Deloitte.

CAS Ratemaking and Product Management Seminar

GLM II

March 21, 2011

A Case Study in Claims Management

Amel Arhab, FCAS, MAAA

March 21, 2011

Antitrust Notice

The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

GLMs in the insurance industry…

Deloitte.

The use of GLMs in Insurance

Traditional Applications

Pricing

- Tiering, schedule plan
- Class plan optimization and optimal scheduled credits/debits

Underwriting

- Enhanced underwriting decision making
- Risk selection, retention strategies, automated underwriting
- Resource allocation, straight-through processing

Emerging Trends

Customer Service

- Queue Prioritization
- Service Offerings
- Resource Allocation

Marketing and Agency Management

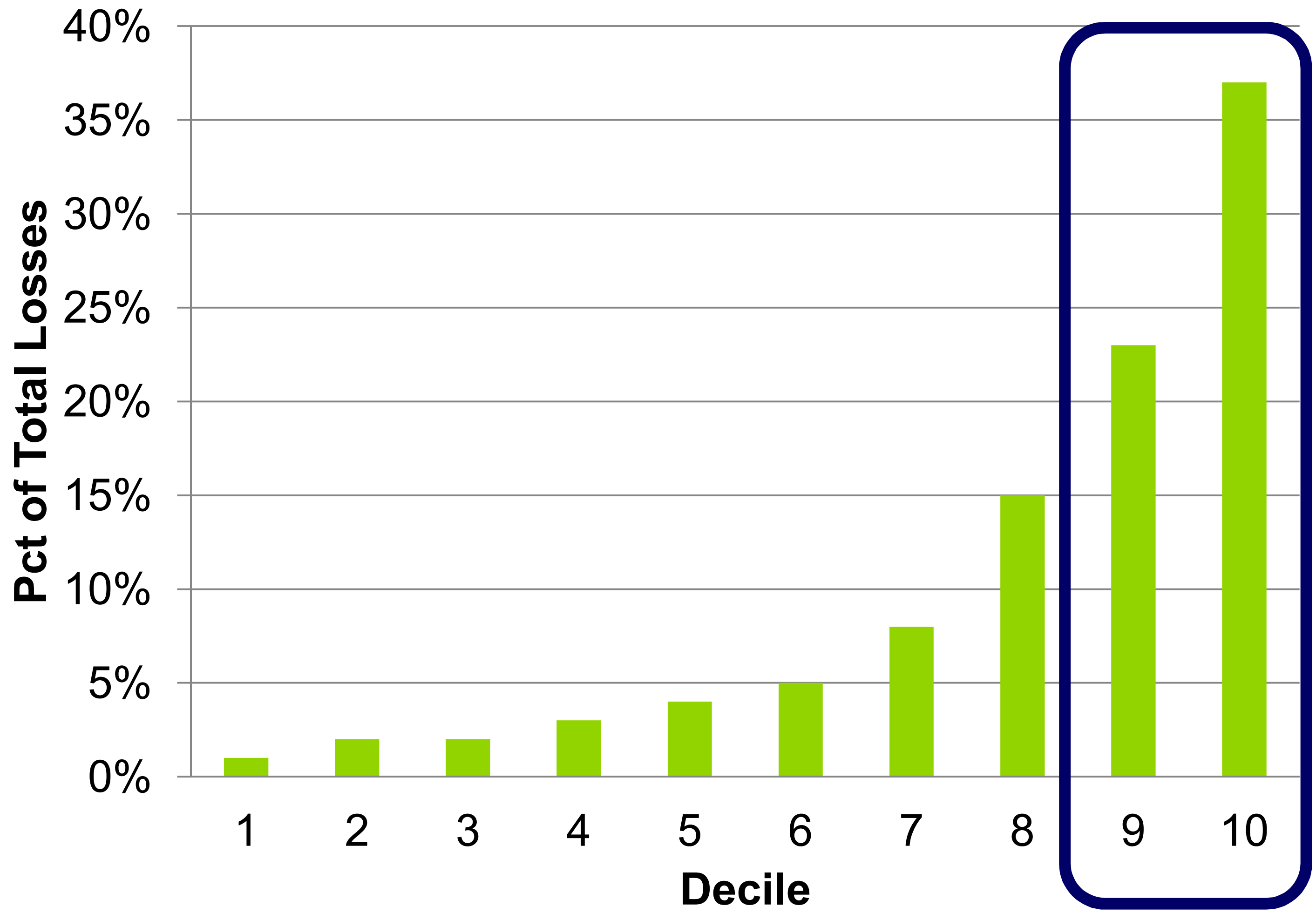
- Targeted Lead Generation
- Cross-Selling Potential
- Agency/Agent Management, Training, Servicing

Claims Management

- Automated Processing and Triage
- Fraud/Salvage/Subrogation Potential
- Duration Improvement and Litigation management

Deloitte.

Rational for GLMs in the Claims Space



The structure of a GLM…

The Structure of a GLM

The general structure of a Generalized Linear Model is:

$$\mu_i = E[Y_i] = g^{-1}\left(\sum X_{ij}\beta_j + \xi_i\right)$$

$$Var[Y_i] = \Phi\, V(\mu_i) / \omega_i$$

GLMs in the context of Claims

What is typically being modeled in claims management?

$$\mu_i = E[Y_i] = g^{-1}(\sum X_{ij}\beta_j + \xi_i)$$

$$Var[Y_i] = \Phi V(\mu_i) / \omega_i$$

- Ultimate severity
- Closing duration
- Propensity for fraud
- Propensity for litigation
- Propensity for salvage/subrogation recovery
- Litigation expenses
- Propensity to explode
- …and more

The link function:

$$\mu_i = E[Y_i] = g^{-1}(\sum X_{ij} \beta_j + \xi_i)$$

$$Var[Y_i] = \Phi V(\mu_i) / \omega_i$$

- The link function "g" is differentiable and monotonic
- Typically, log link functions ($g(x) = \ln(x)$ or $g^{-1}(x) = e^x$) are used for severity, expense, duration to allows for a multiplicative effect
- Logit link function ($g(x) = \ln(x/1-x)$ or $g^{-1}(x) = e^x / 1+ e^x$) for propensity to litigation, fraud, or recovery

Predictive variables:

$$\mu_i = E[Y_i] = g^{-1}(\sum X_{ij}\,\beta_j + \xi_i)$$

$$Var[Y_i] = \Phi\, V(\mu_i) / \omega_i$$

- Predictors will eventually determine how good a model is
- Internal data across many departments can be predictive (claimant, policy information, etc.)
- A wide range of external 3rd party data is also available (geo-demographic, financial, etc.)

The Offset:

$$\mu_i = E[Y_i] = g^{-1}(\sum X_{ij} \beta_j + \xi_i)$$

$$Var[Y_i] = \Phi V(\mu_i) / \omega_i$$

If the effect of a predictive variable is known → don't estimate its β and introduce the offset term.

General examples of offset in the claims space include:

- None
- State effect
- Etc…

The variance function:

$$\mu_i = E\,[Y_i] = g^{-1}\,(\sum X_{ij}\,\beta_j + \xi_i)$$

$$Var\,[Y_i] = \Phi\,V(\mu_i)\,/\,\omega_i$$

Target	Link	Error
Ultimate severity	Log	Gamma/Tweedie
Closing duration	Log	Gamma
Propensity for fraud	Logit	Binomial

GLMs in the context of Claims

Prior weights:

$$\mu_i = E[Y_i] = g^{-1}\left(\sum X_{ij}\,\beta_j + \xi_i\right)$$

$$Var[Y_i] = \Phi\, V(\mu_i) / \omega_i$$

Prior weight are used to assign known credibility to each data point:

- Typically 1
- Can vary by loss year, claim class, etc.
- Can be very creative but be wary of results (validate by cross-sections)

The modeling dataset…

Deloitte.

The Modeling Dataset

Modeling dataset:

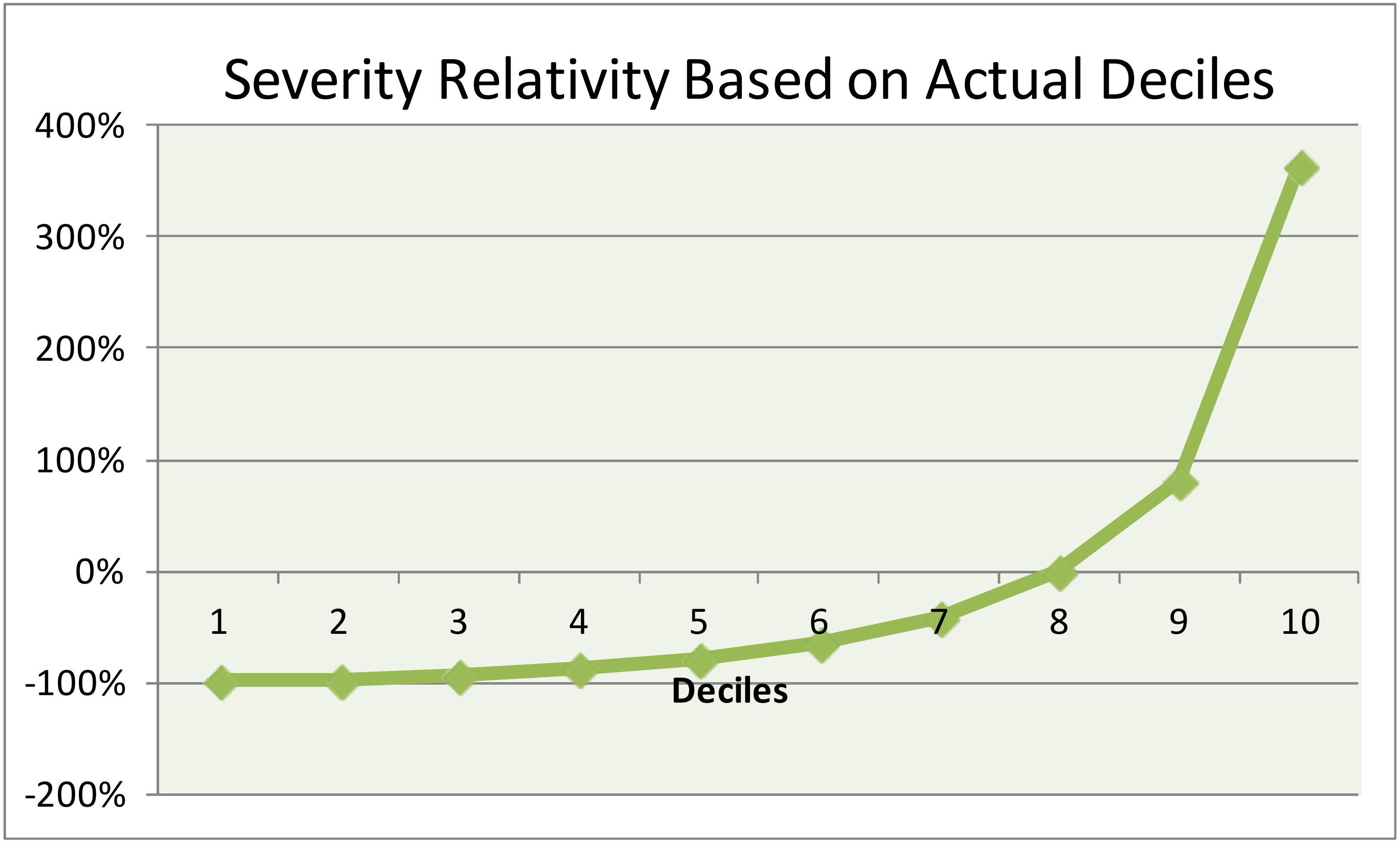
1. Workers' compensation – Lost time indemnity severity
2. Severity was trended and appropriately adjusted
3. Year used 2002-2007
4. Data points used 30,000 closed claims
5. CWOP pay claims were excluded
6. Predictive variables include:
    1. Claims data
    2. Claimant information
    3. Injury details
    4. Employment data
    5. External data

Get to know your response variable…

Distribution of the observed response variable: WC LTI severity

Quartile 4	$1,000,000	← Consider capping
Quartile 3	$23,000	
Quartile 2	$3,500	
Quartile 1	$700	
10%	$400	
5%	$100	
1%	$0	← Consider Zeroing-out / excluding
0% Min	$-20	
Mean	16,000	

A Case Study – Actual Lift



Get to know your predictors…

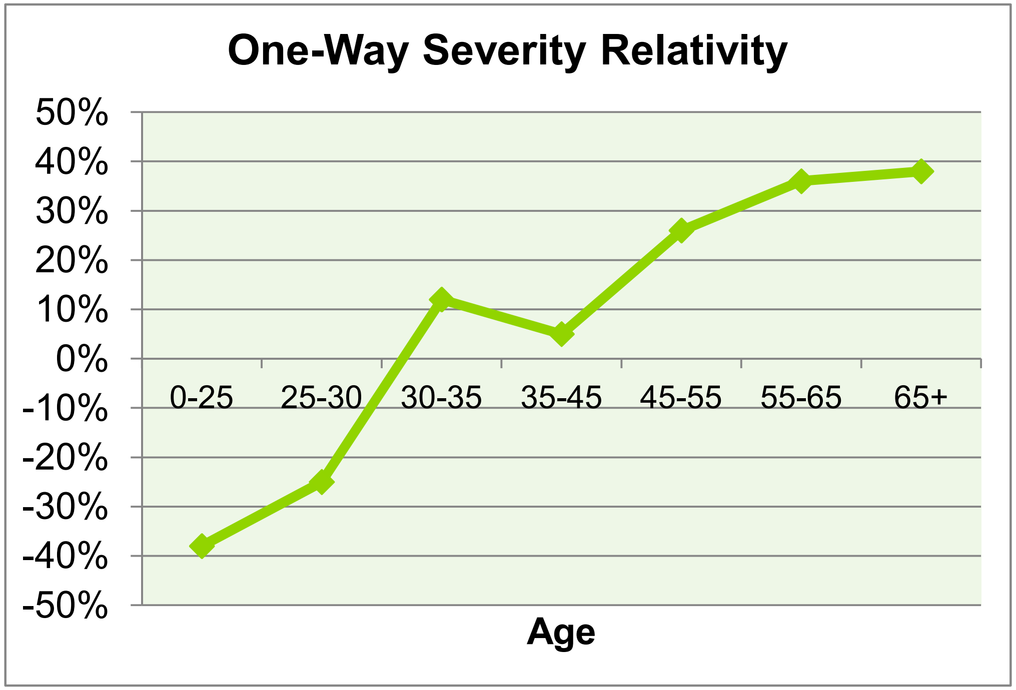
In general, predictors should be tested on the following prior to modeling:

- Variable distribution
- Level of missing values and their meaning
- Variable transformation (grouping, cap max, etc.)

Age	Claim Count
0-25	4,234
25-30	4,266
30-35	5,498
35-45	6,411
45-55	4,514
55-65	3,217
65+	1,748
Missing	112

In general, predictors should be tested on the following prior to modeling:

- Correlation with the response variable
- Correlation with predictors and principal components



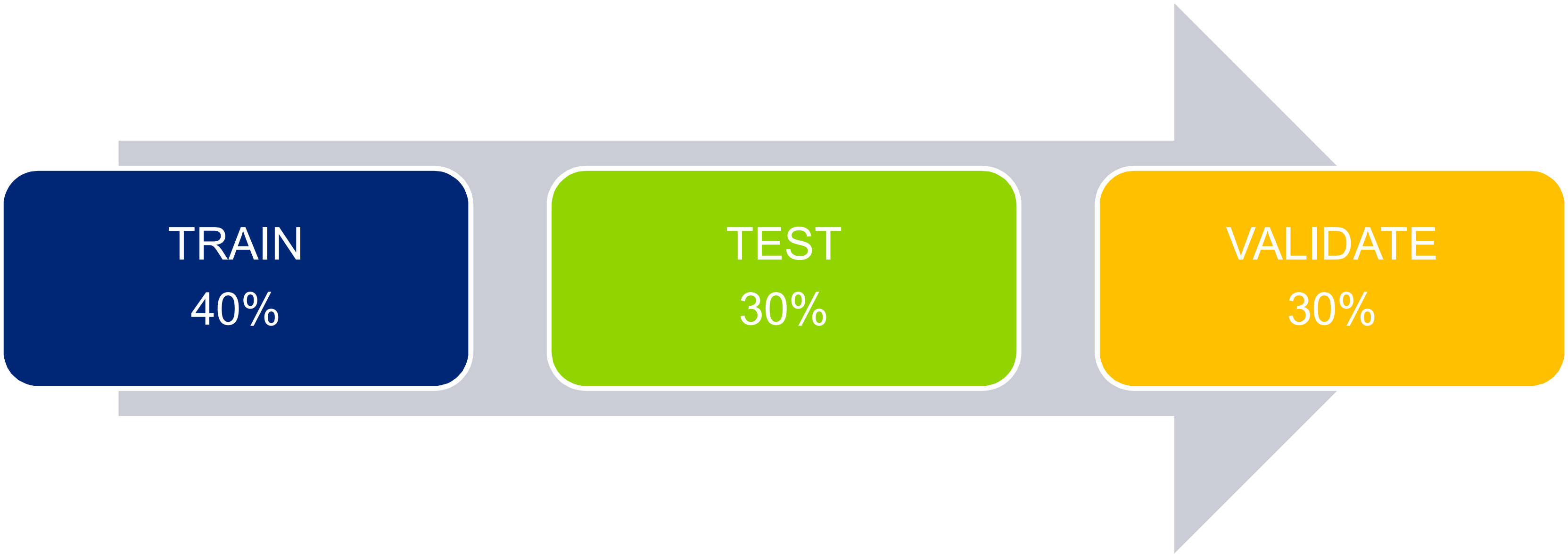
In general, predictors should be tested on the following prior to modeling:

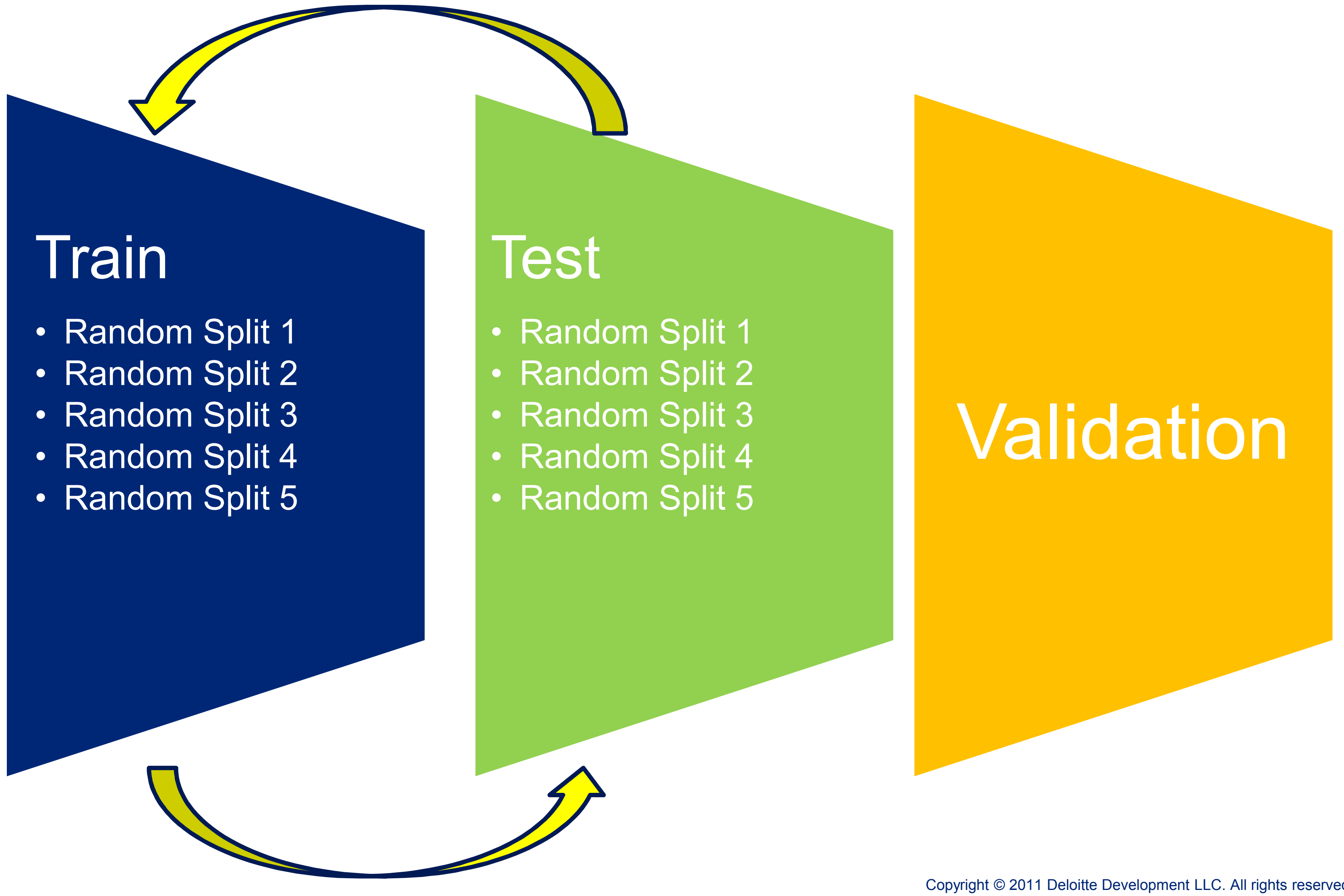
- Business meaning and usability
- Legal and regulatory limitations
- Availability and limitation in production
- Changes over time

Begin the modeling exercise…

Modeling parameters used:

Response Variable	Link	Error	Weight	Offset
Ultimate LTI Severity	Log	Gamma	1	N/A

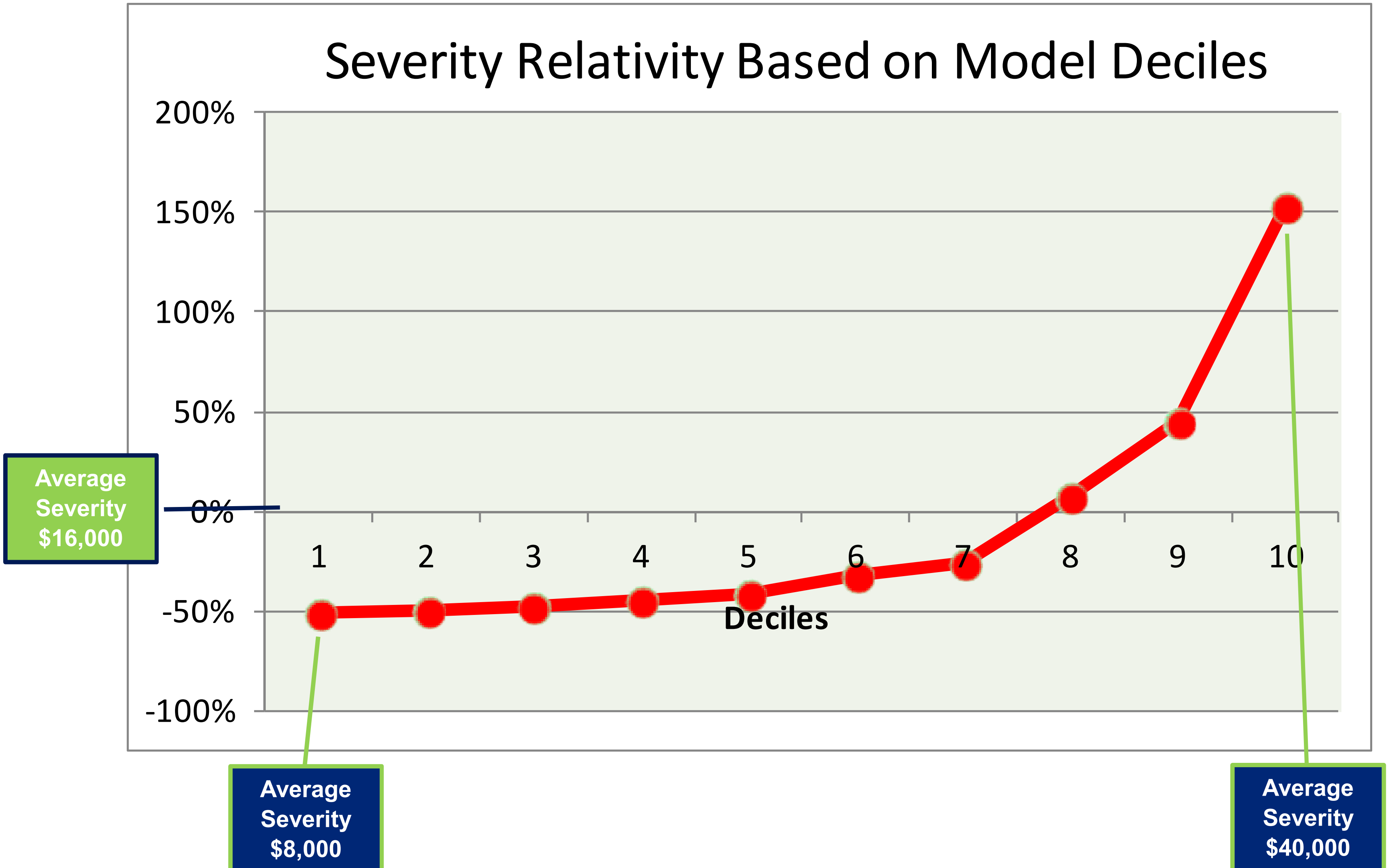



Train
• Random Split 1
• Random Split 2
• Random Split 3
• Random Split 4
• Random Split 5
Test
• Random Split 1
• Random Split 2
• Random Split 3
• Random Split 4
• Random Split 5
Validation

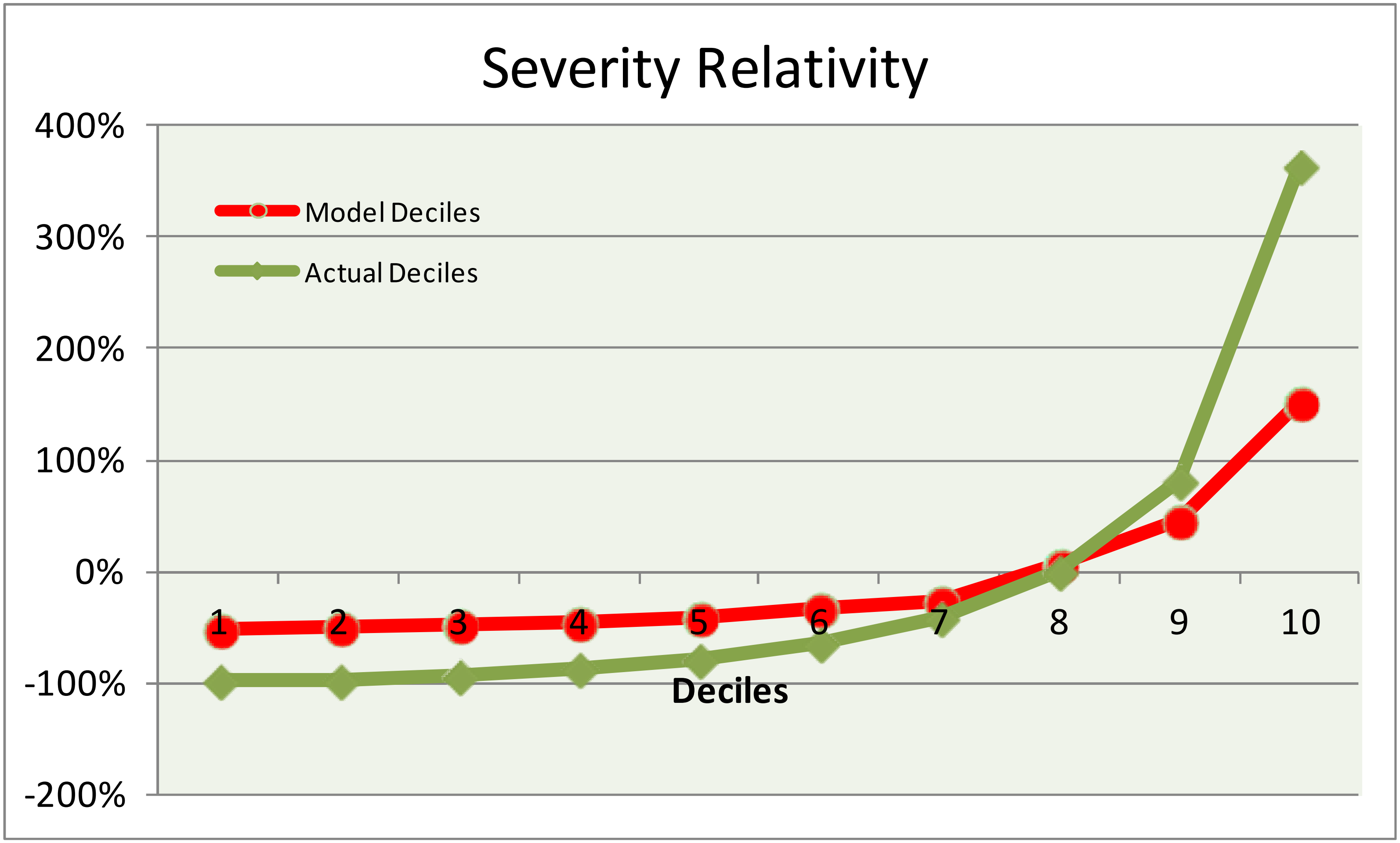
A sample modeling output:

Variable	PE RS 1	ChiSq RS 1		PE RS 5	ChiSq RS 5
X_1 = Age	1.20	74	…	1.19	77
X_2 = Lower Back Injury	0.75	53	…	0.69	53
X_3 = Afternoon Injury	-0.63	48	…	-0.61	47
X_4	0.54	33	…	.55	32
X_5	-0.41	21	…	-0.46	21
…			…		
X_{40} = Missing Indicator for Census Data	0.11	1	…	0.09	2

A Case Study – Validation Lift



A Case Study – Validation Lift



Post Modeling…

Post Modeling

After modeling is complete:

- Implement
- Monitor
- Enhance
- Expand

Questions

Deloitte.

Member of
Deloitte Touche Tohmatsu