# PL-3
# GLM Practical Applications

## 2006 CAS Seminar on Ratemaking

Claudine Modlin, FCAS

Watson Wyatt Worldwide

WWW.WATSONWYATT.COM

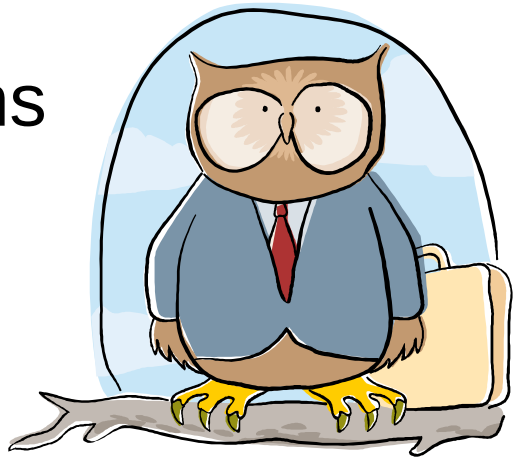
Watson Wyatt Worldwide

# Famous modeling quotes

- ***"Prediction is very difficult, especially if it's about the future."***
  - Nils Bohr, Nobel laureate in Physics
- ***"I have seen the future and it is very much like the present, only longer."***
  - Kehlog Albran, *The Profit*
- ***"A good forecaster is not smarter than everyone else, he merely has his ignorance better organized. "***
  - Anonymous

# If I knew then what I know now….

- Define application; link initiative to success
- Prepare the culture
- Know thy data
- Fit models and interpret
- Translate models to decisions



# If I knew then what I know now….

- Define application; link to success
- Prepare the culture
- Know thy data
- Fit models and interpret
- Translate models to decisions

# Applications

- Ratemaking
- Underwriting
- Marketing
- Retention
- Expense analysis
- Claims management
- Other risk management
- Reserving

# Applications

- Ratemaking
  - revise existing rating factor relativities with multivariate analysis
  - introduce new rating variables or tier structure
  - re-define territorial boundaries or vehicle classification
  - define rating plan that optimizes profit while retaining required volume

# Applications

- Underwriting
  - provide guidelines on debits/credits
  - produce scorecards to automate some elements of risk selection
- Marketing
  - improve direct mail conversion rate for most profitable risks by X%

# Applications

- Retention
  - understand effect of capping rate changes at renewal
  - develop lifetime customer value model
- Expense analysis
  - vary acquisition costs by other criteria
- Claims management
  - develop fraud scorecard
  - advise how TPAs affect claim costs

# Applications

- Other risk management
  - determine which risks to cede
- Sales channel
  - align compensation with expected profitability
- Reserving
  - provide additional method to assist reserving actuaries with ultimate projections

# If I knew then what I know now….

- Define application; link to success
- Prepare the culture
- Know thy data
- Fit models and interpret
- Translate models to decisions

# Prepare the culture

- Discuss early on with other functional areas
  - pricing, underwriting
  - product development
  - IT
  - sales/dist channel
  - marketing
  - legal
  - statistical reporting

- Engage in effective communication

# Prepare the culture

- Determine what's in scope to change and the effect on operations
  - change to existing product or develop new product?
  - acceptance of new variables – esp external
  - tolerance for differences (to current rates or to competitors)
  - in consideration of timelines and operational costs (eg IT)
- Plan to share interim results

# Prepare the culture

- Determine how staffing needs to change
  - skills
  - organizational structure
- Assess hardware/software needs (incl back-up)
- Plan new procedures for checking, peer reviewing and documenting
- Enlist support from upper management

# If I knew then what I know now….

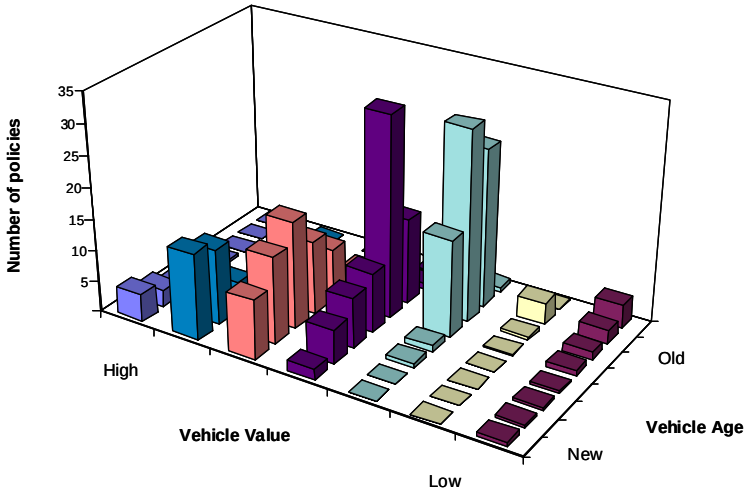
- Define application; link to success
- Prepare the culture
- Know thy data
- Fit models and interpret
- Translate models to decisions

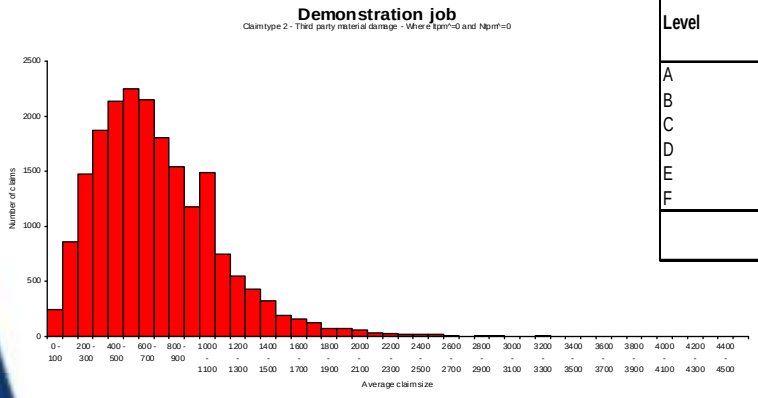
# Know thy data

- Data preparation
  - *to be covered in detail in Rick's presentation*
  - *see my notes in appendix to this presentation*
- Preliminary analysis

# Preliminary analyses

- Size of loss distribution
- One-way results
- Correlation statistics
- Two-way results




Claim type 1 - Third party property damage
Vehicle type (Type)

| Level | Number of records | Exposure | Premium | Number of claims | Incurred losses | Claim frequency | Average cost per claim | Pure premium | Loss ratio |
|---|---|---|---|---|---|---|---|---|---|
| A | 27,661 | 24,757 | 10,584,626 | 1,807 | 8,457,208 | 7.3% | 4,681 | 342 | 79.9% |
| B | 22,089 | 19,777 | 9,623,698 | 1,598 | 6,957,135 | 8.1% | 4,354 | 352 | 72.3% |
| C | 13,768 | 12,334 | 6,305,906 | 1,011 | 4,245,902 | 8.2% | 4,200 | 344 | 67.3% |
| D | 19,662 | 17,592 | 9,382,767 | 1,584 | 6,070,943 | 9.0% | 3,832 | 345 | 64.7% |
| E | 11,235 | 10,076 | 5,676,363 | 982 | 3,262,384 | 9.7% | 3,321 | 324 | 57.5% |
| F | 5,607 | 5,037 | 3,118,064 | 550 | 1,858,753 | 10.9% | 3,379 | 369 | 59.6% |
| | **100,022** | **89,572** | **44,691,424** | **7,532** | **30,852,324** | **8.4%** | **4,096** | **344.44** | **69.0%** |

# Preparing variables

- Categorical most common in insurance
  - sufficient volume needed for model convergence
  - consider several alternative categorizations
- Splines may be useful for continuous risk

# If I knew then what I know now….

- Define application; link to success
- Prepare the culture
- Know thy data
- Fit models and interpret
- Translate models to decisions

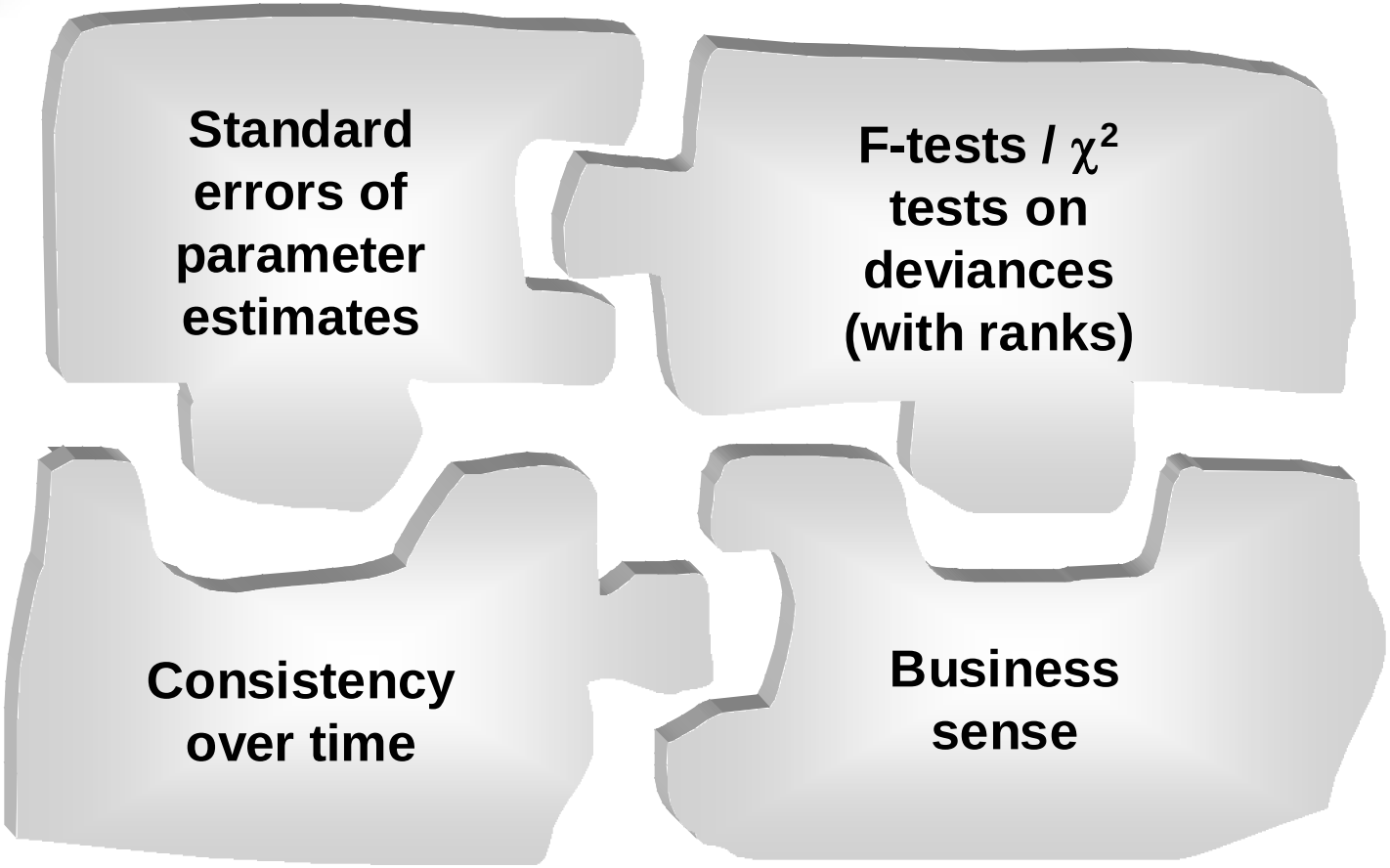
# Non-convergence problems

- Caused by categorization
  - insufficient volume in categorized levels
  - near perfect correlation between variables
- Working with large # of highly correlated factors
  - apply principal components analysis first
  - test families of factors one at a time to find most predictive member (eg # of late pays in 60 days may be most predictive of "late pay" family)

# Censoring large losses

- Sensitivity test
  - super-impose graphical results of same severity model with various large loss thresholds
  - identify where relativities may become volatile
- Consider modeling separately the propensity to have a large claim (must have lots of large claims)

# Model iteration



# Counter-intuitive effects

- Customer behavior in selection (eg deductible, limits)
- Discounts (airbags, fire alarm, defensive driver)
- Traditional effect changes when new variables added (eg territory when add geodemographics)
- Significant effect on a claim type that seems unrelated

# Treatment of counter-intuitive effects

- Live with it
- Investigate with and without factor
- Restrict (offset) model and allow other correlated factors to compensate
- Make judgmental selections

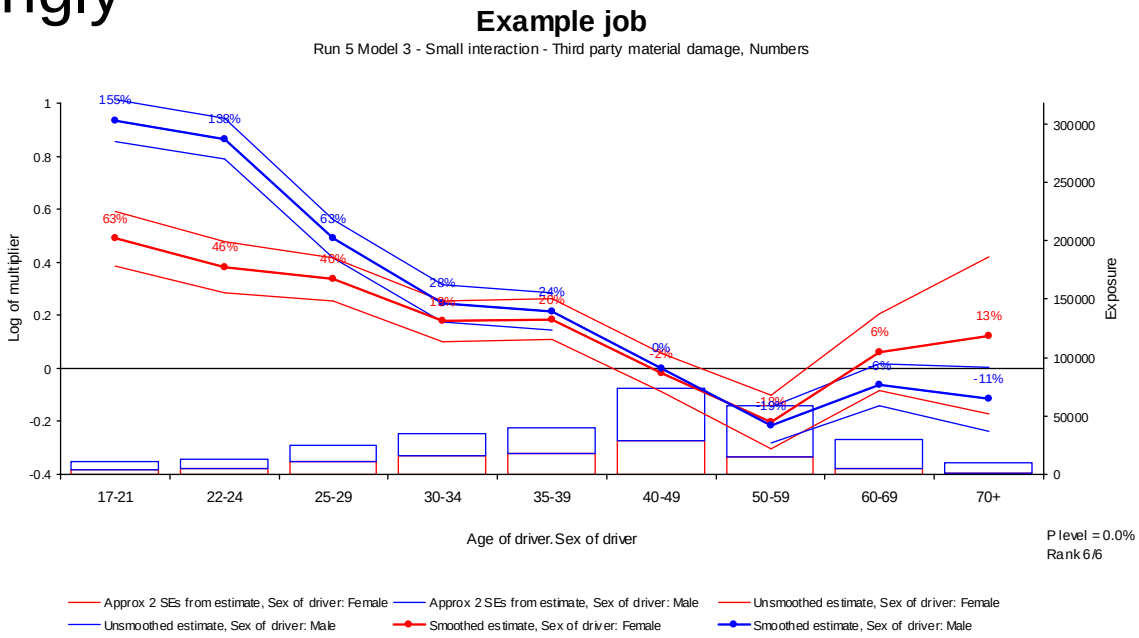
# Missing values

- Address before modeling
  - try to populate (eg can class code identify gender?)
  - remove if negligible
  - limit data to alleviate problem
  - re-categorize if near perfect correlation will cause convergence problems


- Model with and without factor to understand possible distortion on other correlated factors

# Interactions

- Factor effect that varies according to the levels of another factor (eg gender differential varies by age)
- Worth investigating but in practice not always obvious
- Consider volume per cell and re-categorize accordingly



# Interactions

| Group | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Factor | 0.54 | 0.65 | 0.73 | 0.85 | 0.92 | 0.96 | 1.00 | 1.08 | 1.19 | 1.26 | 1.36 | 1.43 | 1.56 |

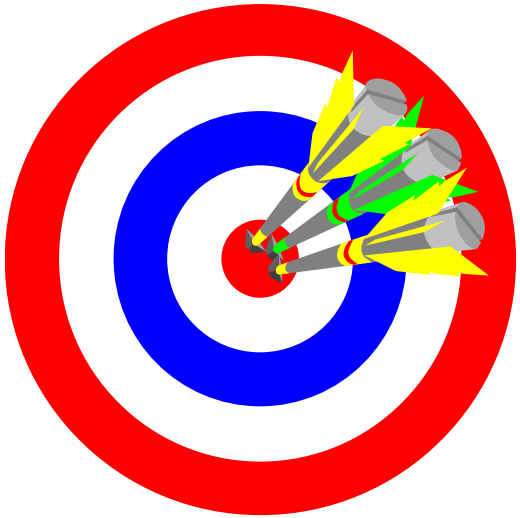
| Age | Factor |
|---|---|
| 17 | 2.52 |
| 18 | 2.05 |
| 19 | 1.97 |
| 20 | 1.85 |
| 21-23 | 1.75 |
| 24-26 | 1.54 |
| 27-30 | 1.42 |
| 31-35 | 1.20 |
| 36-40 | 1.00 |
| 41-45 | 0.93 |
| 46-50 | 0.84 |
| 51-60 | 0.76 |
| 60+ | 0.78 |

| | |
|---|---|
| 1.00 | 1.17 |
| 1.00 | 1.00 |

# Scoring in ratemaking

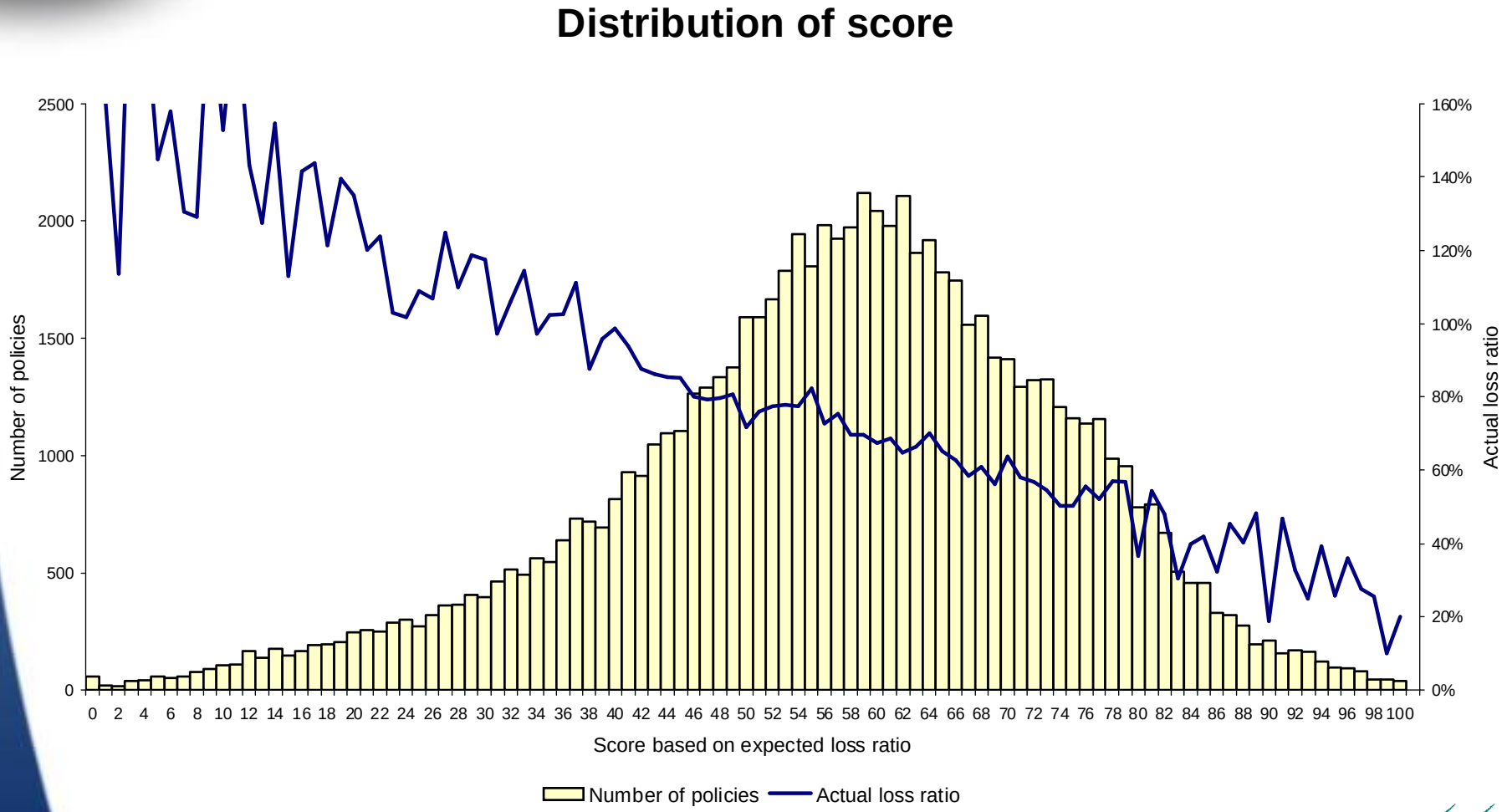
- Derive product of several (multiplicative) factors
- Remove log link function and transform to score range such as (0, 100), thereby preserving risk order
- Categorize this score and re-introduce to GLM

# Profitability scoring



- Construct profitability score based on expected loss ratio
  - risk premium model offset by current premiums
  - banded into discrete bands if desired
- Profitability score can be used to target sections of a portfolio

# Profitability scoring



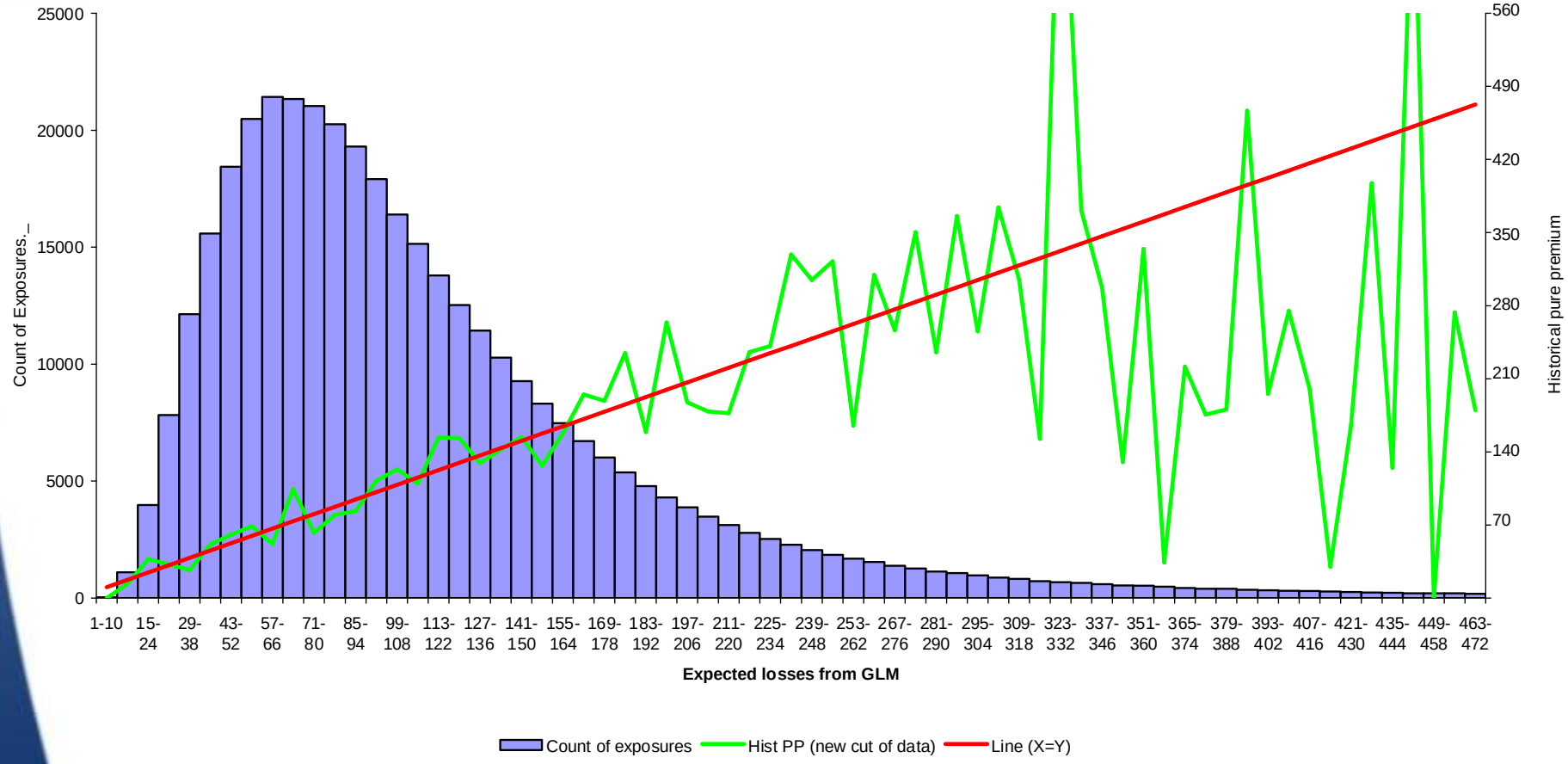
# If I knew then what I know now….

- Define application; link to success
- Prepare the culture
- Know thy data
- Fit models and interpret
- Translate models to decisions

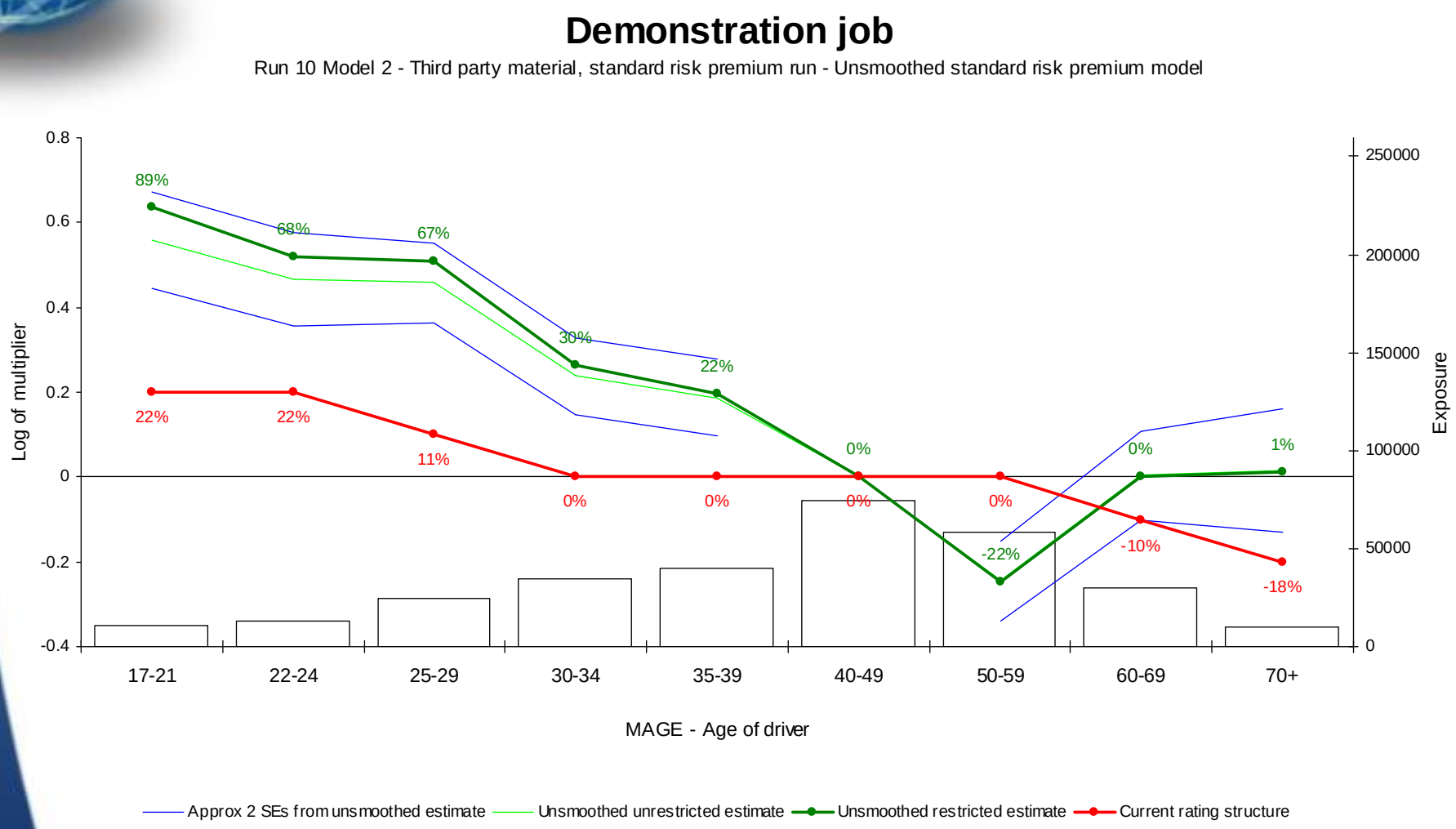
# Translate models to decisions

- Validate model
- Compare to current rates
- Consider competitor rates and profitability
- Consider retention/conversion and price optimization

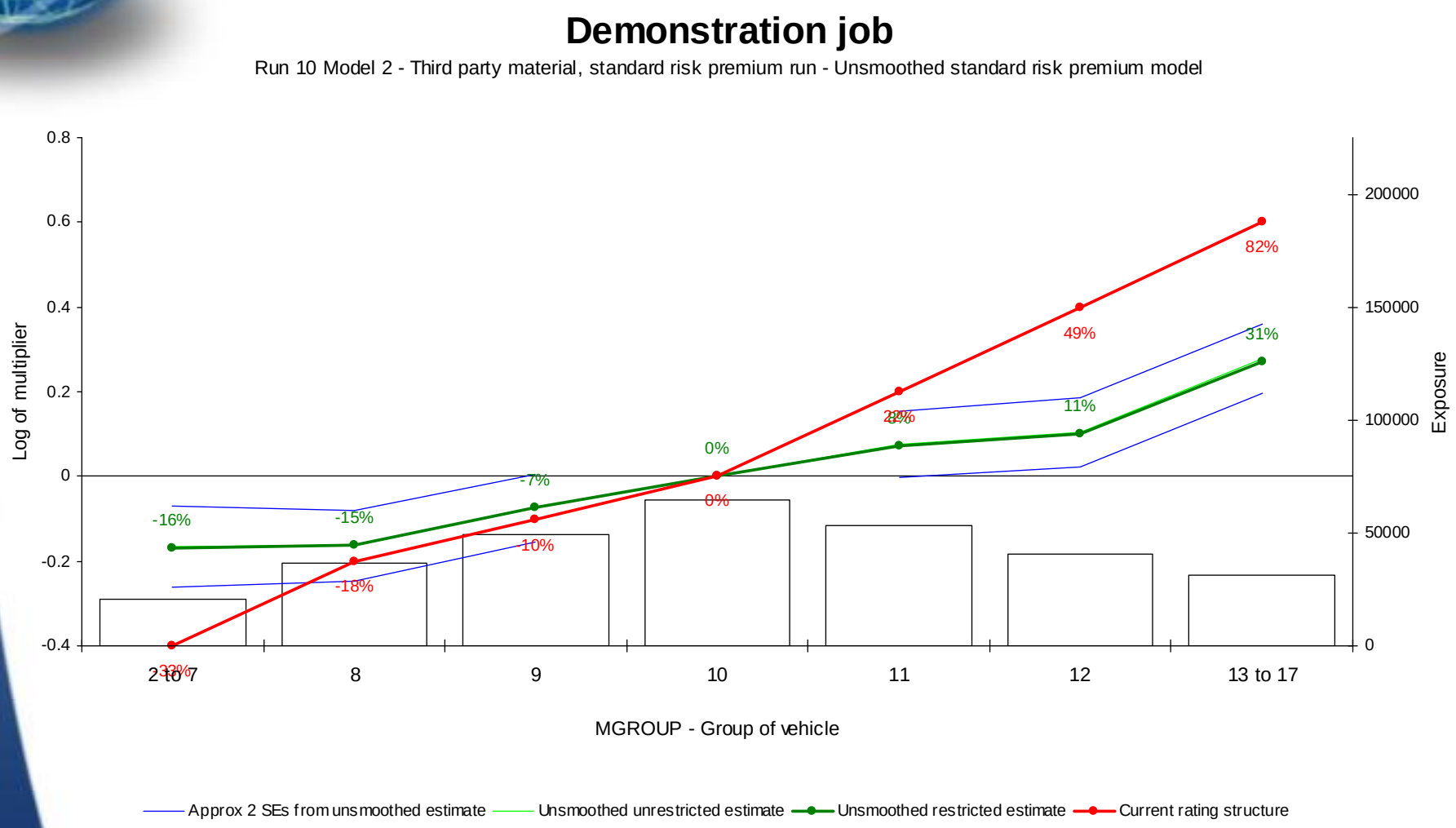
# Model validation



# Compare to current rates (by factor)



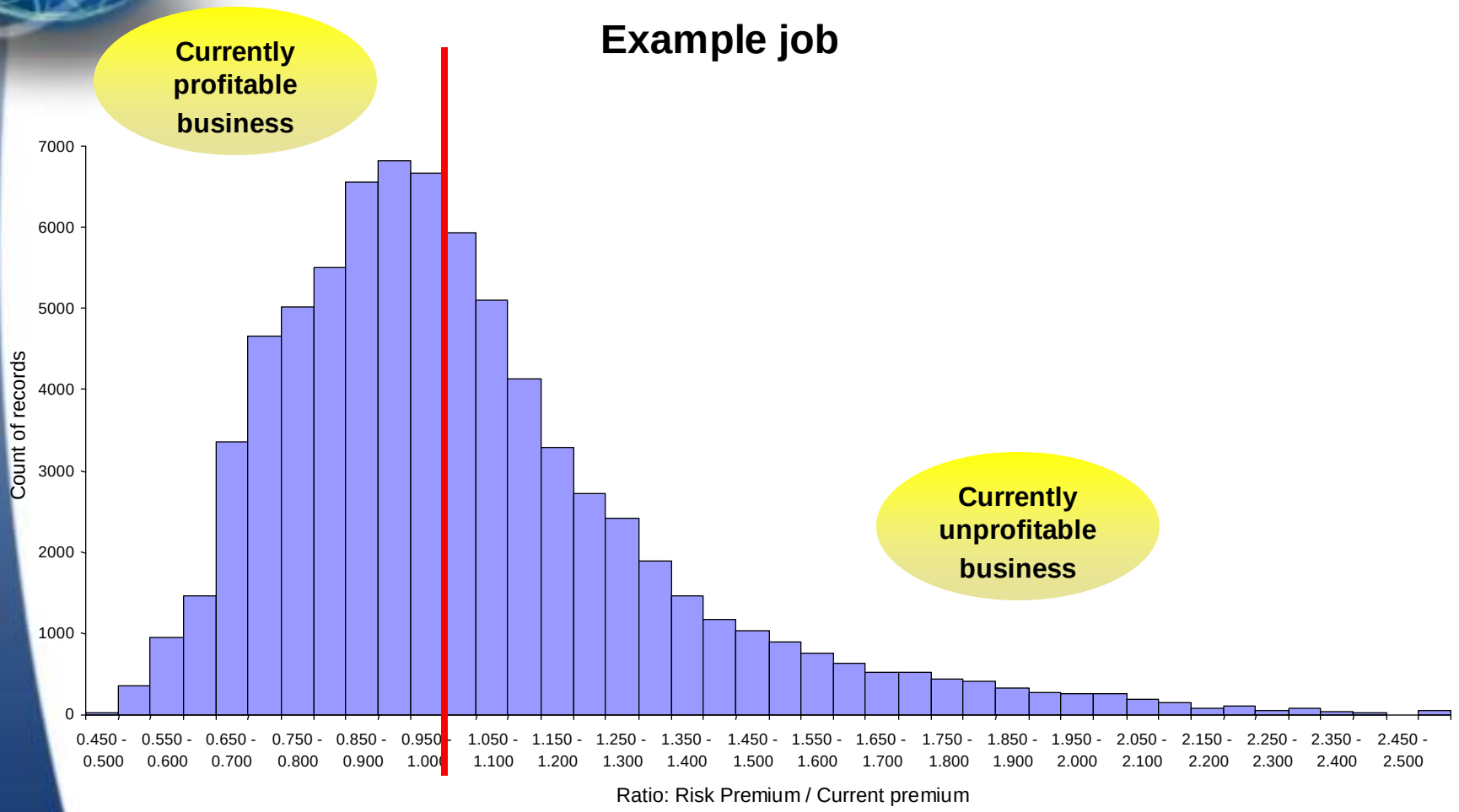
# Compare to current rates (by factor)



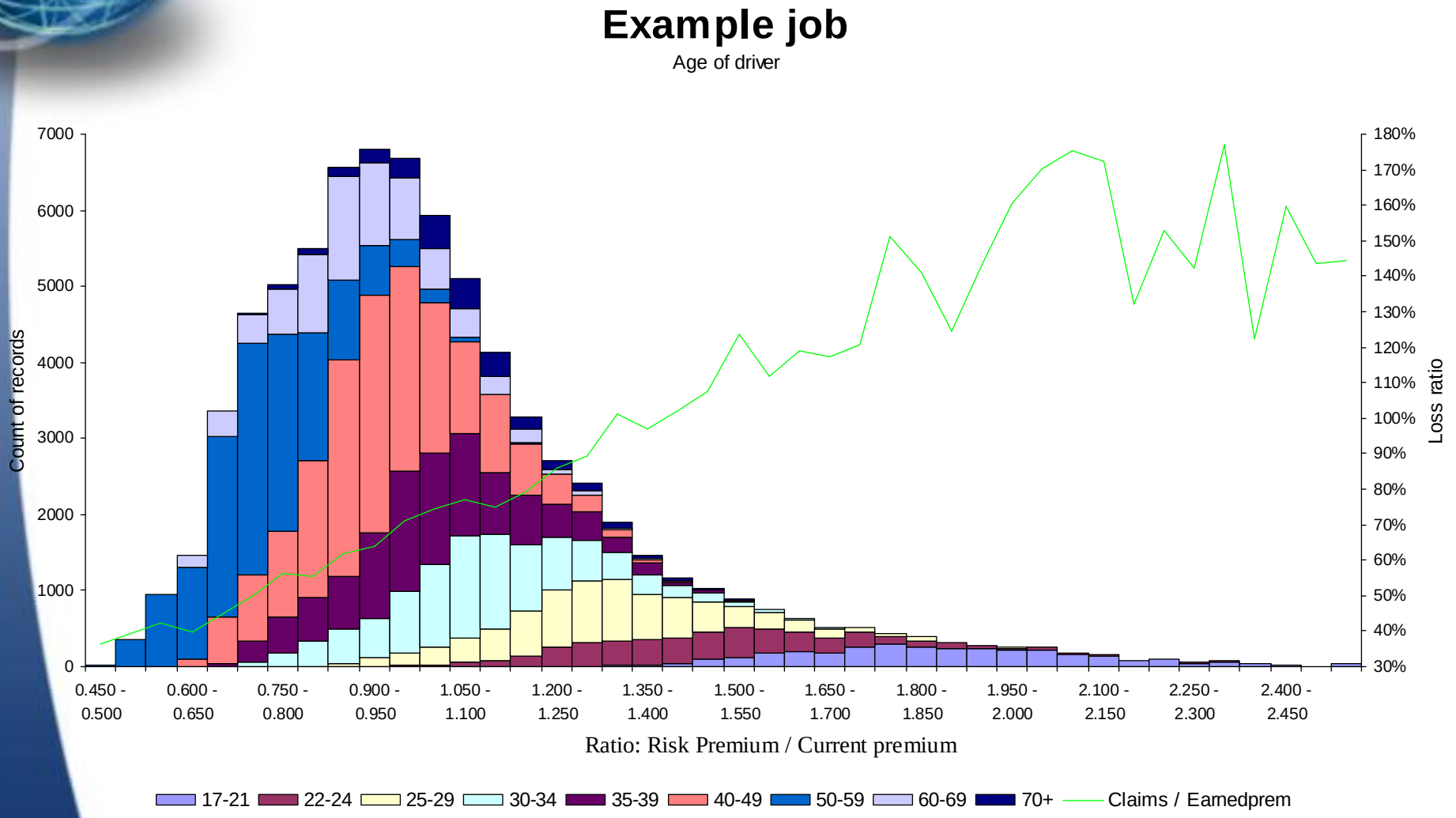
# Compare to current rates (by factor)

- Difficult if factors (or categorization of factors) differ greatly between current and theoretical model
- Difficult if current rates and theoretical model are not both multiplicative
  - can fit proxy model to non-multiplicative piece
  - can fit separate "sub-models", offset each subsequent model by previous model results and iterate

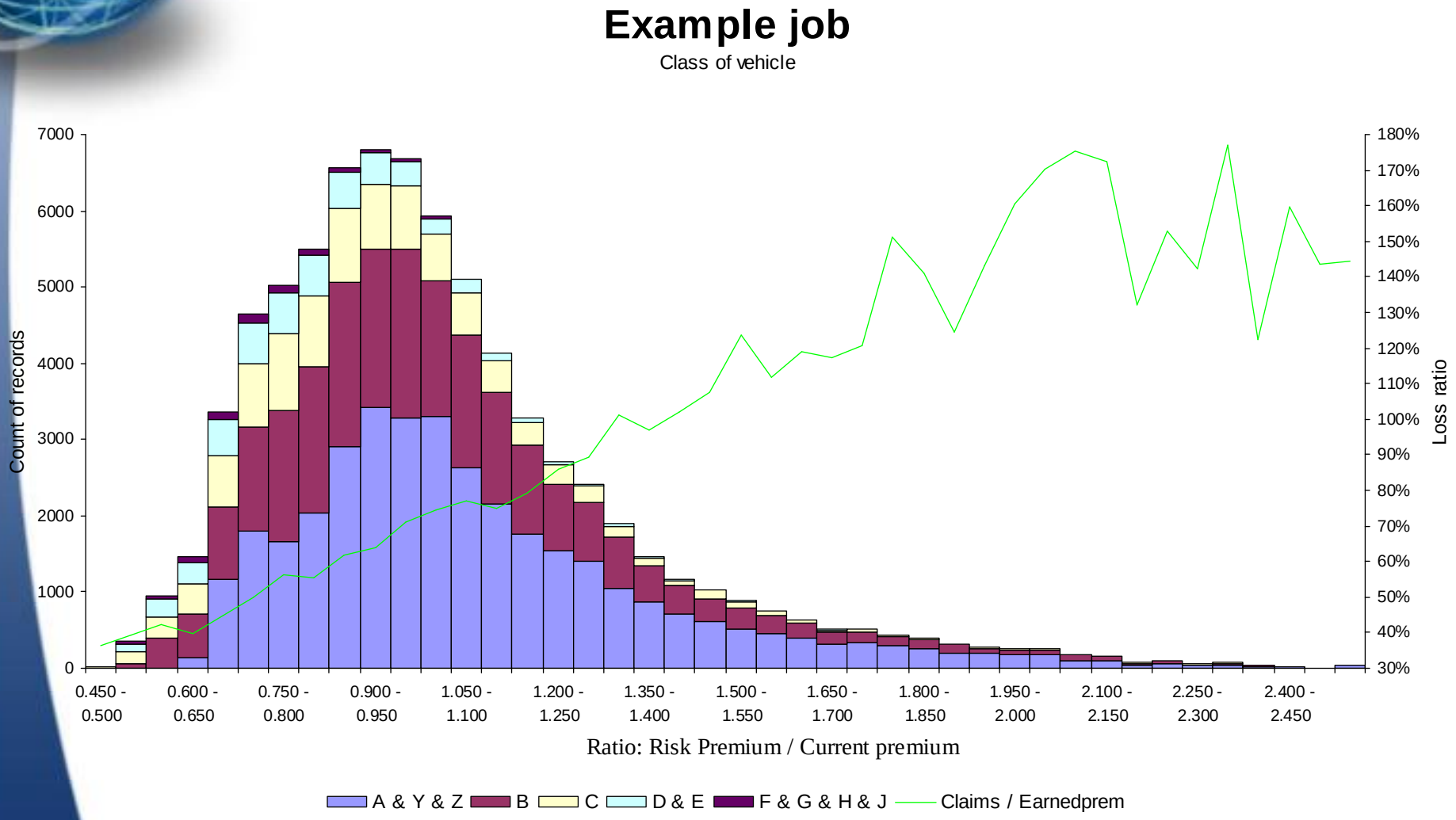
# Compare to current rates (in total)



# Compare to current rates (in total)

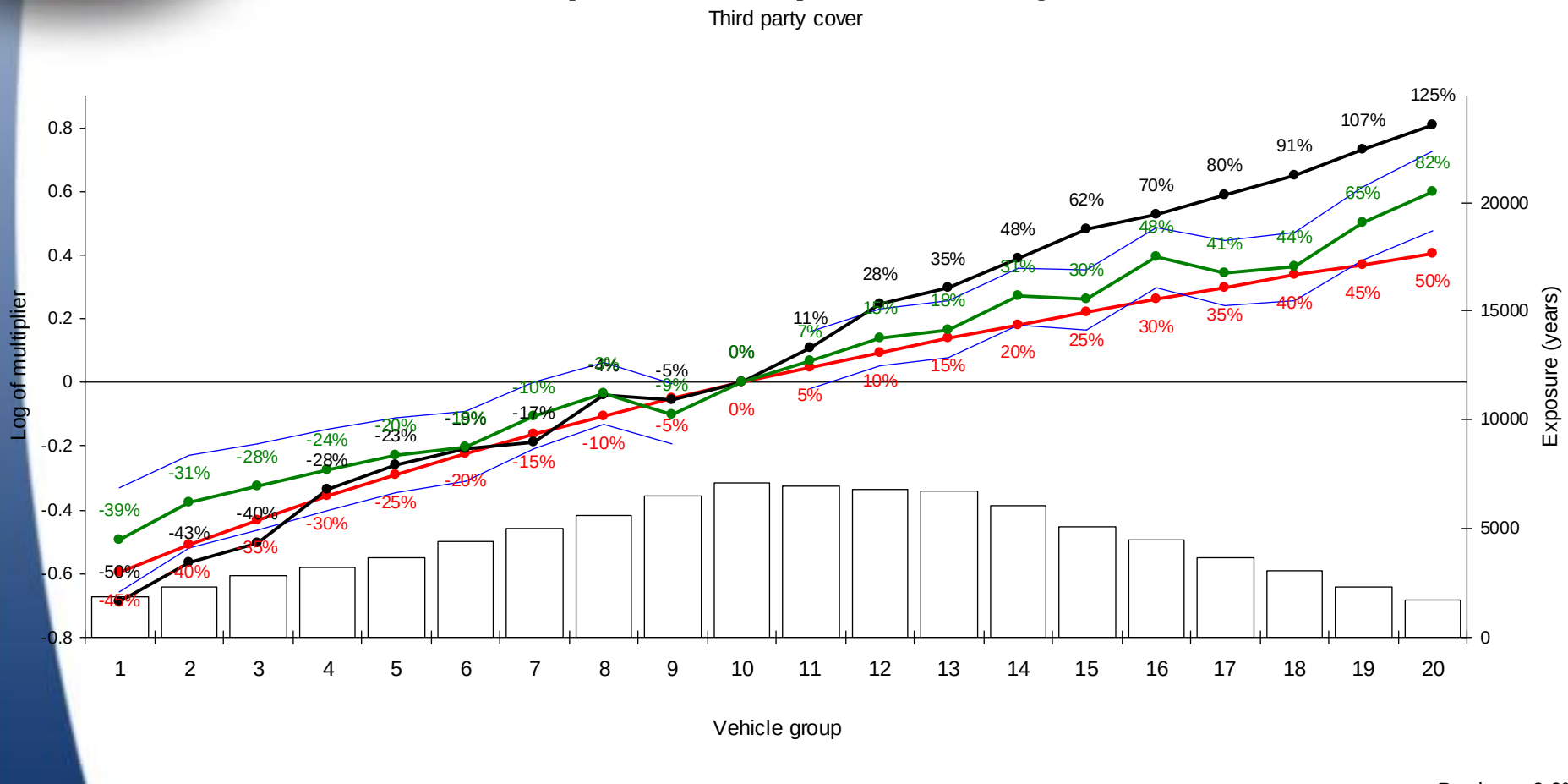


# Compare to current rates (in total)



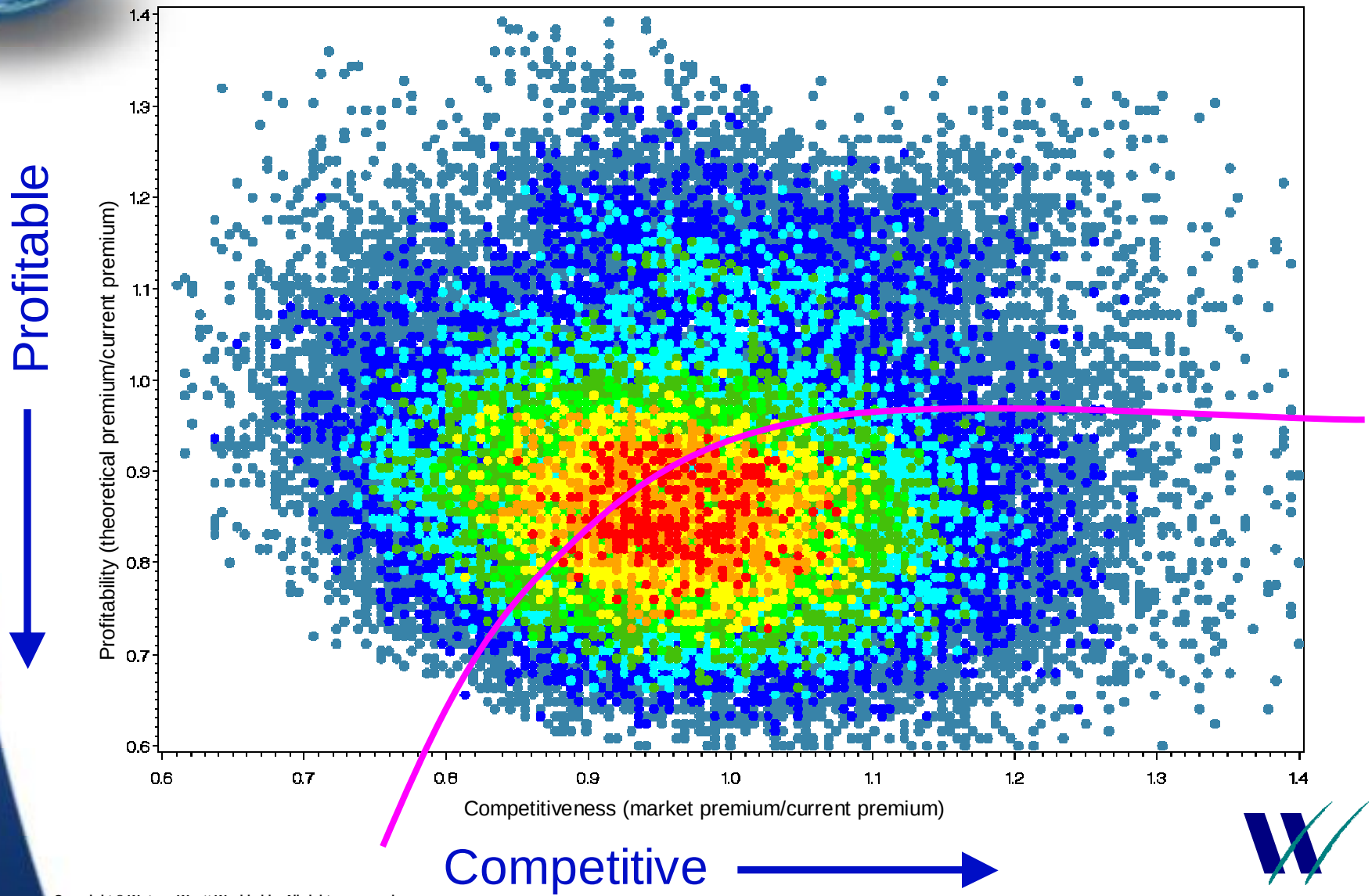
# Consider the competitive position

## Example of competitor analysis




P value = 0.0%
Rank 9/11

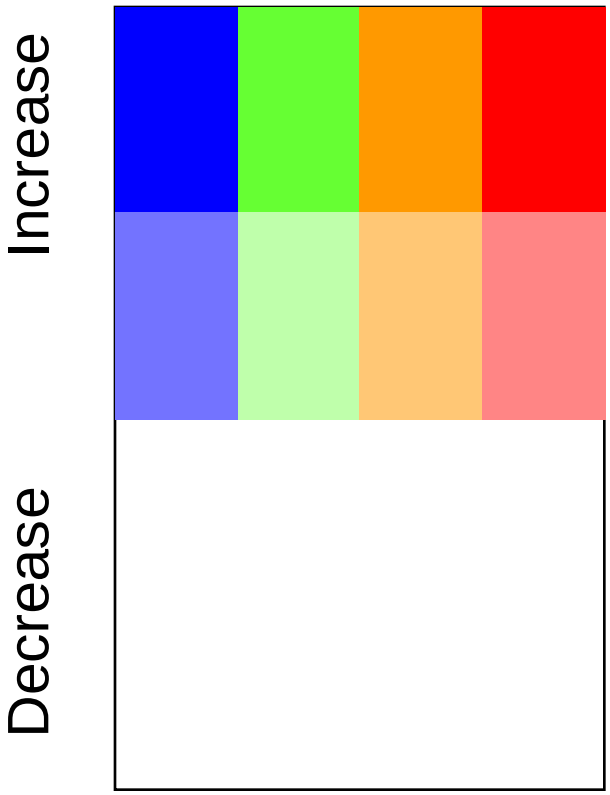
# Consider the competitive position

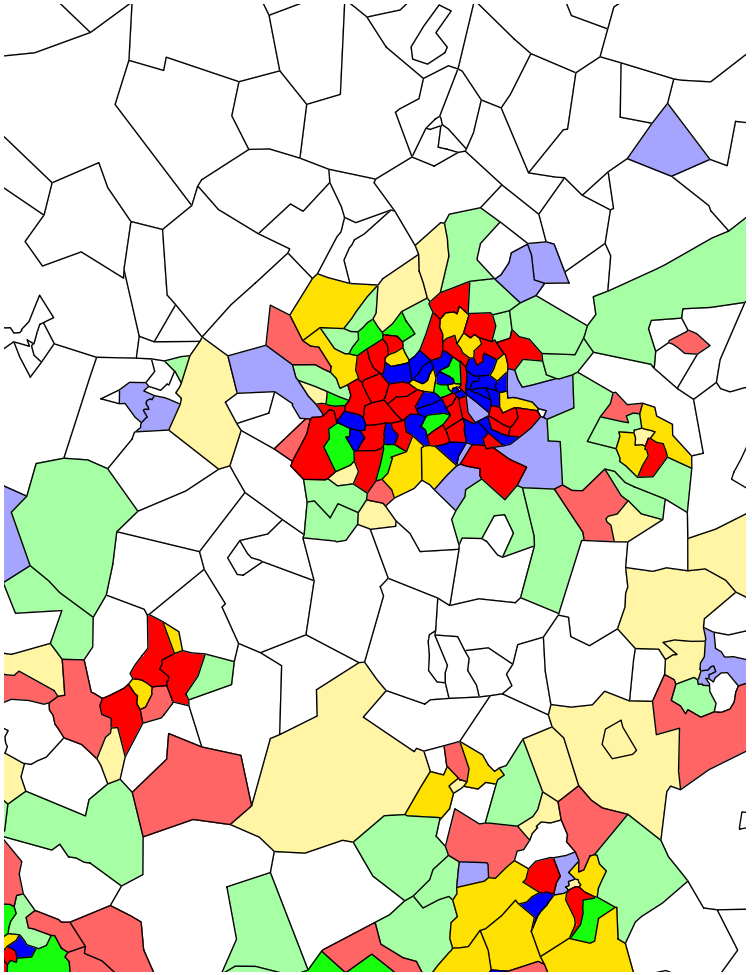


# Comparing model results with existing rates and the market

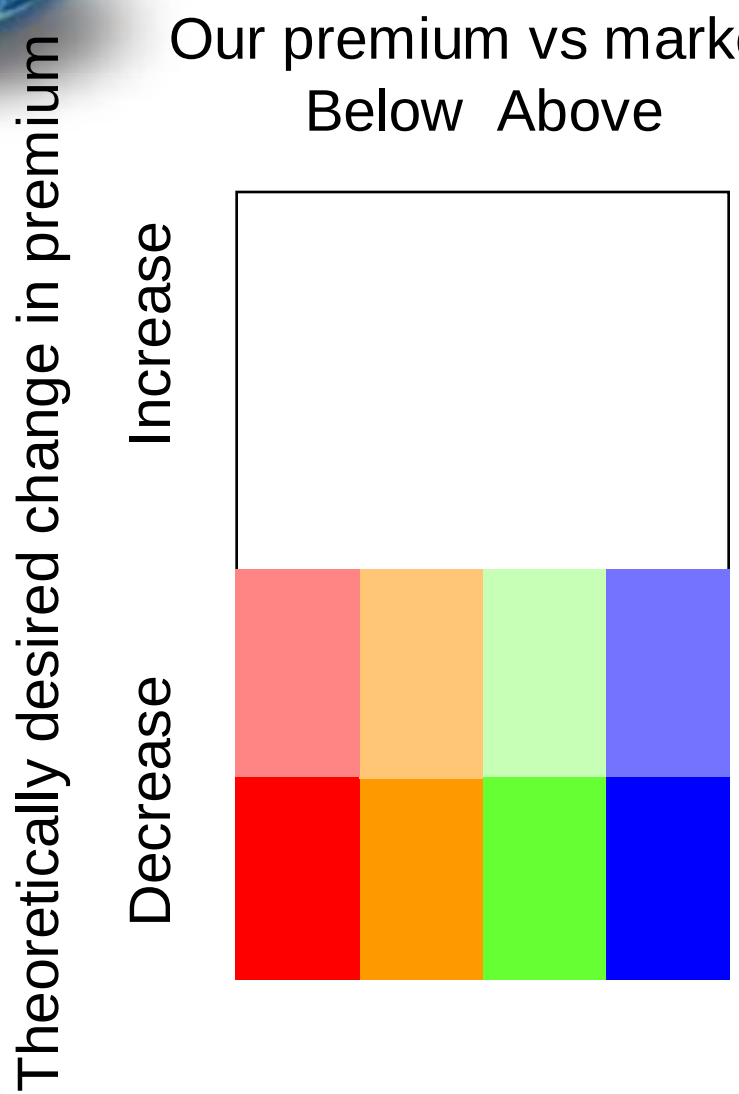
Our premium vs market

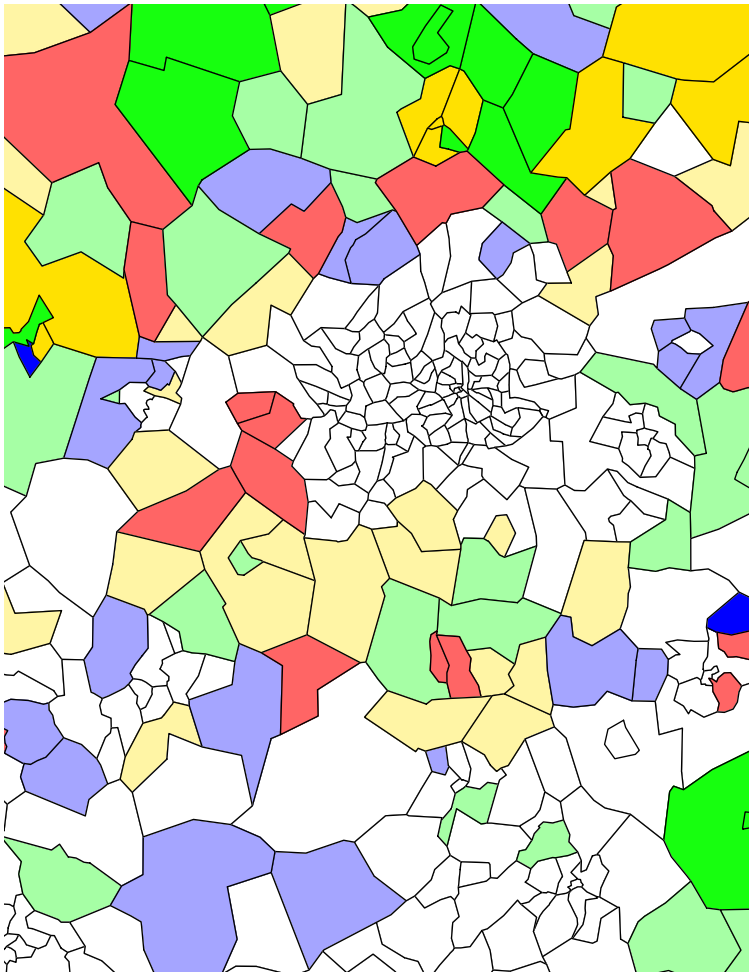
Theoretically desired change in premium



# Comparing model results with existing rates and the market

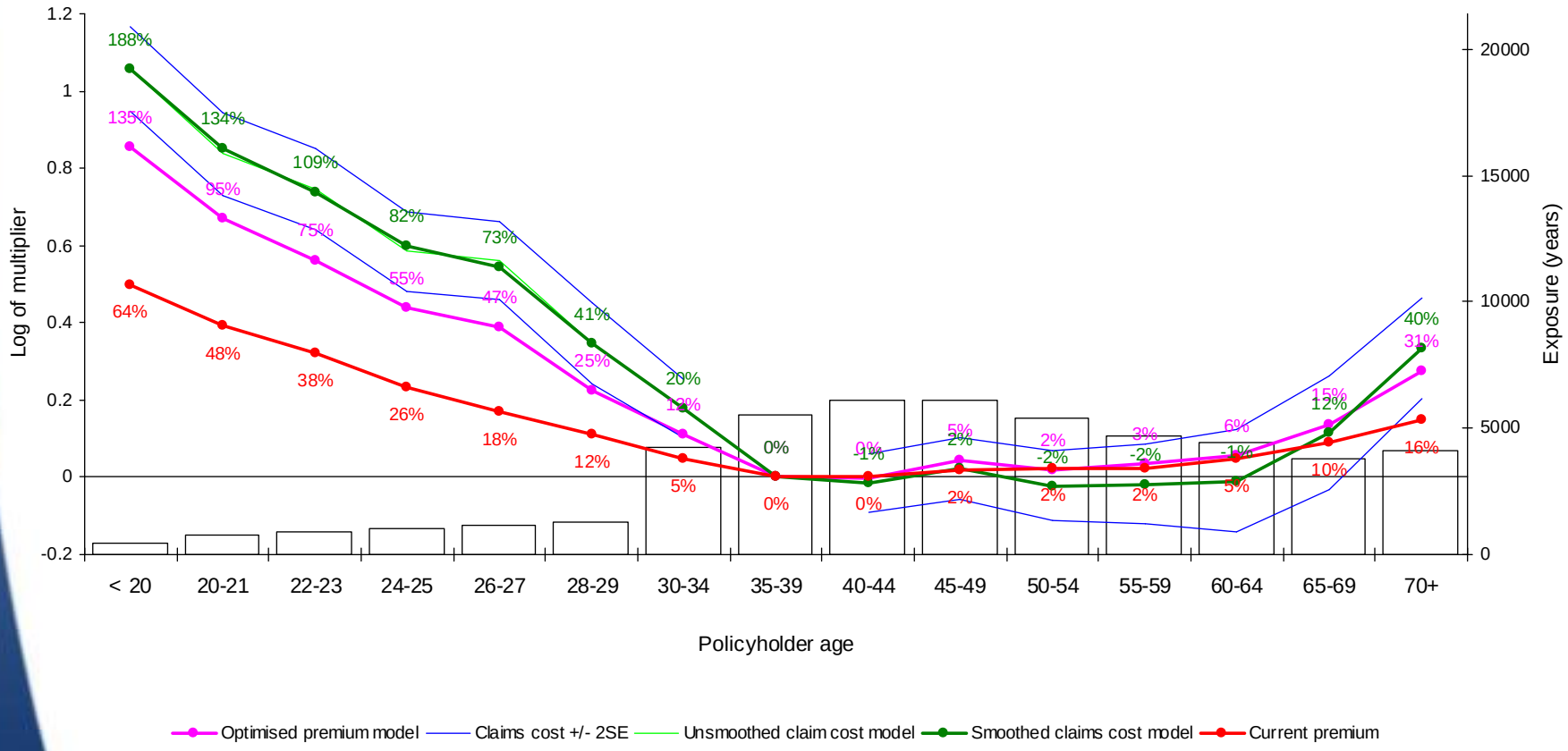


# Price optimization

## Optimized premium

Comparison with claims model and current premium



# PL-3
# GLM Practical Applications

## Appendix on data preparation for ratemaking

WWW.WATSONWYATT.COM

Watson Wyatt
*Worldwide*

# Link policy/claims information

- Identify sources of data (may be various formats)
- Develop plan to link/merge
- Consider volume of data
  - aim for at least 50K earned exposures (depends on # of variables to be analyzed)
  - combine multiple years, states
- Identify claim types to be modeled (sufficient claims volume in each?)

# Define record

- Record:  a risk for a policy period or portion of policy period for which risk has not changed
  - risk:  item, policy, account?
  - method of organization (calendar accident yr, policy yr)
  - treatment of mid-term changes and cancellations
  - treatment of multiple claims for an exposure period

# Gather statistics by claim type

- Key statistics
  - earned exposure
  - incurred claim count
  - incurred losses at most recent reserve estimate (with lag for IBNR)
- Preferred: premium @ current rate level
- Other considerations: ALAE? developed losses? how riders are considered in both premium and loss?

# Gather explanatory variables

- Characteristics at time risk was exposed, categorized according to today's definitions
- Dummy variables (eg time, geography)
- Considerations
  - consistent coding
  - integrity in data collection
  - sufficient claims volume
  - process to procure and append external data

# Calendar-accident or Policy year?

| Policy # | Start | End | Pol year | Gender | Age | … |
|---|---|---|---|---|---|---|
| … | … | … | … | … | … | … |
| 16853165 | 01/21/03 | 01/20/04 | 2003 | M | 45 | … |
| 16853166 | 06/15/03 | 06/14/04 | 2003 | F | 37 | … |
| 16853167 | 11/19/03 | 11/18/04 | 2003 | F | 24 | … |
| … | … | … | … | … | … | … |

| Policy # | Start | End | Pol year | Cal year | Gender | Age | … |
|---|---|---|---|---|---|---|---|
| … | … | … | … | … | … | … | … |
| 16853165 | 01/21/03 | **12/31/03** | 2003 | **2003** | M | 45 | … |
| 16853165 | **01/01/04** | 01/20/04 | 2003 | **2004** | M | 45 | … |
| 16853166 | 06/15/03 | **12/31/03** | 2003 | **2003** | F | 37 | … |
| 16853166 | **01/01/04** | 06/14/04 | 2003 | **2004** | F | 37 | … |
| 16853167 | 11/19/03 | **12/31/03** | 2003 | **2003** | F | 24 | … |
| 16853167 | **01/01/04** | 11/18/04 | 2003 | **2004** | F | 24 | … |
| … | … | … | … | … | … | … | … |

# Other considerations

- Changes during experience period (u/w, legal)
- Splitting records into separate datasets
  - randomly for model validation
  - for modeling separately (eg owners/renters)

# Data cleaning

- Missing values
- Negative or zero statistics
- Illogical combinations
  - claim count with no loss
  - loss with no claim count
  - claim count with no exposure

# PL-3
# GLM Practical Applications

## 2006 CAS Seminar on Ratemaking

Claudine Modlin, FCAS

Watson Wyatt Worldwide

WWW.WATSONWYATT.COM

Watson Wyatt Worldwide