# Claim-Level Triangle Modeling with Structural Drivers

## General liability example

Ming Yi – Senior Analyst – AIG

Email: ming.yi@aig.com

# General Liability Body Injury Loss Model

Model for individual claim data at each point in triangle.

- Information about claims accident date, close date, claim details etc.

Model captures many features better than aggregate triangle

- Better control over the change of mix, which has a big influence on the triangle, which can distort the development factors.
- Includes macro drivers like inflation as variables

Model For Claim severity and for probability of claim closing

- Mainly focused on severity model today.

# General Liability Body Injury Loss Model

Model on body injury for premises

- PD has different severity than BI
- Similarly for products

Model captures many features of historical process

Main model uncertainty feature for risk analysis identified as selection of time period for fitting

- Used 1985 to 2011, but longer or shorter periods have different long-term means
- We could treat this uncertainty in simulation.
- We could first simulate parameters based on data periods and then losses.

# General Liability Body Injury Loss Model

Model based on Generalized Linear Model (GLM) with dependent variable the whole triangle put into a column with further breakout by claim: total paid for closed only

The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a **link function** and by allowing the magnitude of the variance of each measurement to be a function of its predicted value.

AIG

# Final Sum of Payments for Each Claim

Predictive variables can be applied row wise or by diagonal in design matrix – e.g. CPI 3Q 2008 could be for payments in that quarter or acc. date 3Q 2008

- We apply medical inflation by diagonal = calendar quarter of payment

Variables could be categorical or numerical

- Many turn out better categorical, like sales for insured company or report lag for claim
- Group these in size ranges instead of using actual amounts

Constants for each row and each diagonal were fit to residuals after regression for other variables

- Avoids problem of collinear variables
- We call those the unexplained trend
- Not needed for rows, rarely needed for diagonals

# Model Variables

Sales

- How big the company is.
- Inflation adjusted to the most recent accident year.

States

- Grouped by average payment of each state

Report Lag

- Difference between report date and accident date.

Operational Time Or Time From Accident to Close

- Operational time for a claim is the percentage of claims closed before this one.
- E.g., see McGuire 2007: http://actuaries.asn.au/Library/6.a_ACS07_paper_McGuire_Individual%20claim%20modellingof%20CTP%20data.pdf

Industry Major Group; Inflation; SIR Indicator

AIG

# Data Segmentation Process-Open Claims

**Tried to use DFM method to adjust open claims.**

**Problem: Unlike WC, the periodic payment, GL focused on final value. For open claims, final value is not available. However, if use DFM to adjust, brought uncertainty into the model.**

**Decision: Take still open claims out, so only modeled closed claims.**

**Result: Improves Statistical Criteria.**

# Model Selection

Based on penalized likelihood function

For negative loglikelihood, lower is better

Add penalties for number of parameters

- AIC: add 1 for each parameter
- BIC: add ln sqrt of sample size for each parameter
- HQIC: add ln of ln sample size for each parameter
- AIC thought to be to lenient on extra parameters
- BIC maybe too strict
- HQIC is in between

Only use statistically significant parameters and usually AIC, HQIC and BIC tend to give the same conclusions

# Model Selection – Tried a Few

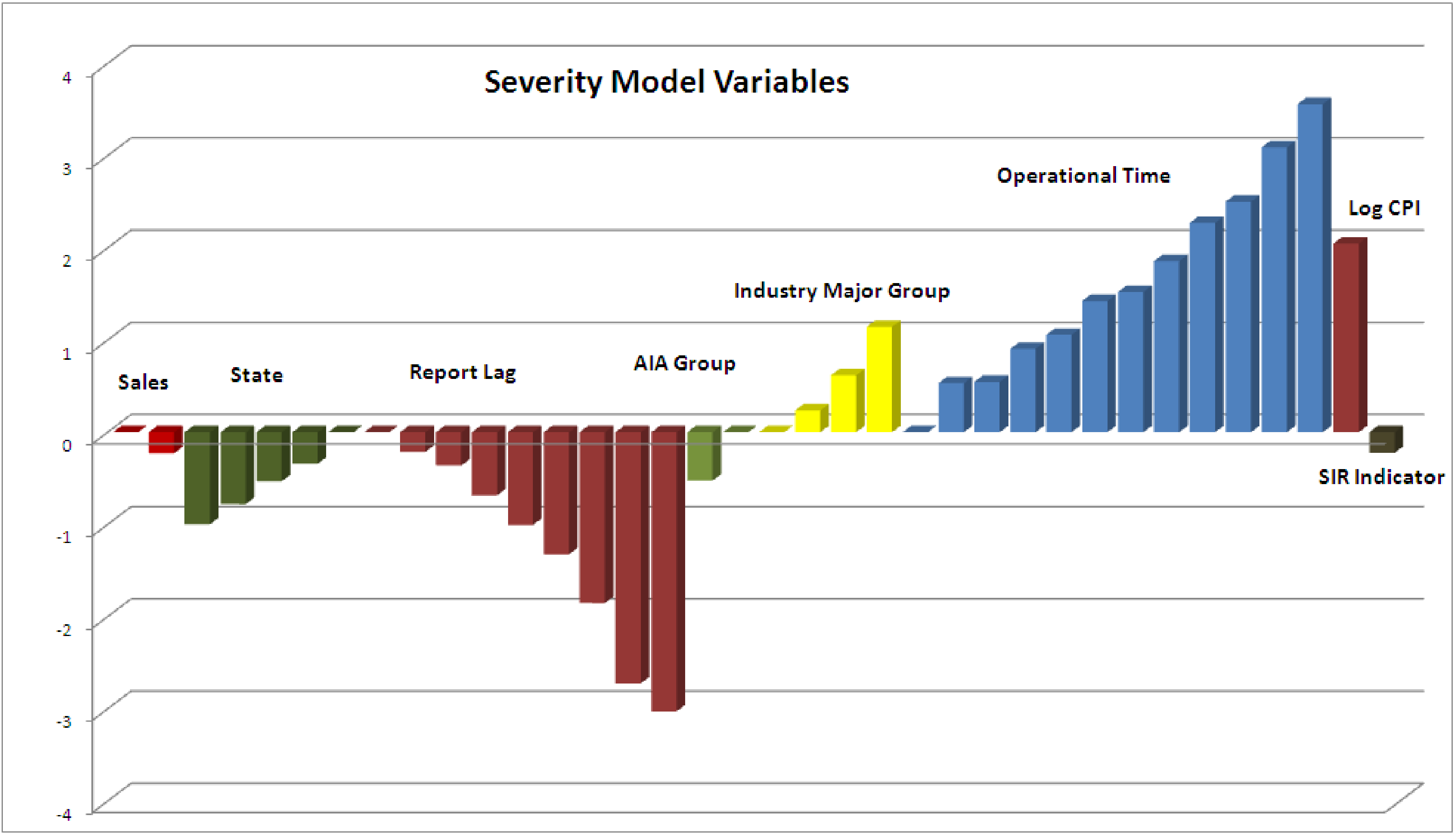
| Link Function | Distribution | AIC(Smaller Is Better) |
|---|---|---|
| Identity | Normal | 72,351 |
| Identity | Gamma | 73,452 |
| Identity | Inverse Gaussian | 74,356 |
| Identity | Normal with Log-Data | 72,182 |
| Identity | Gamma with Log-Data | 74,213 |
| Log | Gamma | 72,698 |
| Log | Inverse Gaussian | 75,245 |

AIG

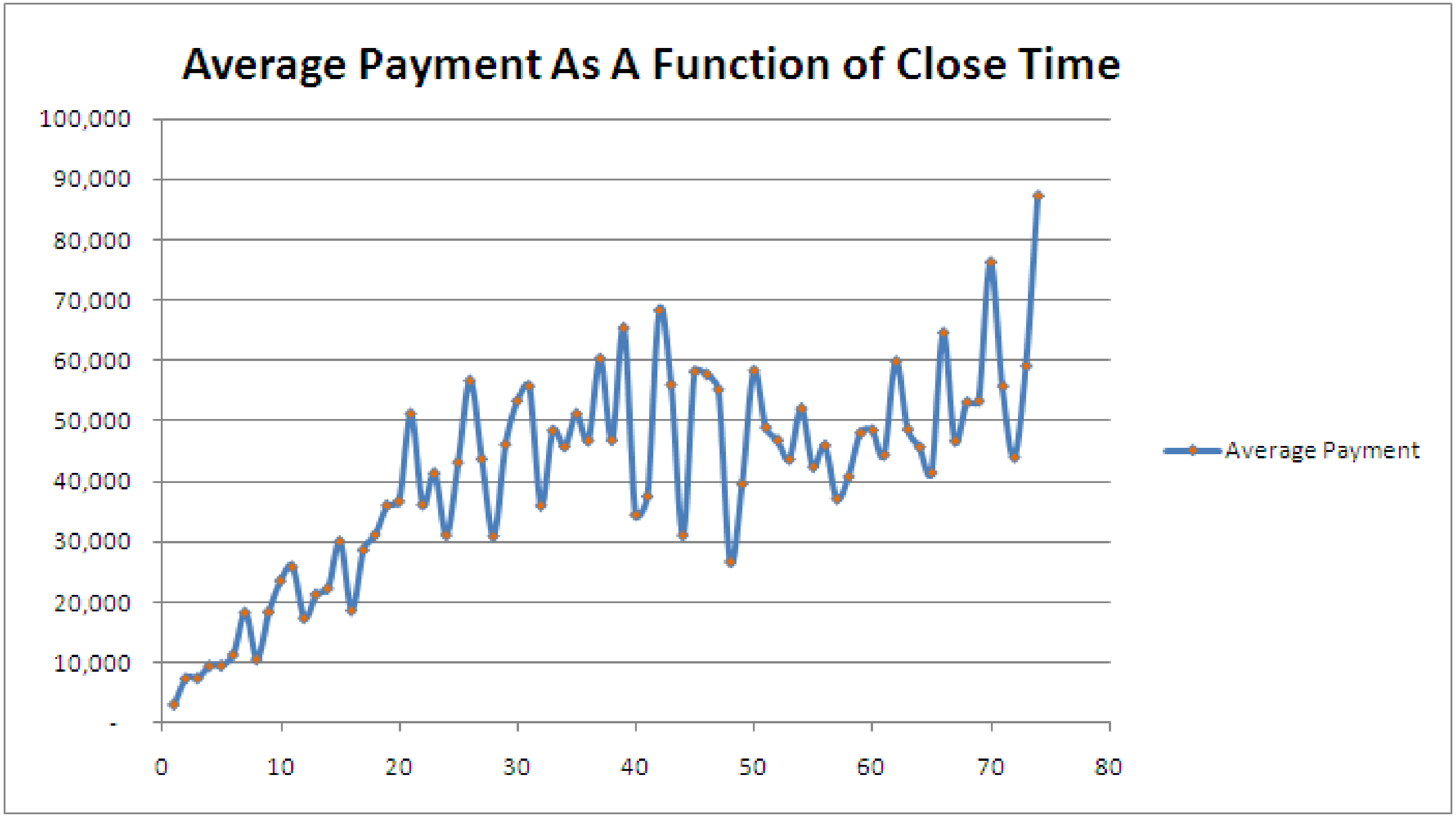
# Analyzing Results

| Source | Num DF | Den DF | F Value | Pr > F | Chi-Square |
|---|---|---|---|---|---|
| Opratn Time | 13 | 16506 | 237.95 | <.0001 | 3093.29 |
| Reportlag | 8 | 16506 | 76 | <.0001 | 607.98 |
| log_CPI | 1 | 16506 | 268.12 | <.0001 | 268.12 |
| IndusMajGrp | 4 | 16506 | 54.63 | <.0001 | 218.51 |
| aia_grp | 1 | 16506 | 171.68 | <.0001 | 171.68 |
| State | 4 | 16506 | 32.58 | <.0001 | 130.32 |
| SIRInd | 1 | 16506 | 23.64 | <.0001 | 23.64 |
| Sales | 1 | 16506 | 20.8 | <.0001 | 20.8 |

# Severity Model Variables



# Model Improvement

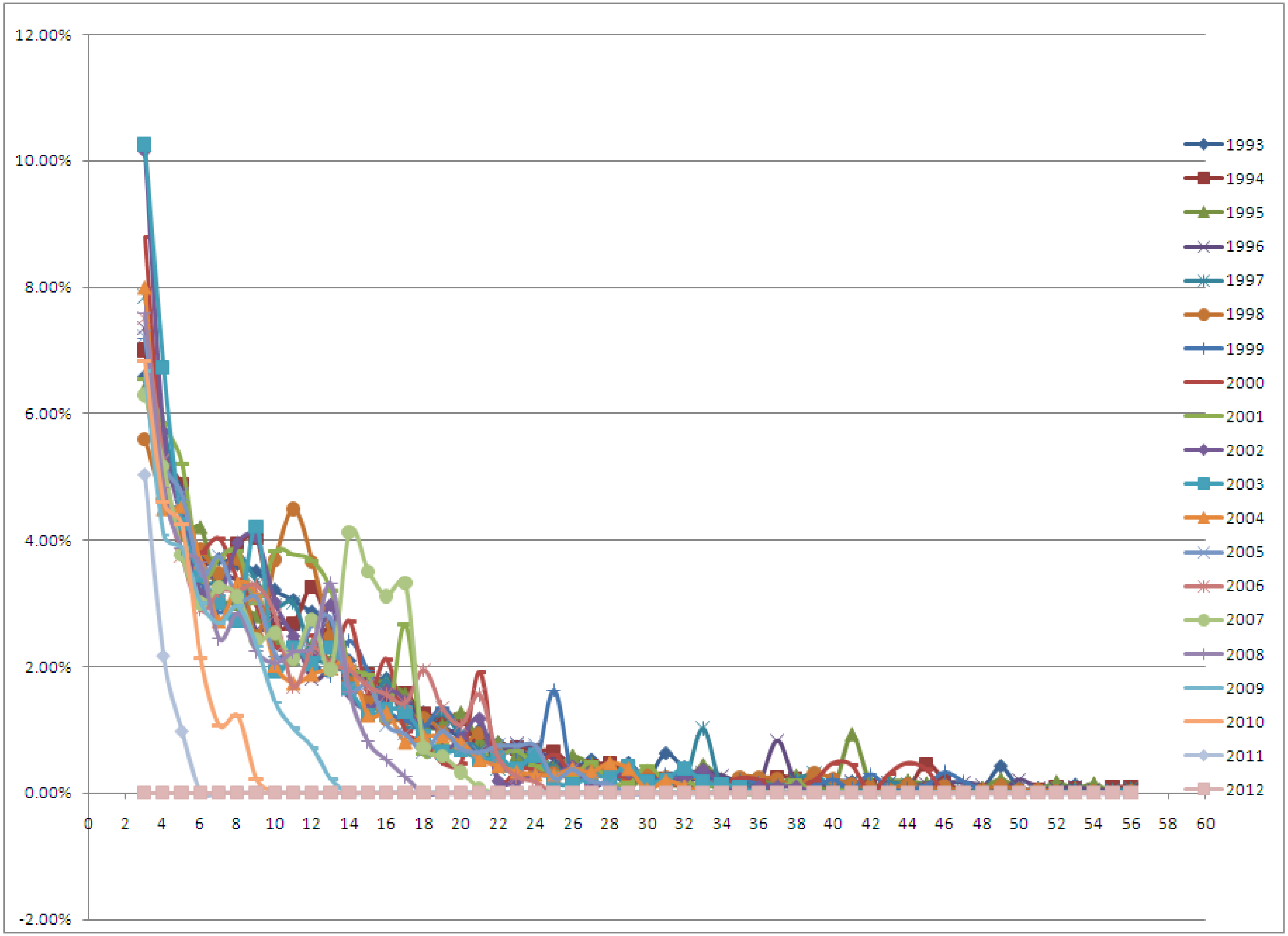


# Model Improvement

- Close Time
- Operational Time

In the triangle, data point from the same column will stay the same for close time, however different for operational time.

# Operational time as a function of Time to Close



# Claim Counts Distribution by Development Factor

- Data point in the first three quarters are distorting, starts from the third quarter.
- Even for the same development quarter, the operational time is different.

| | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|
| 1993 | 10.35% | 8.05% | 7.44% | 5.78% | 5.81% | 5.69% | 5.51% |
| 1994 | 11.17% | 8.45% | 7.76% | 5.70% | 5.44% | 6.27% | 6.42% |
| 1995 | 10.93% | 8.27% | 6.98% | 7.19% | 5.32% | 5.11% | 4.78% |
| 1996 | 12.55% | 8.92% | 8.22% | 5.96% | 5.84% | 5.69% | 4.40% |
| 1997 | 13.21% | 9.77% | 7.70% | 5.72% | 5.64% | 6.34% | 5.64% |
| 1998 | 9.64% | 7.97% | 7.16% | 6.63% | 5.97% | 6.40% | 4.25% |
| 1999 | 12.55% | 9.45% | 7.93% | 5.88% | 4.99% | 5.57% | 5.41% |
| 2000 | 14.85% | 8.61% | 6.40% | 6.40% | 6.80% | 5.67% | 4.99% |
| 2001 | 10.58% | 9.42% | 8.41% | 5.48% | 6.02% | 6.18% | 4.86% |
| 2002 | 17.32% | 9.73% | 7.36% | 5.36% | 5.29% | 6.74% | 6.90% |
| 2003 | 18.09% | 11.86% | 7.51% | 6.08% | 5.33% | 4.80% | 7.43% |
| 2004 | 15.80% | 8.87% | 8.96% | 7.21% | 5.36% | 6.56% | 6.19% |
| 2005 | 13.17% | 9.58% | 8.64% | 6.33% | 6.84% | 5.56% | 5.64% |
| 2006 | 13.61% | 9.47% | 6.81% | 5.24% | 5.79% | 5.79% | 5.95% |
| 2007 | 10.61% | 8.72% | 6.34% | 5.06% | 5.49% | 5.24% | 4.08% |
| 2008 | 14.78% | 9.40% | 7.53% | 7.08% | 4.76% | 5.55% | 4.36% |
| 2009 | 14.52% | 8.88% | 8.43% | 6.55% | 5.83% | 6.42% | 5.06% |
| 2010 | 14.89% | 10.04% | 9.24% | 4.65% | 2.33% | 2.66% | 0.47% |
| 2011 | 8.97% | 3.85% | 1.73% | | | | |


15

# Statistics Comparison

| | Operational Time | Close Time |
|---|---|---|
| **Log Likelihood** | (36,057) | (38,662) |
| **Full Log Likelihood** | (36,057) | (38,662) |
| **AIC (smaller is better)** | 72,182 | 77,400 |
| **AICC (smaller is better)** | 72,182 | 77,400 |
| **BIC (smaller is better)** | 72,444 | 77,696 |

# Advantages

- Controls for changes in mix by state, class, etc. which can distort development factors.
- Links reserves to drivers like lag, inflation, etc. which may or may not give better predictions but provide explanations of reserve changes to management coming out of Wall Street and Federal Reserve.

# Future Improvement

## Claim Probability of Closing Model

For each open claim, probability that it will close in the next year.

Then can simulate when the claim will close.

Then we can apply severity models to simulate claim size.

We will use similar variables

## Generalized Additive Model

Converts categorical variable to numerical.

A model creates some nonlinear function of the variable and uses that as predictive variable instead of grouping.

We did this for operational time.

- **Worked well but couldn't compare goodness of fit.**
- **Can be generalized to fit operational time and time to close simultaneously**
- **Even though collinear, non linear functions will not be collinear**

- Personal View Only
- Any opinions expressed therein are not created, sponsored or endorsed by AIG/CAS.

AIG