FARMERS

BRISTOL WEST

Non-Pricing Generalized Linear Modeling in the Auto Industry

Casualty Actuaries of the Southeast
Boca Raton, FL
March 11, 2008

Derek Schaff, FCAS, MAAA

The Linear Model

- The expected value of the response (dependent) variable Y_i is given as a linear combination of the covariates (independent variables) $X_{1,i}$, ... , $X_{k,i}$, i.e.

$$E(Y_i) = \beta_0 + \beta_1 X_{1,i} + ... + \beta_k X_{k,i}$$

- The actual observed values of Y_i are distributed as normal random variables with mean $E(Y_i)$ and variance σ^2.

Limitations of Linear Model

- The requirement of a normal response variable is restrictive
  - What if only positive values are possible?
  - What if only integral values are possible?
- The assumption of constant variance is not always plausible
  - What if larger expected values imply greater uncertainty?

Generalized Linear Model

- The expected value of the response (dependent) variable Y_i is given as a function of a linear combination of the covariates (independent variables) $X_{1,i}$, ... , $X_{k,i}$, i.e.

$$E(Y_i) = g^{-1}(\beta_0 + \beta_1 X_{1,i} + ... + \beta_k X_{k,i})$$

- The actual observed values of Y_i are distributed as random variables from one of the exponential family of distributions.
  - The exponential family of distributions includes the normal, Poisson, gamma, Tweedie (compound gamma/Poisson) and binomial.

Advantages of Generalized Linear Model

- The link function g allows for more general relationships between the linear predictor and the expected value of the response variable

  - Multiplicative relationship:

    $g(x) = \ln(x)$

  - Restriction of expected values to (0,1) :

    $g(x) = \ln(x/(1-x))$

Advantages of Generalized Linear Model

- The relaxation of the normality of the error structure allows for more general response variables

  - Claim frequency
    - Poisson (variance ~ E(X))
  - Claim severity
    - Gamma (variance ~ $E(X)^2$)
  - Pure premium
    - Tweedie (variance ~ $E(X)^p$, where $1<p<2$)
  - Probability of renewal
    - Binomial

Limitations of Generalized Linear Model

- The error structure is limited to distributions from the exponential family

- The response variables Y_i are assumed to be independent
  - There may be correlations in the response variables, e.g. if the same insured is used in multiple observations of individual claim frequency.

What's Wrong with One-Way Analysis?

- **Suppose the goal is to estimate the impact of gender and marital status on loss ratios.**

EP	Married	Single	Total
Male	$120	$40	$160
Female	$40	$120	$160
Total	$160	$160	$320

Losses	Married	Single	Total
Male	$72	$16	$88
Female	$24	$48	$72
Total	$96	$64	$160

- **One-way analysis concludes:**
  - **Males have a 22% larger loss ratio than females (55% vs. 45%)**
  - **Marrieds have a 50% larger loss ratio than singles (60% vs. 40%)**

What's Wrong with One-Way Analysis?

- Let's examine the two-way loss ratios

Loss Ratio	Married	Single	Total
Male	60%	40%	55%
Female	60%	40%	45%
Total	60%	40%	50%

- Two-way analysis shows that the difference in loss ratios between males and females is entirely driven by the correlation between marital status and gender.

Advantages of GLM's over One-Way Analysis

- GLM's account for all variables simultaneously
  - Multi-way analysis quickly becomes cumbersome as the number of independent variables is increased
- GLM's come with a nice statistical framework
  - One can quantify the error in parameter estimates
  - One can assess the quality of fit of the overall model

Where Can GLM's be Used?

- Many auto insurers currently use GLM methods for pricing
    - The goal of the analysis is typically to calculate the relativities for each rating variable (e.g. age, gender, marital status)

Where Can GLM's be Used?

- GLM's also add value in other departments
  - **Reserving**
  - Marketing
  - Claims
  - Finance
  - Operations

GLM Case Reserves
Background

- Until a feature is paid, it is assigned a case reserve based upon:
  - Coverage (Bodily Injury, Property Damage, etc.)
  - Age of the claim
  - Policy State

- This methodology can lead to biases in reserve triangles
  - Case reserve is not inflation-sensitive.
  - Case reserve is not sensitive to policy limits
  - Case reserve does not take advantage of any feature-specific information

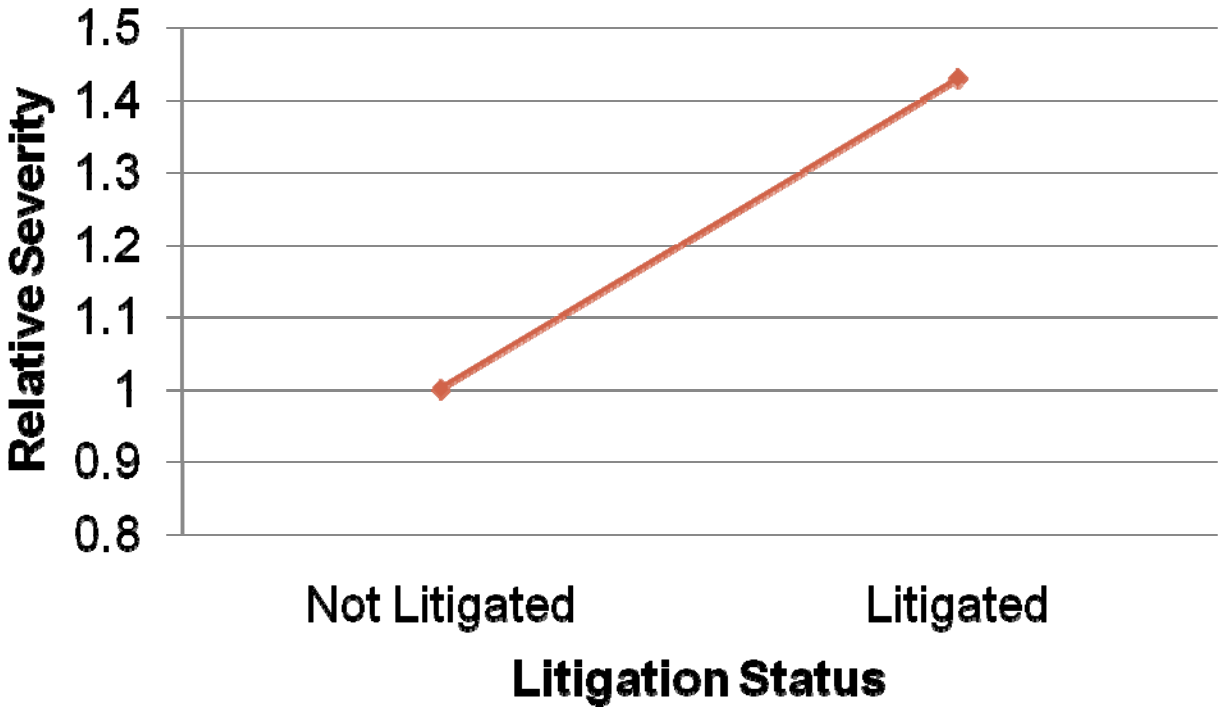
GLM Case Reserves
Model

- Each observation consists of an individual feature which has closed with a nonzero payment
- Response Variable: Paid Loss
- Error Distribution: Gamma
- Some Independent Variables:
  - Accident Quarter
  - Policy State
  - Policy Limits/Deductible
  - Report Lag
  - Settlement Lag
  - Vehicle Value
  - Presence of attorney/litigation
  - Type of Injury (soft tissue, herniated disc, etc.)

GLM Case Reserves

Some Results

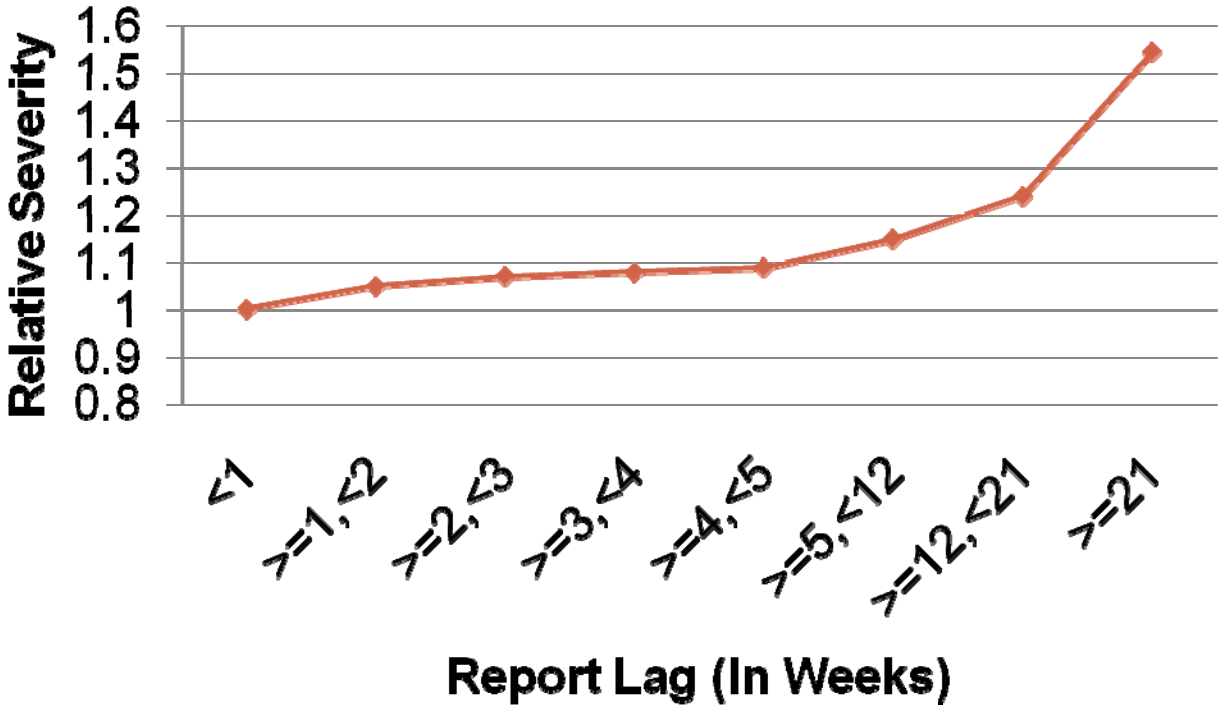
BI Severity is Greater For Litigated Features



GLM Case Reserves

Some Results

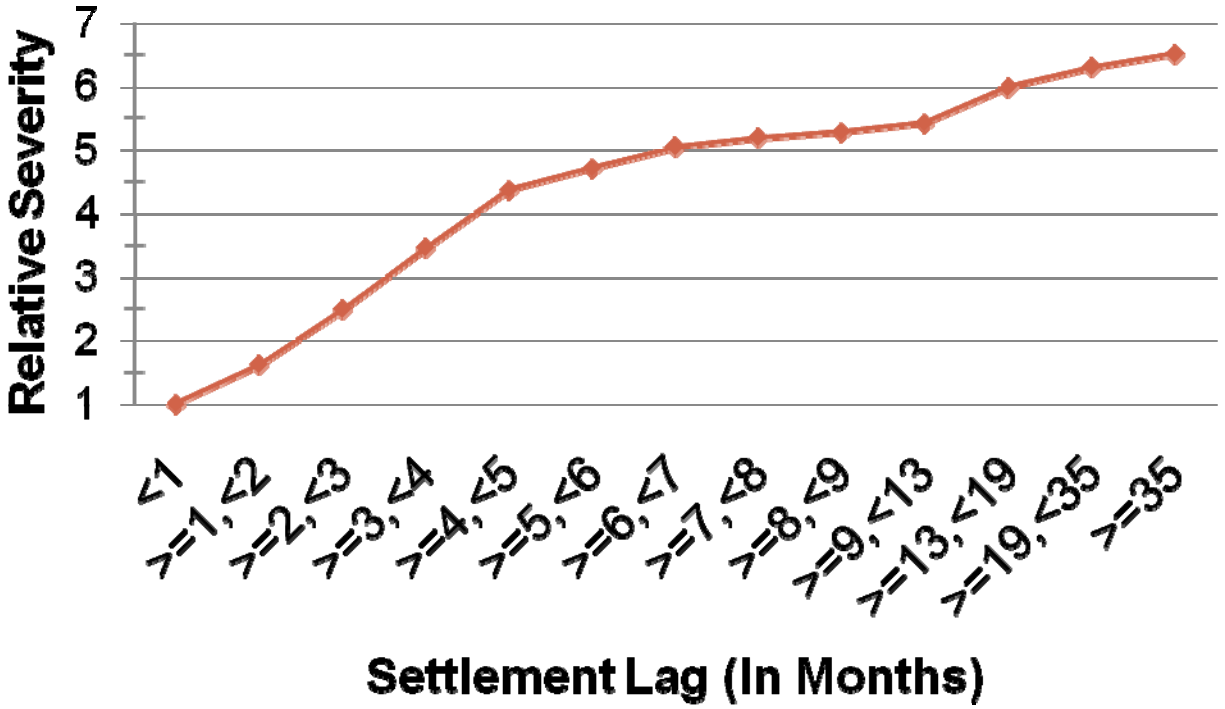
BI Severity is Greater With Long Report Lags



GLM Case Reserves

Some Results

BI Severity is Greater With Long Settlement Lags



GLM Case Reserves
Outcome

- Now the case reserve for each open feature is updated on a daily basis to reflect the individual characteristics of the feature.
  - Helps with bias in triangles due to mix shifts
  - Each individual feature's case reserve better reflects its expected ultimate cost

- Can also use the output from the GLM's to quantify the variance in the estimates and so arrive at an aggregate estimate of variability for the ultimate cost of all known claims.

Where Can GLM's be Used?

- GLM's add value in other departments
  - ~~Reserving~~
  - **Marketing**
  - Claims
  - Finance
  - Operations

Agent Quote Contest
Background

- Marketing has been running a promotion selected states to entice our new Exclusive Agents (EA's) to use our software and start selling our policies.

- Is the promotion effective and worth the cost?

Agent Quote Contest
Model

- Each observation consists of a week/state combination.
- Response Variable: Percentage of previous week non-quoters who quoted for the first time in the given week
- Error Distribution: Gamma
- Independent Variables:
  - Weeks Since Rollout
  - Contest Active Indicator

Agent Quote Contest
One-Way Result

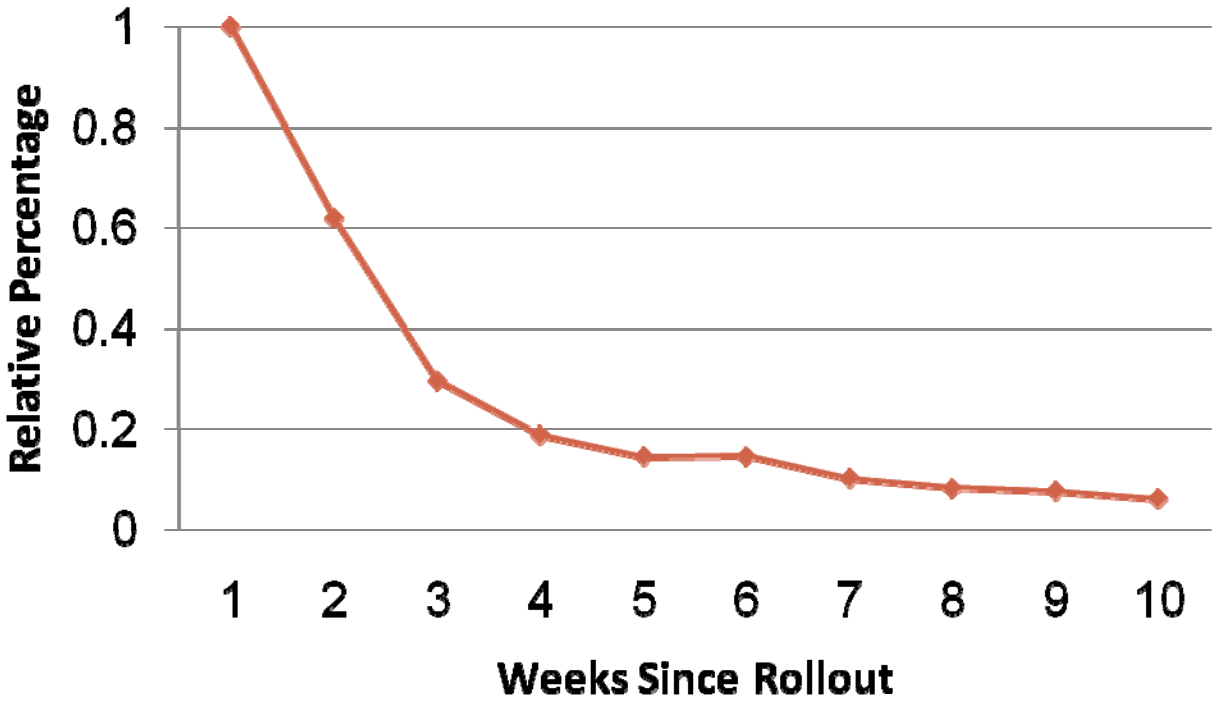
Overall Results Were Poor

	Percentage of Non-Quoters Enticed to Quote
Without Contest	14.2%
With Contest	4.8%

Agent Quote Contest

Multivariate Results

The More Weeks Since Rollout, the Less Non-Quoters Become Enticed to Quote



Agent Quote Contest
Multivariate Results

Once Weeks Since Rollout Have Been Controlled For, the Promotion is Shown to Be Effective

Weeks Since Rollout	Lift of Contest
1	0.7%
2	0.4%
3	0.2%
4	0.1%
5	0.1%
6	0.1%
7	0.1%

Where Can GLM's be Used?

- GLM's also add value in other departments
  - ~~Reserving~~
  - ~~Marketing~~
  - **Claims**
  - Finance
  - Operations

Attorney Involvement
Background

- It is well known that claims with attorney involvement have higher costs on average than those without.

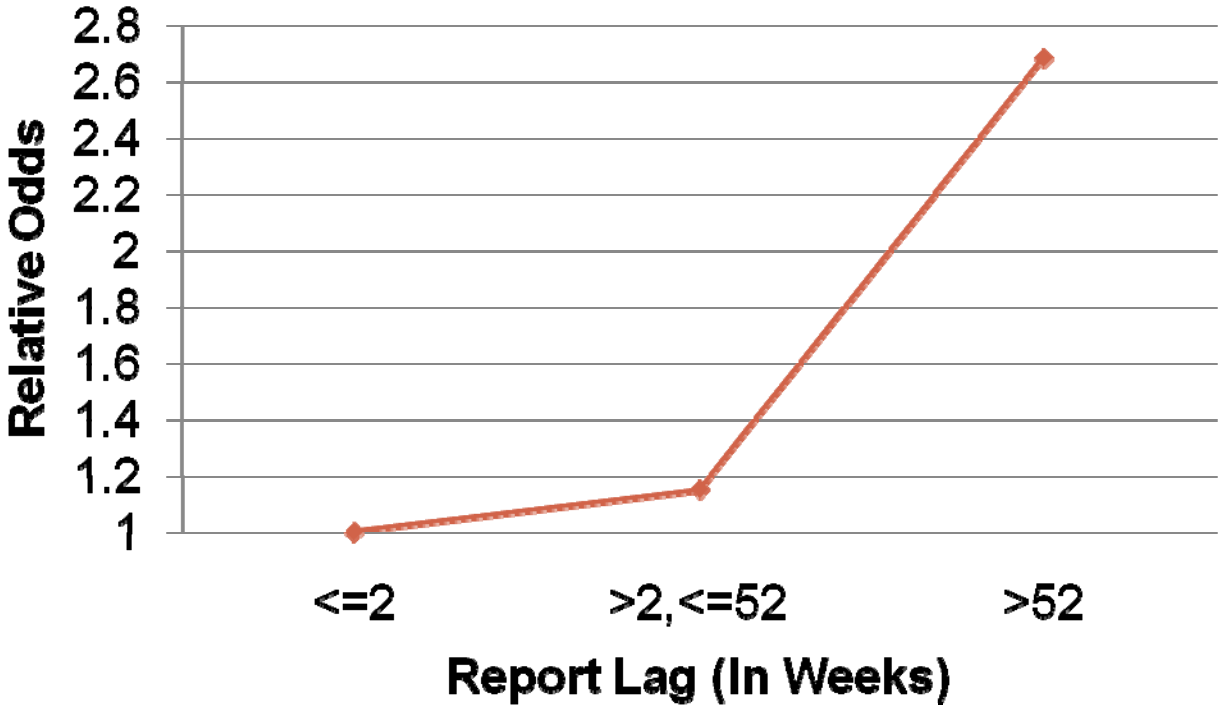
- What drives attorney involvement, and can anything be done about it?

Attorney Involvement
Model

- Each observation consists of a closed Bodily Injury (BI) feature.
- Response Variable: A variable indicating whether or not an attorney was involved
- Error Distribution: Binomial
- Some Independent Variables:
  - Report Lag
  - Settlement Lag
  - Claimant Age
  - Claimant Gender
  - Adjuster Job Title
  - Adjuster's BI Inventory

Attorney Involvement
Some Results

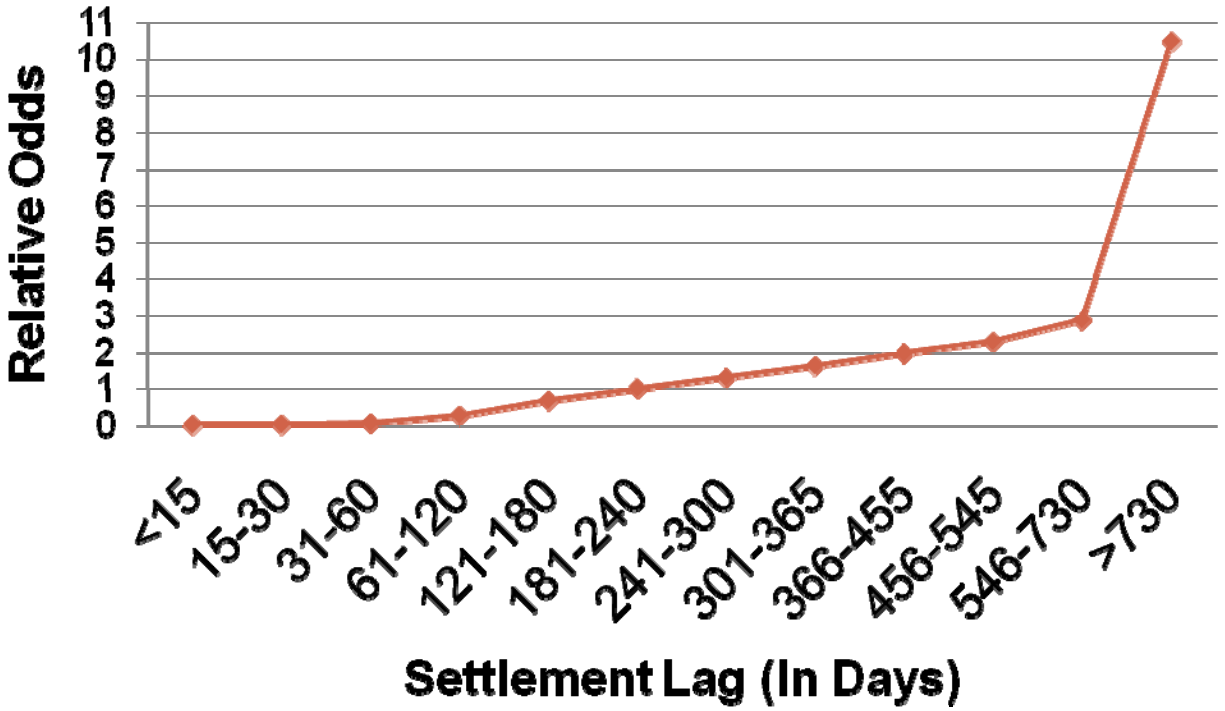
Attorneys are More Likely With Long Report Lags



Attorney Involvement

Some Results

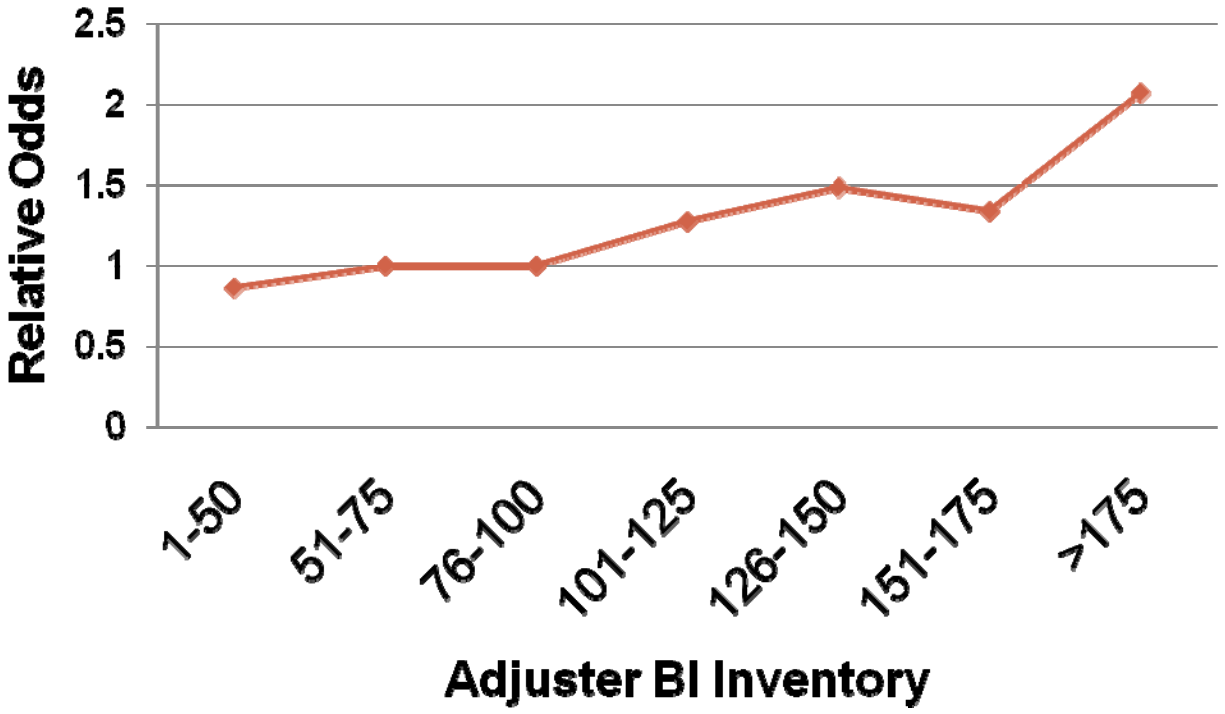
Attorneys are More Likely With Long Settlement Lags



Attorney Involvement

Some Results

Attorneys are More Likely When the Adjuster Has a Large Inventory



Attorney Involvement
Outcome

- Only two independent variables in study were under our control: Settlement Lag and Adjuster BI Inventory
  - Recommended quick settlement when possible to make attorney involvement less likely
  - Recommended limitation of adjuster BI inventory to 125 or fewer features

Where Can GLM's be Used?

- GLM's also add value in other departments
  - ~~Reserving~~
  - ~~Marketing~~
  - ~~Claims~~
  - **Finance**
  - Operations

Collections
Background

- As a part of our collections efforts, we were employing a third party vendor to call policyholders to remind them to pay their premium.

- Are there a subset of policyholders for which the benefit of calling is outweighed by the cost (e.g. policyholders who are very late)? How should we target the collections calls?

Collections
Background

- To assess the effectiveness of collections calls, we need a group of policyholders who have not been called (a control group)

  - Each day a group of policyholders is eligible to be called
  - A small percentage of policyholders were not called on the day they were eligible

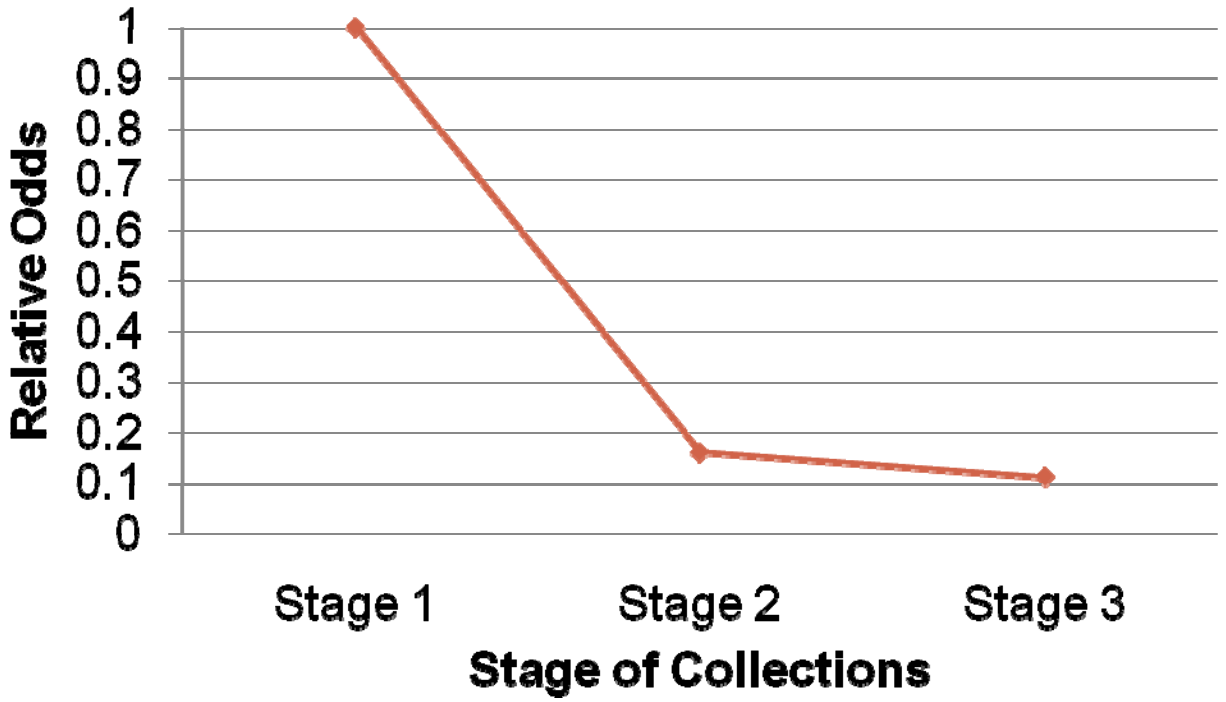
Collections
Model

- Each observation consists of a call-eligible policyholder who is late in paying premium
- Response Variable: A variable indicating whether or not payment followed the day of call eligibility
- Error Distribution: Binomial
- Some Independent Variables:
  - Group (indicating the stage of collections)
  - Policyholder Age
  - Amount Due
  - Whether a call was made or not
  - Policy Mod

Collections

Some Results

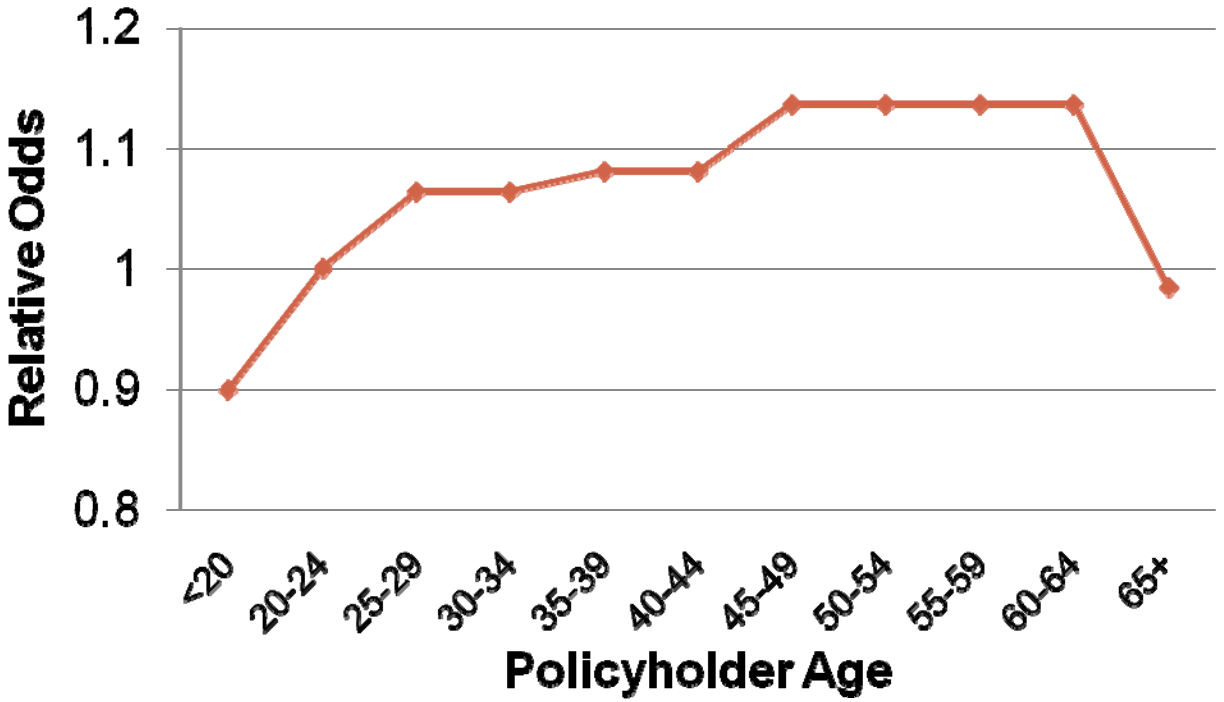
Policyholders at a Later Stage of Collections are Less Likely to Pay



Collections

Some Results

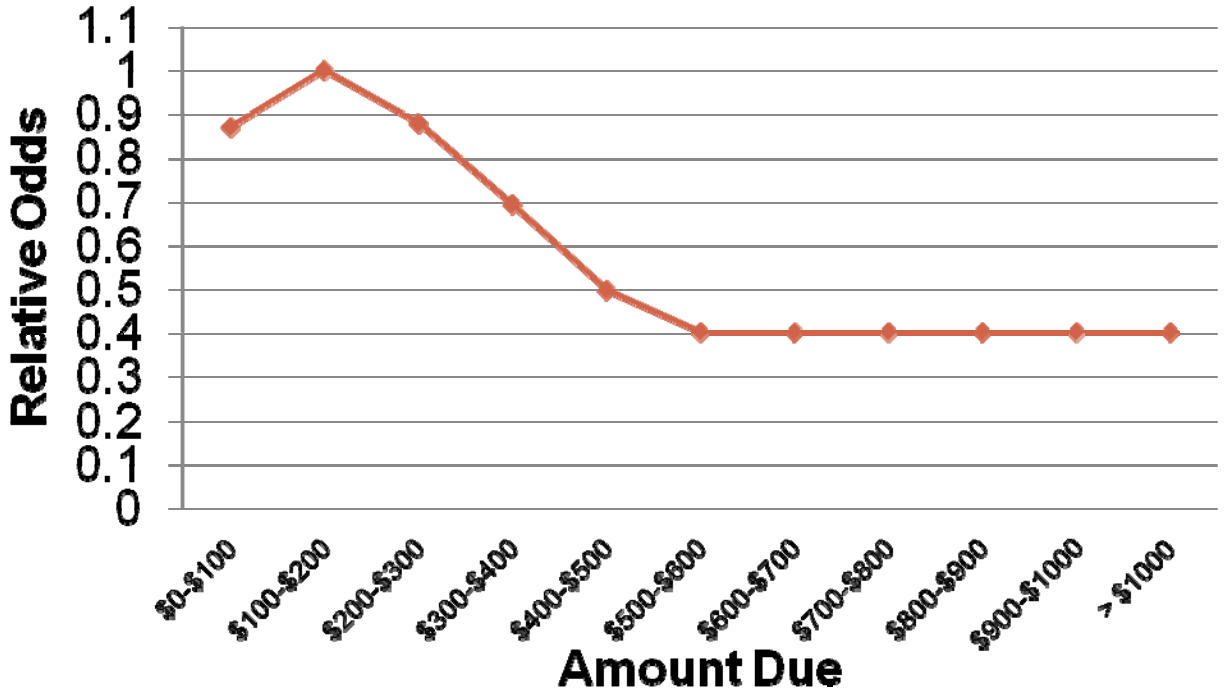
Older Policyholders are More Likely to Pay



Collections

Some Results

Likelihood of Payment Decreases as Amount Due Increases



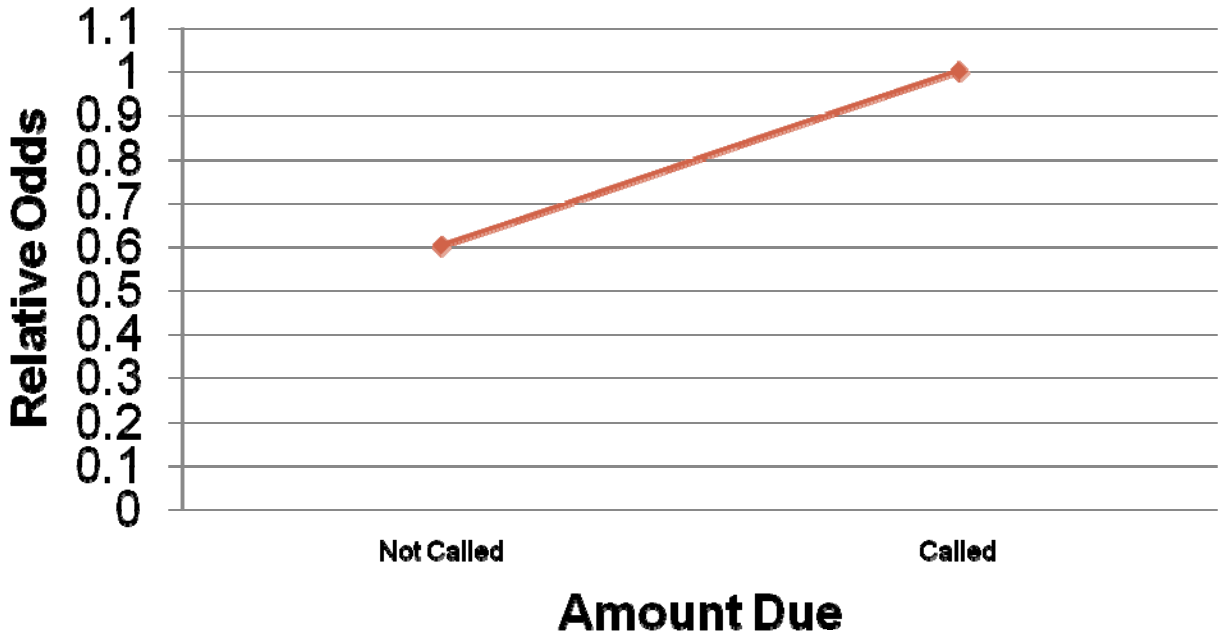
FARMERS

BRISTOL WEST

Collections

Some Results

Calling Increases the Likelihood of Payment



Collections
Outcome

- To decide whether or not to call a call-eligible policyholder, use the model results to calculate the probability of the policyholder making a payment with and without a call being made, and make a call if

  (Prob of Payment w/Call) *(Amt Due) –
  (Prob of Payment w/o Call) * (Amt Due) > (Cost of Calling)

Where Can GLM's be Used?

- GLM's also add value in other departments
  - ~~Reserving~~
  - ~~Marketing~~
  - ~~Claims~~
  - ~~Finance~~
  - **Operations**

CSR E vs. A
Background

- We would like to have a way to evaluate customer service representatives (CSR's), based upon the tendency of policyholders to cancel soon after their contact with the CSR.

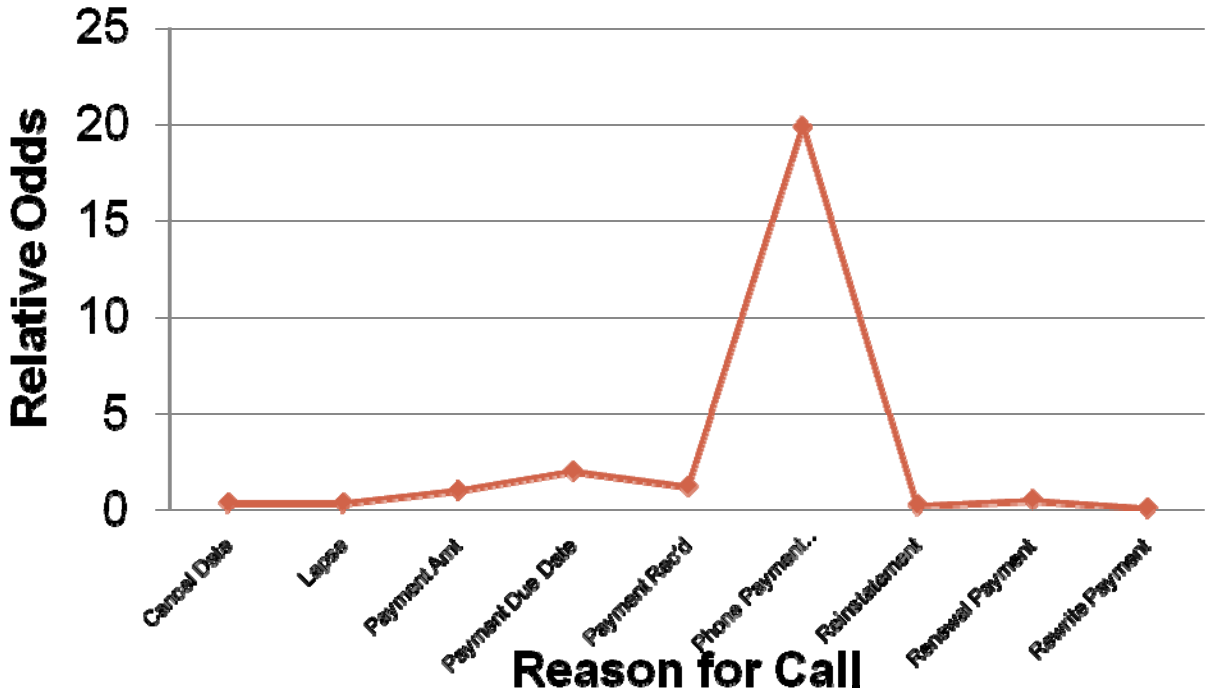
CSR E vs. A
Model

- Each observation consists of a policyholder-initiated call
- Response Variable: A variable indicating whether or not the policy was in force 2 months following the call
- Error Distribution: Binomial
- Some Independent Variables:
  - Reason for Call
  - Policyholder Credit
  - Homeowner Indicator
  - Policyholder Age
  - Whether Policy is New or Renewal

CSR E vs. A

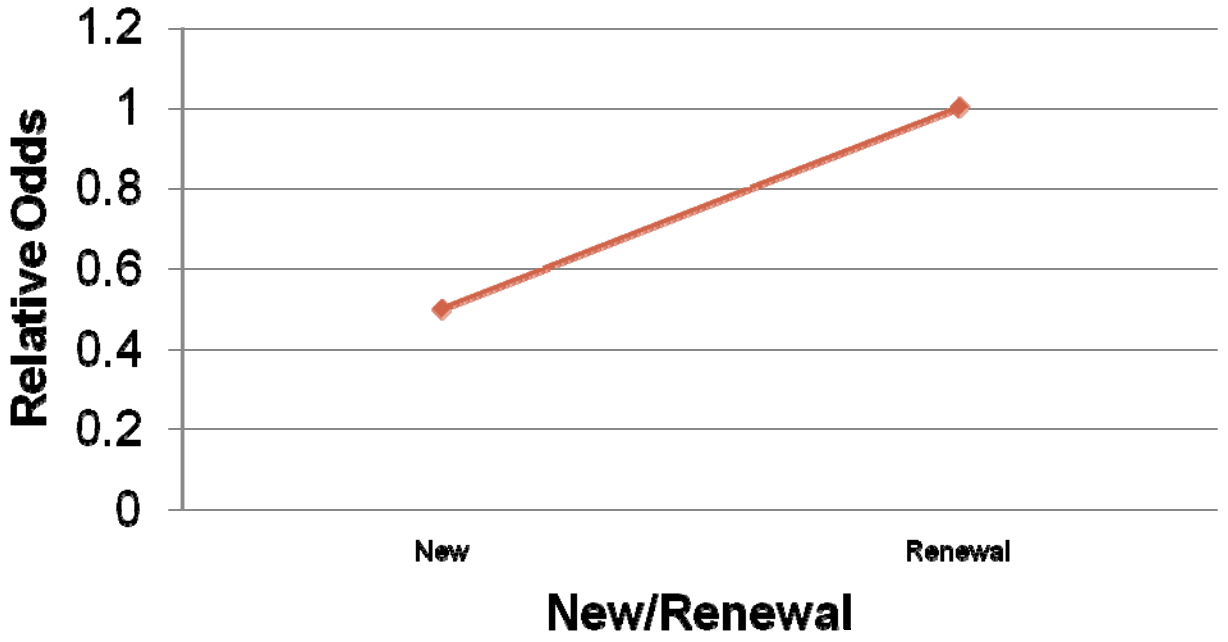
Some Results

Likelihood of Remaining In Force Depends on Reason for Call



CSR E vs. A
Some Results

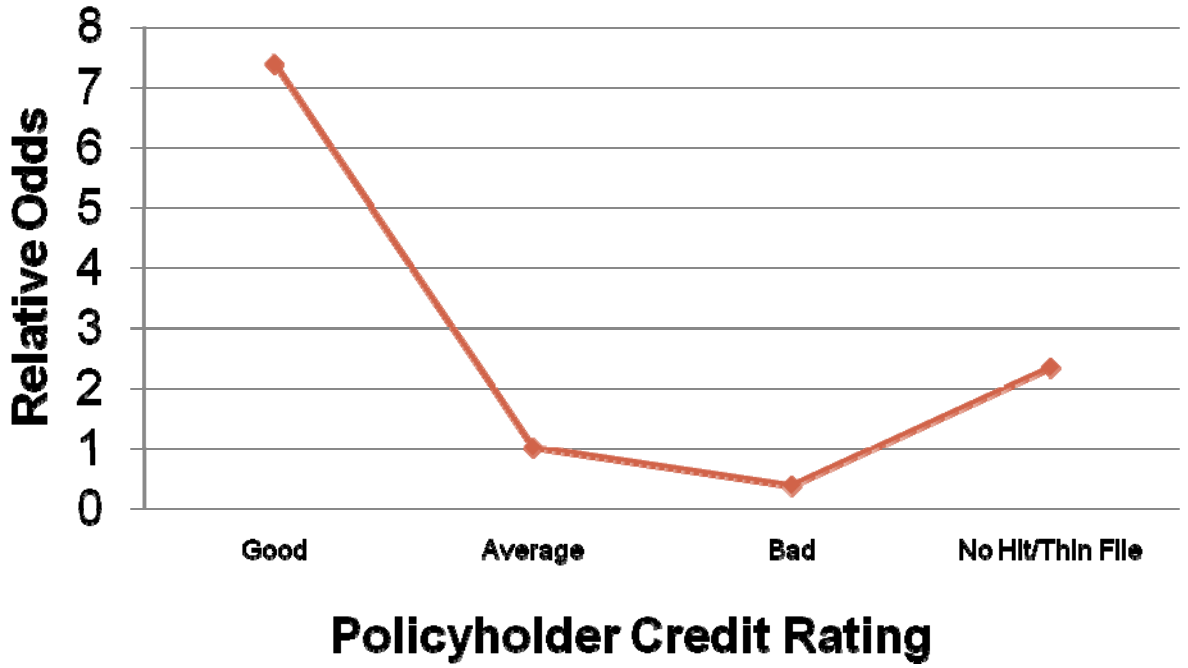
Renewal Policies are More Likely to Remain in Force



CSR E vs. A

Some Results

Better Credit Policies are More Likely to Remain in Force



CSR E vs. A
Outcome

- For each call a CSR receives over a given time period, calculate the expected probability of the policy remaining in force 2 months later.

- Sum these expected probabilities over all calls received. This is the expected number of policies in force 2 months later.

- Compare this expected number to the actual number of policies in force 2 months later. Investigate CSR's who systematically experience actuals less than expecteds.

Summary

- GLM's provide many advantages over the traditional one-way analyses used in other departments to evaluate performance and facilitate decision making.

- It falls to us to gain management buy-in and to explain results to nonactuarial personnel.

Questions?