models that combine GLM and credibility, for example, Nelder and Verrall (1997), Ohlsson and Johansson (2004 and 2006), and Ohlsson (2006). Given the importance of the issue these papers address, why have these models not caught on in actuarial circles (at least not that I am aware)? I might hazard two guesses. First, their math is somewhat complex and perhaps intimidating. Second, their algorithms are iterative and require a nontrivial degree of programming from their users.

There are alternative statistical models, quite similar in theory to Nelder and Verrall and Ohlsson and Johansson, known as Generalized Linear Mixed Models (GLMMs). Statisticians actually developed these models some time ago, but it has only been very recently that popular stat software (like SAS, R, and S-Plus) has been enhanced to provide us with the means to readily estimate these models. Furthermore, it should be noted that models much like GLMMs have even been introduced into the actuarial literature. See, for example, Guszcz (2008), which admittedly introduced these models in a reserving rather than ratemaking setting, but that paper does provide a good introduction and intuition regarding what is going on in the guts of a GLMM, or something much like a GLMM.

I will not argue that GLMMs provide better models than Nelder and Verrall and Ohlsson and Johansson, but the newly available software makes it easier to implement these GLMM models.

1.1 Objectives of This Paper

The objectives of this paper are to:

- Introduce Linear Mixed Effects (LME) models and their generalization to Generalized Linear Mixed Models (GLMMs).

- Show how the LME, at least when applied to simple models, can be solved in closed form and leads directly to shrinkage of random effects towards the mean of the form of Buhlmann-Straub credibility. This motivates at least the hope that a similar shrinkage might be expected from a GLMM, where the math is no longer tractable in closed form.
- Demonstrate the application of a GLMM to a case study in which the hoped-for shrinkage is indeed observed and does indeed approximate the form of Buhlmann-Straub credibility.
- Demonstrate along the way (in Appendix A) SAS code that implements the GLMM.

In one sense, the central point of the paper is Table 4 and Figure 1. These show the shrinkage observed in the case study and the fact that this shrinkage is of form approximately Buhlmann-Straub. The reader who takes nothing else away from this paper should at least keep this Table and Figure in mind as motivation for wanting to learn more about GLMMs as a means of implementing credibility in a GLM setting in a manner very reminiscent of credibility theory they already know.

The reader should also keep in mind what this paper is *not*, and the following comments are offered as a means to managing readers' expectations. What this paper is not is a general review article on the various means by which credibility has been incorporated into a GLM setting. I will not discuss the various ad hoc credibility adjustments to GLM output alluded to earlier, nor Nelder and Verrall, nor Ohlsson and Johansson, nor other more overtly Bayesian or Empirical Bayes methods. I will not discuss the subtle theoretical points in which Nelder and Verrall and Ohlsson and Johansson differ from GLMMs nor examine the differences in results produced by applying ad hoc credibility adjustments vs. Nelder and Verrall vs. Ohlsson and Johansson vs. GLMMs to the case study of this paper. What this paper is intended to say is, "Here is one very interesting way of implementing credibility in a GLM setting. It might or might not be the best from among those methods currently available, but it is certainly promising. It produces credibility-like shrinkage very similar to credibility you, the actuary, are already familiar with. And it has the added advantage of ready implementation via software only recently available."

This paper is also not intended to be a comparison of GLMM implementations in different stat packages. SAS PROC GLIMMIX is likely to be available to many of the readers of this paper, and it happens to be the means I chose to implement GLMMs. But there are also implementations in R, S-Plus, etc., and I don't mean to imply that SAS PROC GLIMMIX is superior to these others.

I leave to another, more energetic and ambitious author the task of writing the general review article that some readers might have been hoping for. Indeed, I would hope that this paper serve as

the impetus for such a review article.

1.2 Prerequisites

It will be assumed that the reader is already familiar with the theory of GLMs and their application to actuarial problems at the level of McCullagh and Nelder (1989), Anderson et al. (2004), and de Jong and Heller (2008).

1.3 Outline of Remainder of This Paper

The remainder of this paper proceeds as follows. Section 2 will introduce the Linear Mixed Effects (LME) model, a simpler cousin of the GLMM, as a means of introducing many features of GLMMs before I discuss their complications. Section 3 will show how, in a very simple case, Buhlmann-Straub credibility emerges directly from the LME model. This is done to motivate the connection between GLMMs and credibility. The LME is generalized to the GLMM in Section 4.

Section 5 presents a case study on live ISO data for an unspecified line of business. By comparing GLM and GLMM runs on essentially the same model form and same data, it is shown that the GLMM introduces a shrinkage of sparsely populated classification variable levels towards the mean. This shrinkage is not seen in the GLM. Furthermore, it is shown that the credibility implied by this shrinkage is very close to the form of Buhlmann-Straub credibility. Section 6 concludes. SAS code implementing the Section 5 case study as well as a discussion of some of the output from that code has been deferred to Appendix A.

2. THE LINEAR MIXED EFFECTS (LME) MODEL

The Linear Mixed Effects model is nothing other than classical linear regression (more correctly, the classical general linear model) with the addition of “random” effects to the “fixed” effects already treated in classical linear regression. The error distribution is assumed normal. Expected values are assumed linear in explanatory variables. In the language of GLMs, the error distribution in the “exponential” family is the normal distribution, and the link function is the identity function.

Especially because it is so central to the understanding of the rest of this paper, more needs to be said about the distinction between “fixed” and “random” effects. The classic fixed effect is a classification variable with relatively few levels, those being the only levels we are interested in. The classic random effect is a classification variable with potentially many levels, only some of which appear in our dataset by design of the sample that produced our dataset from the overall population. Re random effects, the focus is frequently on the variance among the levels rather than on the values

of the levels themselves, which are assumed to have expectation zero. Even when there is interest in the values of the levels of the random effects, the inferential algorithm that predicts those levels must first estimate the random effects variances. It should also be noted that to some extent the distinction between fixed and random effects depends on the context of the study; the very same effect treated as fixed in one study might reasonably be treated as random in another, given the different goals of the two studies.

Consider the following example. Suppose you want to test the relative efficacy of a number of drugs, so your model includes a drug main effect. The levels of that drug main effect test whether some drugs are better than others, better than a control, better than a placebo, etc. in terms of some response that serves as the dependent variable in your model. You run your drug trials at a number of test centers. This suggests that you include test center as another main effect in your model to control for possible test center differences. If you treat test center as a fixed effect, you end up drawing inferences for drug main effects appropriate for those test centers but not validly extendible to medical centers other than those at which you ran the tests. On the other hand, if you drew your test centers relatively at random from a much larger universe of possible centers, and you reflect that fact by treating test center as a random main effect in your model, then you end up drawing drug main effect inferences that can validly be extended to centers other than the ones at which you actually did the tests. Quoted standard errors of the drug effects will be somewhat larger because of the additional uncertainty attributable to the treatment center random effect. For further enrichment re random effects, see Littell et al. (2006) or any one of many texts on random or mixed effects models.

The above describes the classical statistical motivation for random effects, but the actuarial motivation for considering random effects actually differs somewhat from this. If we treat state, or territory within state, as random, is that because we want to extend our model results to states or territories we didn't actually have in our data? No, not usually. If we want our model results to apply to a given state or territory, we usually include that state or territory in our data. But the zero expectation of random effects models creates a natural drift towards zero when data is thin. This shrinkage towards zero looks a lot like credibility with the complement of credibility being the overall mean, and we invoke the statistical machinery of random effects to exploit that shrinkage.

Turning to the description of the LME model, the basic equation is:

$$Y = X\beta + Zu + e. \quad (2.1)$$

Generalized Linear Mixed Models for Ratemaking

This would be the classic linear regression equation (expressed in matrix form) except for the additional random effects term Zu . This is a matrix equation. If there are n observations, Y is an n -vector of the observed values of the dependent variable. X is an n by p matrix, referred to as the model structure (or design) matrix, for the fixed effects. If the model includes an intercept term, there is a column of X consisting of all ones to capture the intercept. For each continuous explanatory variable, there is a column of X consisting of the values for that variable. For each main effect classification variable with m levels, there are m columns of X which are indicator variables for membership in each of the m levels (except that certain full rank parameterizations of the model suppress one of the columns). The indicator variable for the i^{th} level takes the value one if the observation is indeed in that level, zero otherwise. Interaction terms contribute more complex columns to X . β is the p -vector of regression coefficients for the fixed effects. This is not to say that there are a total of p fixed effect variables in the model, only that, taking into account the intercept and the fact that classification variables contribute multiple columns to the structure matrix, it requires a total of p regression coefficients to fully specify the fixed effects part of the model.

Z is the n by q design matrix for the model random effects. The columns of Z are indicator variables for membership in classification variable levels for those classification variables treated as random effects. The q -vector u is the equivalent of β and can be thought of as the vector of regression coefficients for the random effects. The n -vector e is the vector of random measurement errors.

Further structure is imposed by the following assumptions. Both u and e vectors are multivariate normally distributed with expectations 0. The variance of u is a q by q matrix $\text{Var}[u]=G$. The variance of e is an n by n matrix $\text{Var}[e]=R$. The u and e vectors are assumed uncorrelated: $\text{Cov}[u,e]=0$. The structure of G specifies the structure of correlation among the random effects. G is frequently assumed diagonal or block diagonal. Other types of correlational structure among the observations, such as autocorrelated time series or spatial structure, are specified through the structure of the R matrix. The user will most likely specify the structure of G and R , but these matrices may include unknown parameters that have to be estimated as part of the LME algorithm. It is common to speak of G side and R side covariance structure to distinguish between correlation arising through random effects vs. other time series or spatial processes. These G side and R side covariance structures tend to be specified in different places in the model specification syntax.

This model structure gives rise to two relevant distributions. (The reader is forewarned that this

dichotomy will become more important when the LME is generalized to the GLMM in a later section.) The first is the conditional distribution, $Y|u$, of the dependent variable Y conditional on actually knowing the random effects u . This distribution has expectation $X\beta + Zu$ and variance R . The second distribution is the marginal distribution for Y not knowing the random effects, which is the conditional distribution integrated over the random effects. It has expectation $X\beta$ (because u has expectation 0) and variance $V = \text{Var}[Zu + e] = ZGZ' + R$, where Z' denotes the transpose of the Z matrix. Note that the total variance V has G side and R side contributions. In a normal world, where both u and e are multivariate normal, so are both the conditional and marginal distributions for Y , but this result need not extend to the GLMM.

The LME is solvable in closed form via generalized least squares. The estimator for β is BLUE, Best Linear Unbiased Estimator (or EBLUE, Estimated or Empirical Best Linear Unbiased Estimator, if the total variance matrix, V , includes unknown parameters that have to be estimated as part of the solution. For a discussion of BLUE and related terms, see Littell et al. (2006)), and is given by

$$\hat{\beta} = (X'V^{-1}X)^{-1} X'V^{-1}Y. \quad (2.2)$$

The predictor for the random effects is the expectation of the random effects u conditional on the observed Y , is BLUP, Best Linear Unbiased Predictor (or EBLUP, Estimated or Empirical Best Linear Unbiased Predictor), and is given by

$$E[u | Y] = E[u] + \text{Cov}[u, Y] \text{Var}[Y]^{-1} (Y - E[Y]) = GZ'V^{-1}(Y - X\hat{\beta}) \quad (2.3)$$

Discussion of how the unknown parameters in V are estimated would take us too far afield, nor is it necessary for the following argument.

The reader seeking further enrichment re LME models can consult such books as Littell et al. (2006). This book also provides discussion of such standard statistical terminology as BLUE and BLUP, the distinction between estimators for fixed effects and predictors for random effects, the generalization from LME to GLMM, as well as numerous examples of implementations of LMEs and GLMMs via SAS software.

3. HOW BUHLMANN-STRAUB CREDIBILITY EMERGES FROM THE LME MODEL

A key point of this paper is that random effects in LME and their generalization to GLMMs entail a credibility-like shrinkage. This result is exact for LME and particularly easy to see in the following simple example.

Let us assume that class is the only explanatory variable, so we are looking at a one-way ANOVA, treating the grand mean as a fixed effect and the class offsets about the grand mean as random effects. We assume data has been aggregated, so there is only one observation per class, the class i average response y_i . Y is the vector of the y_i . Exposures for class i are w_i . X is the fixed effects design matrix appropriate for an intercept only model, hence a single column of identical ones. β , the vector of fixed effects regression coefficients, is only a 1-vector with the single entry the intercept. The random effects design matrix, Z , has columns that are indicator variables for the various levels of the class variable. If we assume our observations are in class order (first observation in the first class, second observation in the second class, etc.), then Z is just an identity matrix.

We turn next to the structure of the R side and G side variance matrices. R is the variance of the random errors e . We assume the e are independent of one another from class to class, so R is diagonal. Furthermore, we assume the i^{th} class error variance is equal to a proportionality constant known as the within variance, σ_w^2 , divided by the exposure volume w_i . In other words, the error variance declines with increasing volume. So R is diagonal with diagonal elements σ_w^2/w_i . The random effects u are also assumed independent from class to class, so G is also diagonal with diagonal elements equal to the so-called between variance σ_b^2 . Then the total variance matrix $V=ZGZ' + R$ is also diagonal with diagonal elements

$$V_i = \sigma_b^2 + \frac{\sigma_w^2}{w_i}. \quad (3.1)$$

We will not here address the estimation of the unknown within and between variances but treat them for present purposes as known. In fact, with only one observation per class, the within variance may not even be estimable. The reader should also note that by reference to “within” and “between variance” we have slipped into actuarial jargon; to my knowledge “within” and “between variance” are not common statistical terms.

The estimator for the grand mean becomes, with very little algebra, exploiting the many structural simplifications of this particular example,

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y = \left(\sum_i \frac{1}{\sigma_b^2 + \frac{\sigma_w^2}{w_i}} \right)^{-1} \sum_i \frac{y_i}{\sigma_b^2 + \frac{\sigma_w^2}{w_i}}. \quad (3.2)$$

Defining credibility as

$$z_i = \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_w^2}{w_i}} = \frac{w_i}{w_i + \frac{\sigma_w^2}{\sigma_b^2}}, \quad (3.3)$$

equation (3.2) becomes

$$\hat{\beta} = \frac{\sum_i z_i y_i}{\sum_i z_i}. \quad (3.4)$$

So the BLUE estimator of the fixed effects grand mean is none other than the credibility weighted average of the class means.

Turning next to the prediction of the random class effects, we already know by assumption that the unconditional expectations of the random effects vanishes, $E[u]=0$, and we know the total variance $\text{Var}[Y]=V$ is diagonal. We can also show in the present case that the covariance matrix $\text{Cov}[u,Y]$ is diagonal.

$$\begin{aligned} \text{Cov}[u_i, y_i] &= \text{Cov}[u_i, \beta + u_i + e_i] = \text{Var}[u_i] = \sigma_b^2. \\ \text{Cov}[u_i, y_j] &= \text{Cov}[u_i, \beta + u_j + e_j] = \text{Cov}[u_i, u_j] = 0. \end{aligned} \quad (3.5)$$

Collecting these results into the generic BLUP predictor of equation (2.3), the diagonality of the matrices on the right-hand side of (2.3) causes the matrix equation to collapse to a collection of scalar equations in which u_i depends only on y_i .

$$\begin{aligned}
 E[u_i | Y] &= E[u_i | y_i] = \text{Cov}[u_i, y_i] \text{Var}[y_i]^{-1} (y_i - \hat{\beta}) \\
 &= \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_w^2}{w_i}} (y_i - \hat{\beta}) \\
 &= z_i (y_i - \hat{\beta})
 \end{aligned}
 \tag{3.6}$$

So the posterior predictor of the random effect represents a shrinkage of the observed class mean towards the fixed effects grand mean by a factor that amounts to Buhlmann-Straub credibility. (For the reader needing a refresher on Buhlmann-Straub credibility, see chapter 4 of Goovaerts and Hoogstad (1987).) This might give us reason to hope that a similar result might hold at least approximately when LME is generalized to GLMM.

4. GENERALIZATION OF LME TO THE GENERALIZED LINEAR MIXED MODEL (GLMM)

For actuarial applications, the most restrictive assumptions in the LME model are that errors are normally distributed and that expected values are linear in explanatory variables. We already know how much power is gained by generalizing the classical linear model to the GLM and would hope for a similar gain in power on applying similar generalizations to the LME model.

- In the notation of the previous section, the conditional distribution $Y|u$ is now assumed to be in the exponential family rather than normal. Recall that the normal is a special case of the exponential family.
- Rather than assuming the conditional expectation linear in explanatory variables, we assume there is at least a link function g such that the g transform of the conditional expectation is linear: $g(E[Y|u]) = X\beta + Zu$. Note that the identity link is a special case of this assumption.
- We still assume u multivariate normal with mean 0, variance matrix G , and uncorrelated with the random measurement error $Y - E[Y|u]$.

The resulting model is the Generalized Linear Mixed Model (GLMM). Because the normal distribution is a special case of the exponential family, and the identity link is a special case of a more general link function, the LME model is a special case of the more general GLMM.

However, there are a number of important features of the LME model that do not carry over to the GLMM. One of them has to do with marginal distributions. In LME, the conditional distribution $Y|u$ is normally distributed. So is u . The marginal distribution Y , being the integration of $Y|u$ over u , is also normal. This does not always extend to the GLMM. $Y|u$ is distributed in the

exponential family. The random effect u is still assumed normally distributed. But the marginal distribution Y may not even be in the exponential family. Keep that in mind as you interpret GLMM output.

A second complication is that GLMM equations are not usually solvable in closed form. Instead, there are iterative solution algorithms, much as for the GLM. As a further consequence, there is no closed form algebra producing Buhlmann-Straub credibilities as we observed above in the LME case. But we can compare a GLM and a GLMM run on the same data and essentially the same model form to find evidence of shrinkage in the GLMM not present in the GLM. By plotting this shrinkage against measures of volume, we find evidence that the shrinkage is fit closely by credibility of Buhlmann-Straub form. This is demonstrated in the case study of the next section of this paper.

5. A CASE STUDY

5.1 Structure of the Problem

This case study is based on live, not simulated, ISO data. I have masked both line of business as well as names of potential explanatory variables to preserve ISO's intellectual property. The dependent variable being modeled is experience ratio, the ratio of observed losses to expected losses under the current rating plan, the latter denoted as ALCCCL. The data are not at the level of individual risks but rather aggregated into cells defined as crossings on all relevant explanatory variables, producing about 300,000 cells. As a consequence, some cells contain only a few risks; others contain thousands. One might therefore expect choice of proper weights to be important. Classical actuarial reasoning would lead us to expect, just as in the case of loss ratios, that the variance of experience ratios be inversely proportional to volume. In other words, the weights should be some measure of volume of business. I tested a number of possible weights and observed the best weight diagnostics when I used ALCCCL as weights.

Before delving further into detail, we might fruitfully give some brief thought to what it means to model on experience ratios as the dependent variable. The denominator of the experience ratio is in effect the current rating plan. If the current rating plan is entirely adequate, we would expect to see no statistically significant evidence of structure in the experience ratio across the explanatory variable space, so any statistically significant evidence of structure is evidence for changes to the current rating plan, and the model parameters indicate the degree of change.

I limited the case study model to four explanatory variables so as not to swamp the case study with too much detail. The variable of primary interest is a classification variable, CLASS1, with

twelve levels. Some of these levels are sparsely populated, and we will therefore want to treat CLASS1 as a random effect in a GLMM so as to shrink those levels towards the grand mean. There are two other classification variables, COVARIATE1 and COVARIATE2, each with four levels, which we will treat as fixed effects. There is also a continuous variable, COVARIATE3.

By default, SAS encodes classification variable effects in linear models (and their generalizations, such as GLM and GLMM) as contrasts between each level and the last level in the list for that variable. So, for each of CLASS1, COVARIATE1, and COVARIATE2, I have selected a well-populated level to serve as the base level for that variable and recoded it to “9” or “99” to force it to the end of the list of levels for that variable. This means that all contrasts will be expressed relative to stable bases.

Note that we are not modeling frequency and severity separately but rather their joint impact on experience ratio. We therefore need a distribution with positive mass at zero (to capture those cells with no loss) as well as a continuous density on the positive reals to capture cells with loss. It has recently become popular to model such cases with a Tweedie distribution with exponent p between 1 and 2. An exponent $p=1.67$ is a popular choice, and that is what I have chosen for the present case study.

Finally, we assume the ever-popular natural log link so as to yield a multiplicative model.

5.2 The GLM

The GLM model as summarized above was estimated via SAS PROC GENMOD. Further detail re code, etc. is deferred to Appendix A. The resulting parameter estimates from that model are summarized in the following Table 1.

Generalized Linear Mixed Models for Ratemaking

Table 1

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.2963	0.0329	-0.3608	-0.2317	80.96	<.0001
class1	01	1	0.3346	0.0413	0.2538	0.4155	65.78	<.0001
class1	02	1	0.2585	0.1674	-0.0696	0.5865	2.38	0.1225
class1	03	1	0.3056	0.0463	0.2149	0.3963	43.59	<.0001
class1	04	1	-0.1181	0.0774	-0.2697	0.0335	2.33	0.1267
class1	05	1	0.4388	0.1278	0.1882	0.6894	11.78	0.0006
class1	06	1	0.2196	0.0487	0.1242	0.3149	20.36	<.0001
class1	07	1	0.4695	0.0885	0.2960	0.6430	28.12	<.0001
class1	08	1	0.4268	0.0928	0.2449	0.6086	21.15	<.0001
class1	10	1	0.2978	0.0708	0.1591	0.4366	17.71	<.0001
class1	11	1	-0.1779	0.1198	-0.4126	0.0569	2.20	0.1376
class1	13	1	-0.0423	0.1865	-0.4077	0.3232	0.05	0.8207
class1	99	0	0.0000	0.0000	0.0000	0.0000	.	.
covariate1	2	1	0.0704	0.0367	-0.0014	0.1423	3.69	0.0548
covariate1	3	1	-0.0507	0.0840	-0.2152	0.1139	0.36	0.5460
covariate1	4	1	-0.3958	0.1005	-0.5927	-0.1989	15.52	<.0001
covariate1	9	0	0.0000	0.0000	0.0000	0.0000	.	.
covariate2	3	1	-0.1554	0.0531	-0.2595	-0.0514	8.57	0.0034
covariate2	4	1	-0.0778	0.0406	-0.1574	0.0019	3.66	0.0557
covariate2	5	1	0.0617	0.1423	-0.2172	0.3407	0.19	0.6646
covariate2	9	0	0.0000	0.0000	0.0000	0.0000	.	.
covariate3		1	-0.3652	0.0697	-0.5019	-0.2285	27.42	<.0001
Scale		0	670.2088	0.0000	670.2088	670.2088		

Note first that CLASS1 level 99, COVARIATE1 level 9, and COVARIATE2 level 9 all have 0 degrees of freedom (DF) and 0.0000 estimates and standard errors of those estimates, because these are the base levels for their respective classification variables, are therefore pegged at 0.0000, and all other levels are expressed as contrasts off these. Second, some of the CLASS1 levels have much larger standard errors than others. These are the poorly populated levels most in need of credibility treatment. More will be said about the Scale parameter in Appendix A.

5.3 The GLMM

Now it is desired to give CLASS1 a credibility treatment, so the GLMM model as summarized above was estimated via SAS PROC GLIMMIX, treating CLASS1 as a random effect. Again, further detail re code, etc. is deferred to Appendix A, but considerably more detail is provided for the GLMM relative to the GLM, given that the primary focus of this paper is on GLMMs.

Table 2 displays the resulting fixed effects parameter estimates and Table 3 the random effects parameter estimates, specifically for the CLASS1 variable.

Table 2

Solutions for Fixed Effects							
Effect	covariate1	covariate2	Estimate	Standard Error	DF	t Value	Pr > t
Intercept			-0.08888	0.06507	11	-1.37	0.1993
covariate1	2		0.07203	0.03666	312E3	1.96	0.0494
covariate1	3		-0.04689	0.08411	312E3	-0.56	0.5772
covariate1	4		-0.3965	0.1006	312E3	-3.94	<.0001
covariate1	9		0
covariate2		3	-0.1547	0.05310	312E3	-2.91	0.0036
covariate2		4	-0.07826	0.04070	312E3	-1.92	0.0545
covariate2		5	0.05959	0.1421	312E3	0.42	0.6749
covariate2		9	0
covariate3			-0.3643	0.06961	312E3	-5.23	<.0001

These estimates are quite similar to those from the GLM (compare Tables 1 and 2) with the exception of the intercept. This is because CLASS1 is treated as a fixed effect in the GLM, centered about its level 99, and is treated as a random effect in the GLMM, centered about a mean value of approximately 0. The different centering of CLASS1 between GLM and GLMM results in offsetting adjustments to the intercepts in the two models. Standard errors of the fixed effects are also quite similar between the two models.

Generalized Linear Mixed Models for Ratemaking

Table 3

Solution for Random Effects						
Effect	class1	Estimate	Std Err Pred	DF	t Value	Pr > t
class1	01	0.1241	0.06917	312E3	1.79	0.0729
class1	02	0.03040	0.1334	312E3	0.23	0.8198
class1	03	0.09508	0.07121	312E3	1.34	0.1818
class1	04	-0.2898	0.08770	312E3	-3.30	0.0010
class1	05	0.1674	0.1161	312E3	1.44	0.1492
class1	06	0.01150	0.07222	312E3	0.16	0.8735
class1	07	0.2229	0.09499	312E3	2.35	0.0190
class1	08	0.1836	0.09723	312E3	1.89	0.0589
class1	10	0.08142	0.08468	312E3	0.96	0.3363
class1	11	-0.2876	0.1106	312E3	-2.60	0.0093
class1	13	-0.1349	0.1391	312E3	-0.97	0.3322
class1	99	-0.2040	0.06847	312E3	-2.98	0.0029

Note that there is no preferred base level; a parameter is quoted for every level, and the parameters appear to be approximately mean zero.

5.4 Inferred Credibility

We now extract the CLASS1 credibilities implicit in the GLMM by comparing the CLASS1 parameter output from the GLMM to that from the GLM. We do this in the following Table 4.

Generalized Linear Mixed Models for Ratemaking

Table 4

		Inferred Credibility						
		Fixed Effects			Random Effects			(9)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Class1	ALCCL	Class1	exp(effect)	Class1	Class1	exp(effect)	Class1	Class1
		Effect		Relativity	Effect		Relativity	Credibility
01	484,185,185	0.3346	1.3974	1.1417	0.1241	1.1321	1.1399	0.9877
02	16,832,999	0.2585	1.2950	1.0580	0.0304	1.0309	1.0380	0.6545
03	359,748,011	0.3056	1.3574	1.1090	0.0951	1.0998	1.1073	0.9845
04	103,293,336	-0.1181	0.8886	0.7260	-0.2898	0.7484	0.7536	0.8994
05	27,864,645	0.4388	1.5508	1.2671	0.1674	1.1822	1.1904	0.7129
06	324,592,379	0.2196	1.2456	1.0176	0.0115	1.0116	1.0185	1.0506
07	60,941,612	0.4695	1.5992	1.3066	0.2229	1.2497	1.2583	0.8426
08	55,682,170	0.4268	1.5323	1.2519	0.1836	1.2015	1.2098	0.8328
10	108,633,028	0.2978	1.3469	1.1004	0.0814	1.0848	1.0923	0.9190
11	39,019,053	-0.1779	0.8370	0.6839	-0.2876	0.7501	0.7552	0.7742
13	15,101,361	-0.0423	0.9586	0.7832	-0.1349	0.8738	0.8798	0.5542
99	664,914,612	0.0000	1.0000	0.8170	-0.2040	0.8155	0.8211	0.9777
	2,260,808,391		1.2240	1.0000		0.9932	1.0000	

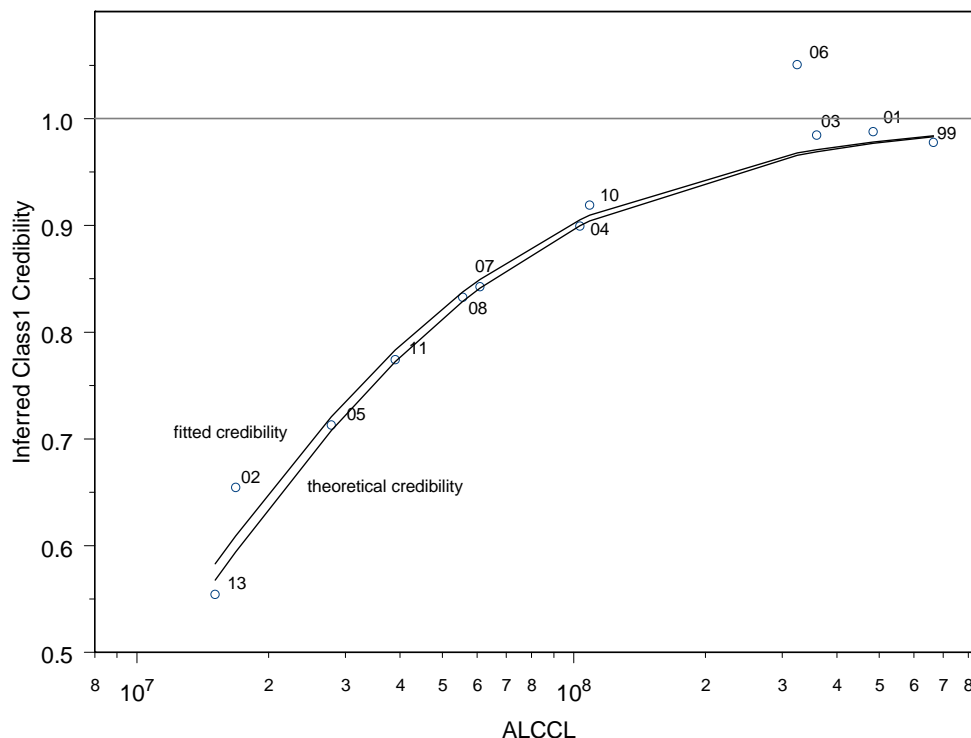
Fixed effect parameter estimates (from GLM) are tabulated in column (3), random effect parameter estimates (from GLMM) in column (6). But these parameters reside in the space of the linear predictor. To put them in the scale of the original observations, we invert the log link in columns (4) and (7). As already noted, the fixed effect parameters are expressed as contrasts to level 99, which was chosen for its volume, and hence stability, rather than for its being relatively centered among the levels. So we would not expect the mean of fixed parameters to be near 0, nor the mean of their exponentials to be near 1, and indeed they are not. Dividing the column (4) exponentials by their mean (weighted on ALCCL) produces relativities relative to a mean relativity of 1 in column (5). Due to the manner in which they were predicted, the random effect parameters are far closer to mean 0, but we still adjust column (7) to a mean relativity of 1 in column (8).

The column (5) and column (8) relativities are now directly comparable. If credibility is implicit in a GLMM, the column (8) random effect relativities should have shrunk towards 1 relative to the column (5) fixed effect relativities. Defining inferred credibility as column (9) = (column (8) - 1)/(column (5) - 1), the evidence is there. Furthermore, plotting these inferred credibilities against

ALCCL reveals evidence of declining credibility with declining volume. (See Figure 1. The two curves will be discussed in the next subsection of this paper.)

Figure 1

Inferred Credibility vs. ALCCL



One anomaly clearly stands out, CLASS1 level 06, for which the credibility is considerably in excess of 1.000. If we examine Table 4, we find that, for level 06, both fixed and random effect relativities are so close to 1.000 that even small errors or distortions in those relativities are magnified in the ratio that defines the inferred credibility. Among possible sources of error in the credibilities could be the fact that the renormalization of relativities to a mean of 1.0000 from column (3) to (5) and from column (6) to (8) in Table 4 introduces correlations among the CLASS1 level parameters not present in the output from the GLM and GLMM.

5.5 Inferred Credibility Is Approximately in Buhlmann-Straub Form

Are the above credibilities in approximately Buhlmann-Straub form

$$c = \frac{w}{w+k} \quad (5.1)$$

where c is credibility, w is a volume measure, and k is a constant? The trick is to determine k . If we assume Buhlmann-Straub form, then we can rework equation (5.1) into the following form:

$$\frac{1}{c} - 1 = \frac{k}{w} \quad (5.2)$$

This suggests we define a dependent variable equal to the reciprocal of our inferred credibilities minus 1 and regress this against an explanatory variable equal to reciprocal ALCCL in a regression through the origin (no intercept). The resulting regression coefficient would be our desired k . Applying this program to our Table 4 results, we find a k of 10.8 million (dollars). This regression is a simple, unweighted one. One could perhaps argue whether a weighted regression would be more appropriate, but this first approximation should suffice.

One can alternatively estimate k from certain parameters in the GLMM output. Appendix B derives a value of 11.5 million, in close agreement with the 10.8 million from the above regression.

Returning to expression (5.1) we substitute ALCCL for w and the two estimates for k , and plot the two resulting curves on Figure 1. The two curves are very close to each other and fit the inferred credibilities quite nicely. One may conclude that the implicit GLMM credibilities, at least for this case, are close to Buhlmann-Straub form.

6. SUMMARY AND CONCLUSIONS

GLMs signal, by quoting large standard errors, uncertain estimates for sparsely populated levels of classification explanatory variables, but they do not also adjust those estimates closer to the mean in response to low credibility. As a consequence, actuaries have desired for some time to introduce credibility into a GLM setting. There have been various attempts, both ad hoc and statistically rigorous, but none appear to have become popular, for reasons not always obvious.

The Generalized Linear Mixed Model (GLMM) provides yet another means of introducing credibility-like shrinkage into a GLM setting. Recently available statistical software, including SAS PROC GLIMMIX as well as new R and S-Plus functions, brings these models within reach of

actuaries.

This paper first introduced the reader to the Linear Mixed Effects (LME) model, a simpler cousin of the GLMM, as a means of introducing issues important for GLMMs but in a less complex environment. It was shown how Buhlmann-Straub credibility falls directly out of the LME math, at least for a simple case. The LME was then generalized to the GLMM, and a case study demonstrated how to use GLMM software and showed that the GLMM preserved shrinkage to the mean in a form at least approximating Buhlmann-Straub credibility.

It is hoped that this paper will give actuaries sufficient knowledge, incentive, and courage to experiment with GLMMs in their next GLM project. New software, such as SAS PROC GLIMMIX, provides them the means to do this.

Acknowledgment

The author wishes to thank a number of reviewers for their helpful suggestions, especially those that resulted in tightening the argument and focusing in on the most important points of the paper.

Appendix A: SAS Implementation of the GLM and GLMM of the Case Study: Additional Detail

I have included this appendix for those readers who would like more detail on how to implement the GLMM of the case study in at least one stat package. This is not to imply that the SAS implementation of GLMMs is better than others, only that SAS is the package I chose. There are implementations of GLMM in R and S-Plus as well as other stat packages.

The reader should recall the model basics. The dependent variable is EXPRATIO, or experience ratio, assumed Tweedie distributed with exponent p equal to 1.67. Explanatory variables are classification variables CLASS1, COVARIATE1, and COVARIATE2, as well as the continuous variable COVARIATE3. CLASS1 is treated as a fixed effect in the GLM and a random effect in the GLMM. All other explanatory variables are treated as fixed effects in both the GLM and the GLMM. The regressions are weighted on ALCCL, a measure of business volume in each record. The link is log.

A.1 The GLM

The SAS code for the GLM is as follows:

```
PROC GENMOD DATA=INDATA;  
  P=1.67;  
  Y=_RESP_;  
  A=_MEAN_;  
  VARIANCE BAR=A**P;  
  DEVIANCE DEV=2*((Y**(2-P)-Y*A**(1-P))/(1-P)-(Y**(2-P)-A**(2-P))/(2-P));  
  CLASS CLASS1 COVARIATE1 COVARIATE2;  
  WEIGHT ALCL;  
  MODEL EXPRATIO= CLASS1 COVARIATE1 COVARIATE2 COVARIATE3/  
    LINK=LOG SCALE=PEARSON;  
RUN;
```

SAS PROC GENMOD does not naturally support the Tweedie distribution, but it does support a facility to allow users to specify their own distributions (by specifying both a variance and a deviance function for their distribution of choice). Lines 2 through 7 of the above code are what specify the Tweedie. The SCALE=PEARSON option in the MODEL statement is also important. The variance law for the Tweedie (which specifies the functional form of the observation variances) is $\text{Var}[y]=\varphi\mu^p/w$, where y is the observation, μ is the fitted value, p is the Tweedie exponent, w is the weight, and φ is the so-called dispersion coefficient. We are here telling the GLM to use a Pearson chi-squared estimator of the dispersion coefficient rather than assuming it equal to 1. The dispersion coefficient is fundamental, because it is the basis for the standard error estimates of the GLM coefficients. The Scale parameter of Table 1 of this paper is the square root of the estimated dispersion coefficient, and, at 670, is certainly quite far from 1.

A.2 The GLMM

Now we want to give CLASS1 a credibility treatment. The following code fits a GLMM to the same data to which we previously fit a GLM, and with as much of the same model form as before as possible, with the exception that CLASS1 is now treated as a random effect.

```
PROC GLIMMIX MAXOPT=50 PCONV=.000015 DATA=INDATA;  
  _VARIANCE_=_mu_**1.67;  
  CLASS CLASS1 COVARIATE1 COVARIATE2;  
  WEIGHT ALCL;  
  MODEL EXPRATIO= COVARIATE1 COVARIATE2 COVARIATE3/ LINK=LOG SOLUTION;  
  RANDOM CLASS1/ SOLUTION;  
  RANDOM _RESIDUAL_;  
RUN;
```

Because a secondary purpose of this paper is to convince the reader that GLMMs are now within reach of actuaries via currently available software, we will spend more time on this code and its

Casualty Actuarial Society *E-Forum*, Winter 2011-Volume 2 20

resulting output than we did on the prior GLM. First, technically, GLIMMIX doesn't fit a maximum likelihood but rather a maximum pseudo-likelihood. This means that, although you still need to specify the variance law of the Tweedie distribution (see line 2 of the code), you do not also need to specify a deviance function. Had you been interested in one of the distributions supported by GLIMMIX rather than the user-defined Tweedie, just as in GENMOD you would have specified that distribution via a `DIST=` option in the `MODEL` statement.

The `MODEL` statement specifies the fixed effects part of the model (and an option to the `MODEL` statement specifies the optional R side of the variance model). The random effects, which determine the G side of the variance model, are specified by the `RANDOM` statements. `PROC GENMOD` automatically gives you tables of parameter estimates and their standard errors, but, if you want those from `PROC GLIMMIX` as well, you have to ask for them via the `SOLUTION` options in the `MODEL` statement (for the fixed effects regression parameters) and in the `RANDOM` statement (for random effects regression parameters).

The `RANDOM _RESIDUAL_` statement is crucial. `GLIMMIX` estimates a dispersion coefficient only for non-user-defined distributions, and even then only for those with a dispersion coefficient in their definition; otherwise, `GLIMMIX` pegs the dispersion coefficient at 1.000 by default. The `RANDOM _RESIDUAL_` statement is the way in which you force `GLIMMIX` to still estimate the dispersion coefficient for user-defined distributions.

Lastly, note the `MAXOPT` and `PCONV` options in the `PROC GLIMMIX` statement. By default, `GLIMMIX` attempts 20 iterations of a certain outer iteration (the fact that there is also an inner iteration will be noted momentarily) before giving up. Furthermore, it determines model convergence when the percentage change of certain parameters from one iteration to the next is less than about 10^{-8} . The case study data was sufficiently volatile that the algorithm was never able to attain this high standard, but, by examining iteration history details, it was found that convergence could be achieved with a slightly relaxed standard of 50 iterations specified by the `MAXOPT` option and a convergence criteria of 1.5×10^{-5} specified by the `PCONV` option.

Before discussing `GLIMMIX` output, I will sketch an outline of what `GLIMMIX` is actually doing when it estimates the model, as this will aid interpretation of subsequent output. Unlike other familiar SAS model-building `PROC`s, `GLIMMIX` does not build models on the original data but rather on pseudo-data. At the beginning of each iteration, it constructs new pseudo-data by linearizing the original data about the expected values from the prior iteration. It then maximizes the pseudo-likelihood on that pseudo-data. It should also be noted that the iteration referred to

here is the outer iteration.

The algorithm doesn't even solve simultaneously for the variance components (the unknown parameters of the total variance matrix V) and the fixed and random effects parameters. Rather, it starts the iteration with a pseudo-likelihood that is a function of variance components, fixed effects, and the dispersion coefficient and is able to adjust out ("profile" out) the fixed effects and dispersion coefficient to produce an objective function that is a function of just the variance components. It then enters an inner iteration to optimize this modified objective function over just the variance components. Armed with estimates of the variance components from the inner iteration it then estimates fixed effects and predicts random effects, then returns to the outer iteration for another pass through, starting with producing the next pseudo-data set, and so on until convergence. Although we will not examine an iteration history table output by our GLIMMIX run, if you were to examine such a table, you would note reference to iterations, restarts, and subiterations, which hints at the structure of inner iterations (subiterations) nested within outer iterations mentioned above.

Turning to the output of the above SAS PROC GLIMMIX code, following a first table that summarizes the dataset, the dependent variable, the assumed distribution, the link function, the weights, and a few other model assumptions, there are tables of additional model dimensions, shown as Tables 5 and 6, that are highly useful for checking that the model estimated is the one the user intended to estimate, and that there hasn't been some misinterpretation through some syntax error.

Table 5

Dimensions	
G-side Cov. Parameters	1
R-side Cov. Parameters	1
Columns in X	10
Columns in Z	12
Subjects (Blocks in V)	1
Max Obs per Subject	312131

Table 5 is crucial for verifying that GLIMMIX correctly interpreted our model specification. We did indeed want one G side parameter to be estimated, the between variance for the CLASS1 random effect. We did indeed want one R side variance parameter, the dispersion coefficient. There should indeed be ten columns in the fixed effects design matrix, one for the intercept term,

four for each of COVARIATES 1 and 2, and one for COVARIATE 3. There should indeed be twelve columns in the random effects design matrix because CLASS1 has twelve levels.

Table 6

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	1
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Table 6 reminds us that the inner iteration does indeed profile out both fixed effects and the dispersion coefficient (Residual Variance), and optimization is only over the remaining G side and R side variance components, in this case, over the single unknown parameter of the between variance of the CLASS1 random effect. Hence the optimization is only over one parameter. Because this parameter represents a variance, it is bounded below by zero, hence the reference to one lower boundary. But it is unbounded above.

Table 7 displays the resulting estimated variance component parameters.

Table 7

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
class1	0.04050	0.02095
Residual (VC)	449268	1137.28

As already noted, only two parameters were requested, the between variance of the CLASS1 random effect and the dispersion coefficient (Residual). The dispersion coefficient is in close agreement with that from the previous GLM. (Recall that the GLM dispersion coefficient is the square of the Scale parameter in Table 1: $670^2 = 448,900$.) The CLASS1 between variance is about .04. Its square root of about .2 (interpretable as approximately 20% because of the log link and because the random effects reside in the space of the linear predictor, in other words, in the logged space) indicates that levels of the CLASS1 variable fall a few tens of percent above and below the grand mean. This seems reasonable.

Following these tables presented above the SAS output provides fixed effects and random effects parameters already presented and discussed as Tables 2 and 3 of this paper. The reader seeking further detail is referred to the SAS PROC GLIMMIX online manual.

Appendix B: Inferring a Buhlmann-Straub k Parameter from GLIMMIX Output

Buhlmann-Straub credibility is of the form of equation (5.1). The k parameter in that equation is frequently written as a ratio of within variance to between variance, which is how it appears in equation (3.3). Can we read from our GLIMMIX output the numbers we would need to estimate within and between variance, and hence the k parameter? Yes, but the reader is forewarned that the following is not a strict derivation but rather a plausibility argument. It should be enough to support the approximate magnitude of the k parameter but not its precise value.

First, if one studies derivations of Buhlmann-Straub credibility (see chapter 4 of Goovaerts and Hoogstad (1987)), one finds that what is referred to as the within variance is actually the proportionality constant in the relationship: observation variance proportional to reciprocal weights. Recall the Tweedie variance law: $\text{Var}[y]=\varphi\mu^p/w$, where φ is the dispersion coefficient, μ the expected value of y , p the Tweedie exponent, and w the weight. Strictly speaking, the numerator of this law is not a constant because μ is not, being a function of explanatory variables. Nevertheless, it might be reasonable to equate the within variance in the Buhlmann-Straub k to $\varphi\langle\mu\rangle^p$, where $\langle\mu\rangle$ is mean expectation.

Next, the variance component indicated by GLIMMIX for the classification variable in question is almost the desired between variance, except that it is measured in the space of the linear predictor, where the random effects live, and not in the original scale of the observations, as needed for the Buhlmann-Straub k . We just have to back-transform via the inverse link function, invoking the following approximation: $\text{Var}[g(x)]\approx g'(E[x])^2\text{Var}[x]$. In other words, the variance of the g transform of x is approximately the variance of x times the square of the derivative of g evaluated at mean x . Here, x is the linear predictor, $\text{Var}[x]$ is the variance component from GLIMMIX, g is the inverse log link, in other words, the exponential. Its first derivative is also the exponential. $E[x]$ is the mean linear predictor. The exponential (the inverse link) of the mean linear predictor is approximately $\langle\mu\rangle$. Then the between variance we seek is approximately $\langle\mu\rangle^2$ times the variance component.

Now, we assumed the Tweedie exponent p to be 1.67. The GLIMMIX output tells us the dispersion coefficient φ is 449,000 and the CLASS1 variance component is .0405. GLIMMIX doesn't tell us, but we know from other checks of our dataset that $\langle\mu\rangle$ is approximately .9, in which

case,

$$\begin{aligned}
 k &= \frac{\text{within variance}}{\text{between variance}} \approx \frac{\varphi \langle \mu \rangle^p}{\langle \mu \rangle^2 \text{ variance component}} \\
 &= \frac{\varphi}{\langle \mu \rangle^{2-p} \text{ variance component}} \approx \frac{449,000}{(.9)^{2-1.67} (.0405)} \approx 11.5 \text{ million.}
 \end{aligned}$$

7. REFERENCES

- [1] Anderson, Duncan, Sholom Feldblum, Claudine Modlin, Doris Schirmacher, Ernesto Schirmacher, and Neeza Thandi, "A Practitioner's Guide to Generalized Linear Models," *CAS Discussion Paper Program*, 2004, pp. 1-116.
- [2] de Jong, Piet, and Gillian Heller, *Generalized Linear Models for Insurance Data*, Cambridge, U.K.: Cambridge University Press, 2008.
- [3] Goovaerts, M. J., and W. J. Hoogstad, *Credibility Theory*, Survey of Actuarial Studies No. 4, Nationale-Nederlanden N. V., 1987.
- [4] Guszcz, James C., "Hierarchical Growth Curve Models for Loss Reserving," *Casualty Actuarial Society E-Forum*, Fall 2008, pp. 146-173.
- [5] Littell, Ramon C., George A. Milliken, Walter W. Stroup, Russell D. Wolfinger, and Oliver Schabenberger, *SAS for Mixed Models*, 2nd ed., Cary, NC: SAS Institute, 2006.
- [6] McCullagh, P., and J. A. Nelder, *Generalized Linear Models*, 2nd ed., Chapman & Hall/CRC, 1989.
- [7] Nelder, J. A., and Richard Verrall, "Credibility Theory and Generalized Linear Models," *ASTIN Bulletin* 27(1), 1997, pp. 71-82.
- [8] Ohlsson, E., and B. Johansson, "Combining Credibility and GLM for Rating of Multi-Level Factors," *CAS Discussion Paper Program*, 2004, pp. 315-326.
- [9] Ohlsson, E., and B. Johansson, "Exact Credibility and Tweedie Models," *ASTIN Bulletin* 36(1), 2006, pp. 121-133.
- [10] Ohlsson, E., "Credibility Estimators in Multiplicative Models," Research Report 2006:3, Mathematical Statistics Stockholm University.

Author's Biography

Fred Klinker (FCAS, MAAA) is an actuarial manager with ISO's Modeling Division. He is involved in a number of collaborative projects with ISO's other major predictive modeling unit, ISO Innovative Analytics, as well as various other efforts to introduce modern predictive modeling techniques, especially GLMs, into standard ISO ratemaking. Much of his career prior to ISO was spent as a research actuary involved in financial and statistical modeling at CIGNA P&C, with brief forays into regulation at the Massachusetts Insurance Department and reinsurance at Zurich Re North America, later Converium.

Reserving in the Age of Obesity

Chris Laws and Frank Schmid

Abstract

Motivation. There is increasing evidence that obesity contributes to the cost of medical care in workers compensation, and that this contribution is significant in magnitude. For instance, a recent study of workers compensation claims of Duke University employees shows that, for the morbidly obese, the medical costs per 100 full-time equivalent employees are nearly seven times as high as for employees of recommended weight. In the following study, the evidence of the contribution of obesity to the medical costs of workers compensation is generalized to a set of claims that comprises 36 U.S. States and nine injury years. Further, it is shown how the cost difference between “obese claims” and comparable “non-obese claims” develops as claims mature—this evidence of the difference in development offers important guidance for both reserving and ratemaking. The study is confined to the effect of obesity on severity—the effect of obesity on claim frequency is beyond the scope of this analysis.

Method. The study makes use of a matched-pairs research framework. Every obese claim in the data set is matched with a non-obese claim. Exact matching applies to all claim characteristics, except age at injury, where proximity matching is employed. The set of matched pairs is then analyzed using a semiparametric Bayesian multilevel model, the nonparametric component of which accounts for the possible nonlinear influence of age. Aside from age, the covariates comprise the injury year, the nature of injury, the U.S. state, and the industry—these four covariates enter the model as random effects. Further, the gender of the claimant and cross-state differences in the legislative environment, as they manifest themselves in mandatory utilization review and mandatory bill review, are accounted for using indicator variables. The model is estimated by means of MCMC (Markov Chain Monte Carlo simulation). The reversible jump concept of Bayesian modeling averaging is used in determining the functional form of the nonparametric component that captures the influence of age.

Results. The study shows that, in the aggregate, obese claims are 2.8 times more expensive than non-obese claims at the 12-month maturity, but this cost difference climbs to a factor of 4.5 at the three-year maturity and to 5.3 at the five-year maturity. Further, the cost difference (at the five-year maturity) is less for females than for males. Mandatory utilization review and, in particular, mandatory bill review significantly reduce the cost difference between obese and non-obese claims.

Availability. The semiparametric multilevel model was estimated using JAGS with R. JAGS (Just Another Gibbs Sampler, <http://www-ice.iarc.fr/~martyn/software/jags/>) is an open-source platform for Gibbs sampling, developed by Martyn Plummer at the International Agency for Research on Cancer of the World Health Organization in Lyon, France. The reversible jump routine was written as a C++ JAGS module. R is an open-source statistical modeling platform (<http://www.r-project.org/>), which is administered by the Technical University of Vienna.

Keywords. Obesity, Multilevel Model, Partial Linear Model, Reversible Jump MCMC, Semiparametric Model, Workers Compensation

1. INTRODUCTION

In July 2009, the Centers for Disease Control and Prevention (CDC) held its inaugural conference “Weight of the Nation,” thus calling to public attention the mounting problem that obesity poses for the health of the American people. A research paper presented at this conference by Finkelstein et al. [4] called the “link between rising rates of obesity and rising medical spending” nothing short of “undeniable.” These authors estimate that obesity accounts for “almost 10 percent of all medical spending and could amount to \$147 billion per year in 2008.” This is about twice the

Reserving in the Age of Obesity

cost estimate of \$78.5 billion that Finkelstein, Fiebelkorn, and Wang [3] had established for the year 1998 in an earlier study.

Obesity may also have significant implications for workers compensation. In a recent study of Duke University employees, Truls, Dement, and Krause [9] found that employees in the highest obesity class, when compared with employees of recommended weight on an FTE (full-time equivalent) basis, filed twice as many claims, had 13 times as many lost workdays, and experienced medical and indemnity costs that were 7 and 11 times as high, respectively.

The Duke University study aimed at estimating the differences in medical and indemnity costs between obese and non-obese employees. To this end, the cumulative payments for every claim were tallied at the end of the study, which is December 31, 2004; for open claims, which amount to 2.8 percent of the total number of claims, estimated reserves were used that had been provided by the competent workers compensation actuary—these details have been confirmed in writing by the corresponding author of the Duke University study, Dr. Truls Østbye. Thus, for the purpose of arriving at total costs, the Duke University study does not rest entirely on observed payments but (for open claims) also includes reserve estimates.

In what follows, we take a different approach than the one pursued by Truls, Dement, and Krause [9]. For one, the nature of our data set is quite different, as will be discussed. But most importantly, it is not our objective to provide a measurement for the difference in ultimate costs between obese and non-obese claimants—such a measurement would have to make use of reserve estimates for open (and potentially re-opening) claims. Instead, we try to shed light on the difference in development between claims from obese and non-obese employees; this way, we are able to provide guidance (on a per claim basis) on the divergence in cumulative payments between these two types of claims for reserving (and ratemaking) purposes.

Because of potential dissimilarities in development between claims of obese and non-obese employees, the difference in the costs per claim may not be apparent early on (e.g., at the 12-month maturity) but may take time to reveal itself. As is shown in this study, the ratio in the medical costs per claim of obese to non-obese claimants indeed develops; whereas this ratio stands at 2.8 at the 12-month maturity, it climbs to 4.5 at the 36-month maturity, and to 5.3 at the 60-month maturity. A possible reason for such dissimilarity in development may be the longer duration of obese claims; another reason may be that the distribution of medical costs, as claims develop, differs between obese and non-obese claims as obesity may raise the likelihood of medical complications. Due to data limitations, no definite statement can be made as to why the ratio of medical costs between

obese and non-obese claims increases over time. Future research on differences in claim duration between obese and non-obese claims may provide an answer to this question.

1.1 Research Context

The Duke University study by Truls, Dement, and Krause [9] is to date the only comprehensive statistical analysis of the effect of obesity on the cost of workers compensation. This study makes use of a longitudinal data set, which was obtained by monitoring a cohort of 11,728 employees of Duke University and the Duke University Health System from January 1, 1997, through December 31, 2004. The cohort was defined by all employees that had at least one HRA (health risk assessment) during this time period; taking an HRA is voluntary and available to employees eligible for health care benefits. (Note that the number of members in the study may have shrunk over time due to employee termination or disability.) The members of this cohort were assigned to body mass index (BMI) categories based on the first HRA they participated in during the time of the study.

At the end of the eight-year time window, the number of claims, the number of work days, and the indemnity and medical costs were tallied for each employee; then, this information was matched up with the BMI category (and other characteristics) of the claimant. There are six BMI categories, ranging from underweight to recommended weight, overweight, and three classes of obesity. The highest level of obesity is class III, which comprises the morbidly obese, identified by a BMI of 40 or higher. The Duke University study finds that for the morbidly obese employees, the medical costs are 6.8 times as high as for employees of recommended weight; at the same time, an employee in this group is twice as likely to have a claim. For obese classes II (BMI of at least 35 but less than 40) and I (BMI of at least 30 but less than 35), the medical costs per employee are (respectively) 3.1 and 2.6 times as high as for employees of recommended weight; the respective multiples for the number of claims read 1.9 and 1.5. (The numbers cited above rest on the bivariate analysis presented in Truls, Dement, and Krause [9], Table 3.)

Another way of presenting the findings of the Duke University study is on a per claim basis. Transforming the medical costs per 100 FTE employees into costs per claim shows that this amount is 3.4 (obesity class III), 1.7 (obesity class II), and 1.7 (obesity class I) times the magnitude recorded for employees of normal weight. The corresponding numbers for indemnity read 5.5 (obesity class III), 3.4 (obesity class II), and 2.9 (obesity class I). As is apparent in the data, on a per claim basis, the percentage difference between obese employees and employees of normal weight is even higher for indemnity payments than it is for medical costs. (Here, too, the numbers rest on the bivariate analysis presented in Table 3 of Truls, Dement, and Krause [9].)

1.2 Objective

The effect of obesity on the medical cost per workers compensation claim is quantified at different maturities, thus showing how the (percentage) cost difference between obese and non-obese claims develops as claims mature. Further, it is shown how the dissimilarity in development between obese and non-obese claims varies with the legislative environment, as such manifests itself in mandatory utilization review (MUR) and mandatory bill review (MBR). Quantifying differences in development patterns across obese and non-obese claims is important for reserving and ratemaking in workers compensation.

1.3 Outline

What follows is an account of how the data set was prepared, followed by descriptive statistics. Section 3 then offers a discussion of the Bayesian semiparametric multilevel model, which is followed in Section 4 by a presentation of the findings for the random effects (injury year, U.S. state, industry, and nature of injury), gender, and age. Section 5 presents estimates of the effects of the legislative environment (MUR and MBR). Section 6 concludes.

2. THE DATA

We use a large sample of workers compensation claims provided by select insurance companies. The data base comprises records from 36 states (AK, AL, AR, AZ, CO, CT, DC, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MD, ME, MO, MS, MT, NC, NE, NH, NM, NV, OK, OR, RI, SC, SD, TN, UT, VA, and VT) and nine injury years (1998-2006). Claims are studied at three different maturities: 12 months, 36 months, and 60 months. Clearly, not all maturities are available for all injury years. For instance, observations at the 36-month and 60-month maturities are available only up to injury years 2004 and 2002, respectively.

For each claim, cumulative medical payments are tallied at each of the chosen maturity dates. The observed cumulative payment at a given maturity is flagged as pertaining to an obese claimant if, at that maturity, there has been a record of a co-morbidity indicator pointing to obesity. Specifically, an observation is flagged as “obese” if the three leading digits of an ICD-9 code serving as a co-morbidity indicator equal 278. We refer to such an observation as an “obese claim.”

It is worth noting that a claim that qualifies as obese at the 36-month maturity may not qualify as obese at the 12-month maturity if the co-morbidity indicator 278 does not show up before the 12-month maturity (but shows up no later than the 36-month maturity). On the other hand, such a

Reserving in the Age of Obesity

claim does not qualify as non-obese at the 60-month maturity, even if the co-morbidity indicator 278 fails to show up again after the 36-month maturity. In summary, a claimant is treated as obese only after actually having been diagnosed as such. There may be instances where a claimant turns obese after the injury; and there may be instances where a claimant is diagnosed as obese only after the claimant's being obese has given rise to medical complications.

For the purpose of proper cost comparison between obese and non-obese observations, claims are dropped once a lump sum payment has been observed among the transactions. For instance, a claim may be included at the 12- and 36-month maturities, but excluded at the 60-month maturity if a lump sum payment was observed later than 36 months (but no later than 60 months) into the duration of the claim. The reason for excluding claims following a lump sum payment is that such transactions represent the present value of a stream of payments that may extend past the maturity date of interest.

The data base offers no information on the BMI, which is a standard measure of obesity (see Truls, Dement, and Krause [9]). As a result, this study does not differentiate between degrees of obesity. At the same time, it can be assumed that the co-morbidity indication identifies the claimant as severely obese, which puts him into one of the higher obesity classes. The small proportion of obese observations in the total number of claims supports this conjecture; for instance, at the 12-month maturity, the proportion of obese observations ranges between 0.1 percent (injury years 1998–2003) and 0.2 percent (injury years 2004–2006). By comparison, the proportion of morbidly obese claimants in the mentioned study of Duke University employees amounts to 4.9 percent.

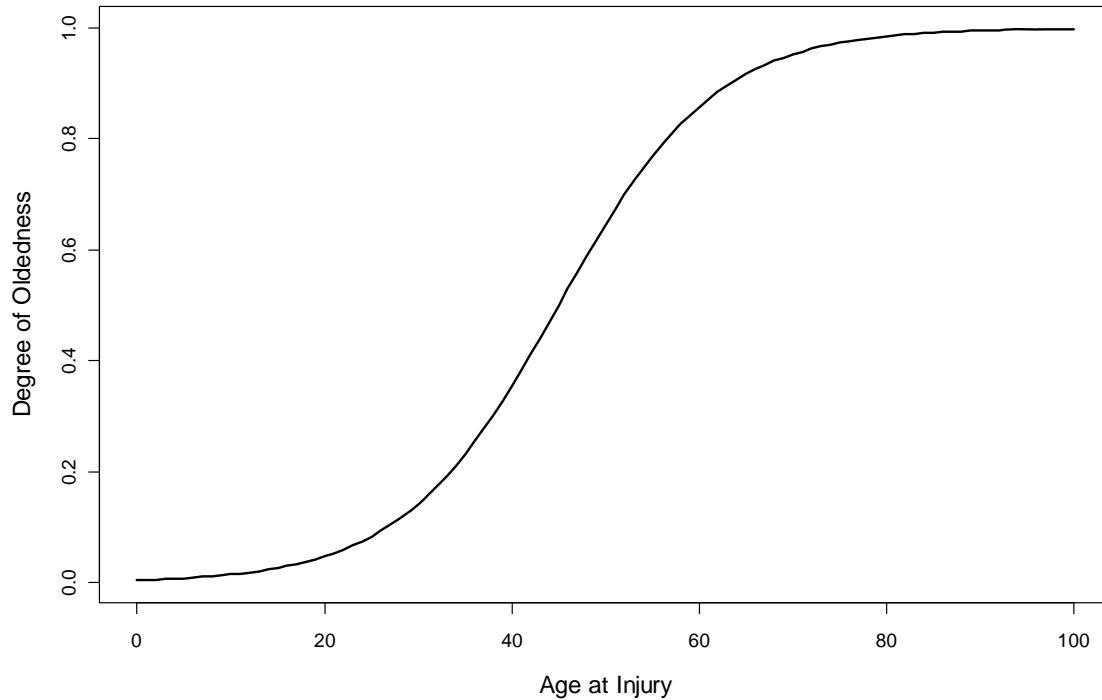
The small percentage of obese claims in the total number of claims, along with the very high total number of claims (3,834,891 claims for the 12-month maturity; 2,956,285 claims for 36 months; and 2,079,225 claims for 60 months) is most suited for a matched-pairs research framework. In such an analytical setting, each obese observation in the data set is matched with a non-obese claim of the same maturity based on injury year, U.S. state, industry, ICD-9 code, gender, and age at injury. Except for age at injury, all matching criteria are categorical in nature, which allows for exact matching (thus obviating the need for propensity matching). At the same time, extending the concept of exact matching to age at injury (which is measured on a scale of whole numbers) leaves many obese claims without matches. Thus, for age at injury, we use proximity matching.

Proximity matching rests on the concept of the nearest neighbor. When exact matching for age is not feasible (for lack of exact matches when using whole numbers for years of age), researchers often resort to matching by age brackets. In matching by age bracket, an obese claim is paired up

Reserving in the Age of Obesity

with a non-obese claim that belongs to the same (for instance) five-year age bracket (subject to being identical in all exact-matching characteristics). A disadvantage of matching by age bracket is that many claims are not matched with the closest neighbor. For instance, a 20-year-old obese claimant may be matched (within the 20-24 age bracket) with a 24-year-old non-obese claimant, but is prevented from being matched with an otherwise identical 19-year-old claimant. Proximity matching avoids this problem by looking for the nearest neighbor. At the same time, it may not be appropriate to have the concept of the nearest neighbor rest on the simple age difference. Because aging is a nonlinear process, it may be preferable to match an obese 25-year-old claimant with a 35-year-old, instead of a 16-year-old. Similarly, matching an obese 55-year-old with a non-obese 35-year-old may be more appropriate than matching this person with a non-obese 74-year-old. For this reason, we use a sigmoid function to create a fuzzy set for old age; the sigmoid function has the form $1/(1+\exp(-\sigma \cdot (h-45)))$, where h is the age at injury and σ was chosen to be equal to 0.12. (For the concept of fuzzy sets see, for instance, Kasabov [7].) Chart 1 shows the fuzzy set for old age (and its complement, young age); the degree of oldness is set to 50 percent at age 45 and then, in an “S-shaped” manner, this degree of oldness approaches zero and 100 percent as the years of age approach zero and 100, respectively. The nearest neighbor of an obese claim to an otherwise identical non-obese claim is defined by the smallest difference (in absolute value terms) in the degree of oldness.

Chart 1: Degree of Oldness, Defined by Sigmoid Function



To summarize, we define as neighbors to a given obese claim the set of non-obese claims that match exactly based on maturity, injury year, ICD-9 code, U.S. state, industry, and gender. In this matching process, a given non-obese claim may be used as a neighbor to more than one obese claim. Among the thus identified set of neighbors, the nearest neighbor is chosen based on the degree of oldness. This nearest neighbor may not be unique because of ties in the oldness distance, in which case we are left with a set of “tying neighbors.” The highest number of tying neighbors (across all maturities) for a given obese claim equals 47; the percentage of obese claims with a single nearest neighbor equals 59.5.

Table 1 displays by maturity and injury year the number of obese claims and the total number of their (potentially tying) nearest neighbors within these categories. In the statistical analysis, for each obese observation, only one of the tying nearest neighbors is chosen, at random. We create multiple (50) data sets, each with random tie-breaking where obese observations have non-unique nearest neighbors. By analyzing these multiple data sets, we are able to provide credible intervals around the parameter estimates that reflect the uncertainty originating in the randomness of the tie-breaking

Reserving in the Age of Obesity

process. (For details on credible intervals, see Carlin and Lewis **Error! Reference source not found.**)

In Table 1, there are two reasons why the number of obese claims (and hence the number of non-obese neighbors) varies by maturity. First, claims are dropped following lump sum payments, as mentioned. Second, claims are added upon obesity showing up as a co-morbidity indicator in a transaction.

Table 1: Numbers of Obese Claims and (Potentially Tying) Non-Obese Nearest Neighbors

Injury Year	Obese	Non-Obese	Obese	Non-Obese	Obese	Non-Obese
	12 months		Maturity 36 months		60 months	
1998	250	537	266	518	271	529
1999	282	565	313	573	304	545
2000	343	802	365	831	364	831
2001	417	1,058	430	1,018	413	953
2002	467	1,108	481	1,106	459	1,049
2003	553	1,290	517	1,157	—	—
2004	630	1,464	656	1,414	—	—
2005	722	1,514	—	—	—	—
2006	836	1,744	—	—	—	—

The concepts of matching by age bracket and matching by proximity of age share the characteristic of matching toward the center of the age distribution. These matching techniques, unlike exact matching (which, for instance, pairs up a 55-year-old obese claimant with a 55-year-old non-obese claimant when using whole numbers), tend to pair up old claimants with claimants of lesser age, and young claimants with claimants of higher age; this is because the claimants on the edges of the age distribution turns are sparse. For instance, within the age bracket 60-64, the number of observations tends to decline with age. Thus, a claimant at the center of this age bracket has a comparatively high chance of being “matched down the age distribution,” as opposed to being “matched up;” this is simply because the number of potential matches within this bracket is higher at lower ages than at higher ages. Conversely, young claimants tend to be “matched up the age distribution.” Note that this matching toward the center (of the age distribution) is not unique to proximity matching but is also characteristic of the traditional approach of matching by age bracket.

Reserving in the Age of Obesity

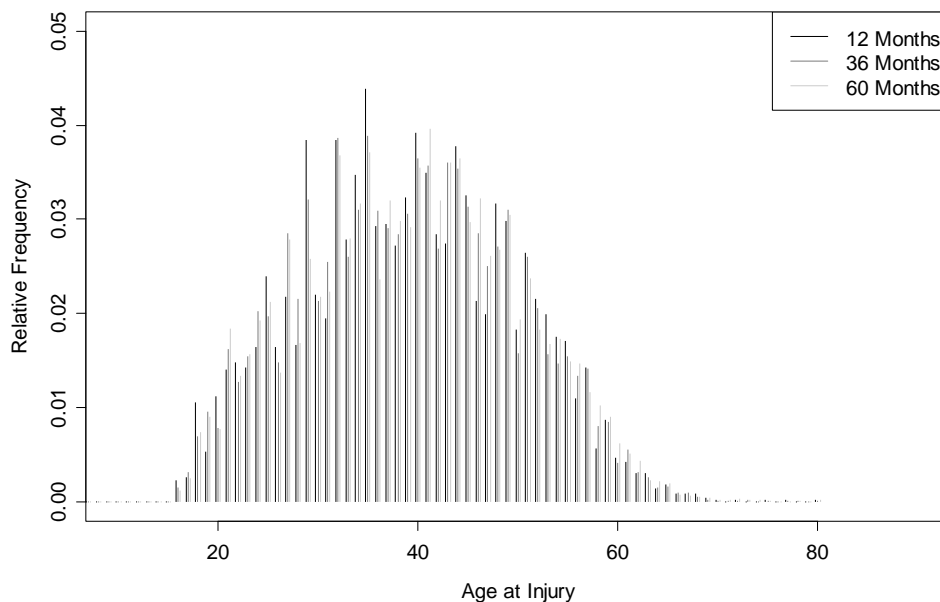
Matching toward the center of the age distribution poses no difficulty for the statistical analysis (because age is included as a covariate in a nonparametric setting, which accommodates potential nonlinearities), but it does affect the interpretation of the regression results for age (and age alone). For instance, if the medical costs of workers compensation claims increase with age, then the influence of age on the cost of obesity is underestimated for young claimants and overestimated for old claimants. As a consequence, the estimated effect of age may be distorted on the edges of the age distribution, taking on an “S-shaped” form. For this reason, the nonparametric regression finding for age has to be interpreted with care.

Exact matching by maturity, injury year, U.S. state, ICD-9 code, and gender is straightforward. Matching by industry relies on an NCCI industry classification; the five industries comprise Manufacturing, Contracting, Office and Clerical, Goods and Services, and Miscellaneous.

There are a total of 1,560 obese observations (or 14.3 percent of the total number of obese observations) going unmatched, because no match is available by maturity, injury year, ICD-9 code, U.S. state, industry, and gender.

Chart 2 offers a distribution of age at injury of the set of studied (obese and non-obese) claims for all three maturities; the observations are pooled over the nine injury years.

Chart 2: Relative Frequency of Claims by Age at Injury and Maturity



Reserving in the Age of Obesity

Chart 3 presents for all studied claims at the 60-month maturity a breakdown of gender by age (thus comprising only injury years 1998-2002); the majority of claimants are male, as expected. Further, Chart 4 details for all three maturities the relative claim frequency by U.S. state. Florida is the most highly represented state, and Rhode Island is the least highly represented. The representation of a state in the data set depends primarily on the size of its labor force, but also on the combined market share of the insurance companies that contribute to the mentioned data base; another contributing factor is the share of the self-insured.

Chart 3: Relative Frequency of Claims at 60-Month Maturity by Gender and Age at Injury

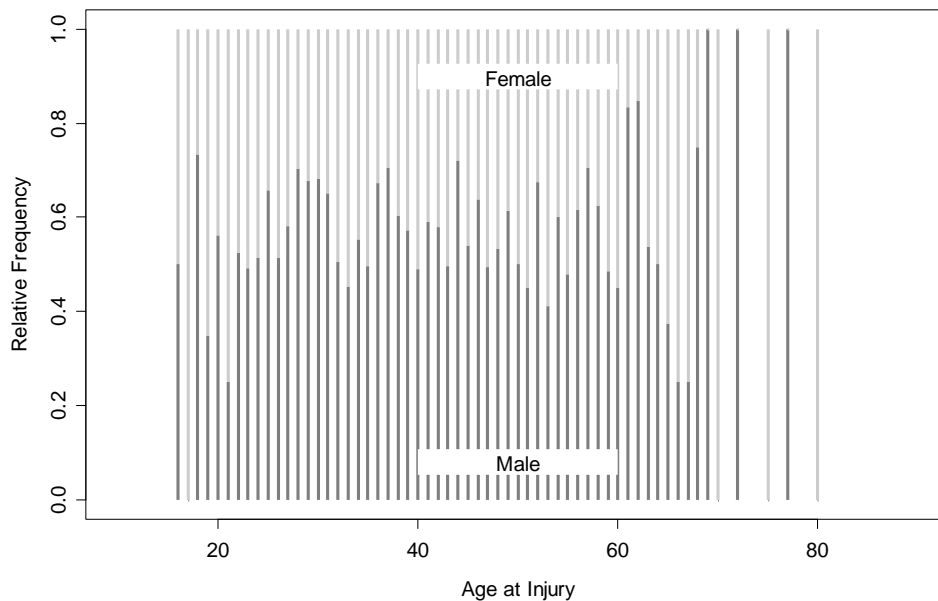
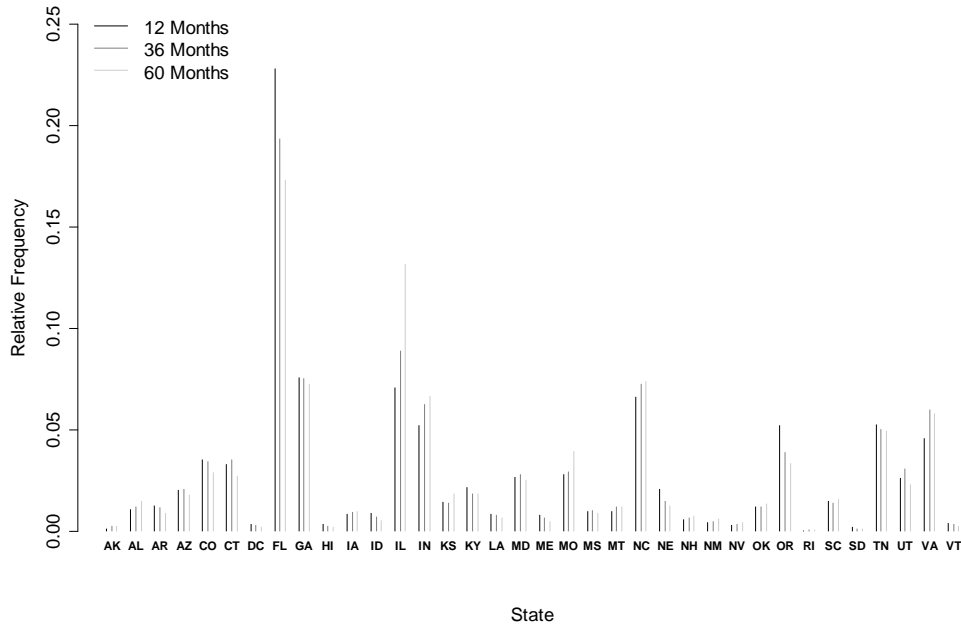
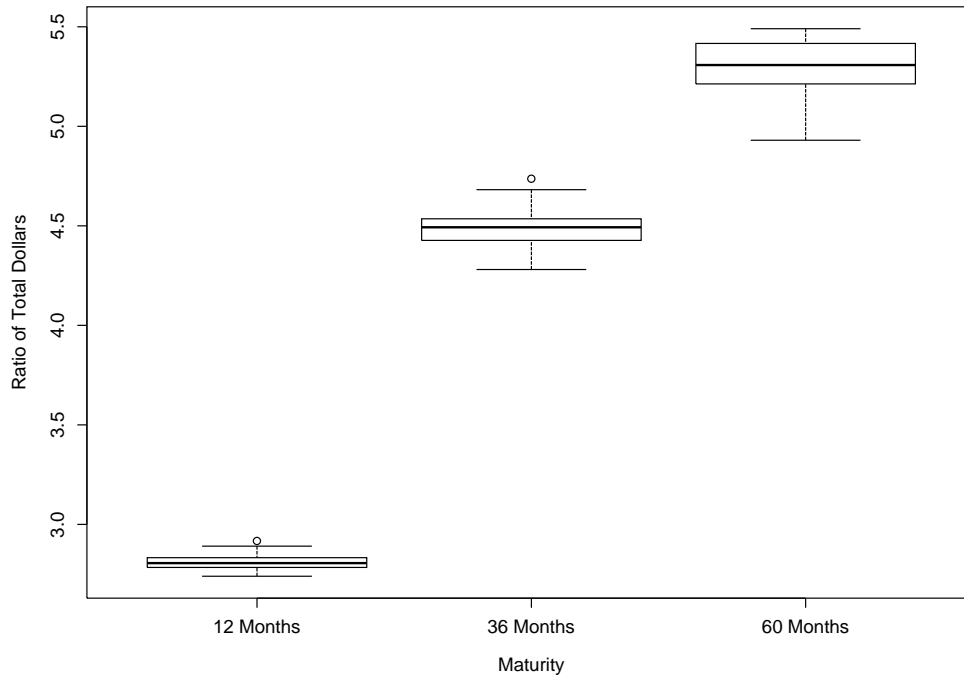


Chart 4: Relative Frequency of Claims by State and Maturity



Finally, Chart 5 offers an estimate of the ratio of the total cost of all obese claims, divided by the total cost of the non-obese claims these obese observations have been paired up with. The chart hosts three box plots, one for each maturity. Each boxplot represents 50 ratios, as generated by 50 randomized data sets. (As discussed, using multiple randomized data sets accounts for the uncertainty that originates in the tie-breaking of equidistant non-obese nearest neighbors when paired up with a given obese claim.)

Chart 5: Box Plots for the Ratios of Total Obese to Total Non-Obese Cumulative Payments for 50 Sets of Matched Claims at Alternative Maturities



In Chart 5, the horizontal line within a given box indicates the median value. The “hinges” of a box represent (approximately) quartiles, meaning that the box comprises about 50 percent of the observations. The distance between the hinges (that is, the height of the box) is called the inner quartile range. The whiskers located above and below the box point to extreme values. Specifically, the whisker above the box represents the highest observed value outside the box that is less than the sum of the third quartile plus 1.5 times the inner quartile range. Correspondingly, the whisker below the box signifies the lowest observed value outside the box that is greater than the first quartile minus 1.5 times the inner quartile range. Data points more extreme than those indicated by the whiskers are plotted as circles and may be considered outliers. For details on boxplots see Chambers et al. [1].

Chart 5 shows that obese claims are 2.8 times more expensive than non-obese claims at the 12-month maturity, but this cost difference climbs to a factor of 4.5 at the three-year maturity, and to 5.3 at the five-year maturity. This divergence in development between obese and non-obese claims has profound implications for reserving, as the added cost of obesity reveals itself only over time. A possible reason for such dissimilarity in development may be the longer duration of obese claims,

although, at this point, due to data limitations, this cannot be confirmed with confidence. Clearly, the Duke University study points to longer durations for obese claimants.

3. THE STATISTICAL MODEL

The purpose of the statistical modeling is to quantify the effect of claim characteristics on the percentage cost difference between obese claims and comparable non-obese claims. Thus, the dependent variable in the statistical analysis is the natural logarithm of the ratio (or log ratio, for short) of the cost of an obese claim to the cost of a comparable non-obese claim. Comparable non-obese claims are identified by means of pair-wise matching, as detailed above.

The statistical model has a semiparametric (or, synonymously, partially linear) structure. Generically, a semiparametric model may be written as

$$y_i = \mathbf{x}_i \cdot \boldsymbol{\beta} + f(z_i) , \quad (1)$$

where y_i is the dependent variable, $\mathbf{x}_i \cdot \boldsymbol{\beta}$ is the parametric, standard linear regression component, and $f(z_i)$ is a smoother that constitutes the nonparametric component. The purpose of the semiparametric structure is to accommodate a potentially nonlinear influence of the covariate z .

The semiparametric model makes use of a Bayesian multilevel (hierarchical) approach, which is estimated using MCMC (Markov Chain Monte Carlo simulation). At the first level, there is the log ratio of obese to non-obese medical costs at a given maturity. At the second level, there are four (non-nested) attributes that are modeled as random effects; these attributes are the injury year, the nature of injury (by aggregated ICD-9 code), the U.S. state, and the industry. The purpose of random effects in multilevel modeling is to shrink the multiple measurements within a second-level category (e.g., within a category in the group nature of injury) toward the weighted mean of group means (e.g., toward the weighted mean of the means of the categories within nature of injury). The fewer observations there are within a category (e.g., the fewer observations there are of a certain nature of injury) and the less precisely these observations are measured, the more the estimated mean of this category is shrunk toward the (weighted) mean of the category means within the group (e.g., the weighted mean of the means of the 22 categories within the group nature of injury). Conversely, the more observations there are within a category and the more precisely these observations are measured, the closer the estimated mean for this category is to its sample mean (see

Gelman and Hill [5]). This concept of shrinkage is closely related to the actuarial concept of credibility, as discussed by Guszczka [6].

For the purpose of statistical modeling (but not for matching), the multitude of observed ICD-9 codes are aggregated into 22 injury categories; such aggregation prevents an undue proliferation of regression coefficients. Table 2 details the aggregation rule. ICD-9 codes not covered in Table 2 are conditions that are typically not related to workplace injuries and illnesses (and, for this reason, are either very rare or do not show up at all in the raw claims data set). Among these conditions are mental disorders, complications of pregnancy, congenital anomalies, and others.

Random effects modeling is available only for categorical variables for which more than two values are observed. Gender, MUR, and MBR can take on only two alternative values and, hence, are modeled as indicator variables. These three indicator variables equal unity if (respectively) the claimant is female, is subject to MUR, and subject to MBR.

Table 2: Injury Category (Aggregated ICD-9 Codes)

Category	ICD-9 Codes	Description
1	001–289.9 and 390–629.9	Diseases other than diseases of the musculoskeletal system and connective tissue and diseases of the nervous system and sense organs
2	320–389.9	Diseases of the nervous system and sense organs
3	710–739.9	Diseases of the musculoskeletal system and connective tissue
4	800–829.1	Fractures
5	830–839.9	Dislocation
6	840–848.9	Sprains and strains of joints and adjacent muscles
7	850–854.1	Intracranial injury, excluding those with skull fracture
8	860–869.1	Internal injury of thorax, abdomen, and pelvis
9	870–897.7	Open wounds
10	900–904.9	Injury to blood vessels
11	905–909.9	Late effects of injuries, poisonings, toxic effects, and other external causes
12	910–919.9	Superficial injury
13	920–924.9	Contusion with intact skin surface
14	925–929.9	Crushing injury
15	930–939.9	Effects of foreign body entering through orifice
16	940–949.5	Burns
17	950–957.9	Injury to nerves and spinal cord
18	958–959.9	Certain traumatic complications and unspecified injuries
19	960–979.9	Poisoning by drugs, medicinal, and biological substances
20	980–989.9	Toxic effects of substances chiefly nonmedical as to source
21	990–995.94	Other and unspecified effects of external causes
22	996–999.9	Complications of surgical and medical care, not elsewhere classified

Reserving in the Age of Obesity

Finally, the influence of age is potentially nonlinear. (Age is the only continuous variable in the model and, hence, the only variable the influence of which may be nonlinear.) Thus, age is modeled in the nonparametric component of the model, which is implemented as a linear spline using reversible jump MCMC. Reversible jump MCMC is a concept of Bayesian model averaging, which is applied to the number of knots of the spline; knots in linear splines are locations where the linear function changes slopes. The model averages over a set of specifications with alternative numbers of knots; the location of these knots are determined by the model. For details on reversible jump MCMC in the context of linear splines, see Lunn, Best, and Whittaker [7].

We use a normal likelihood for the dependent variable (which is defined as the log ratio of obese to non-obese claim costs, by pair of matched claims, as discussed). This likelihood for claim i reads

$$y_i \sim N(\mu_i, \pi) , \tag{2}$$

where μ_i is the expected value and π is the precision. (As is common practice in Bayesian modeling, the notation is in term of precision, which is defined as the reciprocal value of the variance.) We use a gamma prior, $\text{Ga}(1, 0.001)$, for the precision.

The fixed effect of indicator variable j (which, for instance, equals unity for female claimants and zero otherwise) reads

$$\beta_j \sim N(0, 0.001) . \tag{3}$$

The random effect specification for a given set k of indicator variables (such as those that represent the 22 categories of the group nature of injury) reads

$$\delta_{k,m} \sim N(0, \pi_k) , \tag{3}$$

Reserving in the Age of Obesity

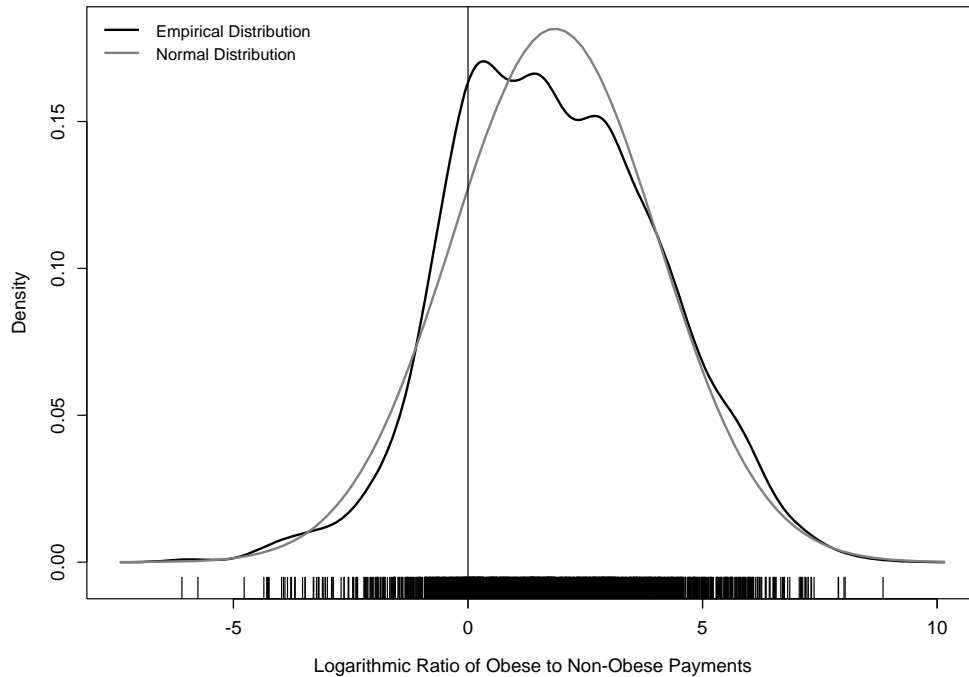
where $\delta_{k,m}$ is the effect of attribute m within group k (e.g., category 20 within the group nature of injury). Again, the prior distribution for the precision reads $\text{Ga}(1,0.001)$. Note that the random effects are draws from a common distribution, and that these effects are centered on zero.

Finally, reversible jump MCMC is implemented using a linear spline on age at injury; the prior for the number of knots is a uniform categorical distribution on the integers in the interval $[0, \text{trunc}(\text{range}(\text{age})/2)+1]$, where age is the set of observed values for age at injury (measured in whole years), “range” determines the difference between the maximum and minimum values, and “trunc” rounds down to the nearest integer. The prior distribution for the location of knots is uniform on the interval defined by the minimum and maximum observed age.

It is worthy of note that in the model outlined above, the (percentage) effect of obesity is allowed to vary with all covariates; this is because the dependent variable is the log ratio of the cost of an obese claim to the cost of a comparable non-obese claim. Thus, instead of postulating a uniform percentage effect of obesity across all claims, the percentage cost impact of obesity is allowed to vary by injury year, nature of injury, U.S. state, industry, gender, and age (and, where applicable, by the legislative environment). This approach is in keeping with the approach taken by Truls, Dement, and Krause [9] in their multivariate modeling of the claim costs of Duke University employees.

Before applying the statistical model, we plot the empirical distribution of the dependent variable (for one of the 50 randomized data sets) against the normal. As Chart 6 shows, the normal likelihood is a fair assumption for modeling the log ratio of obese to non-obese claim costs at the 60-month maturity, in spite of some degree of skewness; the distributions for the shorter maturities are similarly close to normal. (The whiskers at the bottom of Chart 6 indicate the locations of the observations.)

Chart 6: Kernel Density Estimate for a Randomized Set of Log Ratios (for Pairs of Matched Claims) of Obese to Non-Obese Cumulative Payments at the 60-Month Maturity



4. FINDINGS FOR RANDOM EFFECTS, GENDER, AND AGE

We estimate two versions of the model. In a first version, we include as covariates the injury year, the nature of injury, the U.S. state, and the industry; these four covariates are modeled as random effects; further, we include as covariates gender and age at injury, but we do not account for differences (across states and over time) in the legislative environment. In a second version of the model, the findings of which are presented in Section 5, we add indicator variables for MUR and MBR; the explanatory power of these covariates will subtract from the measured random effects of the injury year and the state. That is because without the covariates MUR and MBR, variations in the legislative environment over time and across states are absorbed by the random effects of injury year and state, respectively.

All findings shown in this section pertain to the 60-month maturity. Chart 7 offers a graphical exposition of the estimated random effects of the injury years 1998 through 2002 (which are the only injury years available for the 60-month maturity). The gray dots and whiskers indicate the mean and 80 percent credible intervals for 50 randomized data sets. The black dots and whisker signify

Reserving in the Age of Obesity

the mean and 80 percent credible intervals after aggregating across all 50 data sets. Clearly, there is variation across injury years, and there appears to be a mild time trend in the ratio of obese to non-obese medical claim costs. Remember that, by definition, random effects are centered on zero.

Chart 7: Mean and 80 Percent Credible Intervals for the Random Effect of the Injury Year in a Partial Linear Multilevel Regression on the Ratio of Cumulative Payments of Obese to Non-Obese Claims for 50 Sets of Pairs of Matched Claims at a Maturity of 60 Months

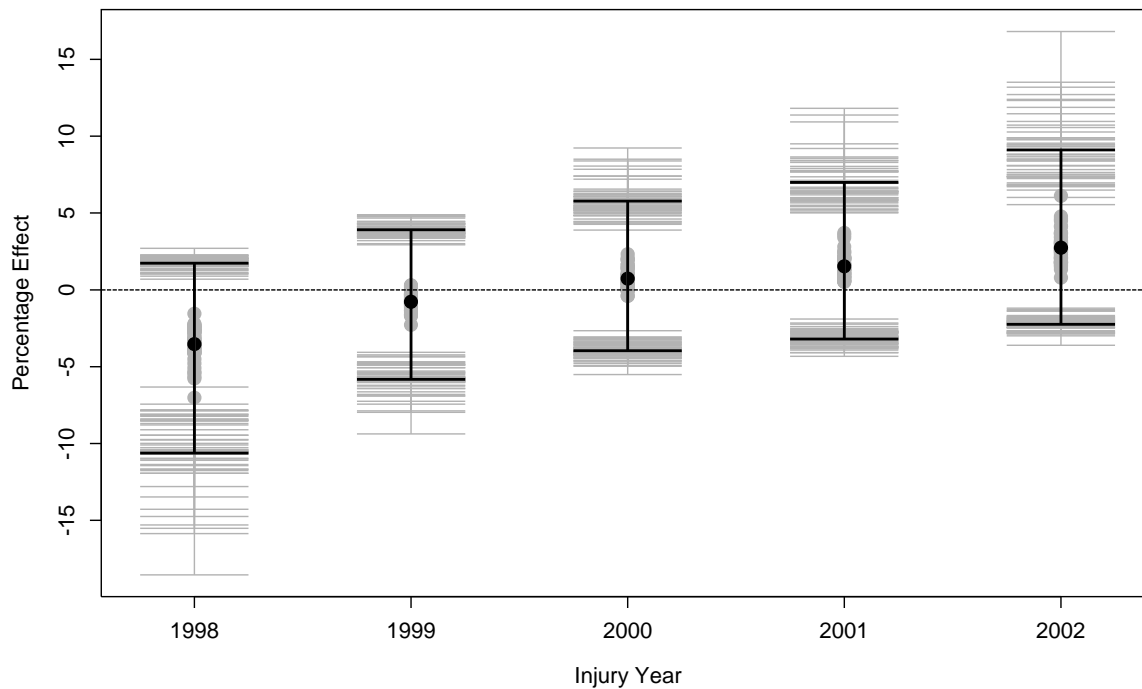


Chart 8 presents the random effect estimates for the 22 injury categories (of which only 19 are observed at the 60-month maturity). The mean of the estimated effect of a given injury category is indicated by the location of the number of the category. The fewer observations there are in a given category and the “noisier” these observations are, the wider the credible intervals. The injury categories with the highest percentage cost contributions to obesity are 7 (intracranial injury, excluding those with skull fracture), 9 (open wounds), 11 (late effects of injuries, poisonings, toxic effects, and other external causes), 14 (crushing injury), 17 (injury to nerves and spinal cord), and 18 (certain traumatic complications and unspecified injuries). At the same time, some of these injury categories are among those with the widest credible intervals; exceptions are categories 9 and 18. The scale of differences among injury categories is quite large. For instance, an estimated displayed

Reserving in the Age of Obesity

effect of 50 percent for a given injury category implies that for this category, the effect of obesity on the ratio of claim costs is 50 percent higher than the average across categories.

Chart 8: Mean and 80 Percent Credible Intervals for the Random Effect of the Injury Category in a Partial Linear Multilevel Regression on the Ratio of Cumulative Payments of Obese to Non-Obese Claims for 50 Sets of Pairs of Matched Claims at a Maturity of 60 Months

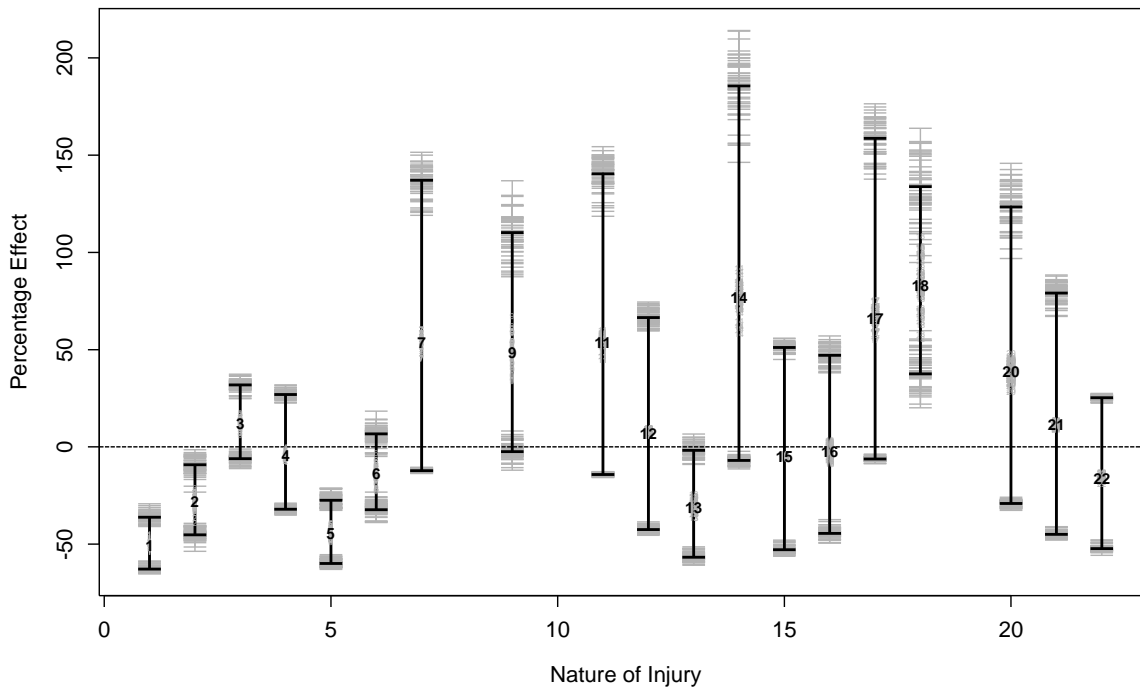


Chart 9 displays the random effect estimates for the 36 U.S. states included in the analysis. Here again, the states with above or below-average estimated effects of obesity also tend to be the ones with large credible intervals. Finally, Chart 10 shows the random effect estimates for the five NCCI industries.

Reserving in the Age of Obesity

Chart 9: Mean and 80 Percent Credible Intervals for the Random Effect of the U.S. States in a Partial Linear Multilevel Regression on the Ratio of Cumulative Payments of Obese to Non-Obese Claims for 50 Sets of Pairs of Matched Claims at a Maturity of 60 Months

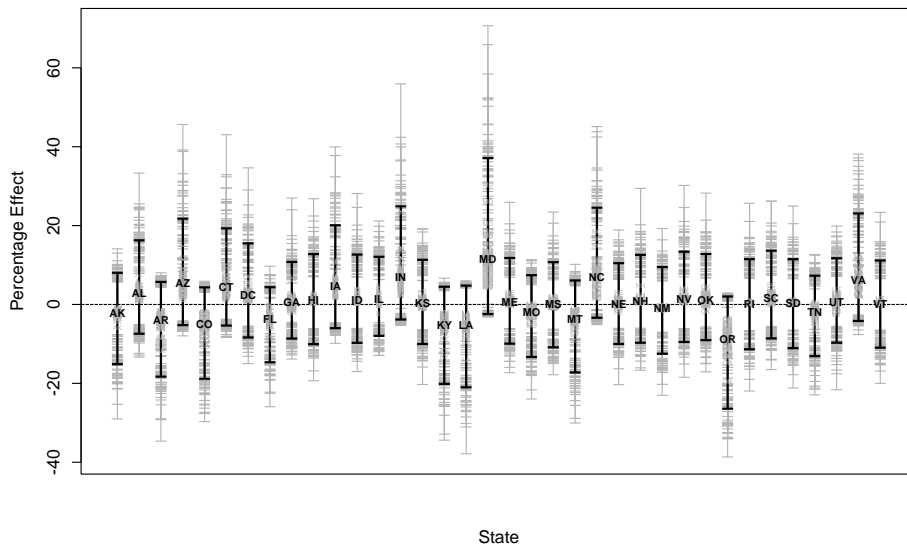
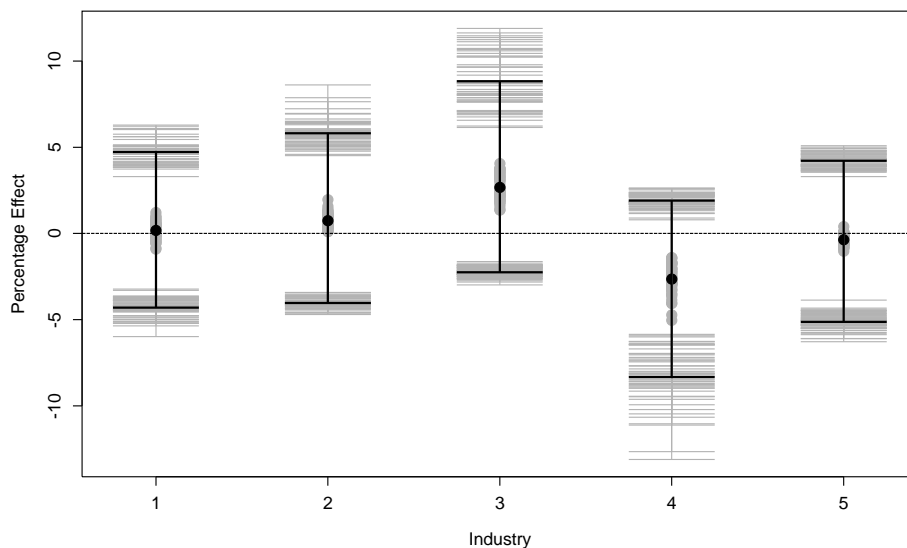


Chart 10: Mean and 80 Percent Credible Intervals for the Random Effect of the Industry in a Partial Linear Multilevel Regression on the Ratio of Cumulative Payments of Obese to Non-Obese Claims for 50 Sets of Pairs of Matched Claims at a Maturity of 60 Months. Industry codes: (1) Manufacturing; (2) Contracting; (3) Office and Clerical; (4) Goods and Services; (5) Miscellaneous



Reserving in the Age of Obesity

The random effects for the injury year and the industry are quite small compared to those of the injury category. This finding is summarized in Chart 11, which offers a comparison of the variances of the four random effects. As shown in this chart, the injury year and the industry have comparatively little explanatory power. There is more variation across states than there is over time and across industries. But the most explanatory power originates in the injury categories.

Chart 11: Boxplots for the Means of the Posteriors of the Variances of the Random Effects

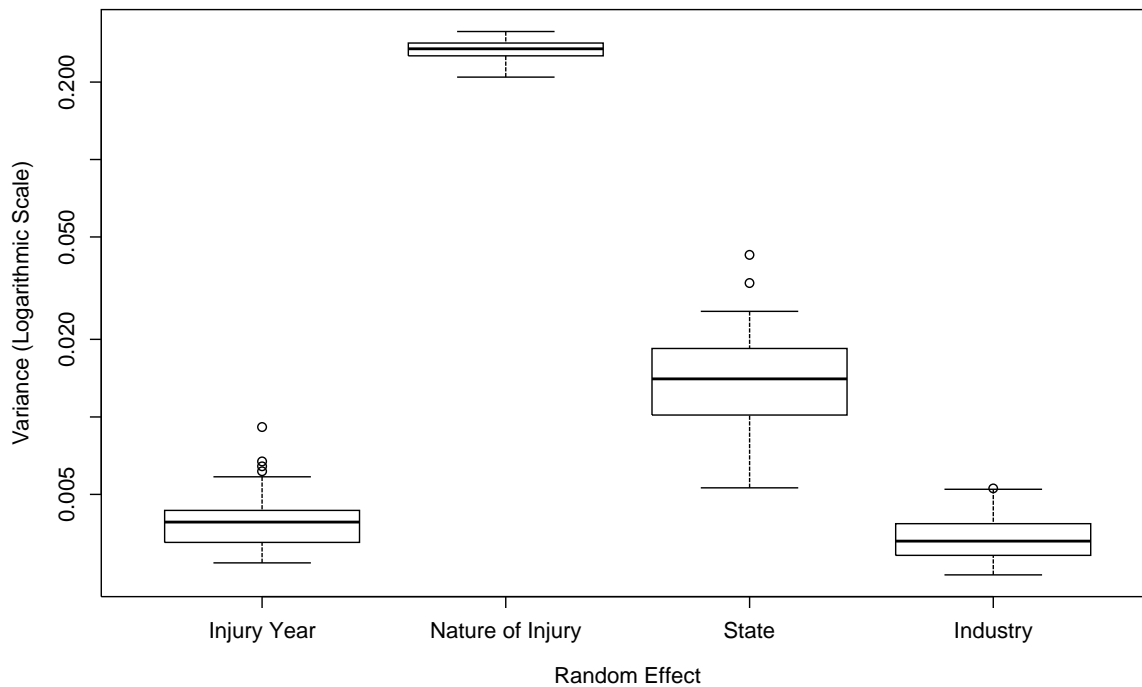
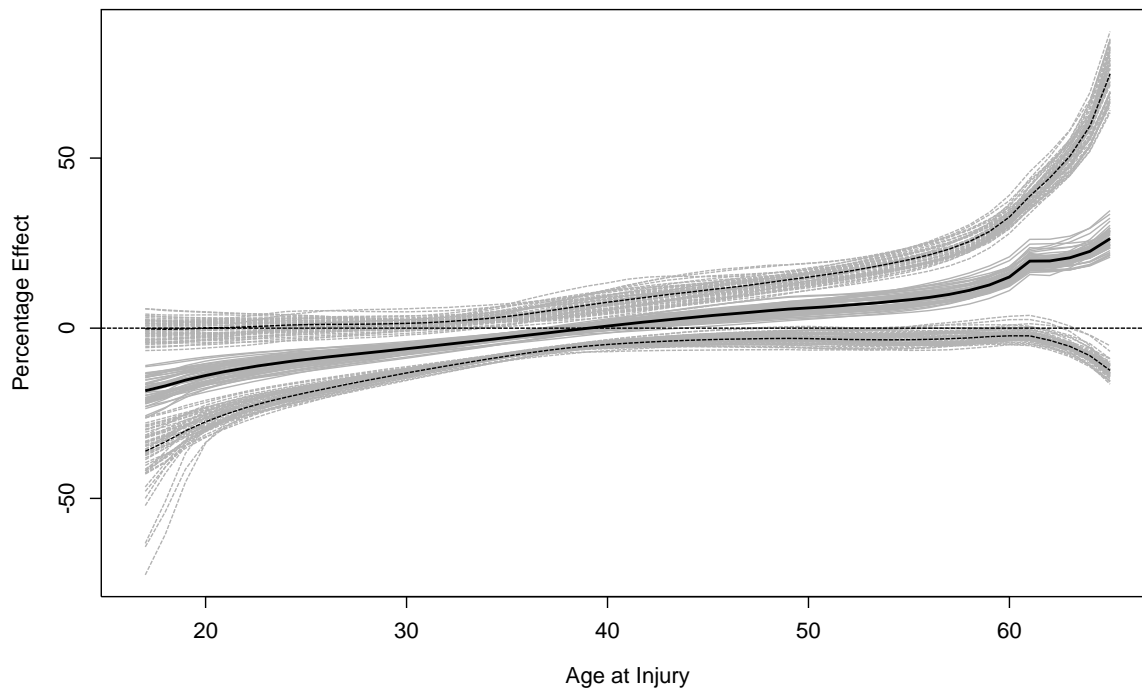


Chart 12 displays the effect of the female gender on the difference between obese and comparable non-obese claims. Here again, the gray whiskers (and “Female” label indicating the mean) pertain to the 50 individual, randomized data sets, whereas the black whiskers (along with the black “Female” label) indicate the results obtained when aggregating over these 50 runs. Although the mean estimate of the female gender is negative, the credible interval includes the zero value. Then again, the negative effect of the female gender on the ratio of obese to non-obese claim costs agrees with the findings of Truls, Dement, and Krause [9] in the Duke University study.

Chart 13: Mean and 80 Percent Credible Interval for the Influence of Age in a Partial Linear Multilevel Regression on the Ratio of Cumulative Payments of Obese to Non-Obese Claims for 50 Sets of Pairs of Matched Claims at a Maturity of 60 Months



5. FINDINGS FOR THE LEGISLATIVE ENVIRONMENT

As discussed, in a second version of the model, we add MUR and MBR as covariates. Information on MUR and MBR is provided by the Workers Compensation Research Institute (WCRI) for calendar years 1997 and 2001. For injury years 2001 and later, the 2001 value applies; for earlier injury years, the 1997 value is used. The covariates are coded as indicator variables that are equal to unity for U.S. states with MUR and MBR (respectively) in place, and zero otherwise.

WCRI [10] puts a state in the “mandated utilization review/management category” if such jurisdiction mandates that “payers review claims for proper medical care utilization...or if the workers’ compensation agency or exclusive state fund conducts utilization review on its own initiative (either for all claims or those that meet certain criteria).” Further, WCRI [10] credits a state with a bill review program if in such jurisdiction “the workers’ compensation agency routinely examines for proper charges all bills or cases that meet explicit criteria, such as a certain number of

Reserving in the Age of Obesity

days lost from work or a dollar limit on medical care costs; the statute mandates the examination of bills by payers; or an exclusive state fund does regular bill review.”

Table 3 details the MUR and MBR categorization of states for the two available years. In 1997, twelve states had MUR, eight states had MBR, and six jurisdictions had both legislative provisions in place. By 2001, the number of states with MUR had dropped to nine, those with MBR had risen to ten, and those with both provisions had decreased to five.

Chart 14 offers boxplots for the estimated 50 values of the influence of MUR, as obtained when analyzing the 50 randomized data sets. Clearly, MUR significantly reduces the ratio in the medical costs per claim of obese to non-obese claimants. For the 12-month maturity, MUR reduces this difference by 15.8 percent. For the 36-month maturity, the effect of MUR grows to 18.8 percent, before reaching 28.5 percent at the 60-month maturity. This evidence attests to the importance of the legislative setting for the medical cost of workers compensation claims.

Chart 15 offers boxplots for the 50 estimated effects of MBR. Although the cost containment effect of MBR is substantial, its influence is weaker than MUR. For the 12-month maturity, MBR reduces the ratio in the medical costs per claim of obese to non-obese claimants by 9.4 percent. For the 36-month maturity, the effect of MUR grows to 16.6 percent, before dropping back to 12.2 percent at the 60-month maturity. Clearly, the maximum cost containment effect is achieved where both of the two legislative provisions are in place: The combined effect MBR and MUR (not shown) at the 12-month maturity amounts to a negative 24.0 percent, before growing to a negative 33.0 percent at the 36-month maturity and reaching a negative 37.4 percent at the 60-month maturity.

Table 3: MUR and MBR by State

State	Year			
	1997		2001	
	MUR	MBR	MUR	MBR
AL	No	No	No	No
AK	No	No	No	No
AZ	No	No	No	No
AR	Yes	Yes	No	Yes
CO	Yes	Yes	Yes	Yes
CT	No	No	No	No
DC	No	No	No	No
FL	Yes	Yes	Yes	Yes
GA	No	No	No	No
HI	No	No	No	No
ID	No	No	No	No
IL	No	No	No	No
IN	No	No	No	No
IA	No	No	No	No
KS	No	No	No	No
KY	Yes	No	Yes	Yes
LA	Yes	Yes	Yes	Yes
ME	Yes	No	Yes	No
MD	No	No	No	No
MS	Yes	Yes	Yes	Yes
MO	No	No	No	No
MT	Yes	No	No	No
NE	No	No	No	No
NV	Yes	Yes	No	Yes
NH	No	No	No	No
NM	Yes	No	Yes	No
NC	No	No	No	Yes
OK	No	No	No	No
OR	No	Yes	No	Yes
RI	No	No	No	No
SC	No	Yes	No	Yes
SD	No	No	No	No
TN	Yes	No	Yes	No
UT	Yes	No	Yes	No
VT	No	No	No	No
VA	No	No	No	No

Source: WCRI [10][11].

Chart 14: Box Plots for the Means of the Influence of MUR in a Partial Linear Multilevel Regression on the Ratio of Cumulative Payments of Obese to Non-Obese Claims for 50 Sets of Pairs of Matched Claims

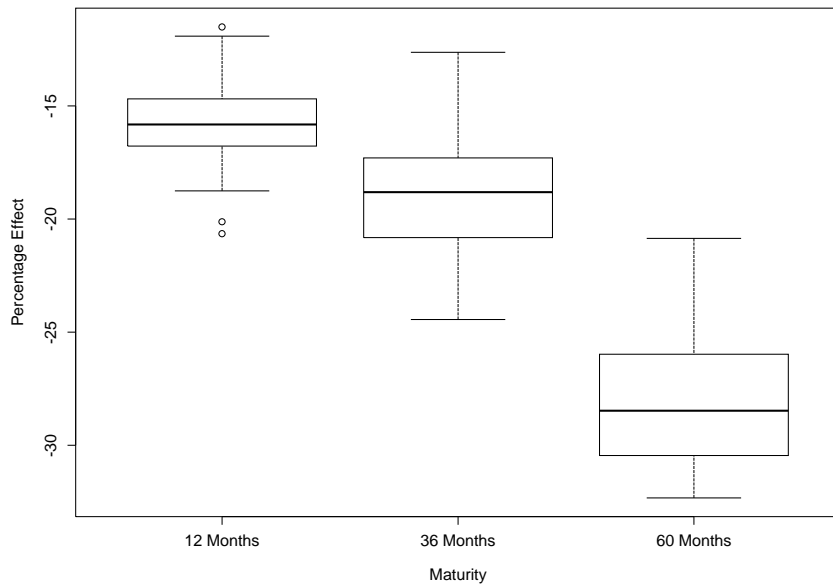
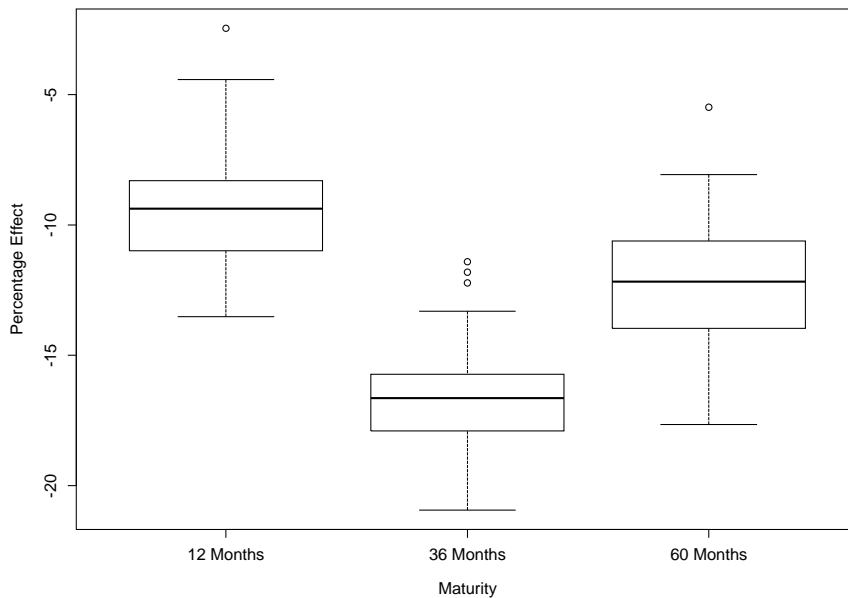


Chart 15: Box Plots for the Means of the Influence of MBR in a Partial Linear Multilevel Regression on the Ratio of Cumulative Payments of Obese to Non-Obese Claims for 50 Sets of Pairs of Matched Claims

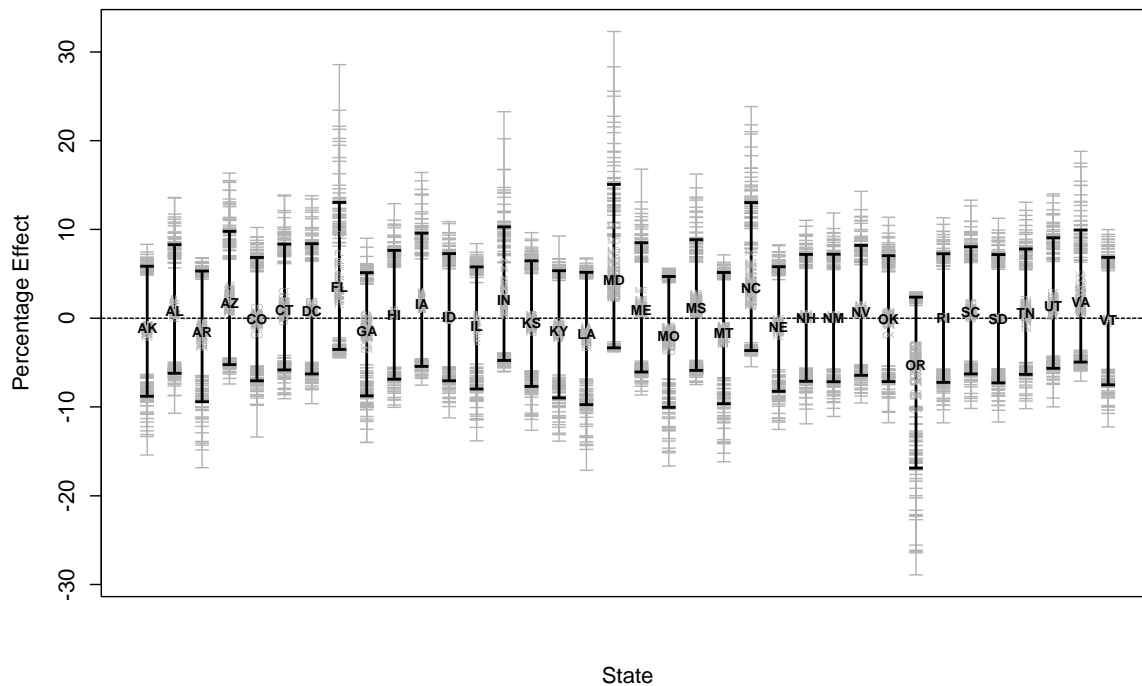


Reserving in the Age of Obesity

As discussed, including MUR and MBR as covariates subtracts from the random effects of injury year and state. This is because some of the variation over time and across states is due to these two legislative provisions. It is therefore of interest to revisit the random effects of states and to investigate how much of the variation across states is due to these two covariates.

Chart 16 displays the random effects for the states when MUR and MBR are included in the regression equation. Clearly, the cross-state variation of the ratio in the medical costs per claim of obese to non-obese claimants is now much smaller than previously. In fact, the variance of the state random effects drops to 0.00579 from the previous level of 0.01406, which implies that MUR and MBR explain about 59 percent the cross-state variation in the percentage cost effect of obesity.

Chart 16: Mean and 80 Percent Credible Intervals for the Random Effect of U.S. States in a Partial Linear Multilevel Regression on the Ratio of Cumulative Payments of Obese to Non-Obese Claims for 50 Sets of Pairs of Matched Claims at a Maturity of 60 Months when Controlling for Differences in the Legislative Environment



6. CONCLUSIONS

Studying a large data set of obese claims, which comprises 36 U.S. states and nine injury years, we are able to provide evidence on the effect of obesity on the medical cost per claim. We showed that

the effect of obesity is substantial, and that the entirety of the effect of obesity reveals itself only over time, as claims mature. Most importantly, we were able to quantify the effect of the legislative environment on the effect of obesity on claim costs.

Our methodology differs in important ways from the Duke University study by Truls, Dement, and Krause [9]. In part, this dissimilarity in approach is necessitated by a difference in the nature of the data set, but also by a difference in objective. Although our data set is considerably larger (as it comprises millions of claims), it is also more limited in the scope of the data items. Most importantly, we have no information on the BMI. Instead of having an objective criterion for obesity (and having obesity differentiated by degree), our data set offers an obesity categorization that relies on the physician's decision to list obesity as a co-morbidity indication.

This difference between subjective and objective categorization of obesity has potential implications for the legislative findings of our model. For instance, it may be argued that the decision of the physician to diagnose a claimant as obese is influenced by the legislative environment, thus causing an endogeneity bias in our measurement of the effects of MUR and MBR. Although we cannot refute such a proposition with confidence, it is important to note that the potential endogeneity is likely to cause the effects of MUR and MBR to be underestimated (instead of overestimated). This is because if the increased scrutiny of MUR and MBR raises a physician's propensity of coding claimants as obese (in order to provide better documentation supportive of the treatment), the implied broader definition of obese claims diminishes the recorded cost difference between obese and comparable non-obese claims in the data set.

Finally, the proportion of claims with obesity as a co-morbidity indicator is comparatively small (0.1 to 0.2 percent) when compared with the proportion of the obese in the workforce. It is likely that obesity serves as a co-morbidity indicator primarily where complications from obesity are highly probable (e.g., when claimants are morbidly obese) or have already materialized. From this perspective, the measured effect of obesity on claim severity may be viewed as an upper bound.

Acknowledgment

Thanks to Uriel Carrasquilla, John Potter, Tanya Restrepo, John Robertson, Harry Shuford, and Martin Wolf for comments and to Chun Shyong for research assistance.

7. REFERENCES

- [1] Carlin, Bradley P., and Thomas A. Lewis, *Bayes and Empirical Bayes Methods for Data Analysis*. 2nd ed., Boca Raton: Chapman & Hall/CRC, 2000.
- [2] Chambers, John M., William S. Cleveland, Beat Kleiner, and Paul A. Tukey, *Graphical Methods for Data Analysis*, New York: Chapman and Hall, 1983.
- [3] Finkelstein, Eric A., Ian C. Fiebelkorn, and Guijing Wang, “National Medical Spending Attributable to Overweight and Obesity: How Much, and Who’s Paying,” *Health Affairs* 22, May 14, 2003, pp. w3-219-w226.
- [4] Finkelstein, Eric A., Justin G. Trogdon, Joel W. Cohen, and William Dietz (2009) “Annual Medical Spending Attributable to Obesity: Payer- and Service-Specific Estimates,” *Health Affairs* 28(5), pp. w822-w831, DOI 10.1377/hlthaff.28.5.w822.
- [5] Gelman, Andrew, and Jennifer Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge, MA: Cambridge University Press, 2007.
- [6] Guszczka, James, “Hierarchical Growth Curve Models for Loss Reserving,” *Casualty Actuarial Society E-Forum*, Fall 2008, <http://www.casact.org/pubs/forum/08fforum/7Guszczka.pdf>.
- [7] Kasabov, Nicola K., *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*, Cambridge, MA: MIT Press, 1996.
- [8] Lunn, David J., Nicky Best, and John C. Whittaker, “Generic Reversible Jump MCMC Using Graphical Models,” *Statistics and Computing* 19(4), pp. 395-408, DOI 10.1007/s11222-008-9100-0.
- [9] Truls, Østbye, John M. Dement, and Katrina M. Krause, “Obesity and Workers’ Compensation: Results from the Duke Health and Safety Surveillance System,” *Archives of Internal Medicine* 167, 2007, pp. 766-773.
- [10] Worker Compensation Research Institute, *Managed Care and Medical Cost Containment in Workers’ Compensation: A National Inventory, 2001-2002*, Cambridge, MA: WCRI, 2001.
- [11] Worker Compensation Research Institute, *Managed Care and Medical Cost Containment in Workers’ Compensation: A National Inventory, 1997-1998*, Cambridge, MA: WCRI, 1997.

Abbreviations and notations

BMI, Body Mass Index
FTE, Full-Time Equivalent
HRA, Health Risk Assessment
ICD-9, International Classification of Diseases, Ninth Revision
MBR, Mandatory Bill Review
MCMC, Markov Chain Monte Carlo simulation
MUR, Mandatory Utilization Review
NCCI, National Council on Compensation Insurance
WCRI, Workers Compensation Research Institute

Biographies of the Authors

Chris Laws is a Senior Actuarial Analyst at the National Council on Compensation Insurance, Inc.

Frank Schmid, Dr. habil., is a Director and Senior Economist at the National Council on Compensation Insurance, Inc.

Towards Multivariate Ratemaking: Claim Frequency Analysis Examples

Hernán L. Medina, CPCU, API, AU, AIM, ARC

Abstract

Motivation. Test how changes in level and distribution of exposures affect different ratemaking models. Actuaries are well aware that loss trend can be distorted by changes in exposure level and business mix. They are trained to recognize situations in which these distortions may arise, and how to adjust for them. Multivariate models are another way of handling these distortions. Using claim frequency as an example, the paper illustrates the design of multivariate analyses resistant to changes in exposure level and business mix.

Method. Simulate data in which the predominant sources of variation are changing exposure levels and changes in the distribution of exposures. Determine indicated trend, development, and classification factors using multivariate and univariate models. Compare the results.

Results. Trend, development factors, and relativity indications from 30 samples having different levels of variation in exposure levels and distribution are obtained by different methods.

Conclusions. Multivariate analyses that incorporate all available information are more robust than other analyses when data have significant changes in exposure levels or changes in mix of business.

Availability.

Input data sets and model outputs are available at www.casact.org.

Keywords. Ratemaking, Trend and Loss Development, Rating Class Relativities, Generalized Linear Models

1. INTRODUCTION

Actuaries began to develop the art and science of property and casualty insurance ratemaking long before computers were invented. At a time when calculations were done with pencil and paper, it was natural to use methods that relied on total sums and averages. When computers were first introduced, storage media were very expensive and processing speeds were relatively slow by today's standards. Thus ratemaking databases were designed to contain totals and averages, and rating systems continued to rely for the most part on univariate analyses based on aggregate data.

Actuaries are well aware of the pitfalls one might encounter using methods that rely on aggregate data. Part of actuarial training is learning to recognize the distortions that might arise, and how these might be corrected or minimized. For example, the CAS' *Basic Ratemaking* textbook indicates that if calendar year data is used to measure loss trend and the book of business is changing significantly in size, the trend can be over or underestimated.¹ An illustration of this situation follows in Table 1.1.

¹ Werner, Geoff and Claudine Modlin, *Basic Ratemaking*, 3rd ed., Arlington, VA: Casualty Actuarial Society, January 2010, p. 113.

Table 1.1

Calendar Year	Earned Car Years	Calendar Year Claims Closed With Payment	Claim Frequency	Years	Trend
2004	198,017	12,504	6.31461	6	3.0%
2005	215,837	13,770	6.37981	5	3.4%
2006	232,026	14,972	6.45273	4	3.8%
2007	225,064	15,304	6.79984	3	3.1%
2008	211,559	14,928	7.05619		
2009	192,520	13,911	7.22574		

Accident Year	Claims Closed With Payment				Development Factors			Ultimate Claims
	Age 12	Age 24	Age 36	Age 48	12 to 24	24 to 36	36 to 48	
2002	5,435	8,696	10,870	10,870	1.600	1.250	1.000	10,870
2003	6,007	9,611	12,013	12,013	1.600	1.250	1.000	12,013
2004	6,726	10,762	13,452	13,452	1.600	1.250	1.000	13,452
2005	7,332	11,733	14,665	14,665	1.600	1.250	1.000	14,665
2006	7,881	12,609	15,764	15,764	1.600	1.250	1.000	15,764
2007	7,644	12,230	15,288	15,288	1.600	1.250		15,288
2008	7,187	11,500			1.600			14,375
2009	6,540							13,080
Selected Age to Age					1.600	1.250	1.000	
Selected Age to Ultimate					2.000	1.250	1.000	

Accident Year	Earned Car Years	Ultimate Accident Year Claim Count	Claim Frequency	Years	Trend
2004	198,017	13,452	6.79336	6	0.0%
2005	215,837	14,665	6.79448	5	0.0%
2006	232,026	15,764	6.79407	4	0.0%
2007	225,064	15,288	6.79273	3	0.0%
2008	211,559	14,375	6.79479		
2009	192,520	13,080	6.79410		

In the example above, development factors are constant across accident year, so we can be reasonably certain of the estimated ultimate claim counts and the trend based on accident year data. Therefore we can conclude the trend based on calendar year data is overstated. In a real-world

situation, however, development factors may be more volatile, and the selection of loss development factors, “introduces some subjectivity into the [accident year] trend analysis.”²

The CAS’ *Basic Ratemaking* electronic textbook explains that the reason for the distortion in the calendar year trend is that as exposure levels change, the distribution of calendar year claims by accident year changes.³ In fact, the effect of exposure level changes on calendar year trend is a special case of a more general phenomenon: the effect of changes in business mix on frequency and severity, which can affect both calendar year trend as well as accident year trend. Consider the example in Table 1.2.

Table 1.2

Area	Accident year	Earned Car Years	Ultimate Claims With Payment	Claim Frequency	Years	Trend
Territory A	2004	220,500	18,820	8.53515	6	3.0%
	2005	231,527	20,353	8.79077	5	3.0%
	2006	243,100	22,011	9.05430	4	3.0%
	2007	255,256	23,803	9.32515	3	3.0%
	2008	268,019	25,745	9.60566		
	2009	281,420	27,844	9.89411		
Territory B	2004	179,500	5,105	2.84401	6	3.0%
	2005	168,476	4,936	2.92979	5	3.0%
	2006	156,900	4,735	3.01785	4	3.0%
	2007	144,744	4,501	3.10963	3	3.0%
	2008	131,981	4,225	3.20122		
	2009	118,580	3,911	3.29820		
Statewide	2004	400,000	23,925	5.98125	6	5.8%
	2005	400,003	25,289	6.32220	5	5.9%
	2006	400,000	26,746	6.68650	4	5.9%
	2007	400,000	28,304	7.07600	3	5.9%
	2008	400,000	29,970	7.49250		
	2009	400,000	31,755	7.93875		

² Werner and Modlin, p. 113.

³ Werner and Modlin, *Basic Ratemaking*, p. 113.

In the example above in Table 1.2, each territory has a 3% trend, but the statewide data shows a trend that is almost twice as high, close to 6%. The reason for this is that the distribution of exposures in the state has been changing. “Distributional changes in a book of business also affect frequencies and severities. If the proportion of risky policies is growing, loss costs will be expected to increase.”⁴

Although the issues above are well known, they are generally handled on an ad hoc basis, and not much has changed in the basic rate review process. Generally, it involves two major steps: (1) determination of the overall indicated rate level change, and (2) determination of indicated classification relativities. Loss development and trend are two of the processes involved in determining the overall indicated rate level change. Thus, a basic rate review often involves at least three databases and systems: (1) loss development database and system, (2) loss trend database and system, and (3) and classification review database and system.

When accident year trends are used in the rate review, the loss development and loss trend processes are intertwined. For example, determining the accident year claim frequency trend typically involves the following steps: developing claim counts to ultimate, calculating ultimate claim frequency for each accident year, and analyzing the trend using a linear or exponential regression model. So the same database could be used for loss development and trend for rate reviews using accident year trend. In most cases, however, the database has been summarized in such a way that it cannot be used to review classification relativities.

From a data management perspective, as well as a business point of view, it is desirable to have a single database as the source for the analyses involved in the rate review process. This helps simplify data quality reviews and helps ensure that the data used in the different analyses balances. This could easily be accomplished. Appendix G of *A Practitioner’s Guide to Generalized Linear Models*⁵ presents several forms of data organization that can be used for generalized linear model (GLM) analysis, as well as their advantages and disadvantages. Using personal auto property damage liability as an example, we will expand one of those forms of data organization into a database that can be used as the source for development, accident year trend, and indicated classification relativity analyses. Furthermore, we will see how to integrate all of these processes into one single model using generalized linear models (GLM) and generalized estimating equations (GEE).

⁴ Werner and Modlin, p. 109.

⁵ Anderson, D., et al., “A Practitioner’s Guide to Generalized Linear Models,” 3rd Ed., CAS Study Note, 2007.

Using a single database for loss trend, loss development, and risk classification requires thoughtful consideration. A company may have some exclusion or adjustments currently used for trend analyses that are not used for loss development or classification analyses, and so on. In a multivariate model, however, you must consider whether it is preferable to adjust the data a priori, or to introduce variables that would control for the factor that would have made an adjustment or exclusion necessary in a univariate analysis. For example, certain vehicle models were recalled in 2010 because of problems involving sudden uncontrollable acceleration.⁶ If a large number of such claims are in the data, one option would be to exclude them from the analysis. Another option would be to leave them in the data and add a control variable to identify these claims in the multivariate model. The coefficient of the control variable would provide the actuary with a way to estimate the effect this unusual event had on the experience. The control variable would be equal to 1 for claims related to the recalled vehicles, and 0 for all other vehicles. If all affected vehicles have been recalled and repaired and no further losses related to this event are expected in the future, then the control variable is set to zero when the model is used to project expected claim counts or losses.

Differences in exclusions or adjustments may arise because different types of data are used for different types of analyses. For example, it is quite common for companies to use calendar year paid claim data for trend analysis, and accident year reported claim data for loss development in personal auto property damage liability rate reviews. Presumably, since different types of data are used for the univariate analyses of trend and loss development, some situation might arise that would make it necessary to adjust the trend data while the loss development data needs no adjustment (or vice versa). If this situation arises in a multivariate context in which loss trend, development and classification factors are estimated simultaneously, an adjustment or control variable would be needed for a model based on paid claim data, but no adjustment would be needed for a model based on reported claim data (or vice versa). As will be shown later, the database can be designed in such a way that it contains both paid and reported claim data. Consequently, it would be easy and advisable to perform two multivariate analyses: one using paid claim data and the other using reported claim data.

1.1 Research Context

We focus on three elements of a basic rate review: loss trend, loss development, and rating class relativities. The actuarial literature on loss trend and loss development generally considers these elements in isolation. An exception involves accident year trends, since the latter require that data

⁶ Bunkley, Nick, and Bill Vlasic, “Carmakers Initiating More Recalls Voluntarily,” *The New York Times*, August 24, 2010.

are developed to ultimate. Even in this case, however, the loss development and loss trend analyses are performed sequentially instead of simultaneously.

Similarly, papers on risk classification tend to consider their subject in isolation. For example, in “A Practitioner’s Guide to Generalized Linear Models” the discussion of loss trend and loss development occurs in Appendix F. The *Guide* suggests using a calendar/accident year method of organization and a dummy calendar year variable as a way to “absorb trends in claims experience that purely relate to time.”⁷ The *Guide* also suggests three options for dealing with loss development:

- Ignoring it — assuming it does not affect the classification factors.
- Including a dummy variable in the model to absorb time-related influences, removing it once the model is finalized, and adjusting the modeled results based on a separate calculation.
- Performing a series of GLM analyses, and comparing GLM relativities based on data at different development periods in order to obtain multivariate development factors.⁸

Styrsky noted that loss trend can be underestimated or overestimated when calendar year data are used in the analysis if the size of the portfolio increases or decreases significantly. He proposed an approach for dealing with this effect by matching each calendar year’s claims by accident year to the exposures that produced them.⁹ Werner and Modlin propose additional solutions to this problem: (1) using econometric models or generalized linear models to measure trend or (2) using accident year data (developed to ultimate) for trend analysis. They note that the loss development process “may introduce some subjectivity” in trend analyses, and state that the use of econometric models and generalized linear models for quantifying loss trends is beyond the scope of the text.¹⁰

Werner and Modlin point out a number of factors that can influence loss trends, such as inflation, technological advances, societal changes, and distributional changes. They suggest we can estimate the effect of distributional changes by looking at the trend in average premium at present rate level (PPR).¹¹ Why do that? The reason is that distributional changes affect both premiums and losses. For example, youthful drivers generally have higher loss costs than adult drivers, and insurers

⁷ Anderson et al., p. 107.

⁸ Anderson et al., p. 108.

⁹ Styrsky, Chris, “The Effect of Changing Exposure Levels on Calendar Year Loss Trends,” *Casualty Actuarial Society Winter Forum*, 2005, pp. 125-151.

¹⁰ Werner and Modlin, pp. 111-114.

¹¹ Werner and Modlin, pp. 8, 81.

generally charge them higher premiums than adult drivers. Thus, if the proportion of youthful drivers in an insurance portfolio increases, both losses and premiums will increase.

As can be seen by the examples and citations above, the effect of changes in exposure level and distribution of exposures on commonly used univariate analysis of loss trend has been well studied and documented. The remaining question is what effect, if any, these exposure changes have on multivariate models.

1.2 Objectives

The objectives of this paper are (1) to illustrate how to expand a driver classification analysis database into a database that can also be used for univariate loss trend and loss development analyses as well as multivariate analyses involving all of these factors, (2) to compare results of using univariate and multivariate models for analyzing ratemaking parameters, (3) to show that multivariate analyses that account for all ratemaking parameters are robust to changes in exposure level and exposure distribution, and (4) to propose a framework for a rate review process completely based on multivariate analyses.

We begin by considering a line of insurance such, as property damage liability, with a relatively simple rating plan involving only territory and driver class. For simplicity, we assume any other rating factors such as anti-lock brake discounts or vehicle symbols are not applicable. We define subjects identified by policy ID and accident year, assuming that each policy insures one driver and one vehicle. Depending on rating manual rules, policies may insure multiple drivers and multiple vehicles. Some rating manuals specify rules for assigning a single driver classification to each vehicle. Other rating manuals assign a weighted average class factor, based on all drivers in the household, to each vehicle. When reviewing the rates and rating factors for a rating manual, the definition of a subject ID should be selected based on the entity to which manual rates and rating factors apply. SAS uses the keyword SUBJECT, but it can handle subjects as well as panels. A panel is a closely related group of subjects such as a household, or all vehicles and drivers insured by one policy, for which observations are expected to be correlated.

We will observe subjects across accident year evaluations, with cumulative claim counts per policy ID and accident year recorded at successive evaluation dates. For example, subject A, identified by policy ID 110000020 and accident year 2004, may have 0 claims as of 12 months, 1 claim as of 24 months, and 1 claim as of 36 months. In contrast, subject B, identified by policy ID 110000020 and accident year 2005, may have 0 claims at all evaluations (12, 24, and 36 months). This form of data organization is an example of longitudinal data, which Molenberghs and Verbeke

describe as the case where “the same characteristic is measured repeatedly over time, and time itself is, at least in part, a subject of scientific investigation.”¹² Please note that we are considering observations of the same policy on two different accident years as two different subjects. We could have considered the subject as identified only by policy ID and tracked the claim counts across both accident years and evaluation dates. However, the evaluations for 2005 as of 12 months and 2004 as of 24 months occur at the same time. Similarly, the 24-month evaluation of 2005 and the 36-month evaluation of 2004 are simultaneous. Having some observations precede each other in time while others are simultaneous makes model parameterization more complicated and beyond the scope of this paper.

Methods for analyzing longitudinal data include generalized estimating equations (GEE) and generalized linear mixed-effects models. We focus on population averaged GEE (PA GEE), which are closely related to generalized linear models (GLM). PA GEE can be thought of as “GLM” in which the variance function includes a covariance matrix that represents the correlation between repeated observations of the same subject or panel. Another difference is that the estimating equations for GLM involve likelihood functions, while GEE use quasilielihood functions. GLM have become standard tools in property and casualty insurance ratemaking. Thus, as we begin to think of insurance data as longitudinal data, it seems natural to use GEE as a tool for analyzing risk classification and time-related effects simultaneously. We will analyze claim frequency trend, claim count development, and claim frequency risk classification factors using SAS PROC GENMOD. We use PA GEE that model the marginal expectation for observations having the same covariate values (time index, evaluation age, territory, and driver class codes). Consequently, even though the inputs are observations from specific policyholders, the model provides information about “average” policyholders.

1.3 Outline

Section 2 of this paper outlines the theoretical background of population averaged generalized estimating equations (PA GEE), and introduces the database organization used as the common starting point for the analysis techniques discussed in this paper. Section 3 presents the results of applying several analysis techniques to simulated data to estimate classification effects (territory and driver class factors) and time-related effects (trend and loss development). The results compared and discussed include accident year claim frequency trend, percentage of cumulative claims closed with payment, and claim frequency relativities by risk classification.

¹² Molenberghs, G., and G. Verbeke, “Models for Discrete Longitudinal Data,” New York: Springer Series in Statistics, Springer Science+Business Media, Inc., 2005, p. 3.

PA GEE analyses involve making initial assumptions about the correlation structure of measurements taken on the same subject at different times. Therefore, two GEE analyses are presented and discussed: one assuming autoregressive correlation AR(1), and the other assuming all measurements related to the same subject are equally correlated — exchangeable correlation. The output of the model is a set of coefficients for the variables in the estimating equation, and a correlation matrix. For examples of correlation matrices output by these models see Section 2.2.1. Since the models use a log link, the exponential of the coefficients of the fully specified models correspond to the annual trend factor, territory and classification relativity factors, and percentage paid (closed with payment) factors.

2. THEORETICAL BACKGROUND AND DATA ORGANIZATION

This section provides a brief description of the mathematical structure of generalized linear models (GLM) and population averaged generalized estimating equations (PA GEE), describes the method of data organization used as the starting point for the analyses described in this paper, and shows how to prepare the data for application of the analysis techniques discussed in the paper. This paper uses only one type of GEE models: PA GEE. There are other types of GEE models, which are beyond of the scope of this paper. For more information, see Hardin and Hilbe.¹³

2.2 Generalized Linear Models

A Practitioner's Guide to Generalized Linear Models defines a GLM in terms of three components:¹⁴

- A random component \mathbf{Y} in which each element y_i is assumed to be independent and a member of the exponential family of distributions, for which the variance is a function of the expected value of Y , a scale parameter, and a weight assigned to each observation.
- A systematic component consisting of a set of explanatory or predictive variables, such as territory and driver classification, represented by a vector \mathbf{X} and a set of coefficients represented by a vector $\boldsymbol{\beta}$.
- A link function g such that

$$g(E[\mathbf{Y}]) = \mathbf{X}\boldsymbol{\beta} = x_1\beta_1 + x_2\beta_2 + \dots + x_n\beta_n \quad (2.1)$$

¹³ Hardin, James W. and Joseph M. Hilbe, *Generalized Estimation Equations*, Boca Raton, FL: Chapman & Hall/CRC, 2003.

¹⁴ Anderson et al., pp. 13, 14.

For example, if we used the natural logarithm as the link function g , then $E(\mathbf{Y}) = \exp(\mathbf{X}\boldsymbol{\beta})$. Thus, if x_i represents whether or not a policyholder resides in territory 1, the relativity for that territory would be given by $\exp(\beta_i)$.

2.2.1 The Independence Assumption

Suppose we are using the latest three accident years (e.g., 2007 to 2009) to evaluate driver classification factors for an insurance portfolio, and each policy insures one driver and one vehicle. Then, if a policyholder has been insured for three years the vector \mathbf{Y} has three entries for this policyholder corresponding to 2007, 2008, and 2009. For purposes of the GLM, it is customary to treat these observations as independent. There is no way to do otherwise, because this is one of the fundamental assumptions of GLM. However, they are likely to be correlated because they are observations of the same subject. Furthermore, the prevalence of safe driver insurance plans, accident and violation surcharges, and merit rating plans suggests that actuaries believe these observations are not really independent. In fact, most actuaries believe a policyholder who has had a claim is more likely to have claim in the future than a policyholder who has had no claims.

2.2 Population Averaged Generalized Estimating Equations

Suppose we observe cumulative claims closed with payment by policyholder by accident year from 2004 to 2009, at 12, 24, and 36 months. Further, assume all claims are closed by 36 months. Then we have 15 observations for each policyholder: three for each of the first four years, two for 2008 and one for 2009. Conversely, we have three missing observations: two for 2009 (the 24- and 36-month evaluations), and one for 2008 (the 36-month evaluation). The 18 total missing and non-missing observations correspond to six years and three evaluation dates for a policy in-force throughout the entire experience period. In this way we can see a policyholder's experience as longitudinal data in which the number of claims is observed at different points in time. We are interested in the relationship between time (accident year and evaluation date) and claim count, as well as the relationship between classification variables (territory and driver class) and claim count.

For the purposes of this paper, we will continue to assume independence across accident years, as is generally assumed when using GLM. Therefore we will define our subjects by policy ID and accident year, once again assuming each policy insures only one driver and one vehicle. It is easy to see that claims closed at different evaluation dates are correlated. For example a policyholder with one claim closed with payment as of 12 months for accident year 2007 will generally have at least one claim closed with payment at each successive evaluation date for that year. The way in which a company codes reopened claims can make this relationship more complicated. An actuary pricing a

book of business would have to understand how reopened claims are coded, and whether or not there has been a change in claim reopening patterns during the experience period. For the purposes of this paper, we assume there are no reopened claims. We will seek only to model the correlation among claims closed with payment at different evaluation dates. A more general correlation structure incorporating correlation across accident years could be formulated, but it is beyond the scope of this paper.

The data for an insurance portfolio observed at subsequent accident years and evaluation dates can be seen as having the random and systematic components of a GLM as well as a correlated component for which the GLM does not account. SAS PROC GENMOD can model the systematic component for data with both independent and correlated observations using the same linear predictor, variance function, and link function as the independent case, but it can also model the correlation structure of the correlated observations.¹⁵

Let \mathbf{Y}_i represent the vector of n_i observations for policyholder i , \mathbf{X}_{ij} the vector of covariates (explanatory or predictive variables) for the j^{th} observation of the i^{th} policyholder, $\boldsymbol{\beta}$ the vector of coefficients, and \mathbf{V}_i the covariance matrix of the n_i correlated observations in \mathbf{Y}_i . Then the GEE model can be specified by the following equations:

$$g(E[\mathbf{Y}_i]) = \mathbf{X}_i \boldsymbol{\beta} = x_{i1} \beta_1 + x_{i2} \beta_2 + \dots + x_{in} \beta_n, \quad (2.2)$$

$$\mathbf{V}_i = \varphi \mathbf{A}_i^{1/2} \mathbf{W}_i^{-1/2} \mathbf{R}(\alpha) \mathbf{W}_i^{-1/2} \mathbf{A}_i^{1/2}. \quad (2.3)$$

Where φ is a dispersion parameter, \mathbf{A} is a diagonal matrix of variance functions $v(\mu_{ij})$, \mathbf{W} is a diagonal matrix of weights, and $\mathbf{R}(\alpha)$ is a working correlation matrix. When no weights are specified by the user, \mathbf{W} defaults to a matrix of 1s, and all observations receive equal weight. When $\mathbf{R}(\alpha)$ is the identity matrix, equation 2.3 reduces to the variance function of the independent case.

2.2.1 Working Correlation Matrix

Six working correlation structures are available in SAS PROC GENMOD: fixed, identity, m -dependent, exchangeable, unstructured, and auto regressive AR(1). In the fixed case, the correlation matrix is specified by the user. The identity is the special case with 1s in the diagonal and 0s elsewhere, and it is equivalent to the independence case. M -dependent means that only m of the observations for a given subject are correlated, and the rest are not. Exchangeable is the case where all observations for a given subject are equally correlated. Unstructured implies that the correlation between any pair of observations is different from, and unrelated to, the correlation between any other pair of observations. The autoregressive structure is more appropriate for observations where

¹⁵ SAS/STAT® 9.2 User's Guide, SAS Institute Inc., 2008, pp. 1984, 1985.

the correlation decays as time elapses. Following are illustrations of the autoregressive and exchangeable correlation structures for a PA GEE model with simulated counts of personal auto property damage liability claims closed with payment observed at 12, 24, and 36 months per policy ID and accident year.

Autoregressive

$$\begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.8323 & 0.6927 \\ 0.8323 & 1 & 0.8323 \\ 0.6927 & 0.8323 & 1 \end{pmatrix}.$$

The autoregressive correlation matrix above would indicate that the correlation between two successive evaluations (12 months and 24 months or 24 months and 36 months) is roughly 83%, while the correlation between the 12-month and 36-month evaluations is about 69%. In contrast, the exchangeable correlation matrix below would indicate that all evaluations have a correlation of roughly 79%.

Exchangeable

$$\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.7916 & 0.7916 \\ 0.7916 & 1 & 0.7916 \\ 0.7916 & 0.7916 & 1 \end{pmatrix}.$$

For a given accident year, the claim count at 36 months is theoretically more correlated with the claim count at 24 months than with the count at 12 months. This would support using an autoregressive correlation structure. Nevertheless, both correlation structures shown above are tested in this paper for comparison purposes.

2.2.2 Missing Values

As mentioned previously, when one observes claim counts for a policyholder by accident year at different evaluation dates, some evaluation dates are missing. In the examples used in this paper, the latest year only has the 12-month evaluation, and the previous year has the 12- and 24-month evaluations. In cases such as this, the GENMOD procedure uses the “all available pairs” method to estimate the moments for the working correlation parameters. This method depends on the “missing completely at random (MCAR)” assumption.¹⁶

The pattern of missing values for a policyholder’s claim counts by accident year and evaluation date is somewhat systematic. For each accident year, either all evaluation dates are present, or they are all missing after some point. This is called a “dropout” missing pattern. It is similar to that of a

¹⁶ *SAS/STAT® 9.2 User’s Guide*, p. 1987.

patient who stops participating in a medical study. Additionally, policies may be written or non-renewed in the middle of the experience period, which creates another source of missing values. A non-renewed policy acts as a dropout. A new policy, on the other hand acts as a drop-in, where the earlier values are missing. Also, as some individuals become older or move, they may become part of another driver class or territory. Finally, insurance companies may not renew policies of people who have had many claims, or they may re-underwrite an insurance portfolio switching policyholders from one class to another if they are found to have been misclassified. Therefore, some of the factors causing missing values are systematic, while others are random. Determining whether or not the pattern of missing values for an insurance book meets the MCAR assumption is beyond the scope of this paper. Readers may refer to section 4.6 of Hardin and Hilbe's *Generalized Estimating Equations*.¹⁷

The data simulations run for this paper contain no missing values other than the ones that would correspond to future evaluation dates for the most recent accident years. The modeling results indicate that the inclusion of evaluation age parameters adequately accounts for the missing evaluation dates. Furthermore, movement of policyholders from one class or territory to another as a result of aging, moving, re-underwriting or non renewal can be seen as a distributional shift in exposures rather than a source of missing values. The data simulations do include samples with significant distributional changes, and the modeling results show that the claim frequency PA GEE models are not affected by distributional shifts in exposure. Therefore we can conclude that the missing values encountered when fitting a claim frequency PA GEE model to an insurance portfolio are not likely to adversely affect the modeling results.

2.3 Data Organization

Many companies have begun to build data warehouses or ratemaking databases with very detailed information including policy effective and expiration dates, driver attributes, vehicle attributes, date of accident, date of report, date closed, amounts paid, amounts in reserve, etc. At the start of a basic rate review, however, separate summarizations are extracted from this database for the trend system, loss development system, statewide indication, and territory and classification analysis review. Appendix G of *A Practitioner's Guide to Generalized Linear Models*¹⁸ presents several forms of data organization that can be used for generalized linear model (GLM) analysis, as well as their advantages and disadvantages. One of these is the calendar/accident year method in which each record has claim counts and loss amounts as of the latest evaluation. A simple expansion of this

¹⁷ Hardin and Hilbe.

¹⁸ Anderson et al.

database is to include separate columns for each evaluation age. This kind of setup can be used as a starting point from which, with very little manipulation, several univariate and multivariate analyses can be performed. The following table illustrates this method of organization for a hypothetical rating plan using only territory and driver classification and insuring only one driver and vehicle per policy. A real database would contain many other attributes identifying the policies, drivers and vehicles insured as well as rating characteristics associated with them.

Table 2.3.1

Policy Id	Accident Year	Territory	Driver Class	Earned Exposure	Claims With Payment Age 12	Claims With Payment Age 24	Claims With Payment Age 36
110000020	2004	1	1	1	0	1	1
110000020	2005	1	1	1	0	0	0
110000020	2006	1	1	1	0	0	0
110000020	2007	1	1	1	0	0	0
110000020	2008	1	1	1	1	1	
110000020	2009	1	1	1	0		

Paid Loss Amount Age 12	Paid Loss Amount Age 24	Paid Loss Amount Age 36	Claims Reported Age 12	Claims Reported Age 24	Claims Reported Age 36	Reported Losses Age12	Reported Losses Age24	Reported Losses Age36
0	25,000	25,000	1	1	1	20,000	25,000	25,000
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
15,000	15,000		1	1		15,000	15,000	
0			0			0		

The example above illustrates a policy for which the territory and driver class have not changed during the experience period. The only missing values are future evaluation dates for the latest two accident years, assuming that all claims have been reported by age 36. Suppose another policy had been written in 7/1/2005, then the earned exposure for that policy in 2005 would be 0.5 and the exposure and claim counts for 2004 would be missing.

Throughout this paper we assume each policy insures only one driver and one vehicle. Therefore, we use policy ID and year as the subject for our PA GEE models. In reality, most policies actually insure more than one driver and one vehicle. Some companies assign a specific driver to each vehicle on the policy, while others use an average driver factor for all vehicles in the policy. Actuaries wishing to use PA GEE models need to be mindful of the driver and vehicle assignment

procedure use in their specific book of business, and they may need to add driver ID or vehicle ID or both to the subject definition. An alternative is to analyze the data in terms of panels including all drivers and vehicles in each insured household.

Another issue that may arise involves claims not related to a specific driver or vehicle. For example, a minor child who is not a driver may be injured as a pedestrian and covered by medical payments. This could be handled in a number of ways: the claims may be excluded, the claims may be coded with a dummy policy ID and the driver and vehicle attributes of the at-fault driver, or they may be coded with a dummy policy ID and the base driver and vehicle attributes. The best course of action would have to be determined by the actuary working on a particular book of business, based on the available information.

Starting with a database structure such as the one above, it is very easy to summarize claim counts for different evaluation dates by accident year to obtain a claim count triangle for chain-ladder development. For details see Appendix C.

For a traditional classification analysis using a GLM with accident year as a dummy variable, the data can be summarized by keeping only the cumulative claims reported as of the latest evaluation date, as illustrated in the Table 2.3.2 below. This leads to one of the types of data organization in Appendix G of the *Practitioners Guide to GLM* in which there is some loss of some information for policies with multiple claims in the same accident year, but this is generally not material.¹⁹ Data such as the one illustrated below will be used for two types of models investigated in this paper: (1) GLM for claims closed with payment as the dependent variable and territory and classification as independent variables, and (2) GLM for claims closed with payment as the dependent variable and territory, classification, and dummy year as independent variables. As will be shown in Section 3.4, the dummy year parameter captures both trend and development effects. Immature year claim counts can be drastically lower than fully mature year claim counts. Trend, on the other hand, tends to be a gradual change. Therefore, modeling the combined effect of trend and development with a single continuous variable can be difficult. For this reason, it is better to use dummy year as a categorical variable.

¹⁹ Anderson et al., p. 109.

Table 2.3.2

Policy Id	Accident Year	Territory	Driver Class	Earned Exposure	Cumulative Claims Closed With Payment As of 12/31/2009
110000020	2004	1	1	1	1
110000020	2005	1	1	1	0
110000020	2006	1	1	1	0
110000020	2007	1	1	1	0
110000020	2008	1	1	1	1
110000020	2009	1	1	1	0

For GLM and PA GEE analyses involving loss development parameters in addition to trend and classification factors, we are interested in the repeated observations across evaluation dates, so we would stack the evaluation dates into one column in order to get one observed claim count per record as illustrated below in Table 2.3.3. Note that the earned exposure needs to be repeated so that the cumulative claims at each age can be associated with the corresponding accident year's earned exposure for the policy.

Table 2.3.3

Policy ID	Accident Year	Evaluation Date	Territory	Driver Class	Earned Exposure	Closed With Payment
110000020	2004	12	1	1	1	0
110000020	2004	24	1	1	1	1
110000020	2004	36	1	1	1	1
110000020	2005	12	1	1	1	0
110000020	2005	24	1	1	1	0
110000020	2005	36	1	1	1	0
110000020	2006	12	1	1	1	0
110000020	2006	24	1	1	1	0
110000020	2006	36	1	1	1	0
110000020	2007	12	1	1	1	0
110000020	2007	24	1	1	1	0
110000020	2007	36	1	1	1	0
110000020	2008	12	1	1	1	1
110000020	2008	24	1	1	1	1
110000020	2009	12	1	1	1	0

The reason for converting from the triangular format in Table 2.3.1 to a stacked format is that the software expects only one dependent variable. Three multivariate models explored in this paper involve the count of claims closed with payment as the dependent variable, and time index (for trend), territory, driver class, and evaluation date as independent variables. The first one is a GLM. The second one is a PA GEE with policy ID and accident year as subject identifiers and

autoregressive working correlation. The third one is a PA GEE with policy ID and accident year as subject identifiers and exchangeable working correlation.

3. CLAIM SIMULATIONS AND RESULTS

This section presents and compares the results of applying different techniques to 30 synthetic personal auto property damage portfolios: 10 scenarios for each of three hypothetical states X, Y, and Z. These portfolios have a very simple classification plan with only three territories and three driver classes. Details of the procedure used to create the portfolios and simulate the claim counts are provided in Appendix A.

The reasons for using synthetic portfolios are: (1) to generate claim databases with parameters known a priori, (2) to eliminate as much as possible random variation of the expected claim frequencies, and (3) to make change in exposure levels and distribution the predominant source of variation. The objective is to gauge the effect of changes in exposure level and exposure distribution on different analysis methods while holding everything else as constant as possible. To achieve this end, we select the following parameters: base claim frequency, annual frequency trend, percentage of claims closed with payment as of each evaluation age, territory relativity, and driver-class relativity. We then use these selected parameters to determine the expected claim frequency for each territory, driver class, accident year, and evaluation age. For example, given the following parameters:

- base frequency = 0.05.
- territory 1 relativity = 1.50.
- driver class 1 relativity = 1.00.
- percentage of claims paid (closed with payment) as of 12 months = 0.50.

We calculate the 2002 expected claim frequency for State X, Scenario 1, territory 1 and driver class 1, at age 12 as: $0.05 \times 1.50 \times 1.00 \times 0.50 = 0.0375$. With a 3% annual trend, the 2003 claim frequency at 12 months would be $0.0375 \times 1.03 = 0.038625$, and the 2004 claim frequency at 12 months would be $0.0375 \times 1.03^2 = 0.03978375$.

Next, we multiply the expected claim frequencies times the corresponding earned exposures in the portfolio to determine expected claim counts for each accident year, territory, and driver class combination. Once we have used the selected parameters to determine the expected claim count for each territory, driver class, accident year, and evaluation age, we select policies at random with replacement up to the number of expected claim counts. We consider a policy not selected to have

zero claims, a policy selected once to have one claim, a policy selected twice to have two claims, and so on. The synthetic portfolios with claim emergence simulations are available for downloading from the CAS Web Site.

State X scenarios assume the annual trend in claim frequency is zero. The main source of variation is changing exposure level from one year to the next. Changes in claim frequency distribution between territories and driver classes are limited to roughly one-tenth of a percentage point. Base frequency is 0.05 (for territory 2 and driver class 1); claim frequency relativities are constant—1.50 for territory 1, 0.80 for territory 3, 2.00 for driver class 2, and 0.75 for driver class 3. Cumulative percentages of claims paid are 50% at 12 months, 80% at 24 and 100% at 36.

State Y scenarios assume a 3% annual trend in claim frequency and increasing exposure from one year to the next. State Y Scenario 1 has essentially the same distribution of exposures across territories and driver class for each accident year as State X scenario 1. The rest of the State Y scenarios show more random variation than State X scenarios in the distribution of exposures among territories and driver classes from one accident year to the next. Base frequency is 0.06 (for territory 2 and driver class 1); claim frequency relativities and cumulative percentages of claims paid are the same as State X Scenario 1.

State Z scenarios assume a 3% annual trend in claim frequency and decreasing exposure from one year to the next. State Z Scenario 1 has essentially the same distribution of exposures across territories and driver class for each accident year as State X scenario 1. The rest of the State Z scenarios have increasing systematic variation in the distribution of exposures among territories and driver classes across accident years. Each accident year, the territory 1 class 1 earned car years decrease while the territory 3 class 3 earned car years increase, and the magnitude of this changes increases from scenario 2 to scenario 10. Base frequency is 0.02 (for territory 2 and driver class 1); claim frequency relativities and cumulative percentages of claims paid are the same as State X Scenario 1.

The following sections compare parameter estimates obtained by different methods for percentage of claims paid (closed with payment) by evaluation age, accident year trend, claim frequency relativities, quasi-likelihood information criterion, correlation matrices, and covariance matrices, where applicable.

3.1 Percentage of Claims Paid (Closed With Payment)

As mentioned earlier, the claim count simulation parameters were selected so the payment pattern would be approximately 50%, 80%, 100% of claims paid by 12, 24, and 36 months,

respectively. Since the number of claims must be a whole number, some deviation from those percentages is to be expected. For example, if the expected claim frequency for a given accident year, evaluation age, territory and driver class is 0.0375 and there are 1,000 earned car years, we can simulate either 37 or 38 claims, not 37.5.

The percentage paid estimate for the chain ladder method is the reciprocal of the age-to-ultimate development factor. Details of the calculation are shown in Appendix C. The percentage paid estimates for the GLM and GEE models are based on the parameter estimates for the levels of the evaluation age. Details are provided in Appendices D and E. The simulations used in this paper assumed stable development patterns. In a real-world situation, changes in claim adjustment patterns, system changes, etc., may cause development factors to change between years. If the change is gradual over several years, a marginal interaction term (based on time index and evaluation age) can be added to the model to account for these changes. If the change is more abrupt, so that accident years after a certain point are different from earlier accident years, a (0, 1) control variable could be introduced to account for the change. An actuary pricing a specific book of business would have to determine an appropriate course of action based on the available information.

The following Table 3.1.1 presents the resulting estimates of percentage of claims paid (closed with payment) by evaluation date using the chain ladder method (CLM), generalized linear model (GLM Full), generalized estimating equations with autoregressive correlation (GEE AR), and generalized estimating equations with exchangeable correlation (GEE Ex). All four methods produced estimates that are close to each other and close to the percentages used to set up the claim payment simulations.

Table 3.1.1

Sample	Percentage of Claims Paid by Age 12				Percentage of Claims Paid by Age 24			
	CLM	GLM Full	GEE AR	GEE Ex	CLM	GLM Full	GEE AR	GEE Ex
X-01	0.49997	0.50000	0.49998	0.49998	0.79998	0.80001	0.79999	0.79999
X-02	0.49999	0.50001	0.50000	0.50000	0.79999	0.80002	0.79999	0.80000
X-03	0.49999	0.50000	0.49999	0.50000	0.80000	0.80001	0.80000	0.80000
X-04	0.49996	0.49996	0.49996	0.49996	0.79997	0.79998	0.79998	0.79997
X-05	0.49999	0.50001	0.50000	0.50000	0.79999	0.80001	0.80000	0.80000
X-06	0.50001	0.50002	0.50001	0.50001	0.79999	0.79999	0.79999	0.79999
X-07	0.50003	0.50001	0.50002	0.50001	0.79999	0.79998	0.79999	0.79998
X-08	0.50002	0.50003	0.50002	0.50002	0.80002	0.80002	0.80002	0.80002
X-09	0.49999	0.49998	0.49999	0.49999	0.80002	0.80000	0.80001	0.80001
X-10	0.50000	0.49998	0.49999	0.49999	0.79999	0.79998	0.79998	0.79997
Y-01	0.50001	0.50002	0.50001	0.50002	0.79999	0.80000	0.80000	0.80000
Y-02	0.50000	0.49999	0.50000	0.50000	0.79999	0.79999	0.79999	0.80000
Y-03	0.50003	0.50003	0.50002	0.50003	0.80003	0.80004	0.80002	0.80003
Y-04	0.50002	0.50004	0.50002	0.50002	0.79999	0.80003	0.80000	0.80001
Y-05	0.50000	0.50000	0.50000	0.50000	0.79999	0.79998	0.79999	0.79999
Y-06	0.50003	0.50000	0.50002	0.50002	0.80001	0.79999	0.80000	0.80000
Y-07	0.50000	0.50002	0.50000	0.50000	0.80003	0.80003	0.80002	0.80002
Y-08	0.49997	0.49996	0.49997	0.49997	0.79999	0.79996	0.79998	0.79998
Y-09	0.50000	0.50002	0.50000	0.50000	0.79998	0.79999	0.79998	0.79998
Y-10	0.50000	0.50002	0.50001	0.50001	0.80000	0.80000	0.80000	0.80000
Z-01	0.49992	0.49988	0.49990	0.49990	0.80003	0.79997	0.80001	0.80001
Z-02	0.49993	0.49986	0.49991	0.49992	0.80001	0.79994	0.80000	0.80001
Z-03	0.49997	0.49997	0.49997	0.49996	0.80014	0.80010	0.80014	0.80012
Z-04	0.50005	0.50000	0.50004	0.50003	0.79999	0.79993	0.79998	0.79997
Z-05	0.50003	0.50009	0.50004	0.50004	0.80013	0.80019	0.80014	0.80015
Z-06	0.50007	0.50003	0.50006	0.50007	0.79996	0.79991	0.79995	0.79996
Z-07	0.49992	0.49983	0.49990	0.49991	0.80000	0.79992	0.79998	0.80000
Z-08	0.49990	0.49992	0.49991	0.49991	0.79997	0.79995	0.79997	0.79997
Z-09	0.50011	0.50007	0.50008	0.50009	0.80018	0.80026	0.80019	0.80021
Z-10	0.50004	0.50005	0.50004	0.50003	0.79997	0.79998	0.79998	0.79997
Average	0.50000	0.49999	0.50000	0.50000	0.80001	0.80001	0.80001	0.80001
Std Dev	0.00004	0.00006	0.00004	0.00004	0.00005	0.00007	0.00005	0.00005
Min	0.49990	0.49983	0.49990	0.49990	0.79996	0.79991	0.79995	0.79996
Max	0.50011	0.50009	0.50008	0.50009	0.80018	0.80026	0.80019	0.80021
Range	0.00021	0.00026	0.00018	0.00019	0.00022	0.00035	0.00024	0.00025

3.2 Claim Frequency Trend

The calendar year trend analysis is based on calendar year data — claim counts are assigned to the year in which the claim was paid. Details are provided in Appendix B. The accident year trend analysis is based on an exponential regression on ultimate claim frequencies, so it depends on the results of the chain ladder method. Details are shown in Appendix C. The accident year trend

estimates for the GLM Full, GEE AR, and GEE Ex models are based on the coefficient for a time index. Details of GLM and GEE models are provided in Appendices D and E.

The samples were intended to simulate changes in exposure level and changes in the distribution of exposures — two issues the basic ratemaking textbook mentions among the ones that can affect the results of univariate trend analyses, and illustrated in the introduction to this paper. Therefore, it is not surprising that the calendar year (Cal Yr) and accident year (Acc Yr) trend estimates deviate from the actual annual trend used to generate the simulated data: 0% for State X and 3% for States Y and Z.

The multivariate methods include a generalized linear model (GLM Full), generalized estimating equations with autoregressive correlation (GEE AR), and GEE with exchangeable correlation (GEE Ex). These methods are resistant to the changes in exposure level and distribution simulated in these samples. The following Table 3.2.1 summarizes the results.

Table 3.2.1

Sample	6-Point Annual Trend Estimates				
	Cal Yr	Acc Yr	GLM Full	GEE AR	GEE Ex
X-01	2.98%	0.00%	0.00%	0.00%	0.00%
X-02	2.97%	-0.04%	0.00%	0.00%	0.00%
X-03	2.96%	-0.02%	0.00%	0.00%	0.00%
X-04	2.96%	0.00%	0.00%	0.00%	0.00%
X-05	3.02%	-0.04%	0.00%	0.00%	0.00%
X-06	3.10%	0.10%	0.00%	0.00%	0.00%
X-07	2.86%	0.05%	0.00%	0.00%	0.00%
X-08	3.02%	0.02%	0.00%	0.00%	0.00%
X-09	2.97%	0.01%	0.00%	0.00%	0.00%
X-10	2.85%	-0.05%	0.00%	0.00%	0.00%
Average	2.97%	0.00%	0.00%	0.00%	0.00%
Std Dev	0.07%	0.05%	0.00%	0.00%	0.00%
Min	2.85%	-0.05%	0.00%	0.00%	0.00%
Max	3.10%	0.10%	0.00%	0.00%	0.00%
Range	0.25%	0.15%	0.00%	0.00%	0.00%

Table 3.2.1, continued

Sample	6-Point Annual Trend Estimates				
	Cal Yr	Acc Yr	GLM Full	GEE AR	GEE Ex
Y-01	3.00%	3.00%	3.00%	3.00%	3.00%
Y-02	3.09%	3.42%	3.00%	3.00%	3.00%
Y-03	2.80%	3.58%	3.00%	3.00%	3.00%
Y-04	3.26%	2.97%	3.00%	3.00%	3.00%
Y-05	3.24%	3.54%	3.00%	3.00%	3.00%
Y-06	3.97%	4.09%	3.00%	3.00%	3.00%
Y-07	2.64%	2.83%	3.00%	3.00%	3.00%
Y-08	2.76%	3.04%	3.00%	3.00%	3.00%
Y-09	3.52%	4.42%	3.00%	3.00%	3.00%
Y-10	2.76%	3.46%	3.00%	3.00%	3.00%
Average	3.10%	3.44%	3.00%	3.00%	3.00%
Std Dev	0.41%	0.51%	0.00%	0.00%	0.00%
Min	2.64%	2.83%	3.00%	3.00%	3.00%
Max	3.97%	4.42%	3.00%	3.00%	3.00%
Range	1.33%	1.59%	0.00%	0.00%	0.00%
Z-01	3.71%	3.01%	3.01%	3.01%	3.01%
Z-02	2.88%	2.47%	2.99%	2.98%	2.98%
Z-03	2.39%	2.28%	2.98%	2.99%	2.99%
Z-04	3.13%	2.39%	2.99%	2.99%	2.99%
Z-05	3.80%	3.33%	3.00%	3.01%	3.01%
Z-06	2.82%	3.00%	3.01%	3.01%	3.01%
Z-07	3.11%	2.07%	3.01%	3.00%	3.00%
Z-08	3.68%	3.04%	3.00%	3.00%	3.00%
Z-09	2.03%	1.62%	3.02%	3.02%	3.02%
Z-10	3.49%	2.29%	3.01%	3.01%	3.01%
Average	3.10%	2.55%	3.00%	3.00%	3.00%
Std Dev	0.59%	0.53%	0.01%	0.01%	0.01%
Min	2.03%	1.62%	2.98%	2.98%	2.98%
Max	3.80%	3.33%	3.02%	3.02%	3.02%
Range	1.77%	1.71%	0.04%	0.04%	0.04%

3.3 Claim Frequency Relativities

This section compares the results of six different multivariate models for claim frequency relativities — four generalized linear models and two generalized estimating equation (GEE) models. The autoregressive correlation model (GEE AR) makes more sense intuitively than the exchangeable correlation model (GEE Ex), since we would expect the correlation between 36-month and 24-month claim counts to be larger than the correlation between 36-month and 12-month claims counts.

Table 3.3.1

Model	Description
GLM 6Yr	Generalized linear model with latest 6 years of data and territory and driver class as independent variables
GLM 3Yr	Generalized linear model with latest 3 years of data and territory and driver class as independent variables
GLM AYC	Generalized linear model with latest 6 years of data, territory and driver class as independent variables, and accident year as control variable
GLM Full	Generalized linear model with latest 6 years of data, territory, driver class, time index, and evaluation age as independent variables
GEE AR	Generalized estimating equation model with latest 6 years of data, territory, driver class, time index, and evaluation age as independent variables and autoregressive working correlation
GEE Ex	Generalized estimating equation model with latest 6 years of data, territory, driver class, time index, and evaluation age as independent variables and exchangeable working correlation

The first two models, which ignore differences across accident year, are less reliable than the last four models — the range of expected values they produce is wider. Additionally, the base frequency (intercept) estimated by these models, which is the expected value across the 6-year or 3-year period, respectively, is understated because these models ignore the fact that the latest two years are not fully developed. The understatement is more pronounced for the 3-year model because two out of three years are not fully developed.

The model with accident year as a control variable (GLM AYC) and the fully specified models (GLM full, GEE AR, and GEE Ex) quite accurately predict the base frequency of 0.05 for State X, 0.06×1.03^2 for State Y, and 0.02×1.03^2 for State Z. The reason for the factor of 1.03 squared in States Y and Z is that a 3% annual trend was assumed in the simulation. Eight accident years were simulated starting with 2002, so to get the base frequency for 2004 we must multiply times 1.03 squared.

Following are the indicated base frequencies and the indicated factors for territories 1 and 3 as well as driver classes 2 and 3. Since the states have different base frequencies, the statistics (average, standard deviation, minimum, maximum, and range) for the intercept are by state. On the other hand, territory and driver class relativities are the same for all states so the statistics are across all 30 samples.

Table 3.3.2

Sample	Indicated Base Frequency (Intercept)					
	GLM 6Yr	GLM 3Yr	GLM AYC	GLM full	GEE AR	GEE Ex
X-01	0.04456	0.03899	0.04999	0.05000	0.05000	0.05000
X-02	0.04458	0.03900	0.05000	0.05000	0.05000	0.05000
X-03	0.04458	0.03900	0.04999	0.05000	0.04999	0.05000
X-04	0.04457	0.03902	0.05000	0.05000	0.05000	0.05000
X-05	0.04454	0.03898	0.05000	0.05000	0.05000	0.05000
X-06	0.04455	0.03894	0.05000	0.05000	0.05000	0.05000
X-07	0.04459	0.03900	0.04999	0.05000	0.05000	0.05000
X-08	0.04461	0.03904	0.05000	0.05000	0.05000	0.05000
X-09	0.04458	0.03904	0.05000	0.05001	0.05001	0.05001
X-10	0.04455	0.03897	0.04999	0.05000	0.05000	0.05000
Average	0.04457	0.03900	0.05000	0.05000	0.05000	0.05000
Std Dev	0.00002	0.00003	0.00001	0.00000	0.00000	0.00000
Min	0.04454	0.03894	0.04999	0.05000	0.04999	0.05000
Max	0.04461	0.03904	0.05000	0.05001	0.05001	0.05001
Range	0.00007	0.00010	0.00001	0.00001	0.00002	0.00001
Y-01	0.05942	0.05405	0.06365	0.06365	0.06365	0.06365
Y-02	0.05939	0.05395	0.06364	0.06364	0.06365	0.06365
Y-03	0.05952	0.05394	0.06365	0.06365	0.06365	0.06365
Y-04	0.05915	0.05322	0.06365	0.06365	0.06365	0.06365
Y-05	0.05987	0.05428	0.06366	0.06366	0.06366	0.06366
Y-06	0.05946	0.05418	0.06365	0.06365	0.06366	0.06366
Y-07	0.05919	0.05355	0.06365	0.06366	0.06366	0.06366
Y-08	0.05942	0.05398	0.06365	0.06365	0.06365	0.06365
Y-09	0.05979	0.05438	0.06366	0.06366	0.06366	0.06366
Y-10	0.06004	0.05444	0.06364	0.06364	0.06364	0.06364
Average	0.05953	0.05400	0.06365	0.06365	0.06365	0.06365
Std Dev	0.00029	0.00038	0.00001	0.00001	0.00001	0.00001
Min	0.05915	0.05322	0.06364	0.06364	0.06364	0.06364
Max	0.06004	0.05444	0.06366	0.06366	0.06366	0.06366
Range	0.00089	0.00122	0.00002	0.00002	0.00002	0.00002
Z-01	0.02039	0.01855	0.02120	0.02121	0.02121	0.02121
Z-02	0.02048	0.01878	0.02123	0.02123	0.02123	0.02123
Z-03	0.02033	0.01845	0.02123	0.02123	0.02123	0.02123
Z-04	0.02049	0.01877	0.02123	0.02122	0.02123	0.02123
Z-05	0.02058	0.01882	0.02122	0.02122	0.02122	0.02122
Z-06	0.02054	0.01887	0.02122	0.02122	0.02122	0.02122
Z-07	0.02048	0.01877	0.02122	0.02122	0.02122	0.02122
Z-08	0.02057	0.01868	0.02123	0.02122	0.02122	0.02122
Z-09	0.02048	0.01870	0.02122	0.02121	0.02121	0.02121
Z-10	0.02055	0.01871	0.02121	0.02121	0.02121	0.02121
Average	0.02049	0.01871	0.02122	0.02122	0.02122	0.02122
Std Dev	0.00008	0.00013	0.00001	0.00001	0.00001	0.00001
Min	0.02033	0.01845	0.02120	0.02121	0.02121	0.02121
Max	0.02058	0.01887	0.02123	0.02123	0.02123	0.02123
Range	0.00025	0.00042	0.00003	0.00002	0.00002	0.00002

Table 3.3.3

Sample	Indicated Rating Factors for Territory 1					
	GLM 6Yr	GLM 3Yr	GLM AYC	GLM full	GEE AR	GEE Ex
X-01	1.49993	1.49999	1.49993	1.49995	1.49992	1.49991
X-02	1.50005	1.50007	1.50019	1.50010	1.50018	1.50024
X-03	1.49988	1.49958	1.50008	1.50005	1.50007	1.50007
X-04	1.49945	1.49933	1.49979	1.49987	1.49979	1.49976
X-05	1.50103	1.50050	1.50011	1.50012	1.50010	1.50009
X-06	1.49981	1.50208	1.50010	1.50010	1.50011	1.50012
X-07	1.49841	1.49927	1.50006	1.49999	1.50005	1.50004
X-08	1.49819	1.49945	1.49982	1.49982	1.49981	1.49978
X-09	1.49848	1.49521	1.49985	1.49982	1.49982	1.49982
X-10	1.49993	1.49660	1.49996	1.50001	1.49995	1.49995
Y-01	1.50014	1.50002	1.50014	1.50010	1.50014	1.50015
Y-02	1.50996	1.52085	1.50002	1.50004	1.50002	1.50003
Y-03	1.48432	1.48797	1.49996	1.49997	1.49996	1.49998
Y-04	1.51082	1.53509	1.50015	1.50015	1.50014	1.50013
Y-05	1.49118	1.49591	1.49992	1.49987	1.49993	1.49994
Y-06	1.49503	1.48626	1.50006	1.50004	1.50005	1.50005
Y-07	1.49821	1.48654	1.50004	1.50005	1.50005	1.50008
Y-08	1.50404	1.50523	1.50006	1.50009	1.50005	1.50004
Y-09	1.49080	1.50031	1.49992	1.49991	1.49992	1.49991
Y-10	1.48157	1.47763	1.50024	1.50023	1.50024	1.50027
Z-01	1.50076	1.50097	1.50076	1.50048	1.50060	1.50054
Z-02	1.49198	1.47865	1.49935	1.49905	1.49924	1.49932
Z-03	1.51716	1.52657	1.49950	1.49934	1.49944	1.49943
Z-04	1.48905	1.46689	1.49996	1.49987	1.49994	1.50001
Z-05	1.48917	1.47332	1.50004	1.50002	1.50005	1.50006
Z-06	1.49382	1.46522	1.49948	1.49946	1.49952	1.49951
Z-07	1.50011	1.47678	1.50008	1.49989	1.50009	1.50018
Z-08	1.50212	1.50813	1.49975	1.49973	1.49977	1.49988
Z-09	1.50589	1.49210	1.50026	1.50002	1.50020	1.50025
Z-10	1.50486	1.50288	1.49970	1.49987	1.49965	1.49964
Average	1.49854	1.49598	1.49998	1.49993	1.49996	1.49997
Std Dev	0.00755	0.01569	0.00026	0.00027	0.00026	0.00026
Min	1.48157	1.46522	1.49935	1.49905	1.49924	1.49932
Max	1.51716	1.53509	1.50076	1.50048	1.50060	1.50054
Range	0.03559	0.06987	0.00141	0.00143	0.00136	0.00122

Table 3.3.4

Sample	Indicated Rating Factors for Territory 3					
	GLM 6Yr	GLM 3Yr	GLM AYC	GLM full	GEE AR	GEE Ex
X-01	0.80010	0.80006	0.80010	0.80002	0.80008	0.80009
X-02	0.79978	0.80021	0.80012	0.80011	0.80013	0.80016
X-03	0.80027	0.80159	0.80020	0.80014	0.80019	0.80017
X-04	0.79964	0.79893	0.79978	0.79982	0.79981	0.79983
X-05	0.79959	0.80026	0.80022	0.80020	0.80024	0.80026
X-06	0.80058	0.79988	0.80002	0.79997	0.80001	0.80002
X-07	0.79895	0.79799	0.80004	0.79988	0.80001	0.80004
X-08	0.80119	0.80236	0.80008	0.80005	0.80011	0.80010
X-09	0.80050	0.79748	0.80002	0.80003	0.80001	0.79999
X-10	0.80284	0.80567	0.80004	0.80007	0.80004	0.80002
Y-01	0.80008	0.80014	0.80008	0.80009	0.80008	0.80007
Y-02	0.80474	0.81198	0.80020	0.80025	0.80022	0.80022
Y-03	0.79996	0.79703	0.80004	0.80004	0.80005	0.80005
Y-04	0.80056	0.81225	0.79999	0.79999	0.79999	0.79998
Y-05	0.79223	0.78969	0.79993	0.79991	0.79992	0.79992
Y-06	0.80515	0.80170	0.80005	0.80013	0.80005	0.80003
Y-07	0.79247	0.79569	0.79997	0.79992	0.79996	0.79995
Y-08	0.81657	0.82273	0.80005	0.80004	0.80006	0.80009
Y-09	0.80232	0.80505	0.80002	0.79998	0.80003	0.80006
Y-10	0.80437	0.81305	0.80009	0.80014	0.80009	0.80007
Z-01	0.80078	0.80073	0.80078	0.80047	0.80073	0.80071
Z-02	0.79741	0.78948	0.79980	0.79989	0.79974	0.79977
Z-03	0.79353	0.78907	0.79999	0.79989	0.80000	0.80003
Z-04	0.80211	0.80804	0.80009	0.80003	0.80003	0.80003
Z-05	0.80568	0.81250	0.79969	0.79969	0.79972	0.79977
Z-06	0.79660	0.79430	0.79915	0.79923	0.79908	0.79895
Z-07	0.79043	0.77850	0.80080	0.80080	0.80080	0.80082
Z-08	0.79660	0.80615	0.79973	0.79970	0.79974	0.79977
Z-09	0.79301	0.79431	0.79983	0.80021	0.79987	0.79988
Z-10	0.80309	0.81325	0.79956	0.79970	0.79949	0.79945
Average	0.80004	0.80134	0.80002	0.80001	0.80001	0.80001
Std Dev	0.00504	0.00897	0.00030	0.00027	0.00031	0.00032
Min	0.79043	0.77850	0.79915	0.79923	0.79908	0.79895
Max	0.81657	0.82273	0.80080	0.80080	0.80080	0.80082
Range	0.02614	0.04423	0.00165	0.00157	0.00172	0.00187

Table 3.3.5

Sample	Indicated Rating Factors for Driver Class 2					
	GLM 6Yr	GLM 3Yr	GLM AYC	GLM full	GEE AR	GEE Ex
X-01	2.00009	1.99986	2.00009	2.00004	2.00012	2.00014
X-02	2.00066	2.00230	1.99982	1.99989	1.99986	1.99985
X-03	1.99854	1.99650	2.00006	2.00007	2.00006	2.00005
X-04	2.00077	1.99887	2.00003	1.99996	2.00005	2.00009
X-05	2.00279	2.00338	1.99985	1.99990	1.99989	1.99991
X-06	1.99924	2.00293	1.99977	1.99975	1.99979	1.99979
X-07	1.99568	1.99215	2.00015	2.00013	2.00014	2.00016
X-08	1.99883	1.99855	2.00006	2.00004	2.00005	2.00007
X-09	2.00018	1.99959	1.99974	1.99969	1.99974	1.99975
X-10	2.00015	1.99842	2.00013	2.00013	2.00017	2.00020
Y-01	2.00006	2.00005	2.00007	2.00002	2.00007	2.00009
Y-02	1.97740	1.97396	1.99992	1.99996	1.99993	1.99993
Y-03	1.98442	1.99056	2.00009	2.00014	2.00011	2.00014
Y-04	2.01840	2.04209	2.00028	2.00028	2.00027	2.00028
Y-05	1.98411	1.99068	1.99987	1.99986	1.99987	1.99987
Y-06	1.99856	1.99691	1.99970	1.99975	1.99971	1.99968
Y-07	1.99981	1.99845	1.99983	1.99983	1.99983	1.99984
Y-08	1.97674	1.96067	1.99988	2.00000	1.99988	1.99985
Y-09	1.97072	1.97999	1.99995	2.00000	1.99995	1.99996
Y-10	1.98134	1.98471	2.00005	2.00008	2.00005	2.00007
Z-01	2.00018	2.00082	2.00018	1.99964	2.00001	1.99996
Z-02	1.99715	1.97901	2.00017	1.99982	1.99999	2.00000
Z-03	2.00068	1.99935	2.00027	2.00033	2.00028	2.00029
Z-04	2.00443	1.99737	1.99930	1.99991	1.99926	1.99908
Z-05	1.97562	1.96000	1.99954	1.99945	1.99958	1.99958
Z-06	1.97755	1.95312	1.99959	1.99968	1.99968	1.99964
Z-07	2.00170	1.99607	1.99935	1.99957	1.99943	1.99957
Z-08	1.97254	1.96864	1.99958	1.99958	1.99950	1.99936
Z-09	1.98454	1.94168	1.99926	1.99915	1.99917	1.99920
Z-10	1.98520	1.97520	2.00101	2.00086	2.00103	2.00110
Average	1.99294	1.98940	1.99992	1.99992	1.99992	1.99992
Std Dev	0.01170	0.01931	0.00035	0.00031	0.00035	0.00037
Min	1.97072	1.94168	1.99926	1.99915	1.99917	1.99908
Max	2.01840	2.04209	2.00101	2.00086	2.00103	2.00110
Range	0.04768	0.10041	0.00175	0.00171	0.00186	0.00202

Table 3.3.6

Sample	Indicated Rating Factors for Driver Class 3					
	GLM 6Yr	GLM 3Yr	GLM AYC	GLM full	GEE AR	GEE Ex
X-01	0.75033	0.75020	0.75033	0.75022	0.75031	0.75035
X-02	0.74927	0.74938	0.74990	0.74997	0.74989	0.74987
X-03	0.74992	0.74999	0.75041	0.75039	0.75040	0.75042
X-04	0.75011	0.74824	0.75026	0.75015	0.75026	0.75031
X-05	0.75116	0.75143	0.75005	0.75002	0.75003	0.75006
X-06	0.75119	0.75154	0.74996	0.74991	0.74993	0.74992
X-07	0.74900	0.74611	0.75025	0.75023	0.75023	0.75024
X-08	0.75114	0.75338	0.75010	0.75009	0.75010	0.75010
X-09	0.75032	0.75210	0.75000	0.75003	0.75001	0.75004
X-10	0.74825	0.74701	0.74994	0.74999	0.74994	0.74994
Y-01	0.75017	0.75013	0.75017	0.75015	0.75017	0.75018
Y-02	0.75196	0.75315	0.75008	0.75004	0.75006	0.75005
Y-03	0.75468	0.76405	0.74997	0.74997	0.74996	0.74994
Y-04	0.75830	0.76694	0.75007	0.75005	0.75006	0.75004
Y-05	0.73462	0.72896	0.75014	0.75010	0.75015	0.75020
Y-06	0.76605	0.77589	0.74993	0.74997	0.74994	0.74993
Y-07	0.75268	0.75035	0.75004	0.75009	0.75004	0.75002
Y-08	0.73668	0.72737	0.74982	0.74983	0.74982	0.74982
Y-09	0.74774	0.73586	0.74995	0.74992	0.74994	0.74994
Y-10	0.74138	0.74617	0.75010	0.75009	0.75011	0.75011
Z-01	0.74969	0.74946	0.74969	0.74972	0.74964	0.74959
Z-02	0.75050	0.75676	0.74997	0.75015	0.75003	0.74995
Z-03	0.74261	0.73244	0.74954	0.74963	0.74956	0.74957
Z-04	0.74387	0.73412	0.74948	0.74961	0.74953	0.74961
Z-05	0.73876	0.74166	0.74960	0.74985	0.74964	0.74954
Z-06	0.73622	0.72883	0.75042	0.75055	0.75044	0.75038
Z-07	0.74471	0.74383	0.75039	0.75055	0.75041	0.75044
Z-08	0.74853	0.74874	0.75010	0.75020	0.75006	0.74996
Z-09	0.74277	0.74448	0.74966	0.74974	0.74971	0.74971
Z-10	0.74335	0.74649	0.75037	0.75014	0.75043	0.75053
Average	0.74787	0.74750	0.75002	0.75005	0.75003	0.75003
Std Dev	0.00659	0.01083	0.00025	0.00023	0.00025	0.00026
Min	0.73462	0.72737	0.74948	0.74961	0.74953	0.74954
Max	0.76605	0.77589	0.75042	0.75055	0.75044	0.75053
Range	0.03143	0.04852	0.00094	0.00094	0.00091	0.00099

3.4 Accident Year as a Control (Dummy or Nuisance) Variable

What happens when we use accident year as a control variable? We get an indicated factor for each accident year that combines trend and development effects. For State X, which has 0% trend, the factors are 1.00 for accident years 2005 through 2007, 0.80 for 2008 (80% of claims paid or 1.25 development factor) and 0.50 for 2009 (50% of claims paid or 2.00 development factor). For States Y and Z, which have 3% trend, the factors are 1.03 for 2005, 1.03^2 for 2006, 1.03^3 for 2007, $1.03^4 \times$

0.80 for 2008, and $1.03^5 \times 0.50$ for 2009. Thus, the trend and development factors could be derived from the accident year parameters. Nevertheless, it would be preferable to model them explicitly as shown in Section 3.3 for the GLM full, GEE AR, and GEE Ex models. Following are the accident year parameters for the GLM AYC model with accident year as control variable.

Table 3.4.1

Sample	Indicated Accident Year Control Factors				
	2005	2006	2007	2008	2009
X-01	1.00016	1.00010	0.99991	0.80016	0.50005
X-02	0.99983	0.99995	0.99983	0.80006	0.49987
X-03	1.00018	1.00013	1.00012	0.80018	0.50005
X-04	1.00032	1.00020	1.00014	0.80019	0.50011
X-05	1.00018	1.00018	0.99977	0.80007	0.49998
X-06	1.00001	1.00009	0.99995	0.79997	0.50011
X-07	1.00022	1.00007	1.00009	0.80001	0.50004
X-08	1.00005	1.00002	0.99983	0.80000	0.50001
X-09	0.99999	1.00018	1.00004	0.79992	0.49997
X-10	1.00032	1.00021	1.00018	0.80012	0.49999
Average	1.00013	1.00011	0.99999	0.80007	0.50002
Std Dev	0.00016	0.00008	0.00015	0.00009	0.00007
Min	0.99983	0.99995	0.99977	0.79992	0.49987
Max	1.00032	1.00021	1.00018	0.80019	0.50011
Range	0.00049	0.00026	0.00041	0.00027	0.00024
Y-01	1.02995	1.06083	1.09256	0.90034	0.57961
Y-02	1.03005	1.06105	1.09284	0.90055	0.57973
Y-03	1.03008	1.06092	1.09273	0.90055	0.57965
Y-04	1.03000	1.06084	1.09261	0.90049	0.57963
Y-05	1.03005	1.06085	1.09274	0.90027	0.57961
Y-06	1.03019	1.06108	1.09301	0.90051	0.57969
Y-07	1.03016	1.06102	1.09268	0.90051	0.57969
Y-08	1.03000	1.06105	1.09291	0.90042	0.57975
Y-09	1.02998	1.06085	1.09248	0.90031	0.57962
Y-10	1.03006	1.06093	1.09275	0.90044	0.57972
Average	1.03005	1.06094	1.09273	0.90044	0.57967
Std Dev	0.00008	0.00010	0.00016	0.00010	0.00005
Min	1.02995	1.06083	1.09248	0.90027	0.57961
Max	1.03019	1.06108	1.09301	0.90055	0.57975
Range	0.00024	0.00025	0.00053	0.00028	0.00014

Sample	Indicated Accident Year Control Factors				
	2005	2006	2007	2008	2009
Z-01	1.03016	1.06112	1.09339	0.90049	0.57971
Z-02	1.03020	1.06086	1.09261	0.89975	0.57907
Z-03	1.02959	1.05993	1.09240	0.89959	0.57929
Z-04	1.02926	1.06157	1.09253	0.89975	0.57948
Z-05	1.03017	1.06099	1.09287	0.90103	0.57993
Z-06	1.02962	1.06147	1.09307	0.90042	0.57999
Z-07	1.03019	1.06115	1.09327	0.90025	0.57953
Z-08	1.02930	1.06042	1.09280	0.90002	0.57956
Z-09	1.02966	1.06144	1.09347	0.90176	0.57975
Z-10	1.03018	1.06104	1.09301	0.90066	0.58002
Average	1.02983	1.06100	1.09294	0.90037	0.57963
Std Dev	0.00039	0.00050	0.00037	0.00067	0.00031
Min	1.02926	1.05993	1.09240	0.89959	0.57907
Max	1.03020	1.06157	1.09347	0.90176	0.58002
Range	0.00094	0.00164	0.00107	0.00217	0.00095

3.5 Quasi-Likelihood Information Criterion

The quasi-likelihood information criterion (QIC) provides a means for choosing between working correlation assumptions for GEE models. A model with a lower QIC is preferable. Based on the QIC results, the GEE with autoregressive correlation fits the synthetic data slightly better than the GEE with exchangeable correlation. As mentioned at the beginning of section 3.3, arguments can be made for using an autoregressive working correlation when the repeated measures are cumulative claim counts at different evaluation ages, so the results are not surprising.

Table 3.5.1

State	Scenario	GEE AR	GEE Ex	Smaller
X	01	368,314.52	368,314.76	AR
X	02	368,739.38	368,739.62	AR
X	03	369,161.31	369,161.55	AR
X	04	369,083.32	369,083.56	AR
X	05	368,673.52	368,673.76	AR
X	06	367,446.39	367,446.62	AR
X	07	367,186.50	367,186.74	AR
X	08	366,262.28	366,262.52	AR
X	09	367,637.24	367,637.47	AR
X	10	368,568.34	368,568.58	AR
Y	01	538,890.02	538,890.25	AR
Y	02	530,240.61	530,240.84	AR
Y	03	531,000.33	531,000.55	AR
Y	04	538,734.71	538,734.93	AR
Y	05	534,610.13	534,610.36	AR
Y	06	536,405.47	536,405.70	AR
Y	07	540,098.23	540,098.45	AR
Y	08	530,921.97	530,922.20	AR
Y	09	540,988.29	540,988.52	AR
Y	10	531,617.72	531,617.95	AR
Z	01	130,980.23	130,980.46	AR
Z	02	125,526.50	125,526.72	AR
Z	03	125,799.49	125,799.72	AR
Z	04	126,938.68	126,938.91	AR
Z	05	125,043.45	125,043.68	AR
Z	06	125,115.15	125,115.37	AR
Z	07	125,787.05	125,787.27	AR
Z	08	126,253.36	126,253.59	AR
Z	09	127,081.81	127,082.04	AR
Z	10	127,521.55	127,521.78	AR

SAS PROC GENMOD also calculates the QICu. This is an approximation to the QIC that can be used to choose between models, but it is not appropriate for choosing between working correlations. The theory of quasi-likelihood functions and the details of the QIC are beyond the scope of this paper. Interested readers are encouraged to consult McCullagh and Nelder’s *Generalized Linear Models*, Hardin and Hilbe’s *Generalized Estimating Equations*, or Pan’s *Akaike’s Information Criterion in Generalized Estimating Equations*. For complete bibliographical information see the references section.

3.6 Covariance Matrices

The REPEATED statement in SAS PROC GENMOD has the options MCOVB and ECOVB. When these options are used, the procedure outputs both the model-based (also called naïve) covariance matrix and the empirical (also called robust) covariance matrix for the model's parameters. If these two matrices are similar, it is a sign that the choice of working correlation matrix is adequate. If they are substantially different, then a different working correlation structure would be more appropriate. The GEE models used in this paper had eight parameters, corresponding to the variables and classification levels listed below. Therefore, each covariance matrix is an 8x8 matrix, and with thirty simulations and two models there are thirty pairs of matrices to compare. Unfortunately, there is no automated way of doing this. It requires visual inspection and judgment.

Table 3.6.1

Parameter	Effect	class	Territory	eval_date
Prm1	Intercept			
Prm2	time_index			
Prm3	territory		1	
Prm4	territory		3	
Prm5	driver_class	2		
Prm6	driver_class	3		
Prm7	eval_date			12
Prm8	eval_date			24

Both the autoregressive and exchangeable working correlations resulted in models where the model-based and empirical correlation matrices were similar. This implies that both the autoregressive and exchangeable working correlations result in models that fit the simulated data reasonably well. As mentioned in section 3.5, however, the autoregressive correlation is preferable both in terms of the QIC results as well as from an intuitive understanding of development factors. The covariance matrices are not listed in this paper, but they can be downloaded from the CAS web site as Excel files.

3.7 Confidence Intervals

SAS PROC GENMOD provides confidence intervals for the parameter estimates for the models discussed in this paper. This output will be available for downloading from the CAS web site. A potential use of these confidence intervals would be to develop risk loads to take into consideration when developing final rates, or as input to an enterprise risk management model, but such topics are beyond the scope of this paper.

3.8 Quarterly Data

For simplicity, the examples and database used in this paper used accident year data. A company may decide to design a database containing accident quarters instead. When working with quarterly data one of the considerations is the effect of seasonality. A common way to deal with seasonality is to work with 12-month rolling averages. This would tend to complicate the calculations needed to produce the input files for GLM or GEE models taking into account trend, development and classification. A more simple solution would be to use control variables to account for seasonal differences in accident quarters. Dickmann and Merz did so in a paper about loss trend.²⁰

3.9 Why Go Back When We Can Go Forward?

This paper has shown that multivariate frequency models incorporating all available information are resistant to changes in exposure level and changes in distribution of exposures. A next step would be to examine the resistance of different models to things that can affect estimates of ultimate claim severity and ultimate losses such as changes in loss payment patterns or changes in reserving practices.

Once we have multivariate estimates of trend and loss development, should we go back and apply them to total losses by accident year to perform a statewide indication? Why not use the estimates of trend, development, territory relativities and driver class relativities to calculate prospective loss costs directly? For example, the parameter estimates (coefficients) for the State Y, Scenario 1 GEE AR frequency model result in the equation:

$$\ln(E[f]) = -2.75436 + 0.4056T_1 - 0.2230T_3 + 0.6932C_2 - 0.2875C_3 + 0.0295t.$$

From which it follows that the expected frequency is:

$$E[f] = 0.06365 \times 1.500^{T_1} \times 0.800^{T_3} \times 2.000^{C_2} \times 0.750^{C_3} \times 1.030^t.$$

Where T_1 and T_3 are variables that take the value 0 or 1 depending on whether or not a policy is from Territory 1 or Territory 3. Similarly, C_2 and C_3 are 0 or 1 depending on whether or not the

²⁰ Dickmann, Kurt S., and James R. Merz, "Consideration in Estimating Loss Cost Trends," *Casualty Actuarial Society Forum*, Winter 2001, pp. 21-60.

driver classification code is 1 or 3, and t is a time index that increases by 1 every year. The parameters corresponding to the percentage of claims paid by 12 months and 24 months have been omitted since we are only interested in the ultimate claim frequency. By picking an appropriate value for the time index we can project the expected frequency appropriate for the future period in which rates will be in effect. A similar equation can be determined for claim severity as well as pure premium. Thus it is possible to obtain four estimates of prospective loss costs:

- Prospective Frequency \times Prospective Severity based on paid data
- Prospective Loss Costs based on paid data
- Prospective Frequency \times Prospective Severity based on reported data
- Prospective Loss Costs based on reported data

4. CONCLUSIONS

By organizing data as illustrated in Section 2.3 we can easily fit univariate and multivariate models for both time-dependent effects, such as loss trend and loss development, as well as classification effects such as territory and driver class.

Univariate models of loss trend can over- or underestimate the trend when there are significant changes in the level or in the distribution of exposures.

Modeling trend and development explicitly is preferable to using accident year as a control or dummy variable.

Multivariate models that incorporate all the available information — differences across accident years such as trend and loss development, and differences among classification groups — are resistant to changes in exposure level and changes in exposure distribution.

Acknowledgments

I would like to thank Kurt Dickmann and Phil Baum for introducing me to basic ratemaking techniques. Thanks are due to Lisa Monard, who allowed me to work on driver classification relativity reviews and vehicle rating factors for liability, medical payments and no-fault, and to Frank Gribbon who recommended reading Stephen Mildenhall's paper on "Minimum Bias and Generalized Linear Models" back in the late 1990s. I am also grateful to Rob Curry for the opportunity to work on predictive modeling, and to Fred Klinker, John Baldan, and members of the CAS Ratemaking Committee for their helpful comments. I am also grateful to my wife, Emelda, for her patience and support as I labored on the simulations and programs that formed the basis for this paper.

Supplementary Material

The synthetic datasets used as inputs for the models in this paper, and the covariance matrices output by the generalized estimating equation models are stored electronically on the CAS Web Site and available for downloading. SAS code is provided in the appendices.

Appendix A – Claim Emergence Simulations

The synthetic data sets used in this paper were created using a two step process: (1) generate an exposure scenario, and (2) generate paid claim counts. The synthetic data consist of 30 scenarios divided into three hypothetical states (X, Y, and Z) with 10 scenarios per state. They are meant to approximate what one might see for a short tail line of business such as personal auto property damage liability.

The objective of the States X, Y, and Z simulations was to test the sensitivity of models to changes in exposure level and changes in exposure distribution. In order to achieve that goal it was necessary to find a claim count generation process that approximated as much as possible the expected claim counts, leaving changes in exposure level and exposure distribution as the predominant sources of variation.

The first step was to generate an exposure scenario. This was accomplished by preparing an input file with total exposures per year, and the percentage of exposures corresponding to each combination of territory and driver class, as shown below for State X, Scenario 01.

Table A.1

Calendar Accident Year	Earned Car Years	Terr 1 Class 1	Terr 1 Class 2	Terr 1 Class 3	Terr 2 Class 1	Terr 2 Class 2	Terr 2 Class 3	Terr 3 Class 1	Terr 3 Class 2	Terr 3 Class 3
2002	160000	0.17	0.08	0.07	0.23	0.14	0.10	0.12	0.05	0.04
2003	176800	0.17	0.08	0.07	0.23	0.14	0.10	0.12	0.05	0.04
2004	198016	0.17	0.08	0.07	0.23	0.14	0.10	0.12	0.05	0.04
2005	215837	0.17	0.08	0.07	0.23	0.14	0.10	0.12	0.05	0.04
2006	232025	0.17	0.08	0.07	0.23	0.14	0.10	0.12	0.05	0.04
2007	225064	0.17	0.08	0.07	0.23	0.14	0.10	0.12	0.05	0.04
2008	211560	0.17	0.08	0.07	0.23	0.14	0.10	0.12	0.05	0.04
2009	192520	0.17	0.08	0.07	0.23	0.14	0.10	0.12	0.05	0.04

In the example above, the objective was to generate claim counts in which the predominant source of variation was the change in exposure level. Hence the distribution of exposures was kept constant across accident years. For other State X scenarios the percentage was allowed to change across years by roughly one-tenth of one percent. So for example, in some years the Territory 1

Class 1 percentage may have been 17.1% while in others it might have been 16.9%. The reason for the small allowed change in exposure distribution was that the main objective of the State X scenarios was to test the effect of changes in overall exposure level. On the other hand, for State Y Scenario 02, the percentage of exposures for Territory 1 Class 1 was allowed to decrease from 15.9% in 2002 to 13.0% in 2009. For States Y and Z, both the level of exposures and the distribution of exposures were allowed to change.

The percentage for a territory and driver class combination was multiplied times the total exposures, to get the subtotal corresponding to that combination. For example, in Table A.1 above, 10% of exposures correspond to Territory 2 Class 3, and the total accident year 2002 earned exposures are 160,000. Therefore, 16,000 exposures correspond to Territory 2 Class 3. The process generated 16,000 records with one earned car-year each. This is not entirely realistic, since for most companies some policies are cancelled midyear. However, midyear cancellations are a small proportion of the book for most companies. The following SAS code excerpt illustrates the process of generating the exposure records.

```
do territory = 1 to 3;
  do class = 1 to 3;
    exposure_percentage = exposure_portion{ territory, class };
    car_years = round( exposure * exposure_percentage , 1 );
    do k = 1 to car_years;
      policy_id = put( territory, z1. ) || put( class, z1. ) || put( k, z7. );
      earned_exposure = 1;
      output;
    end;
  end;
end;
```

For states X, Y, and Z, the next step was to calculate the expected claim counts for each accident year, territory, driver class and evaluation age based on the parameters selected for base frequency, territory relativities, driver class relativities, and percentage of claims paid (closed with payment) at each evaluation age for each accident year. Table A.2 shows the parameters selected for State X, Scenario 01. For example, for Territory 2 Class 3 as of 12 months the expected claim count is $16,000 \text{ exposures} \times 0.05 \text{ base frequency} \times 1.00 \text{ Territory 2 factor} \times 0.75 \text{ Class 3 factor} \times 0.50 \text{ percentage reported as of 12 month evaluation age} = 300$.

Table A.2

Calendar Accident Year	Base Frequency	Terr 1	Terr 2	Terr 3	Class 1	Class 2	Class 3	Age 12	Age 24	Age 36
2002	0.05	1.50	1.00	0.80	1.00	2.00	0.75	0.50	0.30	0.20
2003	0.05	1.50	1.00	0.80	1.00	2.00	0.75	0.50	0.30	0.20
2004	0.05	1.50	1.00	0.80	1.00	2.00	0.75	0.50	0.30	0.20
2005	0.05	1.50	1.00	0.80	1.00	2.00	0.75	0.50	0.30	0.20
2006	0.05	1.50	1.00	0.80	1.00	2.00	0.75	0.50	0.30	0.20
2007	0.05	1.50	1.00	0.80	1.00	2.00	0.75	0.50	0.30	0.20
2008	0.05	1.50	1.00	0.80	1.00	2.00	0.75	0.50	0.30	0.20
2009	0.05	1.50	1.00	0.80	1.00	2.00	0.75	0.50	0.30	0.20

The expected claim counts were stored in a variable called `_NSIZE_`. This is a special variable used by PROC SURVEYSELECT to determine how many records to select from each stratum (accident year, territory, driver class, and evaluation age combination). The amount is rounded to the nearest whole number because claim counts are whole numbers.

```
_NSIZE_ = round( earned_exposure * base_frequency * terr_factor * class_factor
* age_percentage , 1 );
```

The expected claim counts are used to randomly select policies with replacement from each stratum using SAS PROC SURVEYSELECT. Policies not selected are considered to have zero claims, those selected one or more times are considered to have one or more claims. The number of hits determines the number of claims.

```
proc surveyselect data=for_selection ( drop = earned_exposure )
out=work.policies_with_claim
method=urs
sampsiz=work.expected_claim_counts
( index = ( ytcpa = ( year territory class age policy_id ) ) )
;
strata year territory class age;
id year territory class age policy_id ;
run;
```

The 30 synthetic data sets in policy detail, as well as summaries by territory, driver class, accident year, and evaluation age, are available for downloading from the Casualty Actuarial Society's Web Site.

Appendix B – Calendar Year Trend

Calendar year calculations were used in this paper only to illustrate the distorting effect of changing exposure levels on calendar year trend analysis. Only claim frequency trend examples were provided, but the phenomenon occurs for claim severity trend and pure premium trend as well. For more details, the reader should refer to the CAS “Basic Ratemaking” electronic textbook or to Chris Styrsky’s paper “The Effect of Changing Exposure Levels on Calendar Year Trend,” which are listed in the references. To prepare for calendar year trend analysis, diagonals must be subtracted to determine the claims paid during the year. It is also necessary to include only complete calendar years in the analysis. Since calendar year claims may relate to prior accident years, the database may not have enough prior accident years to get a complete calendar year. For example, if the database includes accident years 2002 through 2009 and it takes three years for an accident year to develop to ultimate, then the first calendar year for which complete claim counts can be calculated is 2004. This means that only six complete calendar years can be calculated from the database. This number can be specified in a macro variable to ensure that only complete calendar years are used. See SAS code below.

```
%let state = X;
%let scenario = 01;
%let complete_cal_years = 6;

* calculate difference in diagonals ;
data for_cal_yr_trend;
  set mylib.state&state.&scenario.d;
  retain last_cal_year 0;
  cal_year = year;
  paid_count = paid_count12;
  output;
  if paid_count24 not = . then do;
    earned_exposure = 0;
    paid_count = paid_count24 - paid_count12;
    cal_year = year + 1;
    output;
  end;
  if paid_count36 not = . then do;
    earned_exposure = 0;
    paid_count = paid_count36 - paid_count24;
    cal_year = year + 2;
    output;
  end;
  if last_cal_year < cal_year then do;
    last_cal_year = cal_year;
    call symput('last_cal_year',put(last_cal_year,4.));
  end;
run;

* sum diagonal differences corresponding to each calendar years ;
* include only calendar years with a complete set of differences ;
%let first_year = %eval(&last_cal_year. - &complete_cal_years. + 1);
proc summary nway missing data=for_cal_yr_trend
```

Towards Multivariate Ratemaking—Claim Frequency Analysis Examples

```
( where = ( cal_year not < &first_year. ) );
class cal_year;
var earned_exposure paid_count;
output out=cal_yr sum=;
run;

* calculate claim frequency for each calendar year ;
data cal_yr_freq;
set cal_yr ( drop = _type_ _freq_ );
claim_frequency = paid_count / earned_exposure;
log_claim_frequency = log ( claim_frequency );
time_index = cal_year - &last_cal_year.;
run;

title "Calendar Year Frequency Trend Data, State &state., Scenario &scenario.";
proc print label noobs data=cal_yr_freq split='_';
var cal_year earned_exposure paid_count
    claim_frequency log_claim_frequency;
format earned_exposure paid_count comma9.
    claim_frequency log_claim_frequency 7.5;
label cal_year = 'Calendar Year'
    earned_exposure = 'Earned Car Years'
    paid_count = 'Paid Claim Count'
    claim_frequency = 'Claim Frequency'
    log_claim_frequency = 'Log of Claim Frequency';
run;

* fit exponential regression model;
proc reg data=cal_yr_freq outest=cy_trend;
trend_model: model log_claim_frequency = time_index / noprint; run;
quit;

* calculate annual trend based on model output ;
data freq_factor;
set cy_trend;
time_index = round(time_index, 0.0000001 );
trend_factor = round( exp( time_index ), 0.0000001 );
annual_trend = trend_factor - 1;
format time_index trend_factor 10.7 annual_trend percentn7.2;;
label trend_factor = 'Annual Trend Factor' time_index = 'Time Index Parameter'
    annual_trend = 'Annual Trend';
keep time_index trend_factor annual_trend;
run;

title "Calendar Year Frequency Trend, State &state., Scenario &scenario.";
proc print noobs label data=freq_factor;
run;
```

Appendix C – Chain Ladder Development and Accident Year Trend

To prepare for chain ladder claim count development, the claim counts were summarized by accident year. Next, age-to-age factors were calculated based on the claim count triangle. The average of all years was calculated and used as the selected link ratio. In practice, an actuary might select a different loss development factor based on knowledge of the book of business, changes in claim practices, or other information. However, this is just simulated data, so the only factor affecting the data is random variation.

Towards Multivariate Ratemaking—Claim Frequency Analysis Examples

The age-to-ultimate link ratios are the cumulative product of the selected link ratios, and the percentage reported estimates are the reciprocal of the age to ultimate factors. The SAS code below illustrates this process.

```
%let state = S;
%let scenario = 01;

* calculate claim triangle ;
proc summary nway missing data=mylib.state&state.&scenario.d;
  class year;
  var earned_exposure paid_count12 paid_count24 paid_count36 paid_count48;
  output out=claim_triangle sum=;
run;

* calculate age to age factors ;
data age_to_age;
  set claim_triangle;
  call symput('last_year',put(year,4.)); * macro var to be used for evaluation date ;
  if paid_count24 > 0 then f12 = round( paid_count24 / paid_count12, 0.0000001 );
  else delete;
  if paid_count36 > 0 then f24 = round( paid_count36 / paid_count24, 0.0000001 );
  if paid_count48 > 0 then do;
    f36 = round( paid_count48 / paid_count36, 0.0000001 );
    f48 = 1;
  end;
  format f12 f24 f36 f48 10.7;
  length factor_type $ 16;
  factor_type = 'Age to Age';
  label factor_type = 'Factor Type' f12 = 'Age 12' f24 = 'Age 24'
    f36 = 'Age 36' f48 = 'Age 48';
run;

* calculate average of all available years ;
proc summary nway missing data=age_to_age;
  var f12 f24 f36 f48;
  output out=z_averages mean=;
run;

* calculate age to ultimate link ratios and percentage reported ;
data link_ratios;
  set age_to_age( in = a keep = factor_type year f12 f24 f36 f48 )
    z_averages ( keep = f12 f24 f36 f48 );
  if a then output;
  else do;
    factor_type = 'All-Year Average';
    f12 = round( f12, 0.0000001 );
    f24 = round( f24, 0.0000001 );
    f36 = round( f36, 0.0000001 );
    f48 = round( f48, 0.0000001 );
    output;
    factor_type = 'Age to Ultimate';
    f12 = round( f12 * f24 * f36 * f48 , 0.0000001 );
    f24 = round( f24 * f36 * f48 , 0.0000001 );
    f36 = round( f36 * f48 , 0.0000001 );
    output;
    factor_type = 'Percent Reported';
    f12 = round( 1 / f12, 0.0000001 );
    f24 = round( 1 / f24, 0.0000001 );
    f36 = round( 1 / f36, 0.0000001 );
    f48 = round( 1 / f48, 0.0000001 );
    output;
  end;
run;
```

Towards Multivariate Ratemaking—Claim Frequency Analysis Examples

Ultimate claim counts are the product of claim counts reported as of the last evaluation times the age-to-ultimate factor. The estimated ultimate claim counts are used to calculate the claim frequency for each year. Then an exponential regression is fit to these claim frequencies to determine the claim frequency trend. See SAS code below.

```
data developed_claims;
  set claim_triangle;
  if _n_ = 1 then set link_ratios ( drop = year where = ( factor_type = 'Age to Ultimate' ) );
  time_index = year - &last_year.;
  eval_year = &last_year.;
  eval_date = '12/31/'||put(eval_year,4.);
  select ( year );
    when ( &last_year.      ) do; age_to_ult = f12; cumulative_claims = paid_count12; end;
    when ( &last_year. - 1 ) do; age_to_ult = f24; cumulative_claims = paid_count24; end;
    when ( &last_year. - 2 ) do; age_to_ult = f36; cumulative_claims = paid_count36; end;
    otherwise
      do; age_to_ult = f48; cumulative_claims = paid_count48; end;
  end;
  ultimate_claims = round( cumulative_claims * age_to_ult, 1 );
  claim_frequency = round( ultimate_claims / earned_exposure, 0.0000001 );
  log_claim_frequency = round( log( claim_frequency ), 0.0000001 );;
  label
    time_index = 'Time Index'
    eval_date = 'Evaluation Date'
    claim_frequency = 'Claim Frequency'
    cumulative_claims = "Reported Claim Count"
    age_to_ult = 'Age to Ultimate Development Factor'
    ultimate_claims = 'Ultimate Claim Count'
  ;
  format earned_exposure comma9. cumulative_claims ultimate_claims comma7.0
    claim_frequency log_claim_frequency age_to_ult 10.7;
  keep eval_date year time_index earned_exposure cumulative_claims age_to_ult
    ultimate_claims claim_frequency log_claim_frequency;
run;

title2 'Claim Frequency Trend Analysis';
proc reg data=developed_claims outest=cf_parms;
  model log_claim_frequency = time_index;
  ods select ParameterEstimates;
run;
quit;

data freq_factor;
  set cf_parms;
  time_index = round(time_index, 0.0000001 );
  trend_factor = round( exp( time_index ), 0.0000001 );
  annual_trend = trend_factor - 1;
  format time_index trend_factor 10.7 annual_trend percentn7.2;;
  label trend_factor = 'Annual Trend Factor' time_index = 'Time Index Parameter'
    annual_trend = 'Annual Trend';
  keep time_index trend_factor annual_trend;
run;

data for_exhibit;
  merge claim_triangle ( keep = year earned_exposure paid_count12 paid_count24 paid_count36 )
    link_ratios ( keep = factor_type f12 f24 f36 )
    developed_claims ( keep = time_index ultimate_claims );
run;

title "Claims Closed With Payment, State &state., Scenario &scenario.";
proc print data=for_exhibit noobs label;
  var year paid_count12 paid_count24 paid_count36 factor_type f12 f24 f36 ultimate_claims;
  format paid_count12 paid_count24 paid_count36 ultimate_claims comma8.0 f12 f24 f36 8.5;
```

Towards Multivariate Ratemaking—Claim Frequency Analysis Examples

```
label year = 'Accident Year' paid_count12 = 'Paid Count Age 12'
      paid_count24 = 'Paid Count Age 24' paid_count36 = 'Paid Count Age 36'
;
run;

proc print noobs label data=freq_factor;
run;
```

Appendix D – Generalized Linear Models

Four generalized linear models (GLM) were tested for this paper as shown in Table D.1.

Table D.1

GLM 6Yr	Generalized linear model with latest 6 years of data and territory and driver class as independent variables
GLM 3Yr	Generalized linear model with latest 3 years of data and territory and driver class as independent variables
GLM AYC	Generalized linear model with latest 6 years of data, territory and driver class as independent variables, and accident year as control variable
GLM Full	Generalized linear model with latest 6 years of data, territory, driver class, time index, and evaluation age as independent variables

The first three models use only the latest evaluation for each calendar/accident year. The fourth model (GLM Full) uses territory, driver class, a time index (for trend), and evaluation age (for development) as independent variables and claim count as the dependent variable. In order to include evaluation age in the model, it is necessary to transpose the evaluation age columns into rows, and to create a variable to identify the evaluation age. The data are assumed to reach ultimate value at 36 months. Therefore, a policy that has been in force for the entire experience period has three observations for each mature accident year, two for the penultimate accident year, and one for the latest accident year. The 36-month evaluation is the reference level, so the 12-month and 24-month parameters are relativities to the 36-month or ultimate claim count. This means that they correspond to the percentage paid as of 12 or 24 months respectively. The age-to-ultimate development factor is the reciprocal of the percentage reported. The SAS code below performs the data preparation and model fitting.

```
%let state = X;
%let scenario = 01;
%let first_year = 2004;
%let ref_year = 2007;
%let last_year = 2009;
```

Towards Multivariate Ratemaking—Claim Frequency Analysis Examples

```
%let evals = 3;

* prepare data for claim frequency Generalized Linear Models ;
* for territory and driver class only ;
* for territory and driver class with accident year as control variable ;
data sample_glm;
set mylib.state&state.&scenario.d;
if year < &first_year. or year > &last_year. then delete;
select;
  when ( year = &last_year.      ) paid_claim_count = paid_count12;
  when ( year = &last_year. - 1 ) paid_claim_count = paid_count24;
  when ( year = &last_year. - 2 ) paid_claim_count = paid_count36;
  otherwise paid_claim_count = paid_count48;
end;
Driver_Class = class;
log_exposure = log( earned_exposure );
keep year territory driver_class policy_id earned_exposure
    paid_claim_count log_exposure
;
run;

title1 "Paid Claim Frequency GLM for Territory and Driver Class Only";
title2 "State &state, Scenario &scenario., Six Accident Years";
proc genmod data=sample_glm;
  class driver_class ( ref = '1' ) territory ( ref = '2' )
    / param = ref;
  model paid_claim_count = territory driver_class
    / link=log dist=Poisson offset=log_exposure scale=p;
run;

title1 "Paid Claim Frequency GLM for Territory and Driver Class Only";
title2 "State &state, Scenario &scenario., Three Accident Years";
%let starting_year = &Last_Year. - 2 ;
proc genmod data=sample_glm ( where = ( year >= &starting_year. ) );
  class driver_class ( ref = '1' ) territory ( ref = '2' )
    / param = ref;
  model paid_claim_count = territory driver_class
    / link=log dist=Poisson offset=log_exposure scale=p;
run;

title1 "Paid Claim Frequency GLM for Territory and Driver Class";
title2 "With Accident Year as Control Variable";
title3 "State &state, Scenario = &scenario.";
proc genmod data=sample_glm;
  class driver_class ( ref = '1' ) territory ( ref = '2' ) year ( ref = "&ref_year." )
    / param = ref;
  model paid_claim_count = territory driver_class year
    / link=log dist=Poisson offset=log_exposure scale=p;
run;

* prepare data for Generalized Linear Model ;
* for territory factors, driver class factors ;
* trend factor and loss development factors ;
data sample_glm2;
set mylib.state&state.&scenario.d ( where = ( year not < &first_year. ) );
array paid_cnt {4} paid_count12 paid_count24 paid_count36 paid_count48;
array eval_dates {4} $ ('12' '24' '36' '48');
time_index = year - &ref_year.;
Years = "&First_Year. to &Last_Year.";
driver_class = class;
label
  time_index = 'Time Index'
  log_exposure = 'Natural Log of Exposure'
  driver_class = 'Driver Class'
  paid_claim_count = "Paid Claim Count"
;
do k = 1 to &evals.;
  eval_date = eval_dates{k};
```

```
if k = &evals. then call symput('last_eval',eval_date);
if paid_cnt{k} > . then do;
  paid_claim_count = paid_cnt{k};
  log_exposure = log ( earned_exposure );
  output;
end;
end;
keep Years year territory driver_class earned_exposure
  log_exposure time_index eval_date paid_claim_count
;
run;

title1 "Paid Claim Frequency GLM for State &state, Scenario = &scenario.";
title2 "Territory, Driver Class, Trend and Loss Development";
proc genmod data=sample_glm2;
  class driver_class ( ref = '1' ) territory ( ref = '2' ) eval_date ( ref = "&last_eval." )
  / param = ref;
  model paid_claim_count= territory driver_class time_index eval_date
  / link=log dist=Poisson offset=log_exposure scale=p;
run;
```

Appendix E – Generalized Estimating Equations

The first step in fitting generalized estimating equations (GEE) was to create a modeling sample. The GEE algorithm depends on the definition of a subject or panel. A database with a large number of policies, each treated as a subject or panel, can cause the program to run out of memory on a desktop personal computer. The sampling algorithm shown below first classifies policies depending on whether or not they had a claim reported, even if it was closed without payment. Then it selects the entire six accident year history for the policy. All policies with a reported claim are included in the modeling sample, but only ten percent of the claim-free policies are selected for each territory and driver class combination.

The hypothetical ratemaking database used in this paper has only territory and driver class as rating factors. A real database would have other rating variables. If other variables are used in the rating plan, they should be included in the definition of the strata from which the ten percent samples are taken. They should also be included in the generalized estimating equation, either as predictors or as part of the offset term.

Additionally, sampling weights must be calculated to reflect the original number of observations in the database. The procedure uses the ratio of the original number of observations to the number of observations in the sample. So for policies that had at least one reported claim in the six-year accident history, the weight is 1, and for policies that were claim-free the weight is close to 10. The reason the weight is not always exactly equal to 10 for the claim-free policies is that the original number of observations may not have been a multiple of 10, so ten percent would not have been a whole number. Therefore the nearest whole number of policies had to be selected. SAS procedure SURVEYSELECT was used to select the 10% of claim-free policies. The SURVEYSELECT

procedure calculates sampling weights, but those weights were not used because the input to the procedure was just the list of claim-free policies. Some of them may have had 15 observations (all six-accident years), while other policies may have had less. Therefore, the percentage of policies does not exactly equal the percentage of observations selected, and the weights had to be calculated manually once the entire history for each policy had been selected. The SAS code below performs the sampling procedure and calculates the sampling weights.

```
%let state = X;
%let scenario = 01;
%let first_year = 2004;
%let ref_year = 2007;
%let last_year = 2009;
%let evals = 3;

* prepare data sample containing ;
* * all policies with at least one reported claim ;
* * 1 out of every 10 policies with no reported claim ;

proc sql noprint;
  * find all policies with at least one reported claim in experience period ;
  create table id_with_claim as select unique policy_id
  from mylib.state&state.&scenario.d
  where ( incurred_loss12 > 0 or incurred_loss24 > 0
  or incurred_loss36 > 0 or incurred_loss48 > 0 )
  order by policy_id;
  * sort data by policy id ;
  create table scenario as select *
  from mylib.state&state.&scenario.d
  where ( year not < &first_year. )
  order by policy_id;
quit;

* split data into policies with at least one reported claim and those claim free ;
data with_claim claim_free;
  merge scenario ( in = a ) id_with_claim ( in = b );
  by policy_id;
  driver_class = class;
  if a and b
  then output with_claim;
  else if a then output claim_free;
run;

* determine all combinations of territory, driver ;
* class, and policy id for claim free policies ;
proc sql noprint;
  create table id_claim_free as select unique territory, driver_class, policy_id
  from claim_free
  order by territory, driver_class, policy_id;
quit;

* select 10% of the claim-free policies for each territory and driver class ;
proc surveyselect data=id_claim_free
  out=claim_free_sampled method=SRS rate=0.10;
  STRATA territory driver_class;
run;

* now that we have a list of selected policies for each territory ;
* and driver class, select all the data for those policies ;
proc sort data=claim_free_sampled ( drop = SelectionProb SamplingWeight );
  by policy_id;
```


Towards Multivariate Ratemaking—Claim Frequency Analysis Examples

```
run;
data claim_free_selected;
merge claim_free(in = a ) claim_free_sampled( in = b );
  by policy_id;
  if a and b;
run;

* combined the selected claim free data and the data with ;
* at least one reported claim to create modeling sample ;
data selected;
set claim_free_selected( in = a ) with_claim( in = b );
  if a then with_claim = 0;
  if b then with_claim = 1;
  unity = 1;
run;

* count the number of observations in each stratum in the sample ;
proc summary nway missing data=selected;
  class year territory driver_class with_claim;
  var unity;
  output out=sample_counts sum(unity)=sample_record_count;
run;

* classify all the original data depending on whether
* or not the policy had a reported claim ;
data original;
set claim_free( in = a ) with_claim( in = b );
  if a then with_claim = 0;
  if b then with_claim = 1;
  unity = 1;
run;

* count the number of observations in each stratum in the original data ;
proc summary nway missing data=original;
  class year territory driver_class with_claim;
  var unity;
  output out=original_counts sum(unity)=original_record_count;
run;

* calculate weights equal to the ratio of the number of observations ;
* in the original data to the number of observations in the sample ;
* for each stratum ;
data sample_weights;
merge original_counts ( drop = _type_ _freq_ )
  sample_counts ( drop = _type_ _freq_ );
  by year territory driver_class with_claim;
  sampling_weight = original_record_count / sample_record_count;
run;

* merge sampling weights with modeling sample ;
proc sort
  data=selected( drop = unity )
  out=selected_for_merge;
  by year territory driver_class with_claim;
run;
data modeling_sample;
merge selected_for_merge sample_weights;
  by year territory driver_class with_claim;
run;
```

The next step in preparing the data for GEE is to transpose the accident year evaluation dates into rows. A new variable, `eval_date`, identifies the evaluation date for each record. Furthermore, the policy ID and year date are concatenated into one variable, `policy_id_year`, which will be used to

identify each subject; the evaluation age identifies the repeated claim count observations for each policy-id-year subject. For the latest three years, some of the claim counts are missing because they correspond to evaluation dates that will occur in the future. These records are omitted.

```
* prepare data for Generalized Estimating Equations ;
data sample_gee;
set modeling_sample;
array paid_cnt {4} paid_count12 paid_count24 paid_count36 paid_count48;
array eval_dates {4} $ ('12' '24' '36' '48');
time_index = year - &ref_year.;
Years = "&First_Year. to &Last_Year.";
driver_class = class;
policy_id_year = policy_id || put( year, 4. );
label
  time_index = 'Time Index'
  log_exposure = 'Natural Log of Exposure'
  driver_class = 'Driver Class'
  paid_claim_count = "Paid Claim Count"
  eval_date = 'Evaluation Date'
  policy_id_year = 'Policy Id and Year'
;
do k = 1 to &evals.;
  eval_date = eval_dates{k};
  if k = &evals. then call symput('last_eval',eval_date);
  if paid_cnt{k} > . then do;
    paid_claim_count = paid_cnt{k};
    log_exposure = log ( earned_exposure );
    output;
  end;
end;
keep Years year territory driver_class policy_id_year eval_date
  earned_exposure log_exposure sampling_weight time_index paid_claim_count
;
run;
```

Two GEE models are tested in this paper. The first one uses an autoregressive working correlation structure, and the second one uses exchangeable working correlation. The autoregressive correlation structure assumes the correlation between successive evaluation ages is stronger than the correlation between evaluation ages that are further apart. The exchangeable working correlation assumes the correlation between any two of evaluation ages is the same. Following is the SAS code for these two models. Proc TEMPLATE is used to increase the number of decimal places output for the parameter estimate.

```
proc template;
edit Stat.GENMOD.GEEEst;
define Estimate;
  header = "Estimate";
  format = 10.6;
end;
end;
run;
```

Towards Multivariate Ratemaking—Claim Frequency Analysis Examples

```
title1 "Paid Claim Frequency GEE with Autoregressive Working Correlation";
title2 "For Territory, Driver Class, Trend and Loss Development";
title3 "State &state, Scenario &scenario.";
proc genmod data=sample_gee;
  weight sampling_weight;
  class policy_id_year eval_date
  driver_class ( ref = '1' ) territory ( ref = '2' ) eval_date ( ref = "&last_eval." )
  / param = ref;
  model paid_claim_count= territory driver_class time_index eval_date
  / link=log dist=Poisson offset=log_exposure scale=p;
  repeated subject = policy_id_year
  / withinsubject = eval_date corr=AR corrw mcovb ecovb;
run;
```

```
title1 "Paid Claim Frequency GEE with Exchangeable Working Correlation";
title2 "For Territory, Driver Class, Trend and Loss Development";
title3 "State &state, Scenario &scenario.";
proc genmod data=sample_gee;
  weight sampling_weight;
  class policy_id_year eval_date
  driver_class ( ref = '1' ) territory ( ref = '2' ) eval_date ( ref = "&last_eval." )
  / param = ref;
  model paid_claim_count = territory driver_class time_index eval_date
  / link=log dist=Poisson offset=log_exposure scale=p;
  repeated subject = policy_id_year
  / withinsubject = eval_date corr=exch corrw mcovb ecovb;
run;
```

5. REFERENCES

- [1.] Anderson, Duncan, and Sholom Feldblum, and Claudine Modlin, and Doris Schirmacher, and Ernesto Schirmacher, and Neeza Thandi, "A Practitioner's Guide to Generalized Linear Models," 3rd ed., CAS Study Note, 2007.
- [2.] Hardin, J. W., and J.M. Hilbe, *Generalized Estimating Equations*, New York: Chapman & Hall/CRC, 2003.
- [3.] Horton, Nicholas J., and Stuart R. Lipsitz, "Review of Software to Fit Generalized Estimating Equation Regression Models," *The American Statistician* 53, 1999, pp. 160-169.
- [4.] Lo, Chi Ho, and Wing Kam Fung, and Zhong Yi Zhu, "Structural Parameter Estimation Using Generalized Estimating Equations for Regression Credibility Models," *ASTIN Bulletin* 37(2), 2007, pp. 323-343.
- [5.] McCullagh, P., "Quasi-likelihood Functions," *Annals of Statistics* 11, 1983, pp. 59-67.
- [6.] McCullagh, P., and John A. Nelder, *Generalized Linear Models*, New York: Chapman and Hall, 1989.
- [7.] Molenberghs, Geert, and Geert Verbeke, "Models for Discrete Longitudinal Data," New York: Springer Series in Statistics, Springer Science+Business Media, Inc., 2005.
- [8.] Pan, W., "Akaike's Information Criterion in Generalized Estimating Equations," *Biometrics*, 57, 2001, pp. 120-125.
- [9.] Mildenhall, S., "Minimum Bias and Generalized Linear Models," *Proceedings of the Casualty Actuarial Society* LXXVI, 1999, pp. 393-487.
- [10.] "SAS/STAT® 9.2 User's Guide," 2008, SAS Institute Inc., 2008.
- [11.] Werner, Geoff and Claudine Modlin, *Basic Ratemaking*, 3rd ed., Arlington, VA: Casualty Actuarial Society, January 2010.
- [12.] Zeger, S. L., K.Y. Liang, K. Y., and P.S. Albert, P. S., "Models for Longitudinal Data: A Generalized Estimating Equation Approach," *Biometrics* 44, 1988, Vol. 44, pp. 1049-1060.

Abbreviations and notations

GEE, generalized estimating equations	GLM, generalized linear models
QIC, quasi-likelihood information criterion	QICu, Pan's approximation to QIC

Biography of the Author

Hernán L. Medina is senior research associate at the Insurance Services Office, Inc., a Verisk Analytics Company in Jersey City, New Jersey. He graduated magna cum laude from Saint Peter's College with a Bachelor of Science degree, majoring in Mathematics with a minor in Physics. He also has a Master of Science degree in Mathematics from New York University. He is a Charter Property Casualty Underwriter and a member of the CPCU Society.

Multi-Year Policy Pricing

Benjamin R. Newton, FCAS, RPLU

Abstract: Currently, the pro rata method is widely used to price policy extensions and other changes to policy length. However, there are several other methods that an actuary can use in pricing policy-length changes, such as option and forward pricing, or modeling. There is little discussion in the literature about the relative merits of these different methods, leaving the pricing actuary potentially unsure of which method to use in a given situation. This paper seeks to resolve some of that ambiguity by identifying the features of a contract that would indicate the need to select a certain method over others. The goal of this discussion is to provide the pricing actuary with a framework that can be used to select the most appropriate method for the particular contract that is being priced. The paper then provides simple examples of how these methods might be applied in different situations. Finally, it compares the results obtained using the recommended method to the results from the pro rata method, and points out the potential for pricing insufficiencies when the pro rata method is applied universally. These results should encourage the reader to consider pricing techniques aside from the standard pro rata method.

Availability: The @Risk™ workbook used for example purposes is available on the CAS Web Site at <http://www.casact.org>.

Keywords: Policy extension, pro rata, option, forward, simulation, multi-year, pricing.

1. INTRODUCTION

Policy extensions are likely the most common premium bearing change to a policy. When faced with the task of pricing a policy extension, actuaries will often immediately turn to the pro rata method. For example, one can purchase a physical “Pro Rata Wheel” and there are an abundance of free “Pro Rata Wheel” calculators. This paper will demonstrate that, unfortunately, simply using the pro rata method is often inadequate. In fact, as the paper will show, there is no one method that is sufficient to price all policy extensions. Rather, there are several different pricing methods that are most appropriately used depending on the particulars of the situation. The paper will attempt to guide the pricing actuary through the process of selecting the optimal pricing method by creating a framework detailing which types of policy extensions are most accurately priced by which method and presenting the theory supporting the use of the different methods in these different contexts.

1.1 Research Context

This paper lays out the theoretical underpinnings of pricing policy extensions and provides a framework for selecting the appropriate policy-extension pricing method in a variety of situations. This framework is intended to be used any time a policy duration differs from the standard policy length. Examples of changing policy lengths are six-month policies written on an annual basis, annual policies written on a six-month basis, single-year policies written on a multi-year basis, and

multi-year policies written on a single-year basis, among others. For simplicity all of these various changes to policy length will be referred to in this paper as “policy extensions.”

A brief review of the literature reveals a variety of methods in use today to price policy extensions, including discounts [7], pro rata [5][6][8], option-based approaches [11], and pricing ranges [3]. The idea that there are different methods that should be used to price policy extensions is not inherently wrong; however, the potential source of confusion for the pricing actuary is that these different sources cannot agree on which methods to use when. Even within a single source, a variety of methods can be found such as ISO rules requiring short-rate return of premium, pro rata extension premium, or a combination of short-rate and pro rata return of premium [5].

1.2 Objective

This goal of this paper is to provide the pricing actuary with a framework for selecting the most appropriate method of pricing policy extensions. This framework is based on the premise that it is policy language that largely determines the appropriate pricing method. It is not the goal of this paper to reinvent the wheel, for that reason the examples will remain simple and the discussion will not address expenses. The paper will not delve into significant detail about the technicalities of pricing policy extensions using any particular method. There are many better sources of information on using those techniques in other financial, actuarial, and mathematical literature.

1.3 Outline

Section 2 presents the four ways to price policy extensions (option, forward, simulation, and pro rata) and discusses when each should be used. Section 3 illustrates the pricing implications of writing each type of endorsement and the potential for variation in loss costs based upon different contractual terms.

2. BACKGROUND AND METHODS

The four methods that can be used to price policy extensions are option, forward, simulation, and pro rata. This section provides a brief overview of each pricing method and identifies which method should be used in which circumstances. For further descriptions of the type of policy provisions that would require each of these methods to be used, please refer to Appendix 1.

2.1 Policy Extension Using Option Pricing

An option is a contract that gives the purchaser the right, but not the obligation, to buy a good at an agreed-upon price in the future. In essence, when an insured purchases an option to renew a policy at a fixed price in the future, they are betting that the price of the option is less than the present value of the expected rate increase.

Option-pricing methodology is appropriate when the insured has the option to accept a guaranteed future price or to decline and receive better terms if market conditions are favorable [11]. See Appendix 1 for an example of policy language that would indicate this.

Once it has been determined that the policy language indicates that option-based pricing is appropriate, the actuary can use the following basic formula to price the policy extension. For simplicity, we will assume that the option can only be exercised on the policy expiration date, so that we can use a European call option to price the policy. Using the notation from Hull [4], the value of the option at expiration can be expressed as:

$$\text{Max}(S_T - K, 0) \tag{2.11}$$

Where S_T is the market price of the insurance contract at renewal and K is the guaranteed renewal price. The expected future value of the option at expiration can then be obtained by:

$$E[\text{Max}(S_T - K, 0)] = \int_k^{\infty} (S_T - K) f(x) dx. \tag{2.11}$$

Where $f(x)$ is the probability distribution function of rate change for an insured and $S_T - K$ represents the option value at each rate change. From this formulation it can be observed that the option to renew will expire worthless if the renewal market price S_T is less than the guaranteed price K . Further, if the renewal market price at expiration is greater than K , then the value of the option will be the difference in the market price and guaranteed price K .

This same valuation method can be extended to a multi-period scenario in the event of longer-term renewal guarantees [11].

2.2 Policy Extension Using Forward Pricing

This section will seek to identify which types of contracts should be viewed as a forward contract and will outline the basic formulas necessary to price these contracts. Unlike an option, which is a right to buy or sell without any obligation to do so, a forward contract is both the right and obligation to buy or sell a good at some future time. The important feature of a forward agreement

Multi-Year Policy Pricing

is that the insured and the insurer lock in a price in the future regardless of what the market price might be at the future date. It is appropriate to use forward pricing when the insurer and insured are required to enter into a policy at some point in the future. An example of a forward agreement would be a non-cancelable, automatic policy extension. See Appendix 1 for an example of policy language that would indicate this. If the actuary determines that the policy language indicates that forward pricing should be used, he can price the policy extension as follows.

The value of a forward contract at expiration can be expressed as:

$$S_T - K. \quad (2.20)$$

Where S is the market price of the insurance contract at time T , the renewal date, and K is the guaranteed renewal price. Using the distribution of future changes to policy prices, $f(x)$, the expected future value of the forward contract at expiration can then be expressed as:

$$E[S_T - K] = \int_{-\infty}^{\infty} (S_T - K) f(x) dx \quad (2.21)$$

As this formula indicates, a key difference between a forward contract and an option is that a forward contract allows for potentially positive and negative outcomes for both parties. However, the seller of an option (in this case the insurer) is the only party able to lose money at the option expiration.

2.3 Policy Extension Using Modeling

Any time that policy extensions result in changes to the claim payment process and not just to the collection of premium, it becomes necessary to model these changes. Changes to the claim payment process can occur through alteration of deductibles, limits, or the effective attachment point. Use of a simulated approach allows for the pricing of any possible type of policy once the claim process and contract are understood. Due to the complexity of contracts and claim processes, it is not possible to present a single model that will address all possibilities. Instead, a simple example is presented in Section 3 to demonstrate the importance of building a model appropriate to the contract and the loss process.

2.4 Policy Extension Pro Rata

No discussion of policy extension is complete without the inclusion of the pro rata method. It is the most popular method of policy extension due to ease of computation. Pro rata policy extensions are appropriate when all loss and other expenses vary proportionately with premium or

Multi-Year Policy Pricing

with changes to loss and expense that directly offset the cost of the option or forward contract. Due to the likelihood of changes to loss or expense over an extension period, it is often not the theoretical pricing considerations that cause an actuary to use the pro rata method, but the result of operational considerations.

The calculation of pro rata premium is relatively straightforward and can be expressed as:

$$\text{Policy cost total} = (\text{Policy cost annual} / 365) * (\text{Expiration Date} - \text{Inception Date}).$$

With this simple formulation, each additional day of coverage costs 1/365th of the cost of a one year policy and a two year policy costs double what a one year policy costs.

2.5 Other Costs of Risk Transfer

It is important to consider all costs of risk transfer when pricing multi-year policies and not just changes in premium and losses. Some potential areas for consideration are policy initiation expense, client retention rates, and capital requirements. These are important considerations when pricing multi-year policies. However, an appropriate treatment is outside the scope of this paper.

3. RESULTS AND DISCUSSION

This section will compare the pricing implied by option, forward, and simulation techniques to the pro rata method. The divergence from the pro rata implied pricing will demonstrate the importance of pricing policy extensions in accordance with the contract terms. In order to make these comparisons using easy-to-follow examples, it will be necessary to make some simplifying assumptions.

For policy extensions that can be priced as an option or forward contract, we will use the following assumptions:

- 1) Rate per exposure for all insureds is expected to increase on average 5% per annum.
- 2) Rate change for individual insureds is uniformly distributed from -5% to +15%.
- 3) Exposures are identical in both periods.
- 4) Premium for the first policy year is \$100.
- 5) The discount rate is 10%.

For policy extensions that result in changes to contract terms, frequency, and severity

Multi-Year Policy Pricing

assumptions will be made. Frequency will be assumed to be binomial and severity will be assumed to be lognormal. The assumption of distributions will enable concrete examples and demonstrate the importance of applying a model when appropriate and are not meant to imply that these assumptions are appropriate in all circumstances.

3.1 Policy Extension Using Option Pricing

The cost of issuing a renewal option is determined by the present value of the option at expiration. Using Equation 2.11 and the assumptions laid out in Section 3 above, we can calculate the value of an option with the guaranteed renewal price of \$100 in one year as:

$$E[\text{Max}(S_T - K, 0)] = \int_{100}^{110.11} (x - 100) * \left(\frac{1}{10}\right) dx \quad (3.12)$$

$$E[\text{Max}(S_T - 100, 0)] = 5.625 \quad (3.13)$$

$$\text{Present Value } E[\text{Max}(S_T - 100, 0)] = \frac{5.625}{(1+i)^1} \quad (3.14)$$

$$\text{Present Value } E[\text{Max}(S_T - 100, 0)] = 5.1136 \quad (3.14)$$

Here, the value of the option is greater than zero and the appropriate premium charges for a two-year contract would be as follows:

Price year 1 + Price year 2 + value of option = 100 + 100 + \$5.11 = \$205.11. Here the price of the year 1 contract is known, the price of the second year is guaranteed to be \$100 and the value of the option the insurer sells to the insured is \$5.11. Comparing this to the pro rata method, we see that a pro rata extended policy would generate \$200 over two years or \$100 per year, resulting in a potentially inadequate price.

Not accounting for the value of the option and using pro rata pricing may be acceptable if it strengthens customer relationships, reduces transaction costs, or provides some other benefit to the insurer to offset the decreased premium collection. However, with any potential upside the downside must be considered as well. One potentially significant risk is that option exercise is correlated between insureds and with a hard insurance market. Hard insurance markets are in turn associated with financial distress among insurance companies. Thus, it is easy to imagine a situation in which many policyholders might choose to exercise the guaranteed renewal option simultaneously and at a time of inadequate pricing.

3.2 Policy Extension Using Forward Pricing

The price of a forward contract is calculated from the present value of the forward contract at expiration. Using Equation 2.21 and the assumptions laid out in Section 3 above, we can calculate the value of a forward contract with the guaranteed renewal price of \$100 in one year as:

$$E[\text{Max}(S_T - K, 0)] = \int_{100-95}^{100+1.15} (x - 100) * \left(\frac{1}{20}\right) dx \quad (3.21)$$

$$E[\text{Max}(S_T - K, 0)] = 5$$

$$\text{Present Value } E[\text{Max}(S_T - K, 0)] = \frac{5}{(1+i)^1} \quad (3.23)$$

$$\text{Present Value } E[\text{Max}(S_T - K, 0)] = 4.5455 \quad (3.24)$$

Here, the value of the forward contract is greater than zero and the appropriate premium charges for a two-year contract would be:

$$\text{Price year 1} + \text{Price year 2} + \text{value of forward} = \$100 + \$100 + \$4.55 = \$204.55.$$

Here the price of the year 1 contract is known, the price of the second year is guaranteed to be \$100, and the value of the forward contract the insurer sells to the insured is \$4.55. In this case the value of the forward contract is greater than zero and less than the value of the option contract. At the same strike price, the absolute value of a forward contract will always be between the absolute value of the option contract and zero. Once again, when comparing the forward contract to pro rata extension the pro rata policy generates an inadequate price. The pro rata pricing may be acceptable if other benefits to the insurer exist. Potential benefits mirror those outlined in Section 3.1. However, forward contracts, unlike option contracts, will always be executed at expiration. Thus, the exercise of the forward contract will not be more highly correlated with periods of economic distress for the insurer, but the risk will still exist. The insured will exercise the contract regardless of the underwriting cycle, which should make cash flows more predictable. However, the insurer will not have the ability to quickly increase profitability through rate increases after a period of inadequate pricing.

3.3 Policy Extension Using Monte Carlo Simulation

Pricing a policy extension that changes the limits, deductible, or the effective attachment point should be based upon the present value of future cash flows and often requires a simulation to model. An expeditious way to tackle this problem is through the use of common modeling software

Multi-Year Policy Pricing

or simplified equations, like those that were used for option and forward contract pricing in Sections 3.1 and 3.2. Here, a straightforward model using @Risk™ was built to demonstrate the importance of modeling simple situations.

The policy extension that we will model involves the extension of a single aggregate limit, after a claim has been reported in the first year and will be referred to as an aggregate extension endorsement. To price this extension, the model will assume:

A claim reported in the first year.

65% chance reported claim will become an incurred loss.

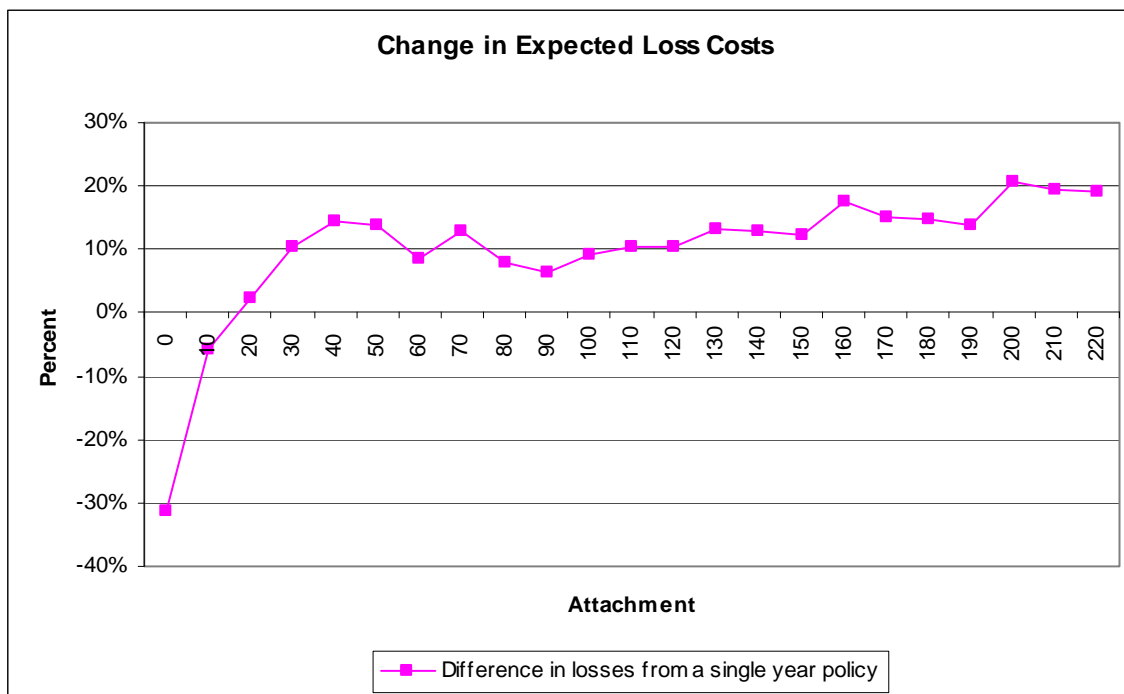
Probability of a claim in the second year will be 2%.

Claim size in year 1 and 2 is log-normally distributed with a mean of \$31.2 million and a median of \$8 million.

The occurrence of claims and the size of claims will be assumed to be independent.

Graph 1 compares the percentage difference in loss costs for a new limit of liability to the extension of the aggregate limit of liability for the entire tower. It is seen in Graph 1 that expected loss costs for the lowest \$20 million in coverage decrease relative to a fresh limit of liability, while exposure for all layers higher on the tower increases.

Graph 1.



Primary carriers are incentivized to pursue this endorsement because of the high-severity and low-frequency nature of claims. With the aggregate extension endorsement the primary carrier is offering less coverage, while carriers higher up the tower, with less pricing control, essentially drop down to a lower attachment point in response to an unrelated second claim. The excess carriers' acceptance of this increased exposure is due to the perception of decreased risk and a weak negotiating position. The model suggests that a pro rata premium charge for this policy extension overcharges for the primary and undercharges for the excess limits.

3.3 Discussion of Results

This paper explores three different methods for pricing policy extensions: option, forward, and modeling. Although modeling and option pricing methods had been used previously by Wacek [11] and Berens [1] to confront aspects of policy extensions, in this paper, these pricing methods were brought together and combined with forward contracts to create a coherent framework. This paper further demonstrates that option, forward, and simulation methods often produce pricing recommendations that stand in sharp contrast to the standard pro rata method of policy extension.

By using these three basic techniques, or a combination of the techniques, an actuary can appropriately price any change to policy length.

4. CONCLUSIONS

Several different methods are available for pricing policy extensions and selecting the appropriate method from among these is essential. In this paper, I have provided some guidance to the pricing actuary regarding when to use option, forward, and modeling methods. As we have seen, option pricing should be used whenever the insured has the option to accept a guaranteed future price or to decline and receive better terms if market conditions are favorable. This paper demonstrates that if the policy language implies that option-based pricing is most appropriate, but the pro rata method is used instead, the result can be a pricing insufficiency. By contrast, forward pricing should be used when the policy language implies that the insurer and insured are required to enter into a policy at some point in the future. I have shown that using the pro rata method in cases that clearly warrant forward pricing can over- or under-price the policy extension, though not as severely as in the case when option pricing is required. Both option and forward pricing methods are specific cases of the generalized modeled result and can simplify the process of estimating the cost of the policy extension. If, on the other hand, the policy extension results in changes to the claim payment process, then these simple pricing methods will be inadequate. In these cases, I suggest that the actuary create a model to determine the appropriate price for the policy extension. A simple example of a model is presented herein and is also compared to pricing using the pro rata method. While the exact results will vary depending on the situation, in the simple modeled example given, we discover yet another case in which the pro rata method suggests an inaccurate price.

Overall, this paper has explored some of the different methods by which policy extensions might be priced. I have also provided several examples in which applying the pro rata method to policy extensions is inappropriate, and results in a potential erosion of the insurer's financial stability. It is the hope of this author that actuaries will be able to use the information provided here to select the most appropriate method for pricing any policy extension that they encounter.

Acknowledgment

I would like to thank Deena Bernett for many suggestions and revisions, the CAS volunteers who reviewed the paper, and many colleagues who provided early feedback on my pricing ideas and this paper. Any errors or ambiguities that remain in this paper are solely the responsibility of the author.

APPENDIX 1

Part 1: Contract Language

The examples below are meant to provide a guide as to where to look in the policy for changes that can affect how to price changes to policy length. It is not a comprehensive list, but instead a starting point for a closer examination of the complex policy language that we work with everyday.

Declarations

Policy length can be changed from the standard policy length. For example, if a policy is normally six months long, the declarations page can change the policy length to one year.

Limits can be defined as per claim, per policy period (aggregate), or some other alternative.

Deductibles can be defined as per claim, per policy period (aggregate), or some other alternative.

Premium can be fixed at policy inception or auditable based upon exposure. Premium can be paid in advance for the entire policy period or paid periodically.

Premium earning can specify that premium is fully earned at inception of the policy period, earned uniformly over the policy period, and subject to a variety of other conditions.

Cancellation provisions can vary from non-cancelable to fully cancelable. Upon policy cancellation, premium can be returned to the policy holder on a pro rata basis, a short-rate basis, or another contractually specified basis.

Definitions

Limits can be defined as per claim, per policy period (aggregate), or some other alternative.

Deductibles can be defined as per claim, per policy period (aggregate), or some other alternative.

Endorsements

Endorsements can alter any or all provisions in a policy form.

Part 2: Policy Examples

The examples below are meant to provide a guide as to how to price a stylized policy. It is not a comprehensive list or a definitive pricing manual, but instead a starting point for thinking about how the claims process interacts with policy language to determine expected changes to loss costs as policy periods change.

Multi-Year Policy Pricing

Option Example

Rate Guarantee

XYZ insurance company agrees to renew policy number 12345678 effective from 1/1/2010 to 12/31/2010 for an additional one-year term at a premium of \$1,000 per auto.

Policy Provisions

Initial policy term: 1/1/2010 to 12/31/2010.

Premium: \$1,000 per auto.

Cancellation: Cancelable at any time with a pro rata return of premium.

Limits: Each policy period has separate limits.

Deductible: Each policy period has separate deductibles

Why is this rate guarantee an option?

The insured has a right, but not an obligation to purchase the second year policy at a predetermined price, which is the definition of an option. For simplicity, the policy provisions were set up to be identical, but this is not a necessary condition for the second policy period to include an imbedded option.

Forward Example

Automatic Policy Extension

XYZ insurance company and ACME Car Driving LLC agree to an automatic extension of policy number 12345678 effective from 1/1/2010 to 12/31/2010 for an additional one-year term at a premium of \$1,000 per auto.

Policy Provisions

Initial policy term: 1/1/2010 to 12/31/2010.

Automatic extension period: 1/1/2011 to 12/31/2011.

Premium: \$1,000 per auto.

Cancellation: Non-cancelable by either the insurer or the insured, except in cases of insurer downgrade below an AM Best “A” rating.

Limits: Each policy period has separate limits.

Multi-Year Policy Pricing

Deductible: Each policy period has separate deductibles.

Why is the automatic extension a forward contract?

The insured has a right and obligation to purchase the second year policy at a predetermined price. Neither party can cancel the contract. The second year of the policy has losses that will interact with the limits and deductible in a manner similar to any single year policy.

A practical concern with this type of forward contract, that might argue for treating as an option is the collectability of premium for the renewal period. One way of addressing this issue is to price the expected future premium in a manner consistent with other types of counter party risk.

Modeled Example

Policy Extension Post Claim

In consideration of the additional premium of \$100,000, it is agreed that the policy period 1/1/2009 to 1/1/2010 is deleted and replaced by 1/1/2009 to 1/1/2011.

This extension of the policy period shall not increase the insurer's maximum aggregate limit of liability for loss under the policy. All the other terms and conditions of the policy remain unchanged.

Policy Provisions

Initial policy term: 1/1/2009 to 1/1/2010.

Premium: \$1,000 per doctor.

Cancellation: Premium for extension fully earned immediately. Non-cancelable.

Limits: Policy period and extension share limits.

Deductible: Policy period and extension have separate deductibles.

Why is a model required to price an extension after a claim?

In this instance, a tower of insurance exists, a claim has been reported and the size of the claim is unknown. Consequently, it is unclear how much of the aggregate limit on the policy or any underlying policy limits will be impaired. In effect, the limit and attachment point for the second year period are unknown at the time the endorsement is written and must be estimated by simulation techniques and expert input.

Multi-Year Policy Pricing

Pro Rata Example

Uniform Exposure to Loss and no Lag before Inception

A workers compensation policy was written from 1/1/2009 to 1/1/2010 and a policy extension from 1/1/2010 to 1/1/2011.

Policy Provisions

Policy Term: 1/1/2009 to 1/2/2010.

Premium: Per dollar of audited payroll.

Cancellation: Cancelable.

Limits: Statutorily unlimited.

Deductible: Per claim.

Why can a policy be priced using the pro rata method?

In this instance, the limits are unlimited and the deductible is per claim. So each day has the same expected loss. Premiums are auditable, which automatically adjusts for any deviations from the constant force of claims over the year. If no change has occurred to the expected loss per dollar of audited payroll, then it would be appropriate to use the same rates for the second one-year period. In essence the second-year period has all the same characteristics as the first year and should be priced the same. It is difficult to find real-world cases where all the factors align to support pro rata policy extensions. Pro rata extensions are rarely theoretically correct, but are likely to continue to be used regularly due to their simplicity for insurers, insureds, and regulators.

Disclaimer

Any examples in this article are for illustrative purposes only and any similarity to actual individuals, entities, places or situations is unintentional and purely coincidental. This material is not intended to establish any standards of care, to serve as legal advice appropriate for any particular factual situations, or to provide an acknowledgement that any given factual situation is covered under any CNA insurance policy. Please remember that only the relevant insurance policy can provide the actual terms, coverages, amounts, conditions and exclusions for an insured.

Supplementary Material

On the CAS Web Site, the @Risk™ worksheet will be available for those who have access to

@Risk™ and without the @Risk™ formulas to those without access to the software.

5. REFERENCES

- [1] Berens, Regina M., “Reinsurance Contracts with a Multi-Year Aggregate Limit,” *Casualty Actuarial Society Forum*, Spring 1997, pp. 289-308.
- [2] Chanes, Jeannine W., “Once is never Enough: Annualization of Per-Occurrence Limits in Three-Year Policies,” Fried Epstein LLP Publications, http://www.fried-epstein.com/assets/data/docs/article_onceNeverEnough.pdf.
- [3] Great American Insurance Company, Non-Profit D&O Rate Filing Form (Arkansas), May 16, 2008, SERFF Tracking Number GRTC-12560311, <http://www.insurance.arkansas.gov/PandC/FilingElectronic/20085/GRTC-12560311.pdf>.
- [4] Hull, John C., *Options, Futures and Other Derivatives*, 6th ed., Upper Saddle River, New Jersey: Pearson, 2005.
- [5] Insurance Services Office, “Virginia Changes in Policy – Cancellation and Nonrenewal,” 2005, Form Number Commercial Auto CA 02 68 12 05, <http://www.scc.virginia.gov/division/boi/webpages/auto/CA02681205.pdf>.
- [6] Kaliski, Alan E., and Richard D. Pagnozzi, “Discussion of Generalized Premium Formulae,” *Proceedings of the Casualty Actuarial Society* LXIII, 1976, pp. 69-71.
- [7] Magrath, Joseph J., “Ratemaking for Fire Insurance,” *Proceedings of the Casualty Actuarial Society* XLV, 1958, pp. 176-195.
- [8] Ross, James P., “Generalized Premium Formulae,” *Proceedings of the Casualty Actuarial Society* LXII, 1975, pp. 50-61.
- [9] Shelter Mutual Insurance Company, Farm Owners Insurance Form (Arkansas), March 21, 2008, SERFF Tracking Number SHEL-125525011, <http://www.insurance.arkansas.gov/PandC/FilingElectronic/20083/SHEL-125525011.pdf>.
- [10] State of California Department of Insurance, Rate Regulatory Advisory Notice 2005-02, “Use of Discounts in Determining Private Passenger Automobile Insurance Rates and Premium,” May 2, 2005, <http://www.insurance.ca.gov/0250-insurers/0300-insurers/0200-bulletins/bulletin-notices-commiss-opinion/upload/Advisory-Notice-2005-02.pdf>.
- [11] Wacek, Michael G., “Application of the Option Market Paradigm to the Solution of Insurance Problems,” *Proceedings of the Casualty Actuarial Society* LXXXIV, 1997, pp. 701-733.

Biography of the Author

Benjamin R. Newton is an actuarial director at CNA Insurance in New York City. He began his career at CNA Insurance, where he has spent time reserving, pricing and underwriting various professional liability lines. Benjamin graduated from The University of Illinois with a BA in Economics and a BS in Mathematics, and is currently pursuing an MBA from Columbia University. He is a Fellow of the CAS and a Registered Professional Liability Underwriter and has volunteered with the CAS since attaining his FCAS. He may be contacted at BenjaminRNewton@gmail.com for discussion of this work or to obtain copies of bibliographical items listed as available on the Internet.

Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

Frank Schmid

Abstract

Motivation. Pricing legislative changes is an integral part of NCCI ratemaking. An increase in the maximum weekly indemnity benefit for temporary total disability claims increases indemnity payments for given injury durations, but, at the same time, these injury durations may increase as well (among claimants affected by the benefit change), thus giving rise to an additional cost effect.

Method. The study makes use of a research framework developed by Krueger [13] (and subsequently employed in several studies) on the effect of changes in the maximum weekly benefit on injury duration. This research framework is a natural experiment, where the treatment effect is measured as the difference in the “post-reform minus pre-reform” differences between treatment and control groups. Two partial linear regression models (generalized additive regression and quantile regression) are used to validate the measured treatment effect. Further, quantile regression is applied to quantify the effect on injury duration of claim attributes such as age and gender.

Results. Using data sets provided by the Oregon Department of Consumer and Business Services and the New Mexico Workers’ Compensation Administration, it is shown that an increase in the maximum weekly benefit of temporary total disability claims leads to a lengthening of the average benefit duration in the group of affected claimants. This increase in the utilization of indemnity benefits contributes to about 30 percent of the total cost impact of the reform.

Availability. Computing and bootstrapping the difference in the “post-reform minus pre-reform” differences between treatment and control groups is explained in detail and is straightforward to execute. The generalized additive model was implemented using the R package *mgcv*, which was developed by Simon Wood (<http://cran.r-project.org/web/packages/mgcv/index.html>). The partial linear quantile regression model was implemented using the R package *quantreg*, which was created by Roger Koenker (<http://cran.r-project.org/web/packages/quantreg/index.html>).

Keywords. Generalized Additive Model, Indemnity Utilization, Injury Duration, Legislative Reform, Quantile Regression.

1. INTRODUCTION

Pricing legislative reforms in workers compensation is an integral component of NCCI ratemaking. A possible manifestation of a legislative reform is an increase in the maximum weekly benefit for temporary total disability (TTD) claims, as has been observed in several U.S. states over the past decades. In a pioneering study, Krueger [13] quantified the impact of such a benefit change for Minnesota. The experimental research framework developed by this author was subsequently applied to analyzing increases in the maximum weekly benefit of TTD claims in Kentucky and Michigan (in 1980 and 1982, respectively; see Meyer, Viscusi, and Durbin [14], and Connecticut (1987; see Gardner [4]); further, Curington [3] employed this approach when quantifying the impact on Permanent Partial Disability (PPD) claims of legislative changes that occurred in New York between 1965 and 1978. In all instances, the researchers established evidence of an increase in

Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

utilization in response to an increase in benefits for the time of absence from work. At the same time, Curington [3] shows that an increase in PPD benefits that apply after the claimant returns to work (while benefits during work absence remain unaltered) shortens the time away from work among severely impaired claimants.

What follows is a study on the effects on injury duration of an increase in the maximum weekly benefit for TTD claims in Oregon and New Mexico. In Oregon, effective January 1, 2002, the maximum weekly benefit rose from 100 to 133 percent of the state average weekly wage. It is shown that this hike in the maximum weekly benefit increased total indemnity payments on TTD claims by 3.82 percent; 31 percent of this increase (or, equivalently, 1.17 percentage points) were due to a utilization increase (as caused by an expansion of the injury duration of TTD claimants whose weekly benefits increased due to the hike in the weekly maximum). In New Mexico, effective January 1, 2000, the weekly maximum benefit increased from 85 to 100 percent of the state average weekly wage. The accompanying increase in total indemnity payments on TTD claims amounted to 4.50 percent, 29 percent of which (or, equivalently, 1.30 percentage points) was due to a utilization increase. The duration/benefit elasticity (defined as the percentage change in benefit duration, divided by the percentage change in the maximum weekly benefit) equals 0.53 for Oregon and 0.43 for New Mexico. Because Oregon and New Mexico display similar elasticities and similar proportions of the cost effect of the utilization increase, only the Oregon findings are going to be discussed in detail. Further, the Oregon data are available in greater number and detail, thus allowing a more comprehensive statistical analysis.

For Oregon, the measured treatment effect is validated using a generalized additive regression model (GAM). Further, to get a more differentiated picture of the increase in benefit duration than a regression on the mean can offer, a partial linear quantile regression model is estimated. This quantile regression approach shows that the increase in injury duration in response to the legislative reform is mostly confined to short durations.

Finally, a partial linear quantile regression model is used to study the effect on benefit duration of age and gender. It is shown that the median injury duration is about log-linear in age within the age bracket 20 through 60; within this bracket, on average, benefit duration increases for every year of age by 1.0 percent for Oregon and 0.72 percent for New Mexico. Further, for Oregon, it is shown that among TTD claims with long durations, the durations of female claimants exceed those of male claimants by about 20 percent.

The generalized additive model was implemented using the R package *mgcv*, developed by Simon Wood of the University of Bath (England); the package is available for download at <http://cran.r-project.org/web/packages/mgcv/index.html>. The partial linear quantile regression model was implemented using the R package *quantreg*, developed by Roger Koenker from the University of Illinois at Urbana-Champaign; this package is available at <http://cran.r-project.org/web/packages/quantreg/index.html>. The statistical software platform R is a GNU project of the Free Software Foundation and is administered by the Technical University of Vienna, Austria.

1.1 Research Context

Studies on the impact of benefit changes on TTD claim durations can be divided into cross-sectional and event studies. In cross-sectional time series and pooled time series cross-sectional studies, differences in legislative provisions across states are modeled in attempts to gauge the impact of these differences on claimant behavior; examples of such studies are Butler and Worrall [2] and Worrall, Butler, Borba, and Durbin [16], and, for PPD claims, Johnson and Ondrich [8]; see also Brooks [1] and, most recently, Guo and Burton [6]. Krueger [13] expresses skepticism about the ability of cross-sectional studies to discern the influence of specific legislative provisions on claimant behavior—this is because of the multitude of cross-sectional variations at any given point in time. As an alternative to the cross-sectional research framework, Krueger [13] suggests using event studies. Instead of focusing on variations across states at a given point in time, event studies home in on variation over time in a given state—the event is defined by the legislative change. In order to isolate the impact of this event, a time window surrounding the reform has to be specified; also, the framework is available only if both a treatment group (claimants affected by the legislative change) and a control group (unaffected claimants) can be identified. If this condition is met, the time window creates a quasi-experimental setting in which the legislative reform can be studied as a natural experiment.

1.2 Objective

The objective of this study is twofold. First, the treatment effect is quantified, both in terms of its expected value and its probability distribution. This treatment effect is broken down into (1) an increase in payments due to the hike in the maximum weekly benefit at given durations and (2) an increase in payments due to lengthened injury durations of the claimants affected by the benefit change. The measured treatment effect is validated using generalized additive and quantile

regression approaches. Second, a partial linear quantile regression model is applied to quantify for the post-reform period the effect on benefit duration of claim attributes such as age and gender.

1.3 Outline

What follows is a presentation of the experimental research framework proposed by Krueger [13]. Section 3 offers a description of the data, which is followed in Section 4 by a presentation of the findings for the treatment effect. Section 5 presents estimates of the treatment effect that are arrived at by means of generalized additive and, alternatively, quantile regression models. The quantile regression model of the effect on benefit duration of age and gender are displayed in Section 6. Section 7 concludes.

2. BACKGROUND AND METHODS

In a seminal study on the effects of an increase in the maximum weekly benefit for TTD claims in Minnesota in 1986, Krueger [13] suggested an experimental research framework where the impact of the benefit change on injury duration is measured as a treatment effect. The author identifies as the treatment group the claimants whose benefits were constrained by the legal weekly maximum both before and after the legislative reform—this group of claimants experiences an increase in weekly benefits equal to the stipulated increase in the maximum weekly benefit. As the control group, Krueger [14][13] chooses claimants whose benefits were unconstrained by the weekly maximum both before and after the reform; thus, the weekly benefits of the control group were unaltered by the reform. The treatment effect, which is defined as the increase in injury duration for the treatment group that is causal to the benefit change, is measured as the difference between the differences in post-reform and pre-reform durations of the treatment and the control groups. Conceptually, the difference between post-reform and pre-reform durations equals the treatment effect plus any change common to all claims; an example of such common effects may be changes in injury duration due to variations in economic activity (possibly related to the business cycle) or due to structural economic change (which may manifest itself in a time trend). In order to eliminate such common effects from the treatment group's difference between post-reform and pre-reform durations, the corresponding difference in duration for the control group is subtracted. The resulting difference in differences delivers the treatment effect.

Chart 1 illustrates the TTD benefit schedule for Oregon during the time window surrounding the benefit change; the legislative reform became effective on January 1, 2002. Up to a pre-injury weekly wage of \$55.56, the weekly benefit equaled 90 percent of that weekly wage. For claimants

Casualty Actuarial Society *E-Forum*, Winter 2011-Volume 2 4

Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

with a pre-injury weekly wage in excess of \$55.56 but not more than \$75, the average weekly benefit equaled \$50. Claimants with a pre-injury weekly wage in excess of \$75 collected the maximum weekly benefit or two-thirds of the pre-injury weekly wage, whichever is lower; the reform raised the maximum weekly benefit from 100 percent of the official state average weekly wage (which was \$645 at the time) to 133 percent. In Oregon, the official state average weekly wage becomes effective on July 1.

The increase in the maximum weekly benefit was not retroactive. This means that for a claimant who sustained an injury before January 1, 2002, the applicable maximum weekly benefit equals 100 percent of the state average weekly wage for the duration of the claim; increases in benefits are confined to the annual increase in the official state average weekly wage.

The treatment (T) group comprises all claimants whose benefits were constrained by the legal maximum both before and after the reform; that is, all claimants that had a pre-injury weekly wage of more than 1.5 times 133 percent of the state average weekly wage. The control (C) group consists of all claimants whose benefits were not altered by the reform (that is, whose pre-injury weekly wage was less than 150 percent of the state average weekly wage) and, at the same time, had a pre-injury weekly wage of more than \$75. The treatment effect, which gauges for the treatment group the change in injury duration that is causal to the benefit change, is defined as “mean injury duration in post-reform treatment group minus mean injury duration in pre-reform treatment group” minus “mean injury duration in post-reform control group minus mean injury duration in pre-reform control group.”

Chart 1: TTD Benefit Schedule on the Day the Legislative Reform Took Effect, Oregon

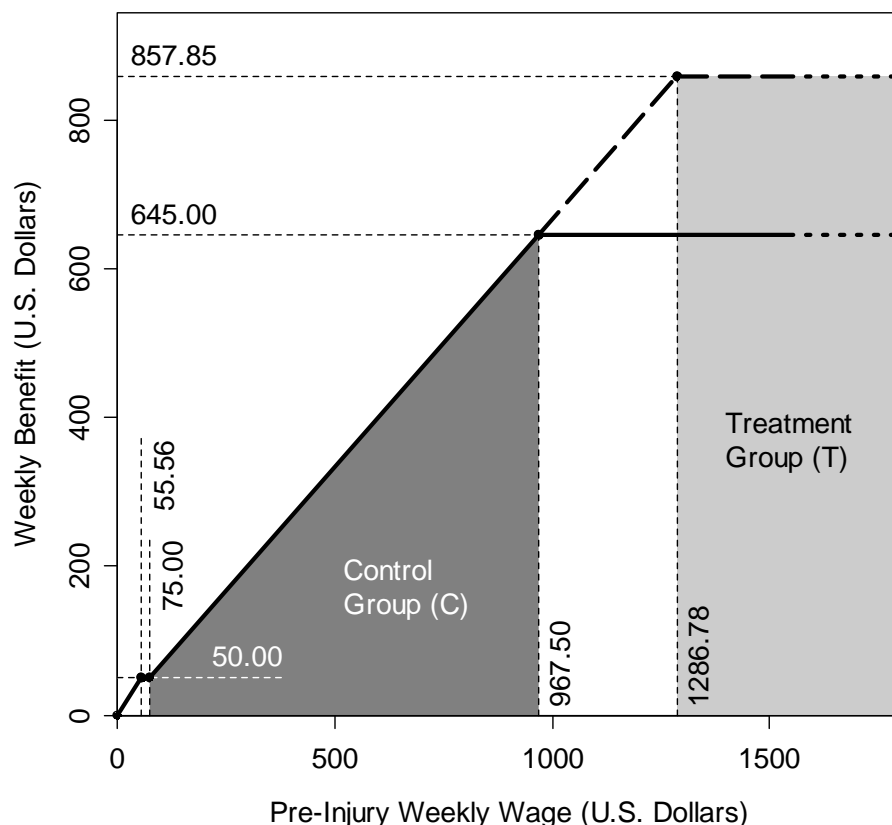
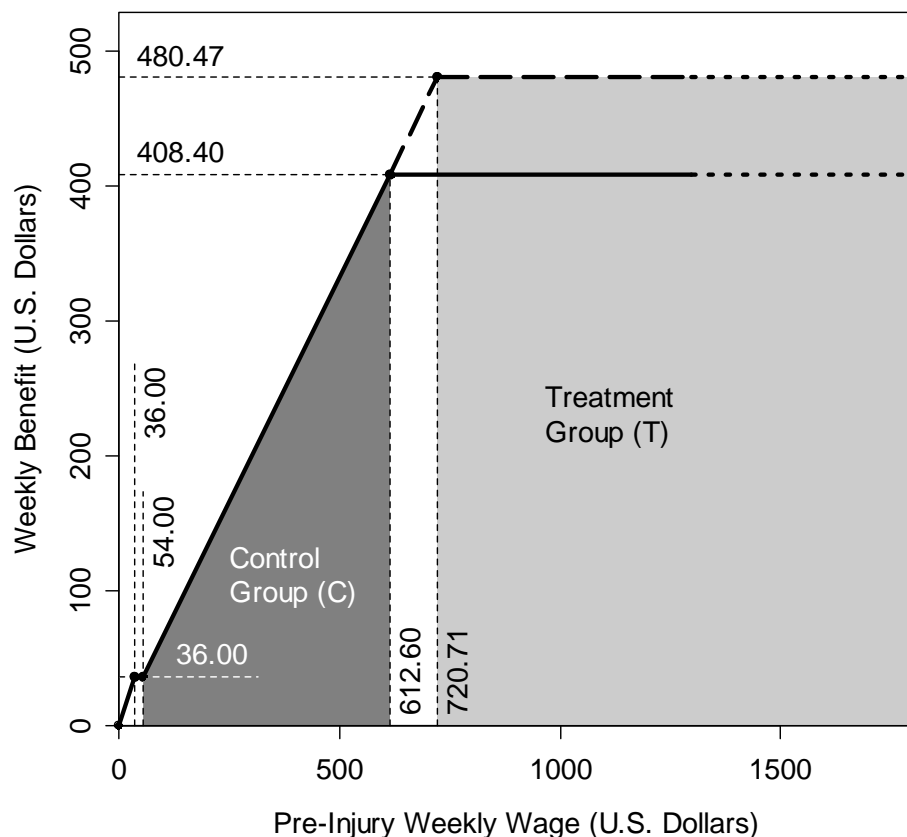


Chart 2 exhibits the TTD benefit schedule for New Mexico that was effective on the day of the legislative reform, which was January 1, 2000. On the same day, the official state average weekly wage experienced its annual adjustment. According to this schedule, up to a pre-injury weekly wage of \$36, the weekly benefit equals 100 percent of that weekly wage. For claimants with a pre-injury weekly wage in excess of \$36 but not more than \$54, the average weekly benefit equals \$36. Claimants with a pre-injury weekly wage in excess of \$54 collect the minimum of the maximum weekly benefit and two-thirds of the pre-injury weekly wage. The reform raised the maximum weekly benefit from 85 percent of the state average weekly wage to 100 percent; this change in benefit level implied a maximum weekly benefit of \$480.47 on the day the reform took effect, up from the \$408.40 that would have applied otherwise. Similar to Oregon, the increase in the maximum weekly benefit in New Mexico was not retroactive.

Chart 2: TTD Benefit Schedule on the Day the Legislative Reform Took Effect, New Mexico



3. THE DATA

Claim records were supplied upon request by the Oregon Department of Consumer and Business Services and the New Mexico Workers' Compensation Administration. The next section offers a description of the Oregon data with a focus on data cleansing and descriptive statistics. This is followed by a brief section on the New Mexico data set.

3.1 The Oregon Data

The Oregon data set comprises all records pertaining to claims that collect lost-time benefits with injury dates between (and inclusive of) January 1, 1999 and December 31, 2004. This way, the data set provides for a 36-month pre-reform window that is followed by a 36-month post-reform

Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

window. The total number of records of award type TTD/TPD (which comprise TTD and Temporary Partial Disability records) equals 98,311 (thus corresponding to 62.52 percent of the total 157,246 lost-time claims records).

Due to reopening (with or without new condition), 731 claim records (or 0.46 percent of the original 157,246 records) were removed from the data. Further, seven TTD/TPD records with injury durations of more than three years were removed due to data quality concerns; six of these claims (of which four are pre-reform) belonged to the control group, whereas the remaining single (pre-reform) claim belonged to the group located between control and treatment groups. (Although Oregon has no statutory limit on the duration of TTD/TPD claims, correspondence with the data provider indicated that claims of this award type, when showing durations of several years, may be of impaired data quality.) In conclusion, all claims in the data set may be considered closed, which implies that there is no problem of right-censoring in the statistical analysis.

Further, for the purpose of data cleansing, we excluded claim records indicative of a claim disposition agreement (CDA); for such claims, there is no breakdown into indemnity and medical costs available. Of claims with multiple closures, we retain only the record with the most recent closure date. We exclude claims where the injury date equals the closure date; such claims may initially have been accepted as nondisabling (medical only), but aggravation later in the life of the claim initiated a TTD claim record.

Benefit and injury durations were measured in weeks of calendar time. The benefit duration was computed as the ratio of total time-loss days for which the claimant received TTD or TPD benefits and the pre-injury number of days the claimant worked during a week. For the purpose of obtaining the injury duration, the benefit duration was adjusted for a waiting period of three days (which comes with a retroactive period of 14 days). This means that for every claim the benefit duration of which is less than two weeks, the injury duration exceeds the benefit duration by three-sevenths of a week.

The data set lumps TTD and TPD claims into a single award type. In order to eliminate TPD claims (and ensure data quality for TTD claims), we judgmentally excluded records where the observed weekly TTD paid falls short of 85 percent of the indicated weekly benefit. The indicated weekly benefit was computed from the reported pre-injury weekly wage based on the applicable benefit schedule.

Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

Finally, to ensure data quality and exclude claims with lump-sum payments, we judgmentally excluded records where the observed weekly TTD paid exceeds 115 percent of the indicated weekly benefit. The final number of TTD claims for the six-year window equals 53,681.

Chart 3 displays a histogram of the pre-injury weekly wage (with a bin size of \$100). For the purpose of this histogram, all observations of the pre-injury weekly wage are inflation-adjusted based on the rate of inflation embedded in the official state average weekly wage effective at the time of the reform. Inflation-adjusted, the minimum pre-injury weekly wage is \$3.96; the maximum equals \$7,469.46; the median and mean values equal \$454.02 and \$516.20, respectively. Clearly, the distribution of the pre-injury weekly wage is strongly skewed to the right.

Chart 4 is a combination of the histogram in Chart 3 and the benefit schedule displayed in Chart 1. The frequency distribution in Chart 4 has a residual bin for a pre-injury weekly wage of \$1,500 or higher. Chart 4 indicates that only a small proportion of the claims are in the treatment group (pre-reform: 350 claims or 0.65 percent; post-reform: 284 claims or 0.53 percent). The control group, on the other hand, is heavily populated (pre-reform: 27,375 or 51.00 percent; post-reform: 22,442 or 41.81 percent). When control and treatment groups are taken together, they add up to 50,451 claims (which amount to 93.98 percent of the total 53,681 TTD claims). The group of claimants located between the control and treatment groups comprises 3.09 percent of claims pre-reform and 2.35 percent post-reform.

Chart 5 exhibits the age distribution of the claimants in single-year age bins. For the purpose of this chart, 166 claims with zero values for the age of the claimant were removed, thus leaving the data set with 53,515 observations. The minimum age in years is 13; the maximum equals 96; the median and mean values equal 38 and 38.1, respectively. Chart 6 displays the relative frequency distribution of gender by age; in this chart, too, 166 claims with zero values for the age of the claimant were removed. Of the 53,515 claimants, 68.9 percent are male.

Chart 3: Histogram of Pre-Injury Weekly Wage (Wage Level at Time of Reform), Oregon

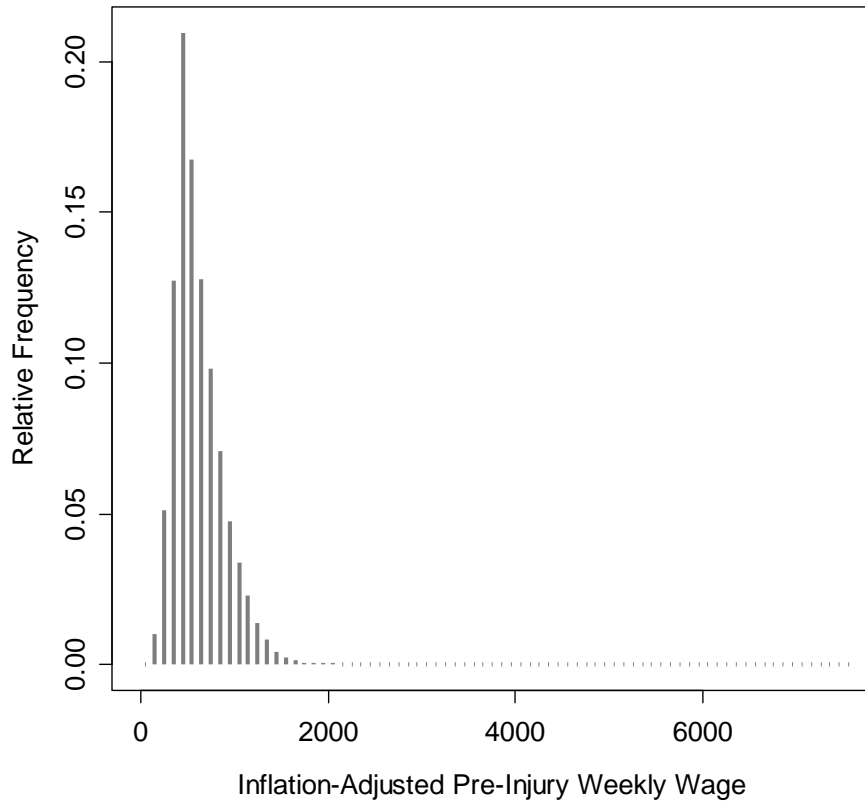


Chart 4: Population of Treatment and Control Groups, Oregon

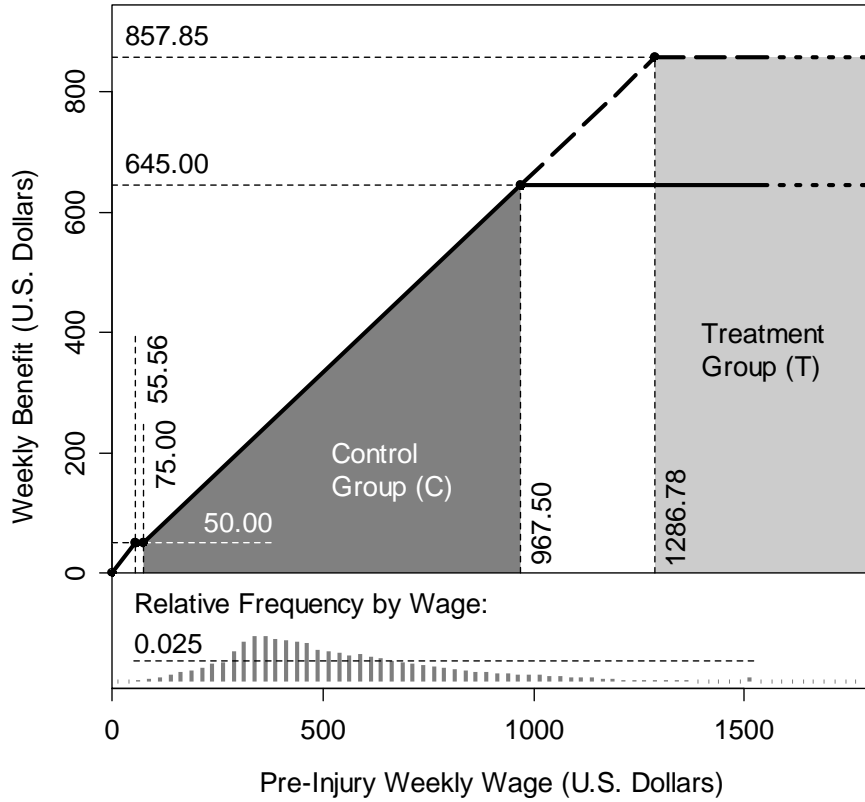


Chart 5: Histogram of Age of Claimant, Oregon

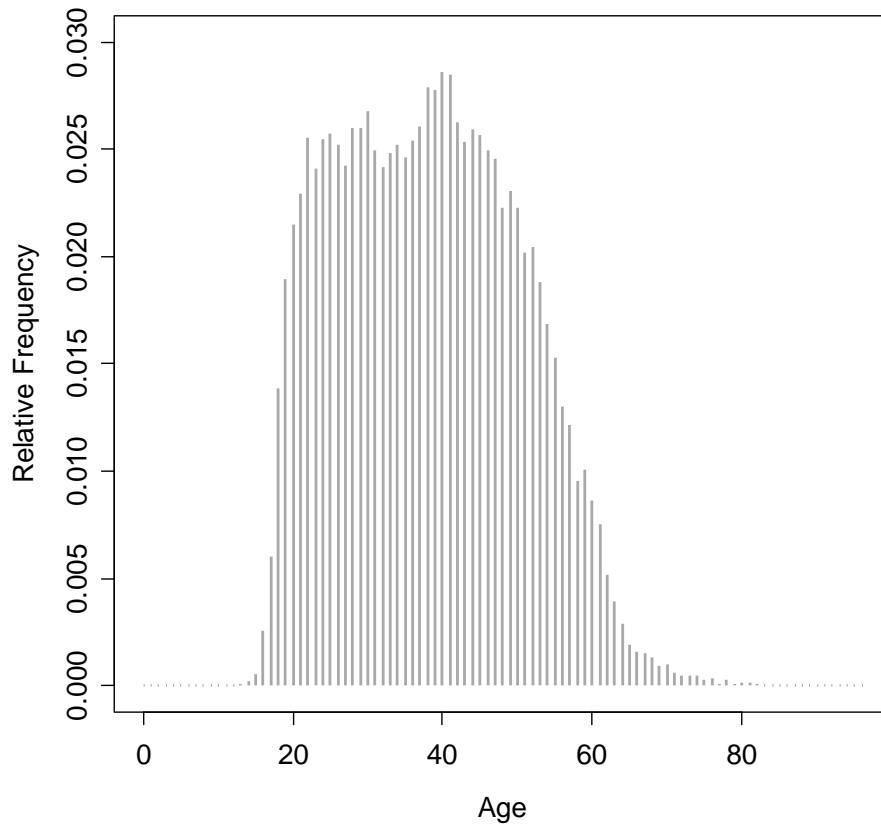
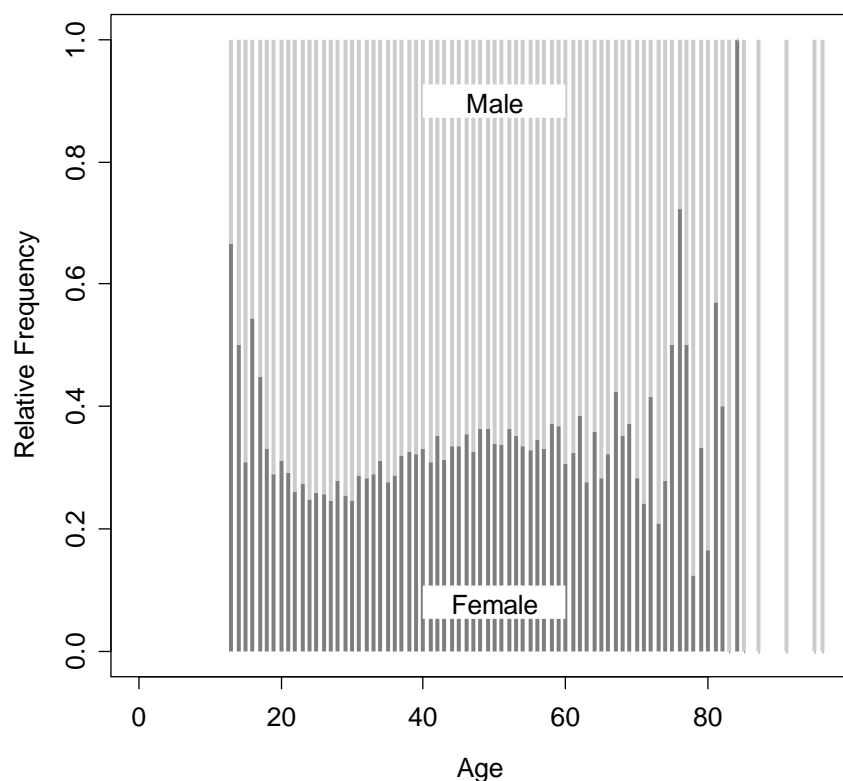


Chart 6: Relative Frequency of Gender by Age, Oregon



3.2 The New Mexico Data

The New Mexico data set comprises all claims that collect lost-time benefits with injury dates between (and inclusive of) January 1, 1997, and December 31, 2002. Just like Oregon, the data cover a 36-month pre-reform window, followed by a 36-month post-reform window.

The data for New Mexico were provided at the level of the claim (unlike the Oregon data, which consisted of claim records and necessitated aggregation where there was more than one record per claim). The lost-time claims in this data set are identified by positive payments in the categories TTD, TPD, PPD, PTD, “Death,” or “Lump sum;” this way, 36,997 claims were identified as collecting lost-time benefits. Of these lost-time claims, there are 2,866 claims (or 7.75 percent) that were categorized as “R” (“Reopened”) or “X” (“Reopened/Closed”).

For the purpose of identifying the set of TTD claims in the population of lost-time claims and

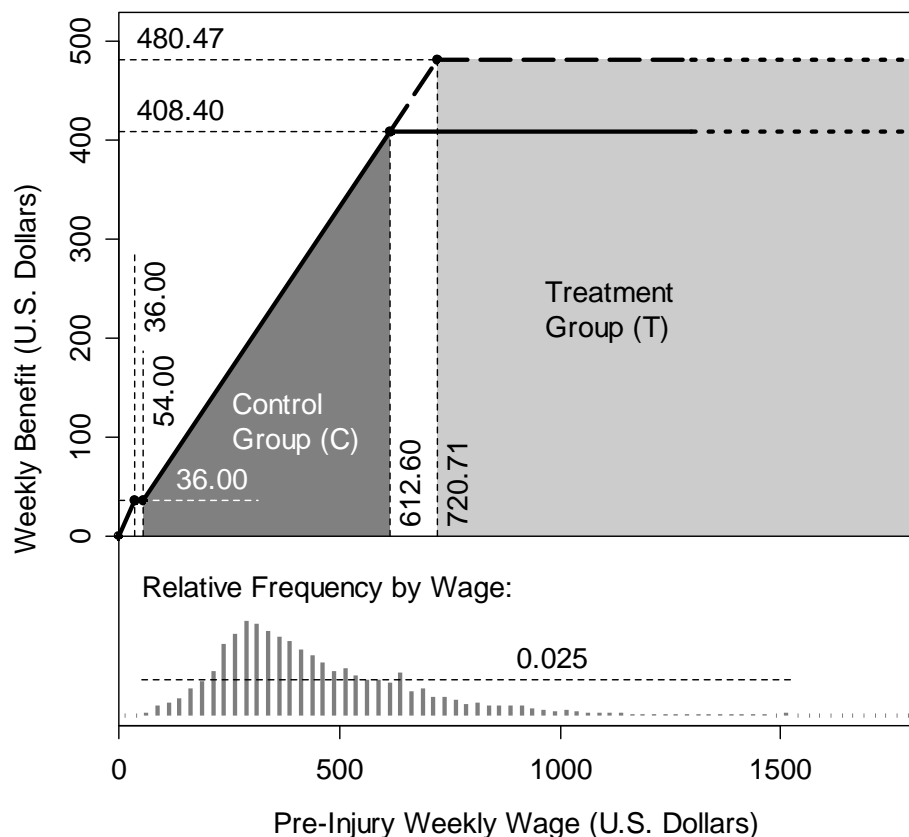
Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

with the intent of cleansing these identified claims, we implemented four rules of exclusion (in the order stated). First, we excluded claims with positive payment entries for TPD, PPD, PTD, “Death” or “Lump sum,” as well as claims with nonpositive payment entries for TTD. Second, we excluded claims with nonpositive entries for TOTAL WEEKS OF LOST TIME and TOTAL DAYS OF LOST TIME. Third, we excluded claims the “Claim Status” of which is not “O” (“Open”), or “C” (“Closed”); this rule excluded claims that were categorized as “R” (“Reopened”) or “X” (“Reopened/Closed”). Fourth, to ensure data quality, we judgmentally excluded claims where the reported TTD duration falls short of 90 percent or exceeds 110 percent of the ratio of TTD PAID and the indicated TTD weekly benefit.

An inspection of the claims shows that all of them can be assumed as closed; such does not necessarily apply to reopened claims, which had been eliminated during the data-cleansing process.

Chart 7 displays the benefit schedule of New Mexico with a histogram that illustrates the wage distribution. As with Oregon, the wage data have been inflation-adjusted to the date the reform took effect, and the histogram has a residual bin for a pre-injury weekly wage of \$1,500 or higher. Compared to Oregon, the proportion of claims in the treatment group is larger (pre-reform: 565 claims or 3.82 percent; post-reform: 831 claims or 5.62 percent), but the control group is again the most heavily populated category (pre-reform: 5,935 or 40.16 percent; post-reform: 6,339 or 42.89 percent). Taken together, control and treatment groups comprise 13,670 claims (or, equivalently, 92.50 percent of the total 14,778 TTD claims). The group of claimants located between the control and treatment groups comprises 3.21 percent of claims pre-reform and 3.86 percent post-reform.

Chart 7: Population of Treatment and Control Groups, New Mexico



4. QUANTIFYING THE TREATMENT EFFECT

In what follows, the treatment effect is calculated for benefit duration (measured in calendar time, as mentioned) and, alternatively, for the amount of benefit payments. As discussed, the benefit duration differs from the injury duration by the waiting period (as applicable). Unless stated otherwise, all findings in this section apply to Oregon.

The treatment effect in benefit duration is computed as the difference between two differences or, when measured in relative (percentage) terms, as the ratio of two ratios. The first difference pertains to the treatment group and equals the mean of the post-reform benefit duration less the mean of the pre-reform benefit duration. The second difference is the corresponding difference in

such means for the control group. Calculated in this way, the treatment effect equals 0.76 weeks or, when the treatment effect is calculated in ratios instead of differences, 17.49 percent. In other words, there is an average increase in the benefit duration among claimants of the treatment group of 0.76 weeks (or, equivalently, 17.49 percent), and this increase can be considered causal to the increase in the maximum weekly benefit. The computation of the treatment effect in weeks and relative terms is detailed in the following two subsections.

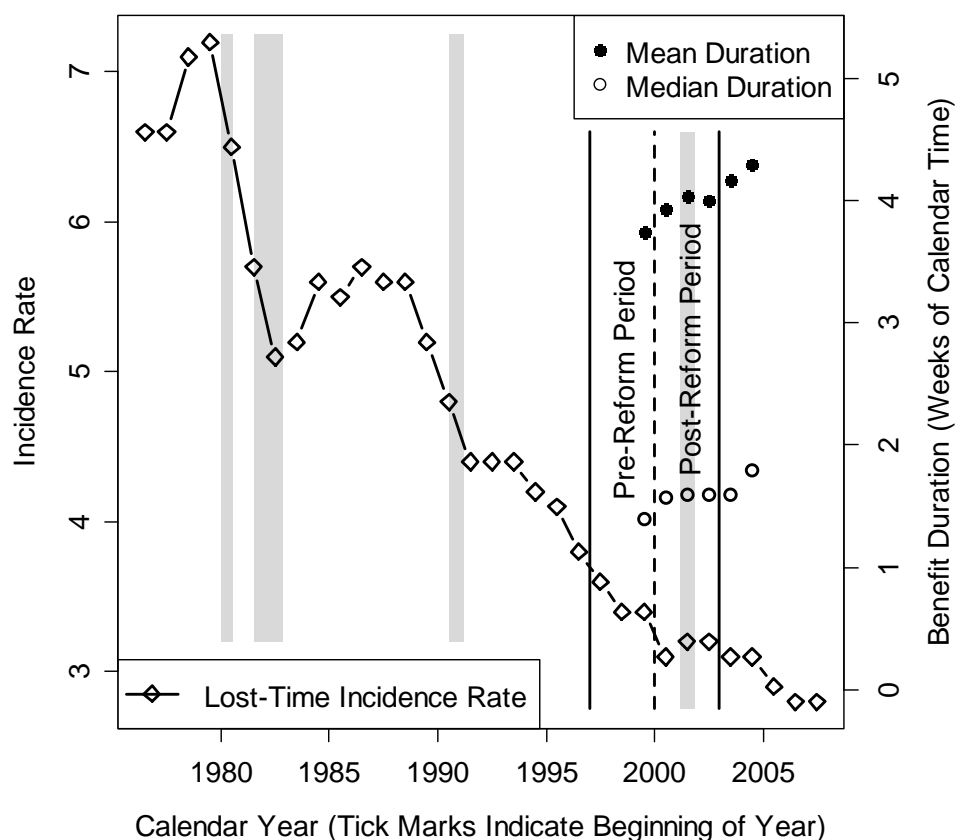
4.1 Treatment Effect in Weeks

To illustrate the importance of employing a control group for the purpose of isolating the treatment effect, Chart 8 displays the behavior of the benefit duration for the pre-reform (1999–2001) and post-reform (2002–2004) time windows; the chart shows the control group only, as the treatment group is affected by the reform. In this chart, economic recessions (as defined by the National Bureau of Economic Research, www.nber.org) are represented by gray bars, which cover the time window between the months that formed the peak and trough of economic activity, respectively; peak and trough are treated as occurring mid-month. Both the mean and the median benefit durations exhibit a positive trend over the six-year time period—this trend is interrupted (median duration) and temporarily reversed (mean duration) during the 2001 recession (peak to trough: February through November).

Studying Chart 8 indicates that it is the economic recovery (as opposed to the economic recession) that (temporarily) disrupts the upward trend in injury duration—the same can be said for the displayed downward trend in frequency, as measured by the Bureau of Labor Statistics lost-time incidence rate (rate of injury and illness cases per 100 full-time workers; cases involving days away from work, job restriction, or transfer). Further research will have to investigate the link between frequency and injury duration, both with regards to their trends and their business cycle behavior.

In order to obtain a probability distribution for the treatment effect, the difference in differences may be bootstrapped, alternatively with and without stratification. In the unstratified bootstrap, 50,451 claims are drawn with replacement from the total of 50,451 claims that comprises (exclusively) the pre- and post-reform periods and the control and treatment groups. Then, the difference in differences (or, alternatively) ratio of ratios is calculated. This procedure is carried out a total of 2,000 times.

Chart 8: Incidence Rate (1976–2007) and Mean and Median Benefit Durations (1999–2004), Oregon



Whereas in the unstratified bootstrap the claims are allocated to their respective group (pre- and post-reform, control and treatment) after drawing, in the stratified bootstrap, the drawing itself is done from the individual groups—the difference in differences (or ratio of ratios) is computed after drawing from each group (with replacement) according to their respective sample populations.

Chart 9 displays the treatment effect in weeks of calendar time—both its mean and its probability distribution are shown. There are two alternative distributions displayed, one being from an unstratified bootstrap and the other from a stratified bootstrap. Chart 10 presents the treatment effect in relative (percentage) terms, again along with bootstrapped probability distributions.

Chart 9: Bootstrapped Change of Benefit Duration in Weeks (of Calendar Time), Oregon

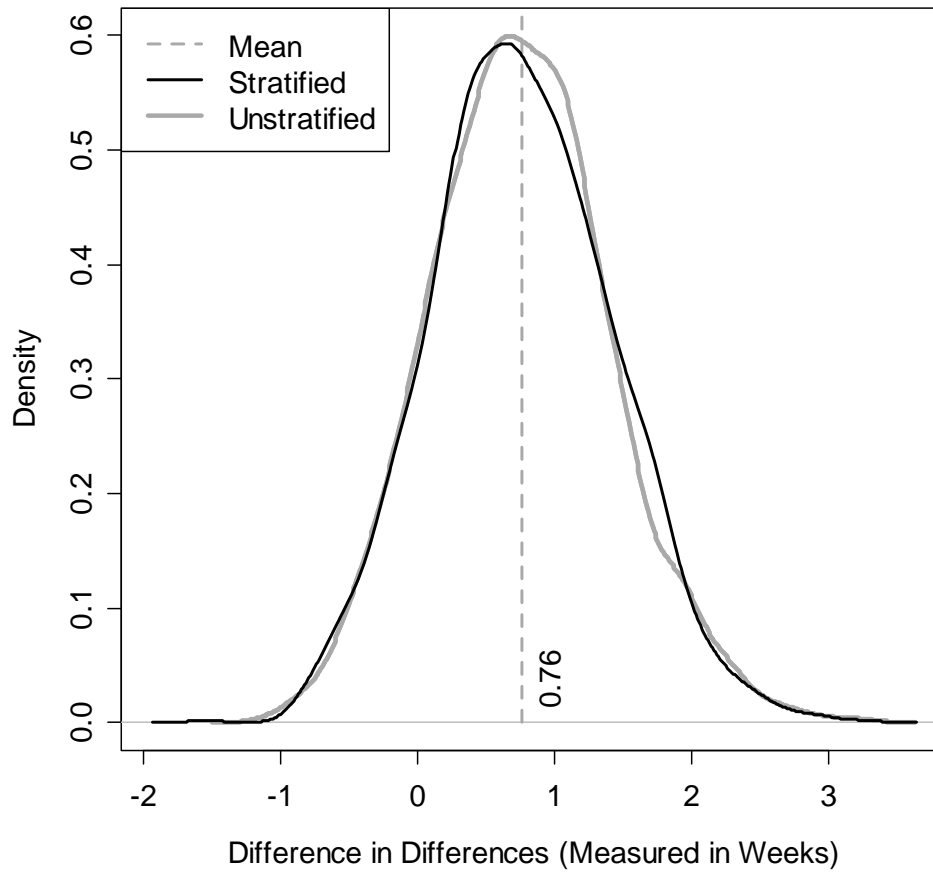
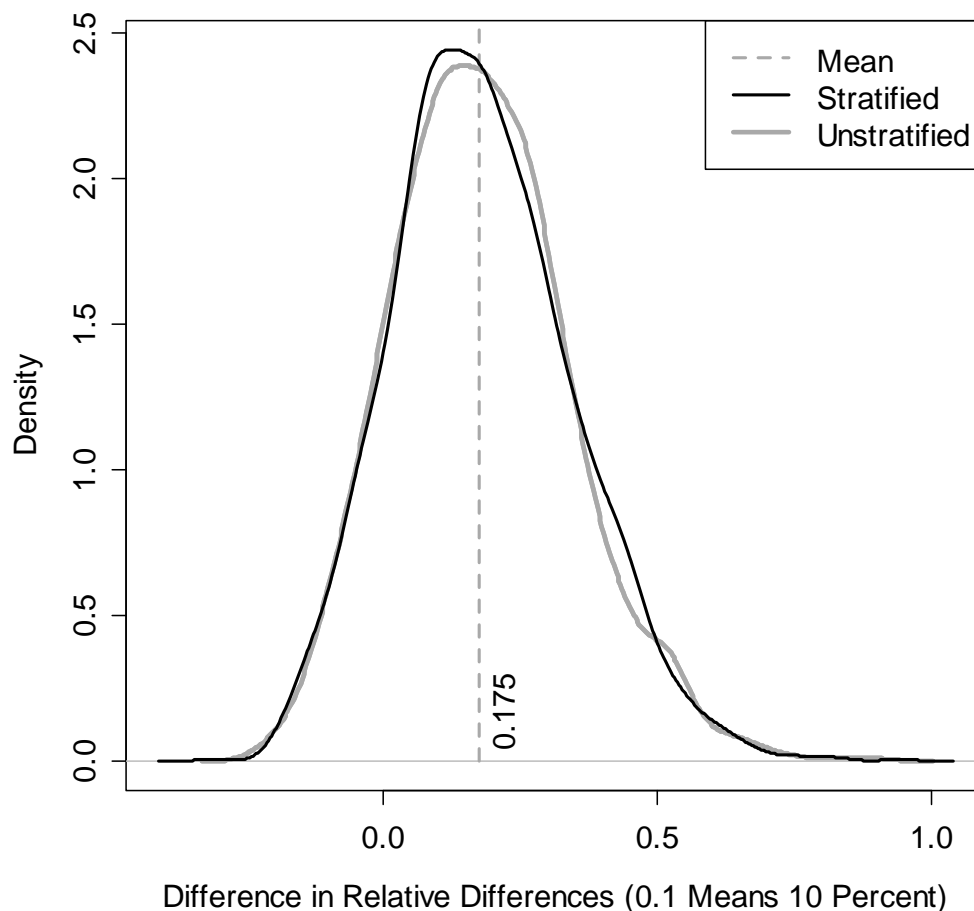


Chart 10: Bootstrapped Relative Change of Benefit Duration (in Calendar Time), Oregon



4.2 Treatment Effect in Relative Terms

Calculating the treatment effect in terms of a relative (percentage) change in the indemnity payments is more complex than calculating the treatment effect in terms of benefit duration. For one, in order to quantify the total dollar impact, the group of claimants located between control and treatment groups (see Chart 1) can no longer be ignored; this group is partially affected by the hike in the weekly maximum benefit. Further, because inflation-adjustment necessitates information on the timing of the payments, the treatment effect in dollar terms is preferably computed using indicated (instead of recorded) benefit payments. Indicated benefit payments are obtained by applying the benefits schedule to the claimant's pre-injury weekly wage, scaled by the computed

Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

benefit duration in calendar time. For the purpose of calculating indicated benefits, all observations of the pre-injury weekly wage are inflation-adjusted to the official state average weekly wage effective at the time of the reform (as mentioned, such normalization also applies to Chart 3 and Chart 4).

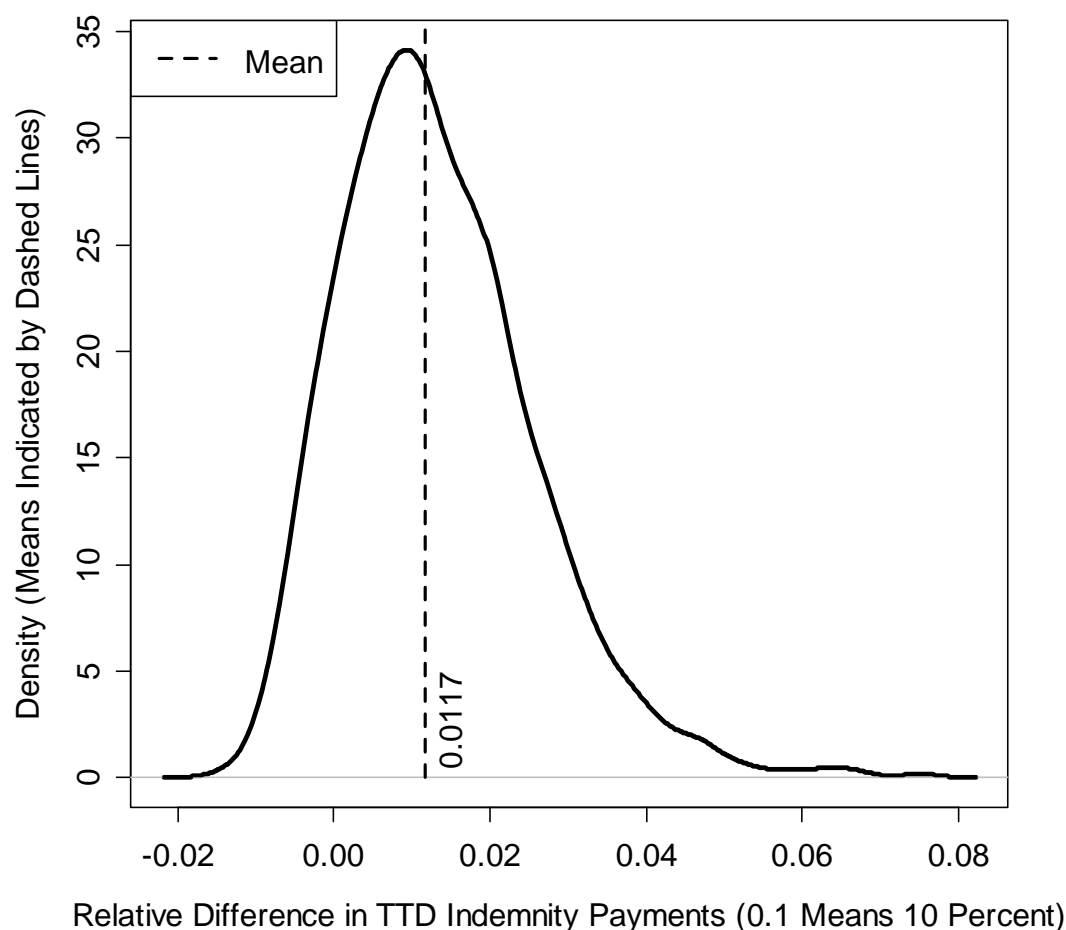
When calculating the treatment effect in dollar terms, the benefit duration of each claim in the pre-reform treatment group is scaled up according to the measured percentage treatment effect on benefit duration. Yet, such cannot be done directly, as benefit durations (in calendar time) of 11, 12, and 13 days are unobservable (because of the 14-day retroactive period). For instance, a claimant with a 13-day injury duration has a benefit duration (in calendar time) of at most 10 days (depending on how many days per week this claimant worked, and on which day the claimant was injured), due to the three-day waiting period; on the other hand, a claimant with a 14-day injury duration always has a benefit duration (in calendar time) of 14 days, due to the 14-day retroactive period. For this reason, the treatment effect (in calendar time) is calculated for injury duration (instead of benefit duration); then, the injury durations of the claimants in the pre-reform treatment group are scaled up according to the resulting relative (percentage) treatment effect. Having obtained the injury durations in such way, the benefit durations (in calendar time) of the individual claims can be calculated (by factoring in the 14-day retroactive period); finally, by applying the benefit schedule, the indicated benefits can be computed from these benefit durations.

Computing the increase in nominal benefits for the treatment group does not suffice for computing the relative (percentage) increase in indemnity payments; this is because the group located between the control and treatment groups (as shown in Chart 1) is also affected by the benefit change. In order to capture this effect, these claims are subjected to a weighted treatment effect (in injury time) when calculating the change in benefit duration. For a given claim, this weight equals the proportion by which the claimant experienced an increase in the weekly benefit due to the hike in the weekly maximum; based on the numbers displayed in Chart 1, the weight equals $(w - 967.50) / (1286.78 - 967.50)$, where w is the pre-injury weekly wage of the claimant after inflation-adjusting this wage to the average weekly wage applicable at the time of the reform.

In conclusion, when scaling up the benefit payments of the pre-reform claims in the treatment group and the group located between control and treatment groups in the way detailed above, then the resulting treatment effect in dollar terms relative to the total indemnity payments equals 1.17 percent. In other words, the lengthening of the benefit duration as caused by the increase in the maximum weekly benefit from 100 to 133 percent of the state average weekly wage produced an

increase in total indemnity payments for TTD claims (within award type category TTD/TPD) of 1.17 percent. Chart 11 displays this effect, along with a bootstrapped probability distribution.

Chart 11: Relative Increase in Payments on TTD Claims due to Treatment Effect, Oregon



Finally, in order to obtain the total effect of the benefit change on payments going to claims of award type TTD, the benefit change itself (at given injury durations) has to be factored in; this effect is in addition to the discussed treatment effect (which quantifies the increase in payments caused by an increase in duration). When adjusting the pre-reform claims of the treatment group and the group between the control and the treatment groups for the increased duration, while at the same time applying the benefit schedule with the increased maximum weekly benefit, the percentage

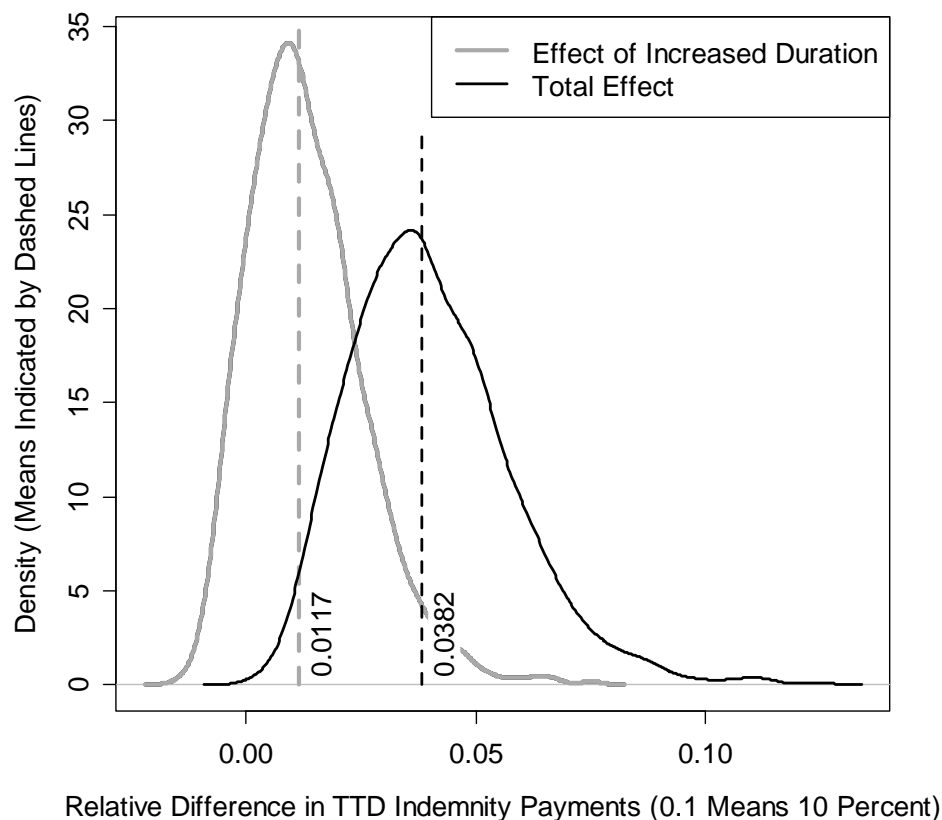
Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

increase in total indemnity payments on claims of award type TTD is 3.82 percent. Thus, about 31 percent of the total effect of the benefit change is due to an increase in utilization (as caused by an increase in benefit duration among claimants affected by the benefit increase). Chart 12 displays the total effect alongside the treatment effect (that is, the utilization effect of increased duration), again with bootstrapped probability distributions.

In the analysis above, the treatment effect was applied to the pre-reform benefit durations before the post-reform benefit schedule was administered to the ensuing post-reform durations for the purpose of calculating the total effect; the resulting percentage of the utilization effect (as caused by the increase in durations) equals 31 percent. An alternative way of breaking down the (same) total effect is to apply first the post-reform benefit schedule to the pre-reform durations before scaling up these pre-reform durations by the treatment effect; when doing so, the percentage effect of increased benefits at given, pre-reform durations equals 2.36 percent, thus delivering a utilization effect that measures 38 percent of the total dollar impact.

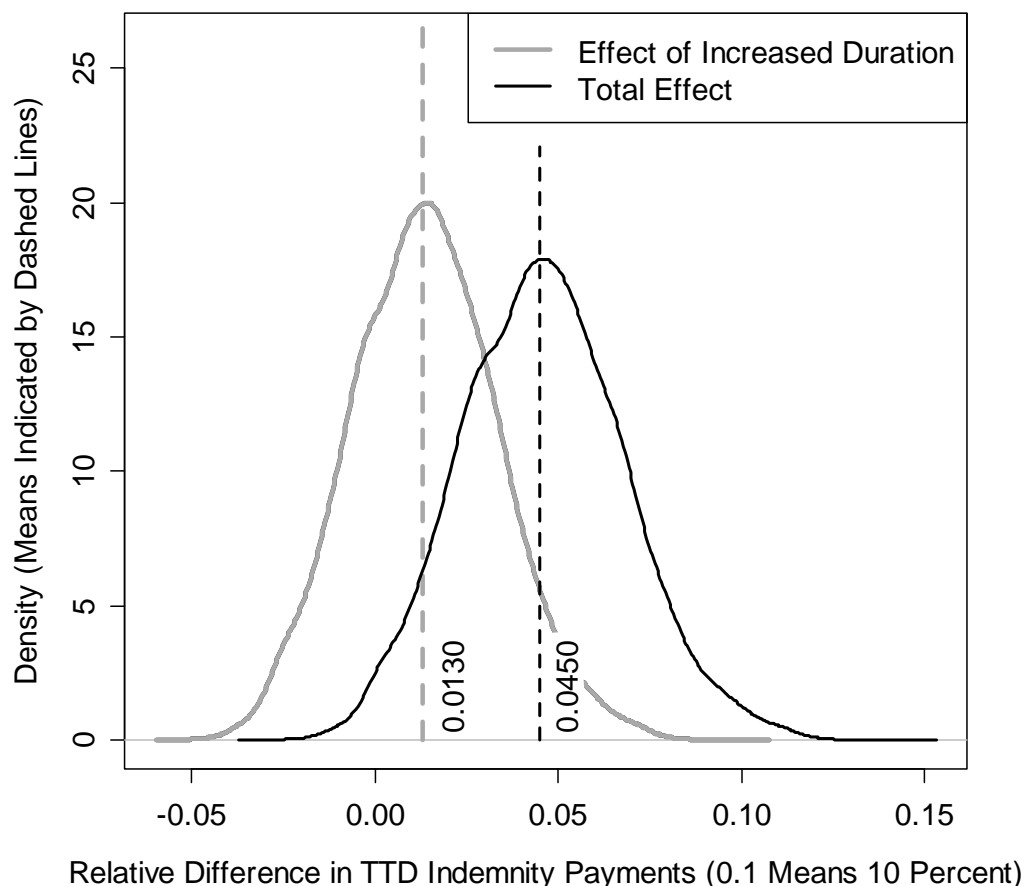
The second approach of breaking down the total effect has the advantage of delivering a ready-to-use formula for arriving at the total effect once the post-reform benefit schedule has been applied to the observed pre-reform durations. Let b be the relative (percentage) increase in costs due to the change in the benefit schedule for observed pre-reform durations, let u be the proportion of the utilization impact in the total effect (as calculated using the second approach), and let $d^{pre-reform}$ and $d^{post-reform}$ be the dollar amounts of pre-reform and post-reform indemnity payments, respectively. Then, we can write: $d^{post-reform} = d^{pre-reform} \times (1 + b / (1 - u))$.

Chart 12: Relative Increase in Payments on TTD Claims: Treatment Effect and Total, Oregon



When calculating the treatment effect for New Mexico, we obtain as the total effect on indemnity payments of the change in the maximum weekly benefit a 4.50 percent increase (compared to 3.82 percent for Oregon). Of this total effect, 1.30 percentage points are due to the utilization increase, which amounts to about 29 percent of the total effect and, thus, is close to the 31 percent established for Oregon. Note that for New Mexico, the waiting period equals 7 days (compared to 3 days for Oregon), and the retroactive period equals 28 days (compared to 14 days for Oregon). Chart 13 exhibits the means and probability distributions of the total impact and the utilization effect, using 4,000 draws for the bootstrapped distributions.

Chart 13: Relative Increase in Payments on TTD Claims: Treatment Effect and Total, New Mexico



Finally, Table 1 summarizes the findings for Oregon and New Mexico. Columns 9 and 10 offer alternative ways of calculating the proportion of the utilization effect in the total effect. For the purpose of adjusting the direct cost effect of a reform (which is obtained by applying the post-reform benefit schedule to the pre-reform durations) using the discussed formula, the values in column 10 have to be used.

Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

Table 1: Summary of Estimated Cost Effects

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Legislative Reform: Percentage increase in maximum weekly benefit	Increase in benefit duration in treatment group (measured in weeks)	Percentage increase in benefit duration in treatment group	Resulting duration/benefit elasticity (column 4, divided by column 2)	Total cost increase in percent	Percentage points of total cost increase that are due to utilization increase at pre-reform benefit levels	Percentage points of total cost increase that are due to increase in benefits at pre-reform duration levels	Percentage of utilization increase in total cost increase (column 7, divided by column 6)	Alternative concept of utilization increase as a percentage of total cost increase (column 6 minus column 8, divided by column 6)
Oregon	33.00	0.76	17.49	0.53	3.82	1.17	2.36	31	38
New Mexico	17.65	0.41	7.64	0.43	4.50	1.30	3.00	29	33

Note: All computations rest on unrounded numbers. Columns (7) and (8) do not add up to column (6) due to the changes not being infinitesimally small. For instance, let $z_0 = x_0 \cdot y_0$, then we can write: $dz \approx x_0 \cdot dy + y_0 \cdot dx$. This relation holds at equality if dx and dy are infinitesimally small.

5. REGRESSION APPROACH TO TREATMENT EFFECT

The “difference in differences approach” to gauging the effect of the increase in the maximum weekly benefit on benefit duration (the “treatment effect”) assumes that changes in economic activity or an underlying trend in benefit duration bear on the control and treatment groups in similar ways. This assumption may be violated if the impact of an economic recession (or the subsequent economic recovery) affects the treatment group (which consists of high-wage earners) and the control group differently. As Chart 8 indicates, for Oregon, the time window prior to the reform was characterized by an economic slowdown, whereas the time window following the reform coincided with an economic expansion. The findings presented in the section pertain to Oregon.

In order to validate the 17.49 percent benefit increase obtained with the “difference in differences” approach, a partial linear regression model is estimated. This model reads

$$y_i = \mathbf{x}_i \cdot \boldsymbol{\beta} + f(z_i) , \quad (1)$$

where y_i is the (natural) logarithm of the benefit duration of claimant i . In this semiparametric model, the vector \mathbf{x}_i comprises the covariates in the linear, parametric component, whereas the smoother $f(z_i)$ models the (single) covariate in the nonparametric part. Here, the covariates in the parametric component are exclusively indicator variables, which represent the claimant’s gender, occupation, injury year, affiliation with the control group, and affiliation with the post-reform treatment group. The covariate in the nonparametric component, the influence of which is allowed to be nonlinear, is the claimant’s age at injury, measured in full years; this covariate was centered. Occupation is categorized into nine major groups based on the U.S. Census Bureau Occupation Codes, as used in the 1990 Census of Population and Housing (see www.census.gov). The reference group consists of male claimants that are employed in a service occupation, sustain a workplace-related injury in 2002 and belong to the post-reform treatment group. As in the “difference in differences” approach, only claims that belong to the control or treatment group are included in the analysis.

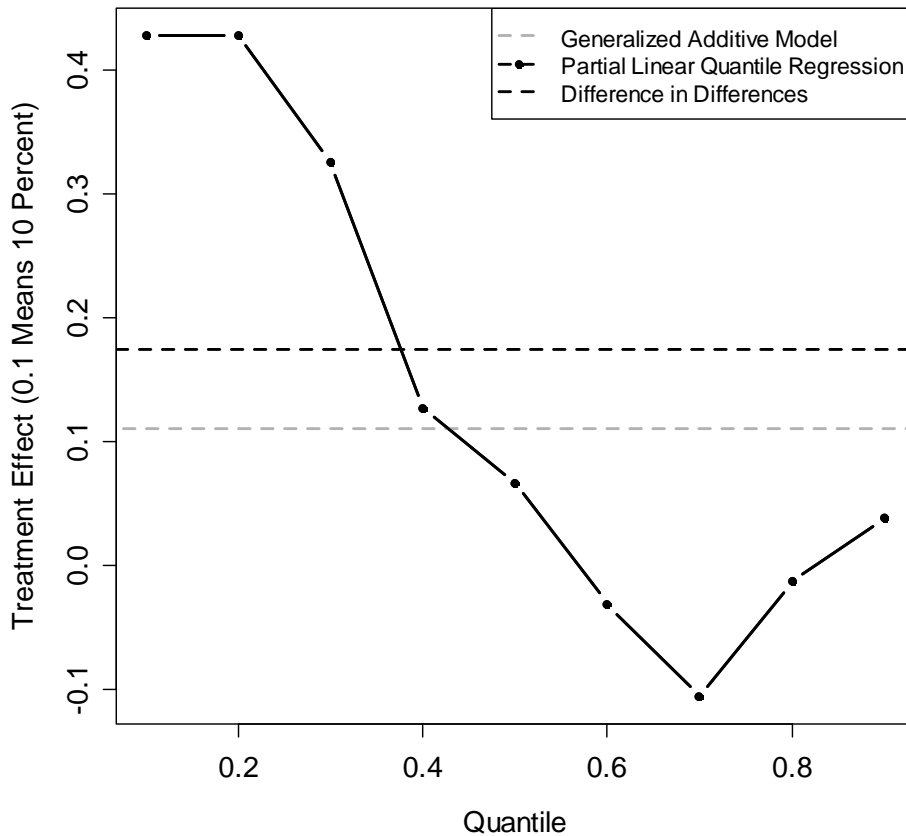
Model (1) is estimated using a generalized additive framework, as provided by the R package *mgcv* (<http://cran.r-project.org/web/packages/mgcv/index.html>; version 1.5-5, May 2011). See the *Casualty Actuarial Society E-Forum*, Winter 2011-Volume 2, page 26.

15, 2009), authored by Simon Wood. The smoother $f(z_i)$ is a thin plate regression spline; see Wood [15] for details. The business cycle effect on injury duration is controlled for by the set of year indicator variables. Variation in the composition of claimants by occupation, gender, or age is accommodated by the respective covariates. The treatment effect equals the regression coefficient of the post-reform treatment group indicator variable, after adjustment for the logarithmic nature of the dependent variable; for details on this adjustment, see Halvorsen and Palmquist [7] (and Kennedy [9], for a discussion of the properties). Chart 14 displays the transformed regression coefficient (as a gray dashed line) alongside the result of the “difference in differences” approach (black dashed line). The treatment effect of 11.1 percent, as obtained from the generalized additive model, is somewhat smaller than the treatment effect of 17.49 percent delivered by the “difference in differences” approach.

Standard regression approaches (such as least squares, iteratively reweighted least squares, or maximum likelihood approaches) typically regress on the mean of the distribution of the dependent variable. Although regression on the mean of the distribution offers important insights into the average effect, it also obscures a possible nonuniform influence across the range of observed benefit durations. Quantile regression, as developed by Koenker and Basset [11], offers a means of uncovering such a possible nonuniform influence. Thus, the partial linear specification of model (1) is reestimated following Koenker, Ng, and Portnoy [12]; in that approach, the smoother $f(z_i)$ rests on total variation regularization. Software for estimating the partial linear model is available as part of the R package *quantreg* (<http://cran.r-project.org/web/packages/quantreg/index.html>; version 2.6, February 5, 2009), authored by Roger Koenker.

Quantile regression minimizes the sum of absolute errors; this problem is solved using linear programming techniques (such as the family of interior point algorithms). If positive and negative errors receive equal weight, then quantile regression quantifies the effect of the covariates on the median of the dependent variable. If, on the other hand, errors are weighted asymmetrically, regression on quantiles other than the median becomes available. For instance, if underestimating the observed value is penalized (at the margin) three times higher than overestimating it, then the solution that emerges is for the 75th percentile (see Koenker [10]).

Chart 14: Regression Approaches to Measuring the Treatment Effect, Oregon



The quantile regression estimates of the treatment effect are displayed in Chart 14. There are estimates for the 10th, 20th, ..., and the 90th quantiles. Whereas the legislative reform increases the benefit duration for very short durations by more than 40 percent, the effect on long durations is essentially nil.

6. BENEFIT DURATION AND CLAIM ATTRIBUTES

Quantile regression, due to its ability to offer a more differentiated picture of the behavior of the dependent variable in response to covariates, is a suitable framework for studying the effect on injury duration of the claimant's age and gender. Then again, when it comes to interpreting these regression results, it is important to remember that the TTD

Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

claims studied here are selected based on their final categorization. Many PTD (and even some Fatal) claims may have been categorized as TTD claims originally; such claims are not included in the analysis, as there is no information available on the initial categorization. Most importantly, claim attributes that are manifestations of the outcome of the event of injury (such as nature of injury or part of body) may be causal to the final categorization into TTD and PTD claims. For this reason, only covariates unrelated to the outcome of the injury, such as age and gender, are chosen in the following analysis. Unless stated otherwise, the findings presented in the section pertain to Oregon.

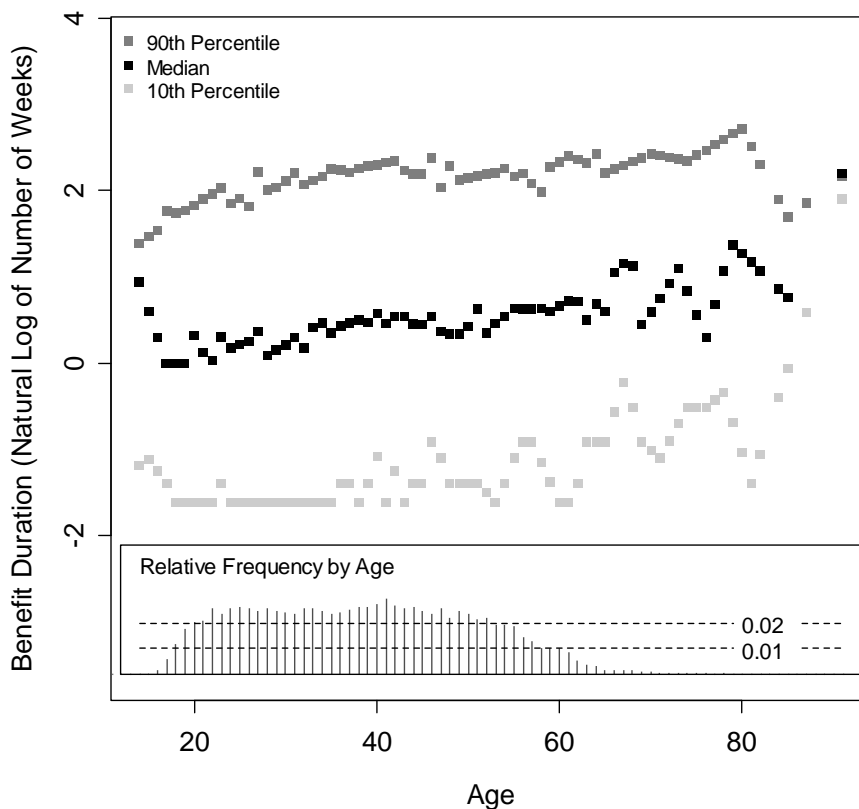
The impact of age is studied for the post-reform period, using model (1). To reduce noise, the data is pooled over the three post-reform years, with indicator variables controlling for differences among those years. A further covariate in the parametric component is gender. Age enters the nonparametric component, as before. The reference group consists of male claimants being injured in 2002.

Chart 15 displays the influence of age on claim severity as obtained for model (1) in a regression on the 10th percentile, the median, and the 90th percentile. There is a symbol for every year of age (where there is an observation for this age), except for the minimum age of 13, which serves as a reference; 18 claims with zero values for age were removed, thus resulting in 24,108 observations for the three-year time period. Note that the displayed effect of age includes the intercept (which is immaterial for the slope of the displayed duration trajectories). Also, at the bottom of the chart, there is a frequency distribution of claims by age; this distribution indicates that the data set is sparsely populated for claimant exceeding age 60; the maximum age is 91.

As Chart 15 shows, for the median, the benefit duration is roughly log-linear in age from the early twenties through the late fifties. An M estimator, applied to the estimated age effect at the median duration within the age bracket 20 through 60 delivers a geometric mean rate of growth per year of age of 1.0 percent. (A similar analysis for New Mexico delivers a geometric mean rate of growth of 0.72 percent.) The estimated age impact beyond the sixties is not meaningful due to the sparse number of claims. At the same time, there is no need to exclude these claims from the analysis, because the smoother adapts to the local environment. The measured age impact is less than the value established by Krueger [13] for Minnesota; using indicator variables for multi-year age brackets, this author finds an impact on the mean duration per one year of age to be about 1.6 percent (see his Table 3; calculated

as $((\exp(0.385)-1)/28.8$ from the regression coefficient for the age group 45–54, relative to the reference group 18–24). The age effects presented by Meyer, Viscusi, and Durbin [14] are elasticities and, due to their nature of being partial derivatives, cannot be generalized to large age intervals.

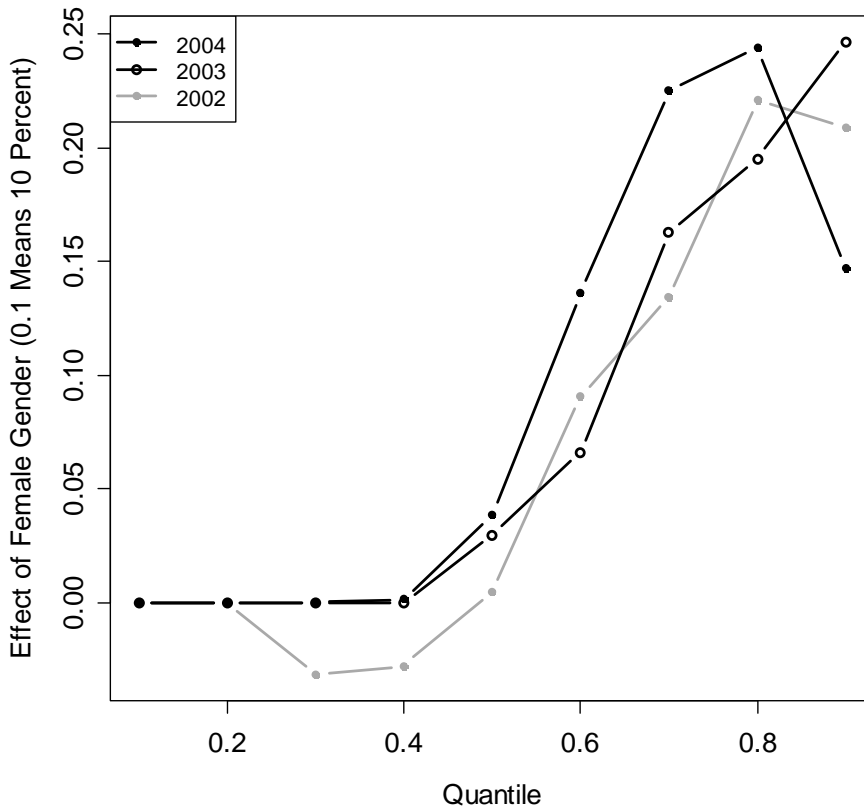
Chart 15: Quantile Regression: Effect of Age on Benefit Duration, Oregon



Finally, the effect of gender on benefit duration is analyzed for the post-reform period, again using the partial linear framework of model (1). Here, every year is estimated in isolation. Gender enters the parametric component of the regression model, along with indicator variables for occupation. Age is again included in the nonparametric component. The reference group consists of male claimants that are employed in a service occupation. Chart 16 presents the influence of the female gender on benefit duration for the 10th, 20th, ..., and 90th percentiles. This chart indicates that there is essentially no difference in

benefit duration between female and male claimants up to the 40th quantile. But for claims of very long benefit duration, female claimants are on indemnity benefits about 20 percent longer than male claimants.

Chart 16: Quantile Regression: Effect of Female Gender on Benefit Duration, Oregon



As mentioned, because claims “self-select” (by final categorization) into TTD, PTD, and Fatal, the effects of claim attributes on TTD benefit duration (and the effect of gender, in particular) have to be interpreted with caution. For instance, it is conceivable that the longer benefit duration of females is related to a higher proportion of PTD claims for males. Put differently, whereas a female claimant may be on benefits for an extended period of time (as a TTD claim) but may eventually return to work, the corresponding male claimant may end up in the PTD category.

7. CONCLUSIONS

Studying temporary total disability claims, the impact of 33 percent (Oregon) and 17.65 percent (New Mexico) increases in the maximum weekly indemnity benefit on benefit duration and the associated percentage increase in indemnity payments were analyzed in a quasi-experimental setting. The effect on benefit duration was measured using a “difference in differences” approach applied to pre-reform and post-reform treatment and control groups. The resulting 17.49 percent (Oregon) and 7.64 percent (New Mexico) increases in benefit duration (the “treatment effects”) translate into duration/benefit elasticities of 0.53 (Oregon) and 0.43 (New Mexico). These values agree with the rule of thumb suggested by Gardner [5], which states that a 20 percent increase in benefits comes with a (minimum) increase in utilization of 10 percent. At the same time, these elasticities are close to the values reported by Meyer, Viscusi, and Durbin [14] for TTD benefit increases of about 50 percent in Kentucky and Michigan; these authors’ elasticities ranged from 0.27 to 0.62, but clustered mostly within the range of 0.3 to 0.4. On the other hand, the elasticities of 0.43 and 0.53 are considerably lower than the value of 1.67 that Krueger [13] established in his study of a 5 percent TTD benefit increase in Minnesota. Then again, Gardner [4], who studied a 50 percent TTD benefit increase in Connecticut, found that for every 20 percent increase in benefits, utilization increases by about 18 percent, thus resulting in an elasticity located between the values established by Meyer, Viscusi, and Durbin [14] and Krueger [13].

For Oregon, the treatment effect was validated using a generalized additive regression model, which yielded a somewhat lower increase of 11.1 percent. Further, using quantile regression on the Oregon data, it was demonstrated that most of this benefit duration increase originates in a lengthening of short durations; long benefit durations are nearly unaffected by the reform. This finding is consistent with evidence established by Krueger [13], Figure 3; further, Curington [3] found that the duration of PPD claims with minor impairment are more responsive to benefit changes than those with major impairment.

An increase in the maximum weekly benefit may give rise not only to longer benefit durations, but also to a higher number of indemnity claims, which would add to the cost of the reform. For instance, Gardner [4] finds in a study for Connecticut that a 50 percent increase in the maximum weekly benefit was associated with an increase in the number of indemnity claims of 5 percent. On the other hand, a recent study by Guo and Burton [6] arrives at the conclusion that the overall benefit elasticity (change in duration and frequency

Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

taken together) is in fact negative, which implies that frequency or duration (or both) drop in response to more generous benefit provisions.

No statement can be made with confidence on how the 33 percent increase in the maximum weekly benefit studied here may have affected the claim count. This is because the data-cleansing algorithms that separate TTD and TPD claims in the Oregon data set, while improving data quality for the study of duration, may adversely affect the validity of the claims count information. Further, the reform of interest, which is the increase in the maximum weekly benefit for TTD claims, was accompanied by an increase in compensation for both scheduled and nonscheduled PPD injuries; this change in PPD benefits may also have influenced the incentive to file claims, apart from the increase in the maximum weekly benefit of TTD claims.

Further research is necessary for a better understanding of the effect of benefit changes on claim counts and of the effect of socioeconomic attributes (beyond age and gender) on the duration/benefit elasticity.

Appendix

A generalized additive regression model (GAM) is a semiparametric (or, synonymously, partial linear) generalized linear model (GLM), specified as the sum of nonparametric and parametric regression components. The purpose of the nonparametric regression component is to accommodate a possibly nonlinear influence of a covariate (or a set of multiple covariates). The estimation of the nonparametric component requires a smoother.

Similar to generalized additive models, partial linear regression models consist of nonparametric and parametric components. Unlike traditional regression approaches (including generalized linear additive models), which quantify the influence of covariates on the mean of the dependent variable, quantile regression models gauge the influence of covariates on quantiles of interest, such as the median.

The design of a quasi-experiment differs from that of a (controlled) experiment in that the assignment to control group and treatment group is not random (but still outside the control of the person conducting the experiment).

The thin plate regression spline is a smoother that finds its analogy in the bending of sheets of metal. A major advantage of the thin plate spline is that it has no free parameters that need tuning. Like other smoothers, the purpose of the thin plate regression spline is to discern a (potentially) nonlinear functional form in the data.

Total variation regularization is a method of “de-noising” data, that is, a way of discerning the underlying structure in data points observed with error. In the semi-parametric quantile regression model discussed here, total variation regularization serves as a smoother.

M estimation is a technique where extreme deviations from the conditional mean of the dependent variables are downweighted. That way, the estimated coefficients are robust to outliers (in the dependent variable). By contrast, least squares regression, which has a quadratic objective function, affords the same weight to all observations.

Acknowledgment

Thanks to NCCI staff: Robert Moss for his leadership in this project, John Robertson and Raji Chadarevian for comments, Anna Elez, Linda Li, Ashley Pistole, and Bruce Ritter for research assistance.

5. REFERENCES

- [1] Brooks, Ward, *A Study of Changes in Frequency and Severity in Response to Changes in Statutory Workers Compensation Benefit Levels*, San Francisco, CA: Workers Compensation Insurance Rating Bureau, 1998.
- [2] Butler, Richard J., and John D. Worrall, (1985) “Work Injury Compensation and the Duration of Nonwork Spells” *Economic Journal* 95, September 1985, pp. 714-724.
- [3] Curington, William P., “Compensation for Permanent Impairment and the Duration of Work Absence: Evidence from Four Natural Experiments,” *Journal of Human Resources* 29(3), 1994, pp. 888-910.
- [4] Gardner, John A., *Benefit Increases and System Utilization: The Connecticut Experience*, Cambridge MA: Workers Compensation Research Institute, 1991.
- [5] Gardner, John A., *Return to Work Incentives*, Cambridge, MA: Workers Compensation Research Institute, 1989.
- [6] Guo, Xuguang, and John F. Burton Jr., “Workers’ Compensation: Recent Developments in Moral Hazard and Benefit Payments,” *Industrial & Labor Relations Review* 63(2), 2010, pp. 340-355.
- [7] Halvorsen, R., and R. Palmquist, “The Interpretation of Dummy Variables in Semilogarithmic Equations,” *American Economic Review* 70(3), 1980, pp.474-475.
- [8] Johnson, William, G., and Jan I. Ondrich, “The Duration of Post-injury Absences from Work,” *Review of Economics and Statistics* 72(4), 1980, pp. 578-586.
- [9] Kennedy, Peter E., “Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations,” *American Economic Review* 71(4), 1982, pp. 801.
- [10] Koenker, Roger, *Quantile Regression*, Cambridge (UK): Cambridge University Press, 2005.
- [11] Koenker, Roger, and Gilbert Bassett, (1978) “Regression Quantiles,” *Econometrica* 46(1), 1978, pp. 33-50.
- [12] Koenker, Roger, Pin Ng, and Stephen Portnoy, “Quantile Regression Smoothing,” *Biometrika*, 81(4), 1994, pp. 673-680.

Indemnity Benefit Duration, Maximum Weekly Benefits, and Claim Attributes

- [13] Krueger, Alan B., “Workers’ Compensation Insurance and the Duration of Workplace Injuries,” NBER Working Paper #3253, 1990.
- [14] Meyer, Bruce D., W. Kip Viscusi, and David L. Durbin, “Workers’ Compensation and Injury Duration: Evidence from a Natural Experiment,” *American Economic Review* 85(3), 1995, 322–340.
- [15] Wood, Simon N., “Thin Plate Regression Splines,” *Journal of the Royal Statistical Society: Series B* 65(1), 2003, pp. 95–114.
- [16] Worrall, John D., Richard J. Butler, Phillip Borba, and David Durbin, “Estimating the Exit Rate from Workers’ Compensation: New Hazard Rate Estimates,” *Technical Paper RD-88-2*, National Council on Compensation Insurance, 1988.

Abbreviations and notations

NCCI, National Council on Compensation Insurance, Inc.
PPD, Permanent Partial Disability
PTD, Permanent Total Disability
TPD, Temporary Partial Disability
TTD, Temporary Total Disability

Biography of the Author

Frank Schmid, Dr. habil., is a Director and Senior Economist at the National Council on Compensation Insurance, Inc.

Mortality Trend Models

Gary G. Venter

Abstract

Every 50 years or so a study of workers compensation mortality patterns is done, generally finding that after medical stabilization – 10 or more years after injury – mortality for seriously injured workers is comparable to that of the overall population. It has been about 25 years since the latest study, so we might be half way to the next one. But in the meanwhile there are trends in population mortality, and these impact loss reserve risk.

Mortality data over time can be arranged in triangles, and models fit to such data are similar to those used in casualty loss development – particularly those that model trends in the three dimensions of calendar year of finalization, age at finalization, and origin year. We fit such models to U.S. population male and female mortality data for death (finalization) ages 55 to 89, with several distributions of residuals. The information matrix is used to estimate parameter standard deviations.

Although there is an extensive literature on fitting these models, most of the papers do not address parameter significance through t statistics, etc. and doing so finds problems with the standard models. One problem is over-parameterization, and a conclusion here is that parameter reduction methods such as smoothing should be used. Other authors have tried this, but a sticky issue is finding parameter reduction methods that actually produce improvements in goodness of fit, as measured by AIC, etc. This is an open problem as far as we know and a direction for future research.

Typically the starting point for the distribution of model residuals is Poisson, but several authors have found that negative binomial fits better. Unfortunately, some of these have misinterpreted the derivation of the negative binomial as a gamma-mixed Poisson to conclude that the negative binomial arises because there are different subpopulations each with different Poisson distributions. But a sum of subpopulations each Poisson distributed is itself Poisson distributed. The mixture becomes interesting when you are drawing at random from a subpopulation whose parameter you do not know. Probably the negative binomial arises from other contagion effects, like weather, disease outbreaks, etc. Unfortunately, these also make residuals across cells not independent, and this effect has been found in other studies as well.

A few alternative ways of parameterizing negative binomial residuals are discussed, and these are also applied to the Poisson-Inverse Gaussian distribution and its generalization, the Sichel. For females the negative binomial fits best but the male data is a bit more skewed than the negative binomial. However the Poisson inverse-Gaussian appears to be too skewed for this data. The Sichel is more flexible, with one more parameter, and fits best.

Further insight into the shifts in mortality over time is provided by fitting Makeham-like curves to each year of death. One implication from this exercise is that male mortality trends at the older ages had a shift in 1988, possibly data related. Probably data older than that is not reliable, or at minimum comes from a different process. The overall conclusion is that more work is needed to come up with reasonable models for mortality trend, with parameter reduction a leading candidate.

For trending, ARIMA models have often been fit to the calendar-year parameters after first differencing for stability. But since the parameters are estimated with error, differencing induces an autocorrelation, so the ARIMA models could be mostly fitting this artifact. Alternatives are discussed.

Keywords: Mortality Risk; Lee-Carter Model; Cohort Effects; Parameter Risk; Model Risk

MORTALITY TREND MODELS

The general categories of process, parameter and model risk are applicable to mortality projection. Model risk is particularly problematic, as it turns out that the better fitting models have aspects that make them questionable for projection purposes. Lee-Carter models with and without cohort

effects with a few distributions of residuals are fit to the population mortality data from the Human Mortality Database (HMD) and are compared based on penalized maximum likelihood.

The models were fit to years of death starting with 1971. Preliminary analysis found different trends for ages of death below 55, due perhaps to reproductive health issues and the impact of HIV during some of this period. Female mortality below age 55 improved dramatically in the 1970s and has changed little since, whereas for males there was a sharp increase in mortality in the 1990s that has since recovered. The oldest age used is 89, as older ages had quite unusual mortality patterns before 1990—mortality reducing with age, etc. These could be data issues. The data available for this study ends with year of death 2006. This resulted in using year-of-birth cohorts 1882 to 1951. The cohort is year of death minus latest attained age at death, so is close to year of birth.

The fits with cohort parameters turn out to be problematic in part because the oldest cohorts have only a few observations, which makes their parameters very responsive to just a few data points, and this in turn creates distortions in other parameters. Adding the data for all years of death 55 – 89 for cohorts 1882 and later, reduces this problem. Another problem with the fits is that in the case of female death rates, the correlations among parameter estimates is high, which reduces the significance of the parameters and leads to questionable values.

Section 1 discusses the models used; Section 2 looks at the fits; Section 3 tries to interpret the parameters; Section 4 address adding more years of death; Section 5 looks at Makeham-like fits; and Section 6 gets to projection risk.

1. MODELS

HMD data comes in the form of number of deaths and number of living, who are considered the exposures to death. These are in cells by year of death and age at death. Subtracting age from year gives the cohort, which is approximately the year of birth, but can be slightly different depending on the time of year that birth and death occurred. Data is also available by actual year of birth but in most models that is considered less important, and cohort is used instead.

Here arrays are taken to have rows for year of death and columns for age at death. The years are 1971 to 2006, and the ages 55 to 89, so the arrays are 36x35, with 1260 elements. The years are indexed by t and the ages by d . The cohort is $t - d$ and is constant along the NW-SE diagonals of the arrays.

The starting point for recent models of mortality is the LC model from Lee and Carter (1992). It models the mortality ratio m , which is deaths divided by exposures, in log form the mean is:

$$\log m_{t,d} = a_d + b_d h_t. \tag{1.1}$$

Here a_d is the base mortality for age d , h_t is the trend level at year t , which generally goes down over time as mortality decreases, and b_d allows different ages to have trend rates that are factors times the overall trend. This is useful in the case of male mortality, for example, where mortality rates for ages 55 to 60 have improved at a greater rate than those for 85 to 89. However, this is where the LC model can run into differences from actual data, as some ages might trend faster or slower for a while but not always.

A popular extension of LC is LC plus cohorts, from Renshaw-Habermann (2006) (RH):

$$\log m_{t,d} = a_d + b_d h_t + c_d u_{t-d}. \quad (1.2)$$

The cohort term u allows for mortality to also vary by year of birth, independently of year of death and age. It is not always clear why it should, but allowing it to seems to substantially improve the goodness of fit of the models. The c factor allows the cohort effect to vary by age; e.g., it might wash out at older ages, or it might be stronger at older ages.

There are some identifiability problems with these models. For instance, increasing every b by a factor and reducing every h by the same factor does not change the fitted values. This is similar for c and u . Here the constraints used for this are to set $b_{1955} = c_{1955} = 1$ and $h_{1971} = u_{1917} = 0$. The cohort 1917 was chosen as it is the last cohort that includes all calendar years. It is also one of the highest mortality cohorts for both males and females. The result of these constraints is that in the LC model a_d is the fitted mortality for $t = 1971$ and h_t is the trend level for age 55. All the other parameters are relative to these. In the RH model, every u is the cohort effect at age 55, where the cohort values are relative to cohort 1917. Traditionally sums of parameters have been constrained as a way to address the identifiability problems, but the approach here eliminates a few parameters, which is necessary to make the information matrix non-singular.

Fitting is done by maximum likelihood estimation (MLE). Denote the exposures in the t,d cell by $E_{t,d}$ and the deaths by $D_{t,d}$. The Poisson model is that $D_{t,d}$ is Poisson in $m_{t,d}E_{t,d}$, where $m_{t,d}$ could come from either the LC or RH model. With mean μ , the log of the Poisson probability at k is $k \log(\mu) - \mu - \log(k!)$. The loglikelihood is then:

$$\sum_{t,d} \{D_{t,d} \log[m_{t,d} E_{t,d}] - m_{t,d} E_{t,d} - \log[D_{t,d}!]\}. \quad (1.3)$$

Two forms of the negative binomial distribution are also fit. The negative binomial has two parameters r and β , with mean $r\beta$ and variance $r\beta(1+\beta)$. But in modeling a whole array of negative binomial variates it is customary to make the mean a parameter and model it with the covariates. In this case the mean would still be $\mu_{t,d} = m_{t,d}E_{t,d}$, as in the Poisson case.

To make the mean a parameter, set $\mu = r\beta$. The two forms arise by either eliminating r by setting $r = \mu/\beta$, or eliminating β by setting $\beta = \mu/r$. Here these are called NB1 and NB2, respectively. Both

have mean μ , but NB1 has variance $\mu(1+\beta)$ and NB2 has variance $\mu(1+\mu/r)$, which are linear and quadratic in μ , respectively. Denoting the log of the gamma function by lgamma , the log of the probability at k for the negative binomial in r and β is:

$$\text{lgamma}(r+k) + k\log(\beta) - \text{lgamma}(r) - \text{lgamma}(1+k) - (r+k)\log(1+\beta). \quad (1.4)$$

The loglikelihoods for NB1 and NB2 can be obtained by substituting μ/β for r or μ/r for β , then $D_{t,d}$ for k and $m_{t,d}E_{t,d}$ for μ , and summing over the observations.

2. FITS

Goodness of fit of different models can be compared using penalized likelihood. The traditional comparison is to start with the negative loglikelihood (NLL) and add a penalty. Here the traditional criteria divided by 2 are used, as these are more directly related to the NLL, but the standard names are retained. Thus the Akaike Information Criterion (AIC) uses a penalty of 1 for each parameter. If N is the sample size (number of observed cells), the Bayesian Information Criterion (BIC) uses a penalty of $\frac{1}{2} \log N$ for each parameter. There is some feeling among information theorists that the AIC is too lenient on extra parameters, but the BIC is too punitive. The Hannan-Quinn Information Criterion (HQIC) is intermediate. It gives a penalty of $\log \log N$ for each parameter. It turns out that most of the conclusions are the same for each criterion, so until a difference arises, only the BIC will be used, but HQIC will be the fallback if there is a difference. For $N = 1260$, the penalty is about 3.57 per parameter. Thus an extra parameter has to improve the NLL by that much to be justified.

LC and RH Poisson models were fit to male and female mortality. For both datasets, the RH model fit quite a bit better than LC. The RH NB1 and NB2 models were then fit. Table 1 shows the NLL for each model and the improvement in NLL required to meet the BIC requirement for the extra parameters from the model above it. After the parameter constraints there are 35 a parameters, 34 b parameters and 35 h parameters, so the LC model has 104 parameters. In the RH model there are 34 c parameters and 69 u parameters, for cohorts 1882 to 1951, ex 1946. Thus it has 207 parameters. The negative binomial versions have yet one more parameter.

Table 1. Fit Comparisons

Model	NLL		Parameters Added	BIC Needed NLL Improvement	NLL Improvement	
	Female	Male			Female	Male
LC Pois	13670	14868				
RH Pois	9598	10081	103	368	4072	4787
RH NB1	8798	8996	1	3.6	800	1085
RH NB2	8748	8972	0	0	50	24

The LC model fits considerably worse for males, with an NLL 1198 higher than for females. The RH model fits much better for both, with an improvement in NLL of 4072 for females and 4787 for males, compared to an improvement of 368 required to justify the extra parameters according to BIC. The difference between males and females is narrowed to 483, so RH is an even more substantial improvement for males. Adding the extra parameter for NB1 also significantly improves both fits, and NB2 is a bit better yet. The β parameter for NB1 is about 2.5 for females, and 3 for males, so the variance for each cell is 3.5 to 4 times the cell mean, compared to equal to the mean for Poisson. That is a substantial difference, and with variances that big it is no wonder the Poisson fit is not as good. The r for NB2 is about 8600 for females and 7500 for males, which for this data translates to variances of 2 to 7 times the mean, with the higher ratios going to the larger cells.

The best NB NLLs for LC were 9444 for females and 9652 for males, so Poissonness is the bigger culprit for LC – Poisson than the lack of cohort parameters. Still the NB RH model is better than NB LC for females by 696 and for males by 680, which are still well above the BIC need of 368, although nowhere near the NLL improvements of 4000+ for the Poisson models.

To give a visual impression of the fits, the empirical and modeled values of log m are graphed for a few years of death by age at death for the Poisson models. The graphs do not look much different for the negative binomial models, and in fact the parameters are not that different either. The advantage of the negative binomial models is more in the error distributions than in the fitted means. Essentially the cells with higher variance are not penalized as much in the likelihood functions for being different from their means, so the fit gets better for the smaller cells. This is not enough to be very noticeable in the graphs, however.

Figures 1 and 2 show the female data and fits. The mortality rates increase by age and this is close to a linear function for the log rates. In a graph of the rates for several calendar years of death, most

of the vertical range is taken up by this increasing trend, which makes it difficult to see the differences among the calendar years. To look at goodness of fit, the linear trend by age is not so critical, so in the graphs this is eliminated by subtracting age at death / 11 from each log mortality rate, essentially rotating the graph to make the lines roughly horizontal. This makes all the vertical range available to compare the actual and fitted rates for the various calendar years. A constant of 10.5 has been added to make the resulting numbers start near zero on the vertical axis. Rates have been declining over time, so the most recent calendar year is at the bottom of the graph. The dotted lines are the data, and the solid lines are the model.

Figures 3 and 4 are similar for males, but more years are able to be shown as the trends are greater for males, which separates the years a bit. Also, since the male mortality rates are higher, the rotated rates start around 0.45 instead of zero.

Figure 1. Rotated Graph of LC Female Mortality Rates

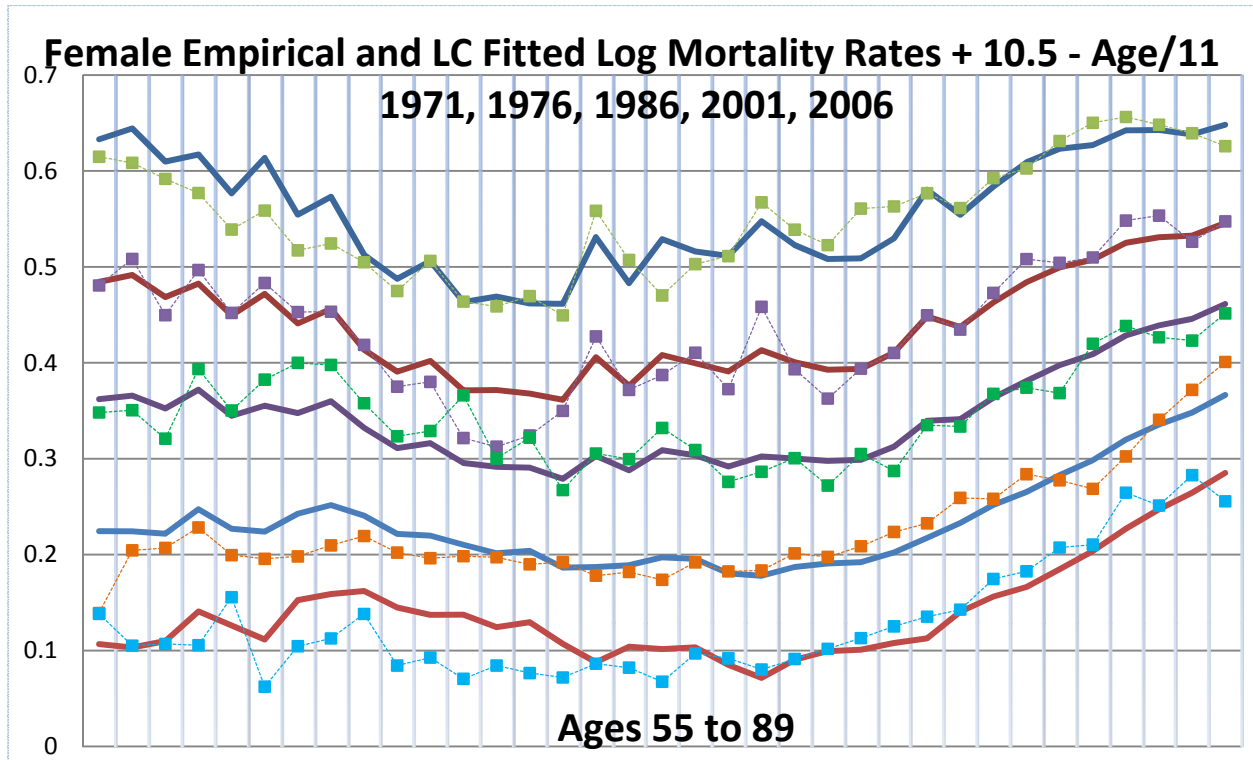


Figure 2. Rotated Graph of RH Female Mortality Rates

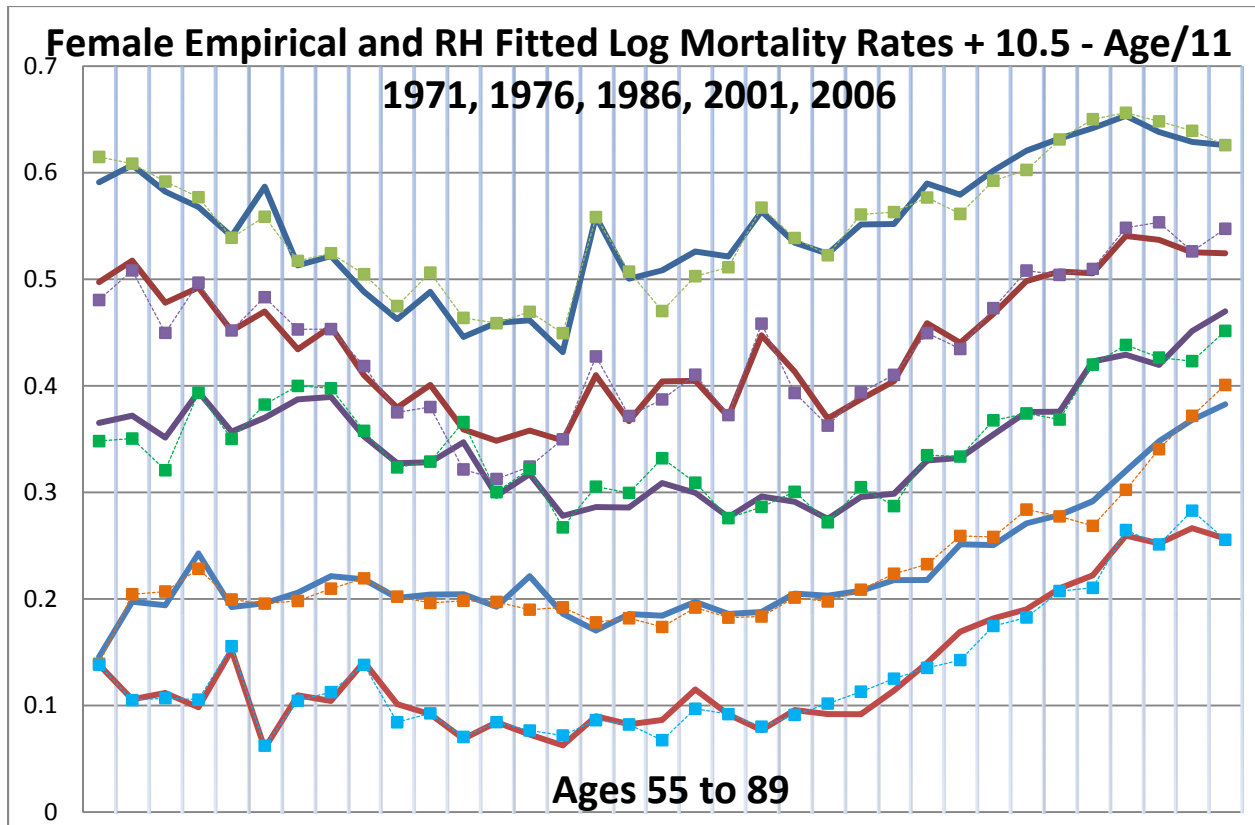


Figure 3. Rotated Graph of LC Male Mortality Rates

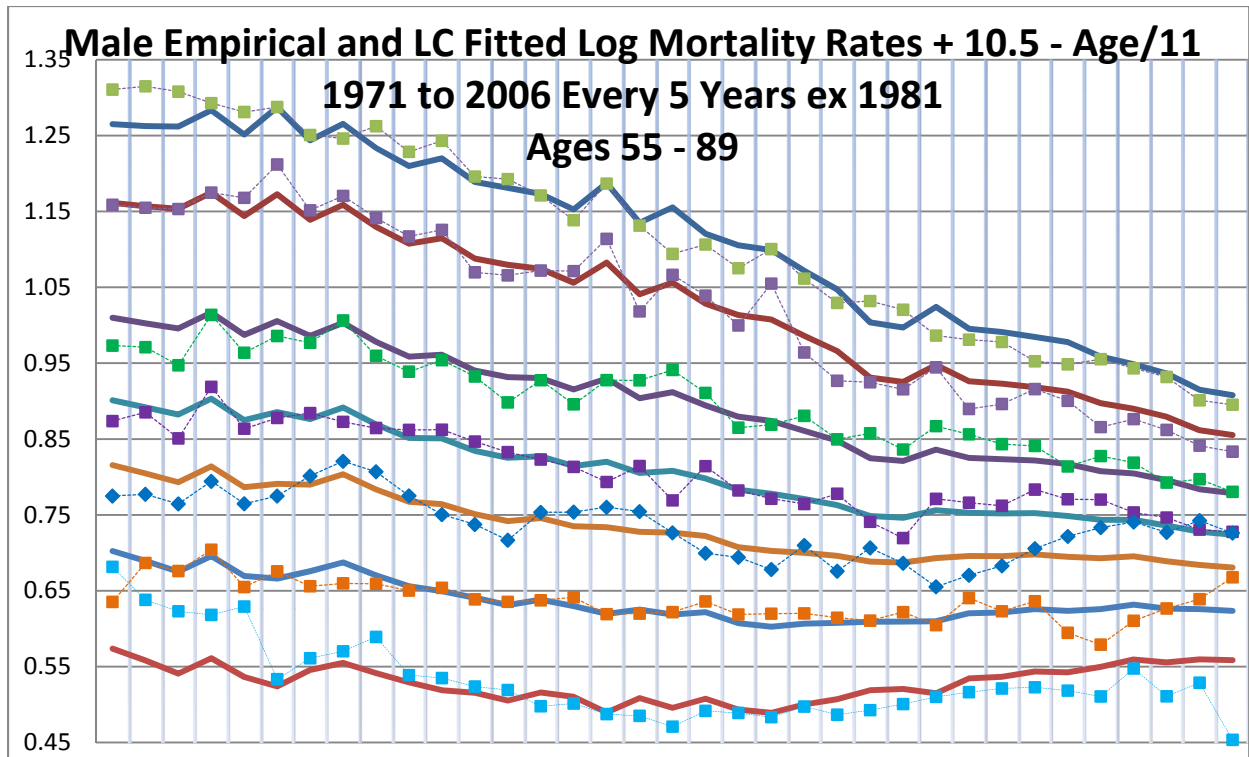
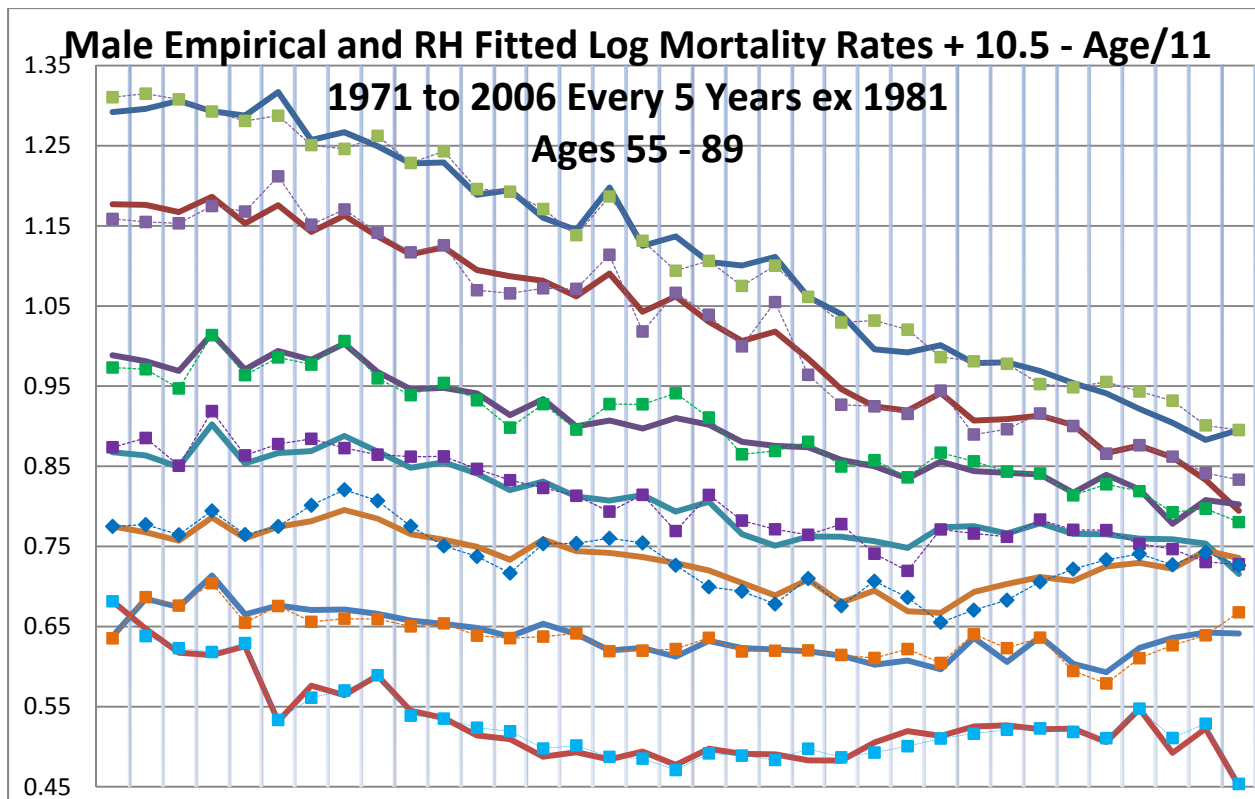


Figure 4. Rotated Graph of RH Male Mortality Rates



For males, the downward trend from year to year is less at the older ages. The LC model can handle this by having lower b factors for the older ages. When the LC model misses, it seems to be mostly for the youngest and oldest age groups. By adding in cohort parameters, the RH model can account for many of these effects. However, some of this is suspicious, as some of the cohorts have few observations in the data. Hence the modeled mortality at age 55 is higher in 2006 than in 2001, which follows the data, but that in itself does not establish 1951 as a high-mortality cohort.

For females, the shape of the graph is somewhat different than for males, and there is not so much diminishing trend at older ages. The LC model does not have enough flexibility to capture the changes in shape, and some of the years have long strings of significant errors of the same sign. The RH cohort-effects are able to adjust for a lot of this, but again this is sometimes because of cohorts with only a few observations, such as age 55 for 2001 and 2006.

The RH model provides better fits both in graphical tests and penalized MLE. Although the fits are worse for males, the difference narrows considerably for the better-fitting models. The negative binomial versions are much better than the Poisson, with the NB2 a bit better-fitting than the NB1. There are some plausibility problems with the RH model, however.

3. INTERPRETING THE PARAMETERS

The best-fitting NB2 parameterizations are used in this section. Does the fact that NB2 fits better than NB1 have any implications? NB2 is the form that comes from mixing a Poisson by a gamma distribution. This arises in experience rating, for instance, if each policy is Poisson-distributed, but there is a gamma distribution of Poisson means across the population. Taking a policy at random, its claims are conditionally independent given its Poisson mean, but unconditionally correlated due to the common Poisson mean. This is a way of modeling non-independent claims, or contagion.

It is tempting then to argue that the population as a whole is a mixture of groups with different mortality, due to different lifestyles, access to medical care, etc., and that is the source of the contagion observed. However that is a different kind of mixture. The population as a whole consists of all the groups taken together, not one drawn at random. The sum of independent Poisson distributions is itself Poisson, so the mixture argument does not explain contagion at the level of the entire population. Moreover, the number of deaths is the sum of Bernoulli processes and would be binomial, not Poisson, if there were not already some source of contagion to begin with.

There are factors affecting mortality rates for the population as a whole, such as weather, flu outbreaks, etc., that make deaths not independent. This could be the principal source of contagion at the population level. The NB1 model makes the variance about 4 times the mean for each cell in the data, whereas for the NB2 model it ranges from about 2 to 7 times the mean, with the factor larger for the larger cells. The fact that NB2 fits better suggests that the contagion events hit the larger cells harder. That is, the ages with the greatest number of deaths also have the greatest increases in deaths when adverse conditions arise.

Figure 5 graphs the a_d parameters, which represent the base log mortality rate by age, before application of trends and cohort effects, for males and females. Male mortality is higher than female at all ages, but that does not show with these parameters. The calendar-year parameters and cohort parameters interact with these so in themselves they are not that meaningful.

Figure 5. a_d Parameters

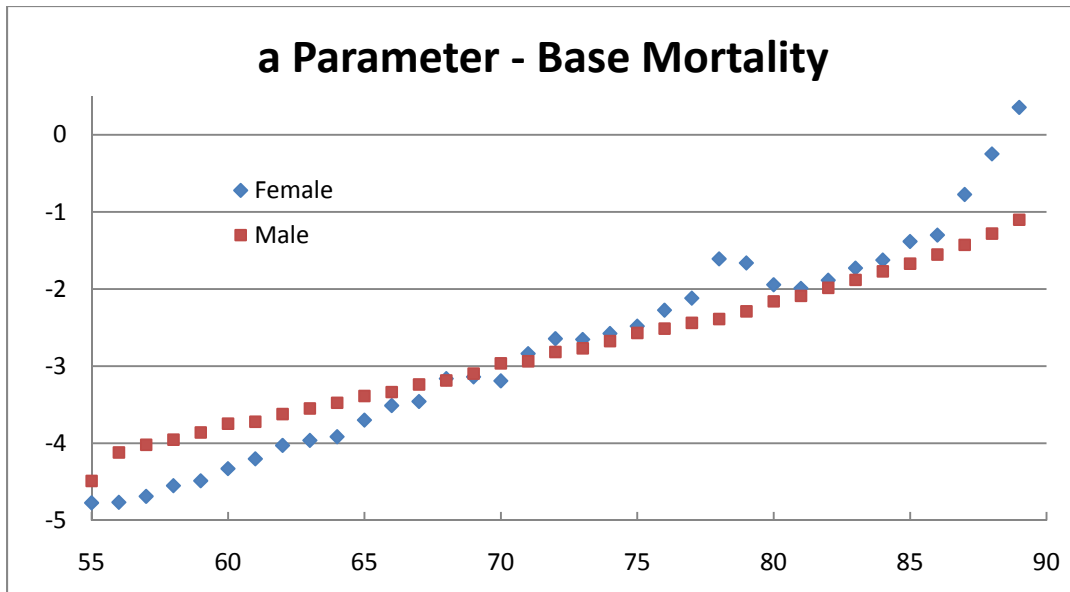
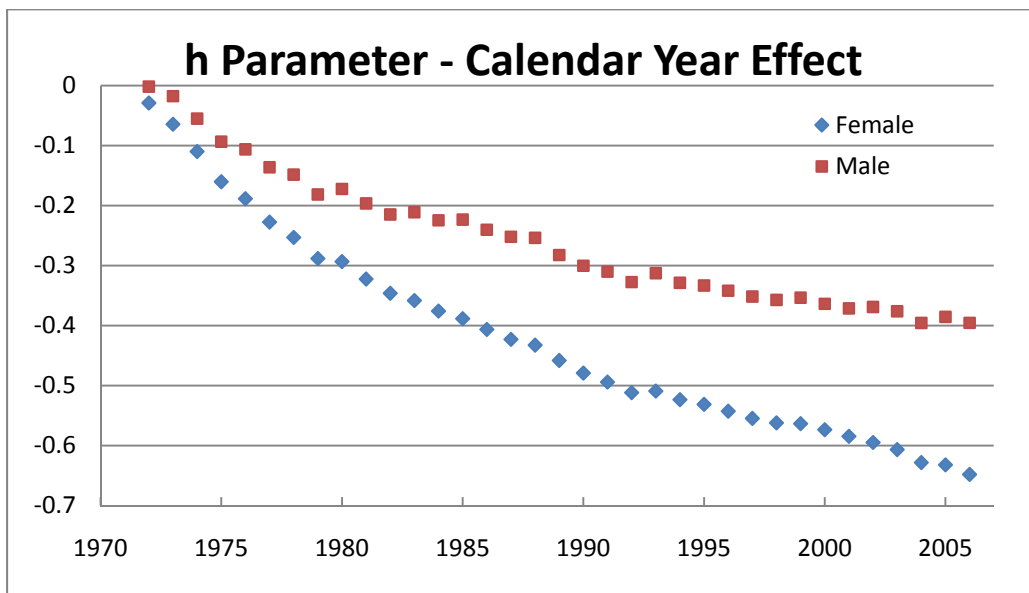


Figure 6 shows the calendar-year trends as reflected in the h_t parameters. These were forced to be zero at year 1971. The female levels are trending faster, even though at the younger ages in the range the male rates are dropping faster. This is a distortion of the calendar-year parameters due to interaction with other parameters.

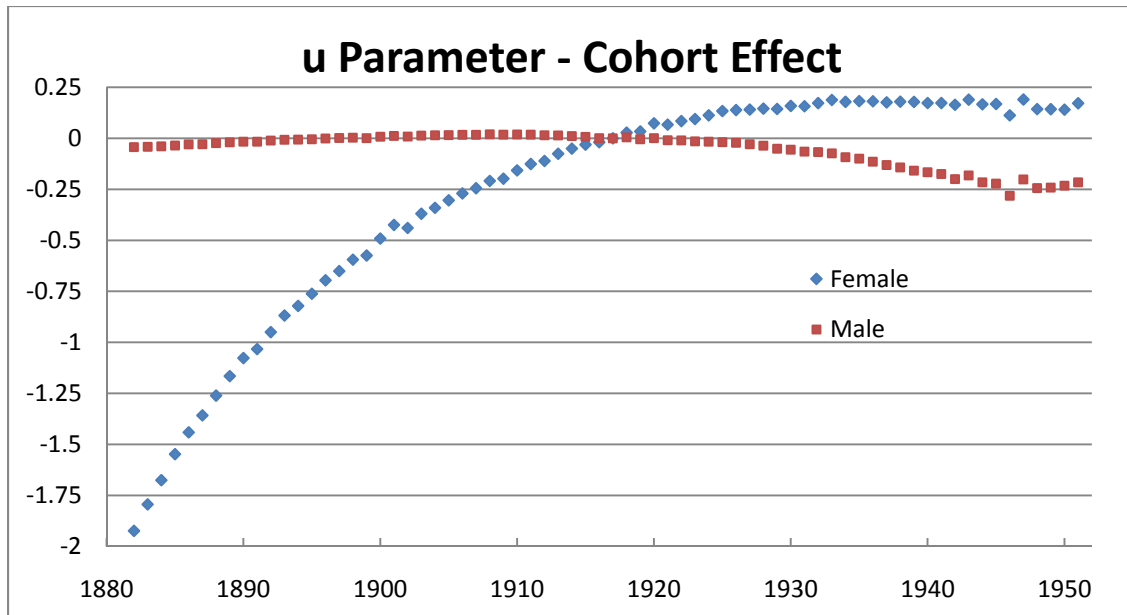
Figure 6. h_t Parameters



The b_d parameters, which modify the overall trend to produce an age-specific trend, are shown in

Figure 7. The surprise here is the accentuation of the trend effect for male octogenarians, whose trend has actually been less than for other ages. In the model the cohort effects offset this effect to match the data. The last cohort that affects ages 87, 88 and 89 is 1899, and cohorts prior to 1900 do not get into this dataset at ages less than 72, leaving room for the parameters to adjust themselves to produce the best possible fit at older ages without affecting younger ages. This raises questions, however, about the applicability of the parameters beyond this data range.

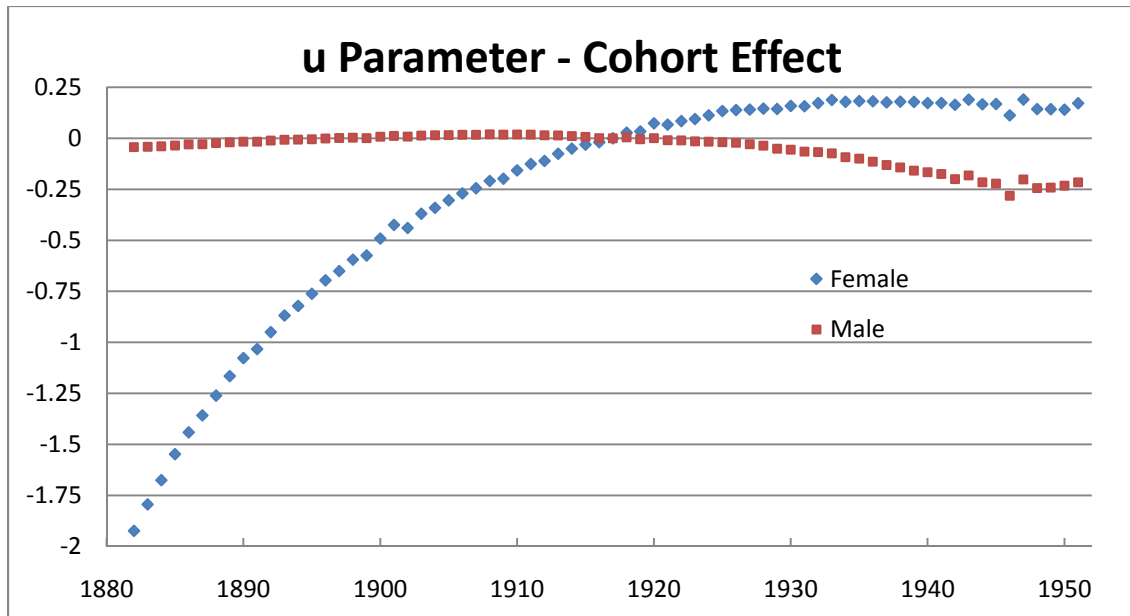
Figure 7. Trend Age Modifiers b_a



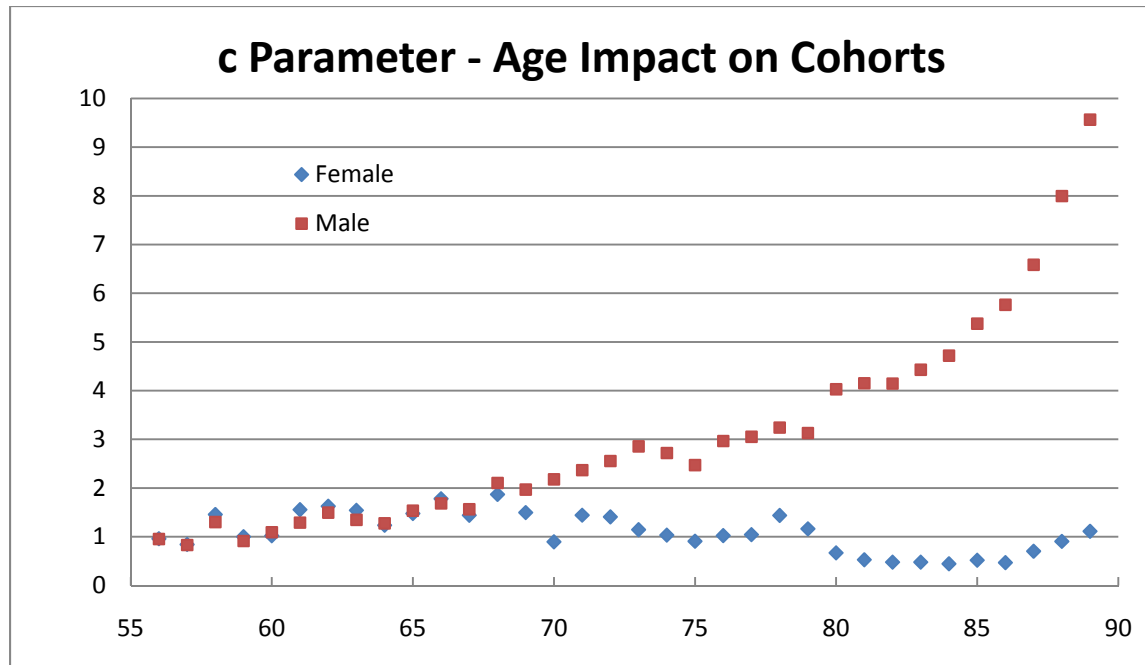
The cohort parameters u_{t-d} are graphed in Figure 8. This includes the 1917 cohort that is forced to zero. The sharply lower mortality for females in the earlier decades is not realistic. It produces an increasing trend across cohorts, which partially offsets the trend across calendar years.

Although it is much milder, an increase in mortality is seen in the first 30 or so cohorts for males. A possible explanation is that these reflect a selection effect. It was more unusual for men and women born in the 1880s to reach higher ages, so those who did were a select group more likely to live even longer. The cohort parameters are intended to represent a mortality differential for the cohort that would apply at all ages, but because not all ages are in the data, the most it could represent is a conditional differential, conditional on having lived long enough to get into the data in the first place. Such an effect would be further enhanced by the fact that the oldest cohorts only appear in the oldest ages in this data.

Figure 8. Cohort Parameters u_{t-d}



Finally the c_d parameters, which are by-age modifiers to the cohort effects, are in Figure 9. These are forced to unity at age 55, and stay near there for a year or two before moving to the vicinity of 2 for ages in the late 60s. From there the female parameters get generally lower for the older ages, indicating a reduced impact of the cohort effects at later ages. This partially offsets the much greater trend in cohorts for females. For the males, the parameter accelerates in the 80s, ending near 10. This appears to be part of the fine-tuning of the fit for higher ages made possible by the cohort parameters with few supporting observations.

Figure 9. Cohort Effect by Age Parameters c_d 

The observed slower trend in mortality at older ages for males is modeled in this RH parameterization by an underlying higher trend at older ages (b parameters), offset by a starting group of cohorts who had lower mortality rates to begin with (u parameters), the effects of which increased sharply at older ages (c parameters). The LC plus cohorts model for female mortality is an even stranger combination of offsetting effects.

4. FIXING THE FITS

One of the problems with the fits above is the sparsity of data in many cohorts. Another is the high correlation among parameters in the female model, which is discussed later. For the recent cohorts, the only way to add data is to wait. For the older cohorts, however, there is data available. Extending the data to include all calendar years of death for cohorts 1882 to 1915 with death ages 55 to 89 is possible, and this increases the number of observations to 35 for each such cohort. This would be expected to give better estimates for those cohort parameters u_{t-d} , but also for the c_d parameters that modify the cohort parameters for age effects, and indirectly on all the other parameters, which may have less flexibility to fit to random fluctuations in the data. The original data will be referred to as the partial data and the expanded set as the full data.

The per-parameter penalty $\log(N)/2$ for BIC goes up to 3.763, with 1855 observations, from 3.569 for the 1260 observations in the partial data. The full data need calendar-year parameters h_t

starting with 1937 (assuming zero at 1936) instead of from 1972, so the models have 33 more parameters. There are no additional a, b, c or u parameters. Thus the LC model now has 137 parameters, and the RH negative binomial models have 242 parameters. The full sequence of models above was fit, but now the NB1 fits better for males. To resolve this, more distributions were fit for the RH model. The NB3 is intermediate between the NB1 and NB2; the Poisson-inverse Gaussian (PiG) is similar to the negative binomial, but is more skewed; and the Sichel is a three-parameter generalization of the PiG, which can be more or less skewed than the PiG but not less skewed than the NB. These distributions are discussed further in Appendix 1. The results are:

Table 2. Triangle Fit Comparisons

Model	NLL		Parameters Added	BIC Needed Improvement	BIC Improvement	
	Female	Male			Female	Male
LC Pois	21,726	24,047				
RH Pois	15,452	16,630	103	388	6274	7417
RH NB1	13,176	13,567	1	3.8	2276	3063
RH NB2	13,172	13,576	0	0	4	-9
RH NB3	13,163	13,568	0	0	9	-1
RH PiG	13,172	13,567	0	0	-0.4	0.015
RH Sichel	13,172	13,565	1	3.8	0	2.2

Again the RH model provides a tremendous improvement in the Poisson fit, as does moving from Poisson to negative binomial. The NB3 is the best fit for females, but the NB1 is the best NB for males. The difference between the NB models is that VM, the variance/mean ratio, is fixed at $1+\beta$ for the NB1, is $1+\mu/r$ for the NB2, and is $1+(\mu/r)^{1/2}$ for the NB3. For females the cell means range from 6000 to 44,000. With the fitted parameters, this gives VM of 5.1 for NB1, 2.4 to 11.1 for NB2, and 3.4 to 7.4 for NB3, which gives the best fit. For males the NB1 VM is 6.1. Another version of the NB discussed in Appendix 2 fits slightly better with a range for VM of 5.3 to 6.7, but uses an additional parameter which does not give enough better fit to justify it.

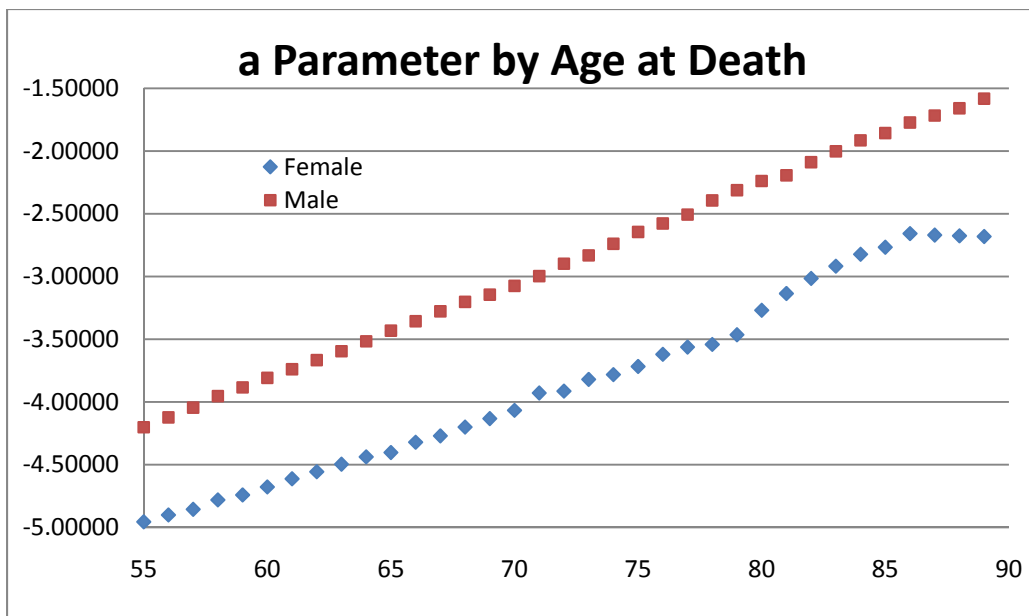
The improvements shown for the last three models are from the better of NB1 and NB2. The PiG and Sichel models also have 1, 2 and 3 versions like the NB. For females, the 2 version of the PiG was found to be slightly worse than the NB2, indicating that the additional skewness was not helpful. The corresponding Sichel has the NB2 as a limiting case, but otherwise has higher skewness than the NB2 with the same mean and variance. The fact that it did not give any improvement over the NB2 suggests that, if anything, less skewed distributions may fit better for females.

For males the PiG, version 1, was very slightly better than the NB1. The Sichel fit even better with an intermediate skewness. However, the improvement in NLL is problematic. At 2.2 it is less

than the 3.8 required by the BIC, but better than 1, which the AIC requires, or 2.0, which the HQIC calls for. There is a good deal of literature suggesting that BIC is too stringent in rejecting parameters. Burnham and Anderson (2004) make a strong push for AIC and the small sample AIC, based on the idea that the sample is not generated from the model being fit, but rather the model is a fairly compact representation of a more complex process. For a sample size of N and p parameters, the small sample AIC penalizes the NLL by $Np/(N-p-1)$. With $N = 1855$, the additional penalty for the 243rd parameter over the 242nd is 1.32. Thus the AIC, HQIC and small sample AIC all support the additional parameter for the Sichel distribution in this case. Thus it will be taken as the best-fitting model.

The parameters shown below are from the best-fitting Sichel model for males and NB3 model for females. It appears that the full data helps with the male model but does not solve the problems with correlation in the female model.

Figure 10. Base Mortality—a Parameter



The a parameters in Figure 10 look reasonable for males but strange for females, especially the decline for the oldest ages. The calendar-year h parameters in Figure 11 also appear reasonable for males but trend upward for females. In this parameterization for females, the downward mortality trend over time ends up as a cohort trend, partially offset with an opposing calendar-year trend.

Figure 11. Calendar-Year h Parameters

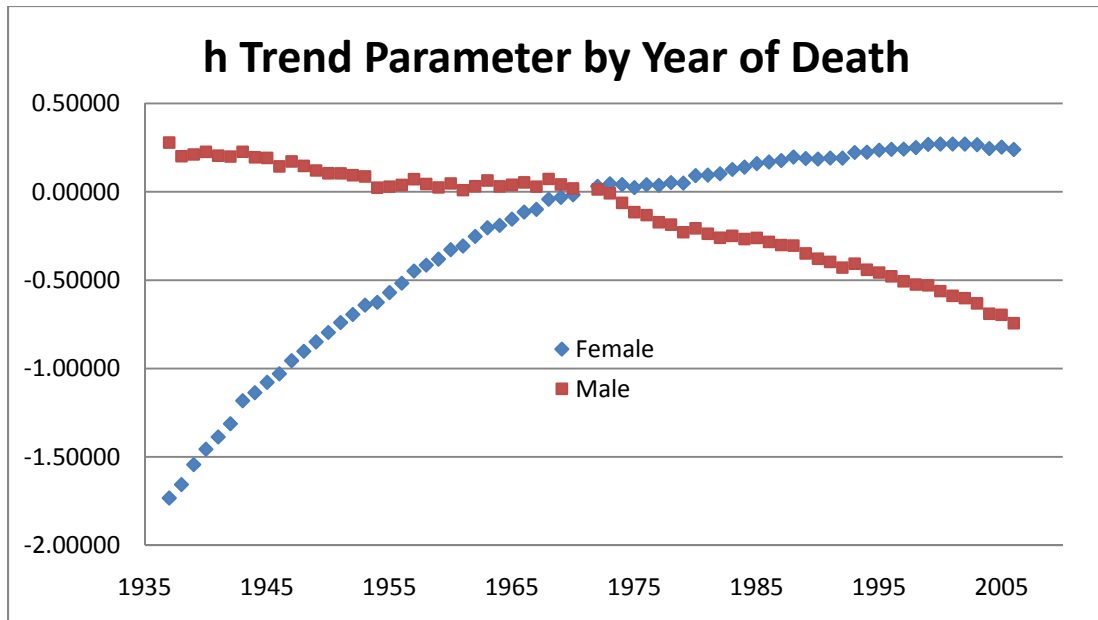
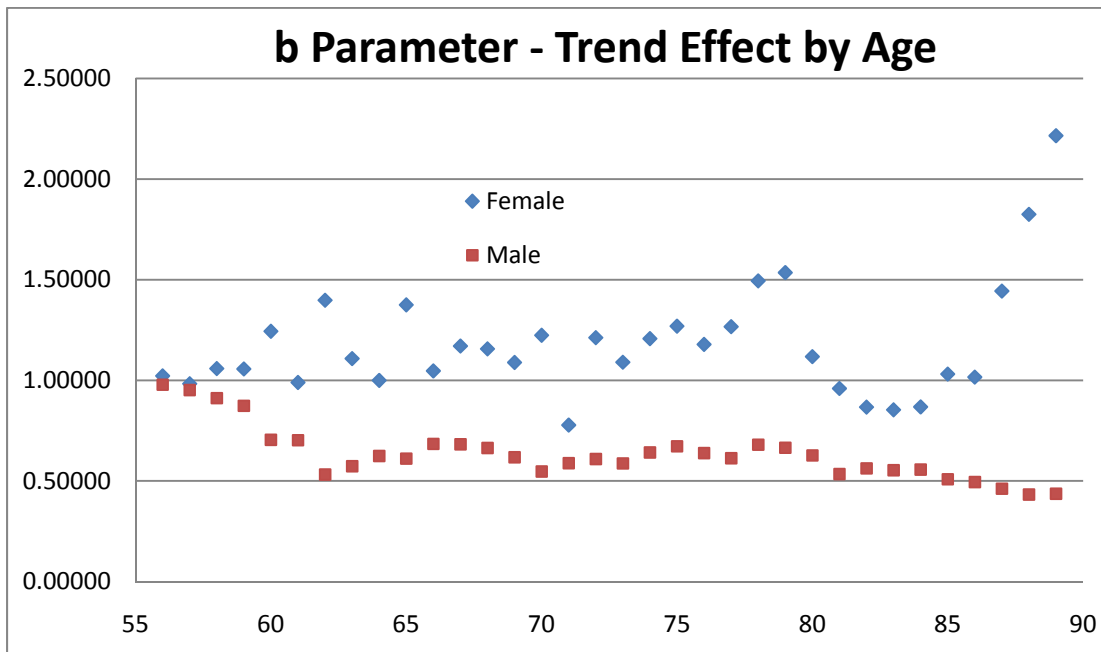


Figure 12. Trend Effect by Age at Death – b Parameters

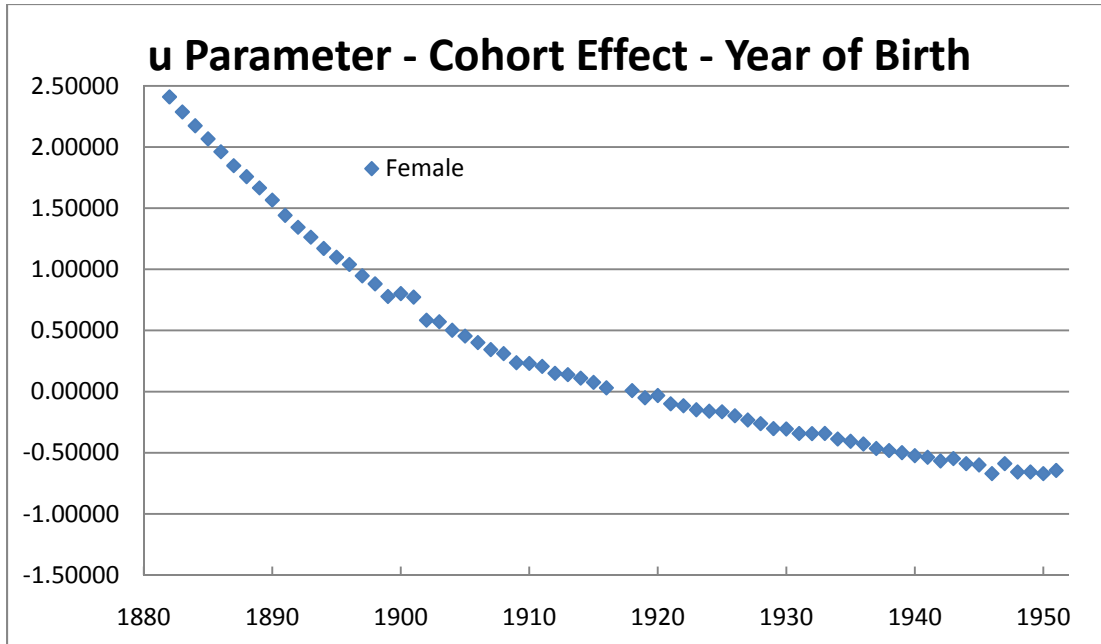


For the b parameters in Figure 12, the sharp upward movement at the oldest ages for females probably has something to do with the lower base mortality at the corresponding points.

The female cohort trend is in Figure 13, which shows a sharp downward trend in mortality in the

direction of later years. This overwhelms the upward trend by calendar year to produce an overall downward trend in mortality, which matches the data, but is not intuitive as an explanation of the data. The h and u parameters with the full cohorts for females are mirror images of what they are for the partial cohorts. This is discussed further below.

Figure 13. Female Cohort Effect u



The male cohort parameters are on a completely different scale and so are graphed separately in Figure 14. The full and partial cohort parameters are consistent for males and so can be graphed together. The effect of conditioning on attaining various ages is clearer in the partial cohorts, where the conditioning is on progressively older ages, peaking in about 1910. There is a similar but much smaller effect in the full cohorts, perhaps due to a changing significance on the fact of attaining age 55. In both cases, there is an increase in the mortality in the most recent cohorts, but this is based on very few data points. Also for the male model, the c parameter, set to 1 at age 55, stays that low only for a few ages then goes to much higher values at older ages, as shown in Figure 15. Thus, the higher cohort parameters for the latest cohorts are getting relatively low c parameters applied, and are not likely to remain so low when more data comes in, with higher c parameters.

Figure 14. Male Cohort Effects for Partial and Full Cohorts

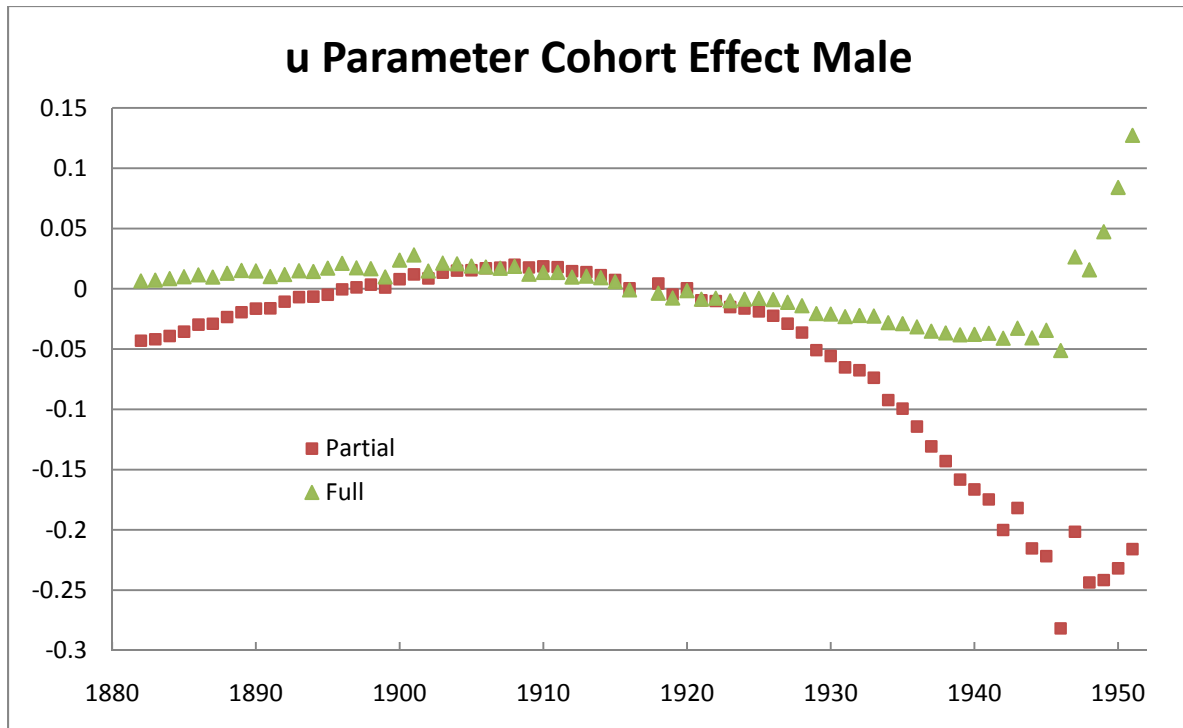
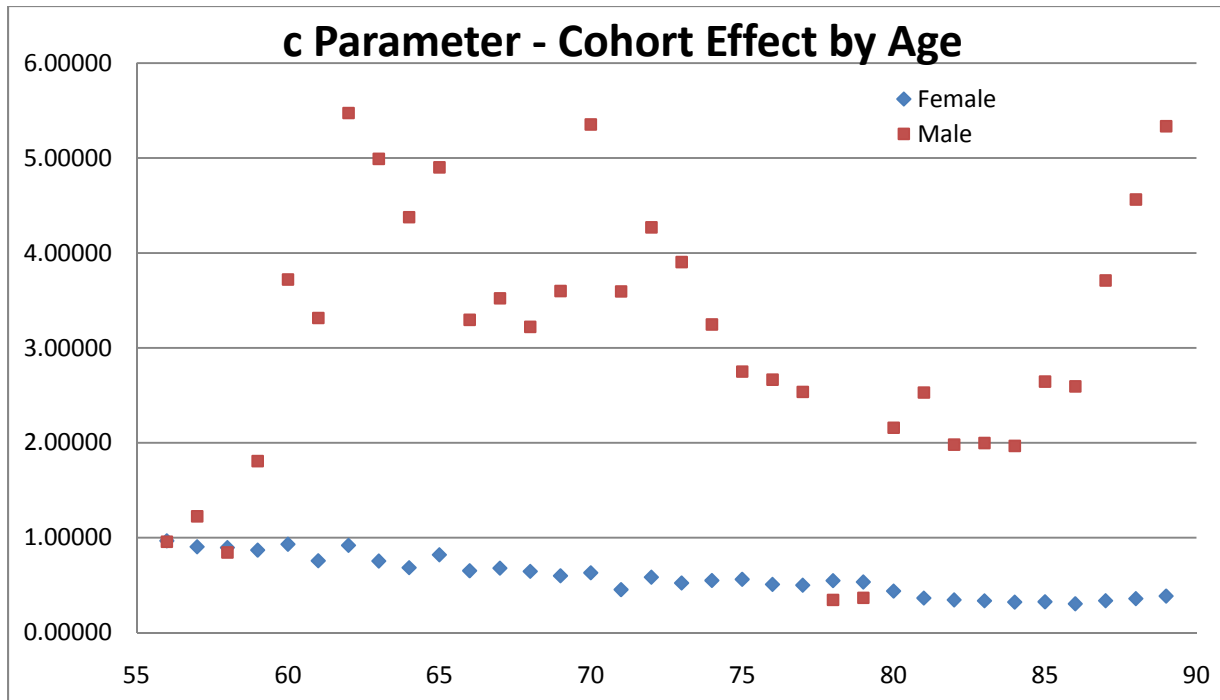


Figure 15. Age Impact on Cohorts—c Parameter



An idea of the statistical significance of the parameters can be gained by estimating the parameter covariance matrix as the inverse of the Fisher information matrix from the MLE estimation. Recall that this is the matrix of all 2nd partial derivatives of the NLL. This yields parameter standard deviations and so t-statistics for each parameter and also covariances, and so correlations, among parameters.

For the male triangle parameters, virtually all the parameters had t-statistics with absolute values above 2. The few exceptions are parameters very close to zero, usually near points that were forced to be zero. With 242 parameters, there are over 25,000 correlations, so they are not printed here. However, the averages of the absolute value of the correlations by type of parameter (excluding parameters with themselves) are shown in Table 3 for males and Table 4 for females.

Table 3. Average Absolute Value of Correlations by Parameter Types—Males

	a	h	b	u	c
a	52.7%	31.9%	36.1%	43.7%	25.3%
h	31.9%	35.6%	22.3%	45.9%	37.3%
b	36.1%	22.3%	42.7%	27.3%	22.0%
u	43.7%	45.9%	27.3%	79.0%	67.3%
C	25.3%	25.3%	22.0%	67.3%	67.4%

Table 4. Average Absolute Value of Correlations by Parameter Types—Females

	a	h	b	u	c
a	98.9%	98.2%	9.4%	97.6%	34.1%
h	98.2%	98.9%	8.4%	98.4%	31.6%
b	9.4%	8.4%	40.9%	8.5%	30.5%
u	97.6%	98.4%	8.5%	97.9%	31.3%
c	34.1%	34.1%	30.5%	31.3%	56.5%

The extremely high correlations among the a, h and u parameters in the female model make the individual parameters highly questionable. There could be many local maxima of the likelihood function, and there is no guarantee that the parameters found are a global maximum. Even if they are, the correlations make the parameter values unstable. In fact, the partial and full datasets gave oppo-

site but similarly offsetting directions for the female calendar-year and cohort trends.

This shows up in the t-statistics as well, which are near 1 in absolute value, so not significant, for all the h and u parameters in the female model.

Moving to the full cohorts then appears to improve the male model, which has reasonable parameters and correlations among parameters, as well as significant t-statistics. For the female model, the high correlations (which, though not shown, are similar for the partial cohorts) make the fit problematic.

Usually when there are high correlations, the solution is to leave out some variables. But the greatly improved fit of the RH model over the LC model appears to rule out omitting the cohort parameters. Parameter reduction through smoothing would still leave quite problematic parameter values as well. One option may be to keep the cohort parameters but not the calendar-year parameters, making the trend a purely cohort matter. It does not seem likely that this would give a good fit, but it might be worth trying.

Another option would be to set the base mortality parameters as the average or some weighted average of the mortality rates for each age in the full data. This was actually Lee and Carter's initial recommendation. This would give the other parameters less opportunity for mischief. A similar approach could be to use a parameterized curve, like Makeham or splines, for the base mortality. Yet another possibility might be to multiply the cohort and calendar-year parameters, and then apply a single age parameter to the product. This type of model is used extensively in casualty loss reserving, but has had mixed results (informally communicated) in mortality studies.

5. MORTALITY CURVES

The raw mortality rates for each year of death are somewhat noisy, and so cannot be readily compared graphically. However fitting mortality curves, like Makeham curves, to each year smoothes the data and lets the trends stand out more clearly. Here a generalized Makeham (GM) function is fit to the raw death rates, although fitting to force of mortality is more typical. Richards (2008) discusses some such generalizations, based on earlier work by Beard (1959) and Perks (1932). Using a curve to fit the a_d parameters requires a log transform, and the form used here takes 4 parameters α , β , θ , γ :

$$a_d = \theta + \log[(1 + \alpha\beta^d)/(1 + \gamma\beta^d)]. \quad (5.1)$$

Fitting such curves with four parameters to the log death rates in each year 1971 – 2006 results in the use of 144 parameters, compared with 104 for LC and 207 for RH with partial cohorts. Using the best-fitting negative binomial, the following values of the NLL were produced:

Table 5. Comparative Fits Including GM

Model	NLL		Parameters Added	HQIC Needed Improvement	NLL Improvement	
	Female	Male			Female	Male
LC	9444	9652				
GM	9482	9553	40	79	-38	99
RH	8748	8972	63GM, 103LC	124GM, 202LC	696 LC	581 GM

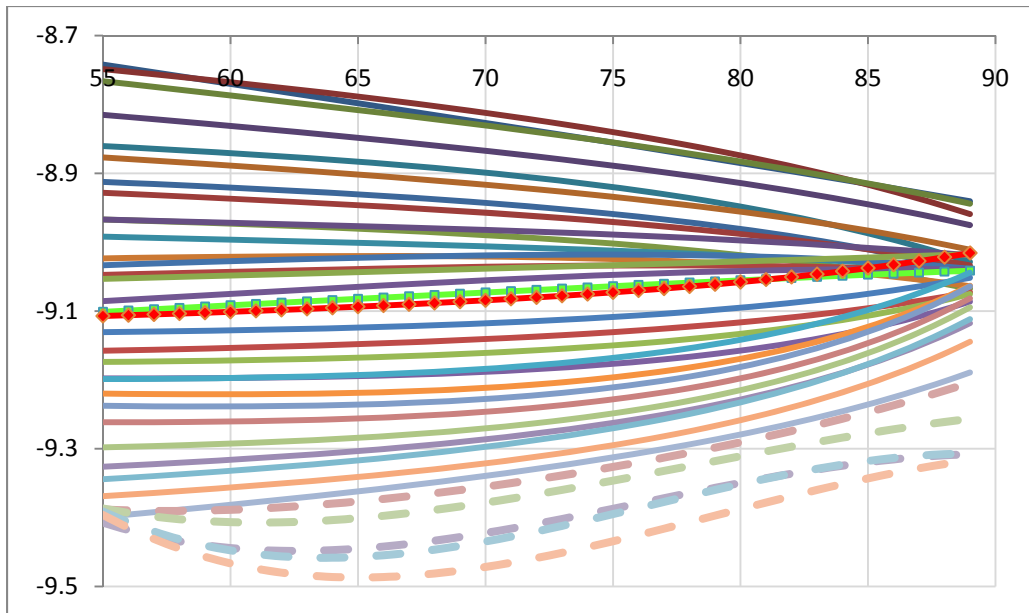
The goodness of fit test here is the HQIC, which requires an improvement in NLL of $\log(\log(\text{sample size}))$ for each extra parameter, where here the sample size is 1260, requiring an improvement of 1.96555 per parameter. This is intermediate between the AIC and BIC. As can be seen, every model fit the female data better than the male data, and the RH model gave the best fit to both data sets, even though it is of dubious interpretation here. The generalized Makeham curve fit better than LC for the males, where the mortality curve was changing more over time, but LC fit better for females.

Nonetheless, for both males and females, the curves provide continuous versions of the mortality functions for each year which are smooth enough to show all years on a chart, thus providing some insight into what the changes in the mortality functions have been.

The male curves in Figure 16 (with age/12 subtracted) actually divide into three periods. First for 1971 until 1987, which is the light line with the square markers, the curves are straight or downward-curving. Then starting in 1988 (dark with diamond markers) the curves bend upward. Until around 2001 or 2002 (first dotted curve) the mortality at age 55 is steadily improving, but the improvement at the other end of the curves is slower and sometimes non-existent. Then somewhere around 2000 to 2002 the improvement at age 55 stops and the improvement at the older ages accelerates. The last three years show a different shaped curve from the earlier years.

The changes in shape show why LC has problems fitting this data, but the fact that the biggest changes were at the ends of the lines shows why RH can give a big, albeit artificial, improvement in the fit. The graph suggests that projecting future changes in longevity has a high degree of uncertainty involved. Should you just project the last five years, or from 1988 on, or average improvements in mortality over all the data? This could make quite a difference, especially at some ages.

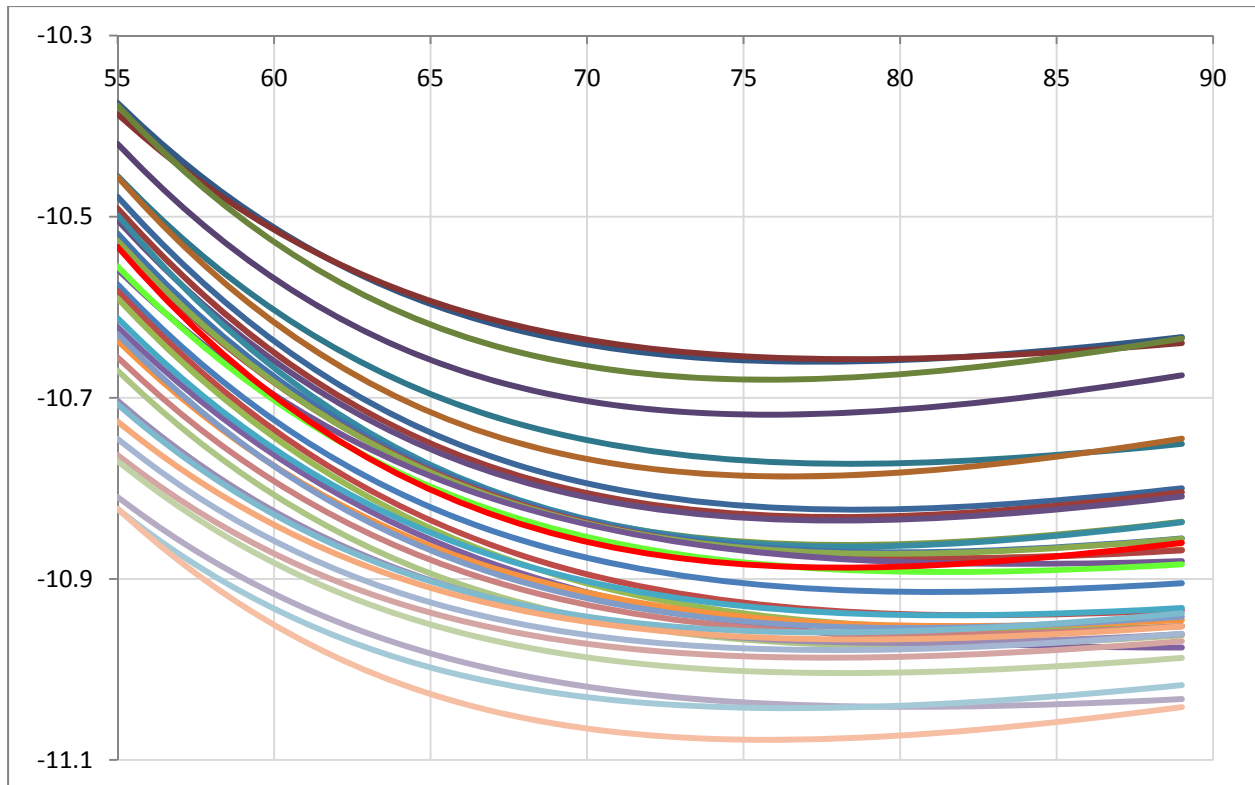
Figure 16. Generalized Makeham Male



The recent lack of improvement at age 55 is particularly problematic. That could be related to the recent reduced access to health care in the US for people under 65. If so, you would expect it to eventually improve over time as access improves. At the other end of the curve, it might be reasonable to assume that the older ages will improve at the same rate as most of the curve, as there seems to be a trend in that direction over quite some time. Nonetheless this is an assumption imposed on the projection process and thus adds to the projection uncertainty.

The generalized Makeham model did not fit as well as LC for females, but the fits in Figure 17 still provide some insights. Here age/10 was subtracted to remove the upward trend. It is apparent that there has not been as much change in the shapes of the curves as in the male model. What does stand out, however, is variation in the rate of mortality improvements across the age groups. For instance, for ages 75 and above, there were fairly long periods with very little improvement in mortality, punctuated here and there with years of substantial improvement. Ages 65 and below, on the other hand, had much more steady generally small improvements. As with the male data, there has been little improvement at age 55 in the latest few periods. Also since about 2000 there has been somewhat similar year-to-year improvements in the male and female graphs, even by age.

Figure 17. Generalized Makeham Female



As with the male model, this graph brings out some problems in projecting future trends. Can you assume the greater improvement in the last 5 or 6 years will now continue? Would a time-series model with highly fluctuating rates of improvement be better at the older ages? Perhaps in both genders it would be appropriate to calculate trends under different assumptions then include all the scenarios, with selected weights, in the overall longevity improvement uncertainty model.

6. PROJECTION RISK

Projection risk can be calculated for a particular dataset of annuitants, which is not what is available here, but some general observations on how to carry out such a calculation using LC and RH models are presented.

To begin, the calendar-year trend levels have to be projected. Standard time-series methodologies produce ever-widening ranges as the trend continues. However, here there is another wrinkle, as the h parameters being trended are estimated parameters, and so are observed with error. An area of regression studies is errors-in-variables models, which has a number of potential methods. If the variances of the h parameters have been estimated and they are relatively constant, then a simple reasonable simulation of a future level could assume that same variance, and first simulate the future

levels with errors and then simulate actual future levels from there using that variance.

A number of mortality modelers have used AR1 models for the annual difference in levels, usually with a negative autocorrelation, to project the trend. Most of these do not take into account the errors-in-measurement issue, however. For an independent series measured with a constant error variance, differencing induces an autocorrelation of -50 percent, arising from the same error having opposite signs in consecutive observations, so the AR1 model may be distorted by the induced autocorrelation. Other mortality projection studies have used the Kalman filter, which recognizes measurement errors, to project the levels, but the simple Kalman filter is based on a random walk, which can have too much autocorrelation. An alternative to AR1 and the Kalman filter is state-space models, which provide common generalizations of both.

If projections are needed for cohorts not in the study, then trending of cohorts also has to be considered. Even the use of the recent cohort parameters should take into account their potential measurement errors, perhaps with a state-space model.

Parameter uncertainty can be implemented by simulating the parameters from the covariance matrix from the Fisher information matrix, which gives an estimate of the covariance matrix of the parameters. Asymptotically the parameters have a multivariate normal distribution with this covariance matrix, so they can be simulated using the normal copula, Cholesky decomposition, etc. However, even though the error distributions are asymptotically normal, they may not be normal for a finite sample, and other distributions could be used to simulate parameter risk, perhaps gamma, which is the exact error distribution for some models, and approaches the normal asymptotically. Other distributions that approach the normal could also be used. One criterion is that the normal should not be used if there is too much probability that a parameter that has to be positive could be simulated as negative from the normal.

Once a routine is in place to simulate parameters and to trend the h and u parameters, the number of deaths can be simulated from the negative binomial or Sichel distribution. If a routine to do this is not available, probably simulating from a transformed gamma with the same first three moments would not be too far off.

Model risk is a more difficult issue. The RH-Sichel model appears fairly reasonable for the male data, but the cohort parameters for the last several cohorts are questionable, being based on few observations. Parameter uncertainty would be large for such parameters. Perhaps using the models but including extra parameter uncertainty for model risk would give usable results.

7. SUMMARY AND FUTURE DIRECTIONS

The Lee-Carter model allows only highly constrained shifts in the shape of the mortality curve over time, and adding cohort effects gives much better fits. However these are found to generate new problems, such as potential over-fitting, instability for projections, and highly correlated and insignificant parameters. Also, the negative binomial fits better than the Poisson, which has been seen before and is likely to be a standard result. The best form for the NB is not consistent, however, and may differ for different datasets, depending on how contagion actually applies. For males, the Sichel distribution is better still.

Model risk is an issue, since the RH model can fit well at the ends of the age range using cohort parameters based on few observations. Using full cohorts can reduce this possibility at the older ages but not at the youngest ages. Also the RH parameters can be highly correlated, as in the female model, suggesting that some other model should be found, possibly by reducing the number of parameters.

Projections of mortality risk under current methodologies are thus likely to be unreliable. But better-fitting models are not likely to solve this problem as the RH model fits extremely well. Perhaps other models can be found with fits intermediate between LC and RH but with more parameter stability than RH.

ADDENDUM

Now 2007 data is available, and some of the recent trends are continuing. Mortality for ages 65+ continued to improve compared to 2006, but for ages in the mid-50s, the lack of improvement continued. Whether this is just a random fluctuation or some underlying trend, such as obesity or reduced access to medical treatment, is yet to be established.

REFERENCES

- [1.] Beard, R. E. 1959. "Note on some mathematical mortality models." In: *The Lifespan of Animals*, G. E. W. Wolstenholme and M. O'Connor (eds.), Little, Brown, Boston, 302-311.
- [2.] Burnham, K.P., and Anderson, D.R. 2004. "Multimodel Inference: Understanding AIC and BIC in Model Selection." *Sociological Methods Research* 33: 261–304.
- [3.] Hannan, E., and Quinn, B. 1979. "The Determination of the Order of an Autoregression." *Journal of the Royal Statistical Society B* 41: 190–195.
- [4.] Lee, R.D., and Carter, L.R. 1992. "Modeling and Forecasting U.S. Mortality." *Journal of the American Statistical Association* 87: 659–675.
- [5.] Perks, W. 1932. "On some experiments in the graduation of mortality statistics." *Journal of the Institute of Actuaries* 63: 12-40.
- [6.] Renshaw, A.E., and Haberman, S. 2006. "A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors." *Insurance: Mathematics and Economics* 38: 556–570.
- [7.] Richards, S.J. 2008. "Applying Survival Models To Pensioner Mortality Data." Presented to the Institute of Actuaries, 25 February 2008, <http://www.actuaries.org.uk/sites/all/files/documents/pdf/sm20080225.pdf>.
- [8.] Rigby, R.A., D.M. Stasinopoulos, and Akantziliotou, C. 2008. "A Framework for Modelling Overdispersed Count Data, Including the Poisson-Shifted Generalized Inverse Gaussian Distribution." *Computational Statistics and Data Analysis* 53: 381–393.

APPENDIX 1. COUNT DISTRIBUTIONS

The negative binomial distribution has two parameters, r and β , with mean $r\beta$ and variance $r\beta(1+\beta)$. In the full data there are 1855 cells, and when the negative binomial is used, each cell has a value of r and β . The mean $\mu = r\beta$ is the value given by the RH model, but how r and β vary across cells depends on how the model is set up. In the NB1, it is assumed that every cell has the same value of β , so the ratio of variance to mean is $1+\beta$ for every cell. In the NB2, every cell is assumed to have the same value of r , with β set to μ/r , which gives variance to mean ratio $1+\mu/r$, which is higher for the cells with higher means. However there are many other ways the parameters can vary across cells. For instance, suppose there is a constant q for all cells, with r and β given by $r = q\mu^{1/2}$ and $\beta = \mu^{1/2}/q$. Then the mean is still $r\beta = \mu$, and the variance to mean ratio for a cell is $1 + \mu^{1/2}/q$. This is what is called the NB3 in the text. Its variance/mean ratio is still higher for the larger cells, but not by as much as in the NB2.

This can be generalized to the NB p distribution, which adds a parameter p to control the variance/mean ratio. It sets $r = q\mu^{1-p}$ and $\beta = \mu^p/q$. The mean is again $r\beta = \mu$, but now the variance to mean ratio for a cell is $1 + \mu^p/q$. The value of p can be found by MLE. For males, the resulting value of p is 0.2, but the NLL is not enough better to justify the additional parameter by any of the information criteria. For females, the p is 0.53, but again this did not improve the NLL enough to justify the extra parameter. It might be argued that the NB3 already has an extra parameter of $p = 1/2$, but this is a bit ambiguous as the parameter is not free to be fit. In this case the NB3 fits the female data by enough better to justify an additional parameter.

When fitting a single NB distribution to a dataset, all of these forms are the same. The difference comes when fitting a number of distributions to a number of cells where a common relationship of variance and mean is desired. The NB p forms discussed here by no means exhaust the possible such relationships. In general, if the variance/mean ratio desired is $1+G(\mu)$, just set $r = \mu/G(\mu)$ and $\beta = G(\mu)$. For instance, $G(\mu) = q \log(\mu)$ might work in some cases, possibly even for the male data in this paper.

The Poisson—inverse Gaussian (PiG) distribution can be derived analogously to the NB as a Poisson mixture, but now the Poisson parameter is mixed by the inverse Gaussian instead of the gamma. Again it has 1, 2, 3 and p versions, etc. The inverse Gaussian is 50 percent more skewed than the gamma with the same mean and variance, and the PiG inherits this greater skewness, although not by the same ratio. The third central moment divided by the mean is the 3rd moment analogue of variance/mean for count distributions. For the negative binomial this is $1+3\beta+2\beta^2$, while for the PiG it is $1+3\beta+3\beta^2$. For $\beta=5$, which is fairly typical in the fits here, that gives 66 for the NB and 91 for the PiG, both of which would have variance/mean = 6.

The Sichel distribution is a generalization of the PiG and is a Poisson mixed by a generalized inverse Gaussian. It can be more or less skewed than the PiG but not less than the NB, which is a limiting case. It uses the modified Bessel function of the second kind (sometimes called the third kind), $K_\nu(t) = \frac{1}{2} \int_0^\infty x^{\nu-1} \exp\left[-\frac{1}{2}t(x + x^{-1})\right] dx$.

The Sichel distribution with parameters r , β and ν can be most readily expressed with two auxiliary parameters c and s , with $c = K_\nu(r)/K_{\nu+1}(r)$ and $s^2 = 1+2\beta c$. The probability function at j is:

$$p_j = \frac{(r\beta c)^j K_{j+\nu}(rs)}{s^{\nu+j} j! K_\nu(r)}$$

This is a reformulation of the version given in Rigby et al. (2008). Their parameters can be mapped from these by taking $\alpha = rs$, $\sigma = 1/r$, $\mu = r\beta$, and $c = 1/c$.

The PiG is just the case $\nu = -1/2$, for which $c = 1$. The Sichel mean is still $r\beta$ and the variance is $\mu(1+h)$ with $h = 2\beta c(\nu+1) + \mu(c^2 - 1)$. For the PiG, this simplifies to $h = \beta$. The Sichel ratio of the third central moment to the mean is $\mu_3/\mu = 1 + 2\mu(\beta c - h) + 3h + 2\beta c h(\nu+2)$.

The ν parameter can be very different than $-1/2$, and for the male data here was estimated as 2155. The β parameter was close to 6, and r was set at μ/β . The resulting third moments were usually intermediate to those of the PiG and NB.

APPENDIX 2. FITTING NOTES

With several distributions to be fit, routines were sought that did not use derivatives of the NLL or could use numerical derivatives. The R package `subplex` uses an efficient form of the simplex algorithm, and was found useful in getting rapid improvement in the NLL from initial guesses. However it seemed to have difficulty in final convergence, often ending up in a region where the NLL was changing very slowly but was not near a minimum. Running `subplex` two or three times with default settings usually helped a good deal.

From there the `optim` routine in the Stats package was found to be useful in proceeding more toward a minimum. The `optim` option used most often was BFGS with `gr=NULL`, which takes fast approximate numerical derivatives of the NLL to find the best direction for improvement. Usually it would start off with only small improvements, but usually ended up finding a region where more rapid improvement was possible, then slowing down again near to convergence. Relative and absolute convergence criteria of $1e-17$ and $1e-12$ were used, which may be beyond machine precision. However the routine would converge, although usually not to a true minimum.

The next step was to define a gradient function of the parameters using numerical derivatives from the `numDeriv` package. This is a slower but more accurate gradient, and using BFGS with it always improved the fit. The problem is that the convergence is defined by the NLL not changing

much, which does not always end up with all derivatives very close to zero. Since the 2nd derivatives at the minimum are needed for the information matrix, it seemed a good idea to make the derivatives reasonably close to zero. For this the routine `dfsane` from the `BB` package was found helpful. In perhaps 50 iterations it could find points close to the optim parameters but with a reduction of 2 or 3 orders of magnitude in the largest (absolute) derivatives. It usually produced only very small changes in the NLL from what `optim` had yielded, however.

For the Bessel functions, the base R package function does not work with high values of the index (say $\nu > 1500$). There is a Bessel package available for Windows in R-Forge. It has a function `besselK.nuAsym` that does work for large values of the index, but not for small values. It needs an additional package `Rmpfr`, which is available on CRAN.

There are recursive formulas for the PiG and Sichel probabilities, but these are awkward at best for probabilities for tens of thousands of events.

The parameter constraints that force some parameters to be zero or one are different from much of the literature, which uses constraints on the sums of parameters. However doing it this way helps guarantee that the information matrix is not singular, which is necessary for its inversion.

(Yilu Zhang and Lina Ma helped research the R methodology for fitting distributions used here.)