Modeling Lessons From Across the Pond - Insights from Decades of Deregulated Modeling

Duncan Anderson MA FIA
Watson Wyatt LLP

WWW.WATSONWYATT.COM

Watson Wyatt Worldwide

Agenda

- External data
- Geographical spatial analysis
- Product features of modeling interest
- Retention modeling / price optimization

Agenda

- External data
  - geodemographic data
  - geophysical data
  - vehicle data
  - banking data
- Geographical spatial analysis
- Product features of modeling interest
- Retention modeling / price optimization

Geodemographic data

- Often designed for selling detergent
- Attaches to post code / zip code therefore easy to use at point of sale
  - can be hidden in a territory definition
- Marketing segment types often not predictive
- Underlying data often more interesting
- Simple measure of urban density often predictive

Example of effect of urban density on EU country #1 homeowners theft frequency

Real GLM output cannot be disclosed in handouts

Graph shown in presentation showed strong multivariate effect of urban density

Effect of density varies - homeowners example from EU country #1

- Effect of increasing density on risk:

	Frequency	Severity
Theft	↑	↑
Fire	↓	↑
Acc. Dam.	↓	↑
"Weather"	↓	↑&↓
"Other"	↑	↑

Excludes some covers such as subsidence and flood

Example of urban density
Auto, EU country #2 - Theft frequency

Real GLM output cannot be disclosed in handouts

Graph shown in presentation showed strong multivariate effect of urban density

Example of urban density

Auto, EU country #2 - Property Damage frequency

Real GLM output cannot be disclosed in handouts

Graph shown in presentation showed multivariate effect of urban density which is significant but weaker than the effect for auto theft claims

Geodemographics can be rather related!

	R1	R2	R3	R4	G1	G2	G3	G4	G5	G6
R1										
R2	11%									
R3	32%	3%								
R4	17%	7%	58%							
G1	8%	2%	57%	16%						
G2	8%	2%	53%	15%	49%					
G3	7%	3%	44%	14%	33%	33%				
G4	5%	4%	21%	8%	30%	30%	30%			
G5	3%	2%	31%	6%	36%	35%	34%	31%		
G6	8%	2%	65%	16%	37%	35%	31%	29%	34%	
G7	8%	2%	65%	16%	36%	34%	30%	30%	34%	71%

Cramer's V for a selection of standard rating factors (R1, .., R4) and geodemographic factors (G1, …, G4)

Example of geodemographic factors
Homeowners, EU country #3

Real GLM output cannot be disclosed in handouts

Graph shown in presentation showed strong multivariate effect of a geodemographic factor related to average life-stage of an area

Example of geodemographic factors
Homeowners, EU country #3

Real GLM output cannot be disclosed in handouts

Graph shown in presentation showed strong multivariate effect of another geodemographic factor

Example of geodemographic factors
Homeowners, EU country #3

Real GLM output cannot be disclosed in handouts

Graph shown in presentation showed strong multivariate effect of factor related to average type of building in the area

Agenda

- External data
  - geodemographic data
  - geophysical data
  - vehicle data
  - banking data
- Geographical spatial analysis
- Product features of modeling interest
- Retention modeling / price optimization

Geophysical data

- Available in many countries
- Particularly helpful for elements of claim for which own claims experience is not credible / predictive, including
  - flood
  - subsidence
  - storm
  - etc
- Norwich Union mapping
- Examples of data available in UK:
  - Flood risk
  - Soil type / subsidence risk
  - Windstorm risk
  - Frost risk
  - Theft risk (police data / geodemographic data)
  - Earthquake risk
  - Average building type (susceptibility to winds etc)

Examples of geophysical data EU country #3 - homeowners

Real GLM output cannot be disclosed in handouts

Graph shown in presentation showed strong multivariate effect of a weather related geophysical data item

Examples of geophysical data EU country #3 - homeowners

Real GLM output cannot be disclosed in handouts

Graph shown in presentation showed multivariate effect of another weather related geophysical data item

Agenda

- External data
  - geodemographic data
  - geophysical data
  - vehicle data
  - banking data
- Geographical spatial analysis
- Product features of modeling interest
- Retention modeling / price optimization

Vehicle data

- Links to license/registration plate available in some EU countries, particularly common in UK
- Benefits include
  - faster quotation process
  - accurate factors
  - more factors
- In UK, DVLA also holds personal data which is not disclosed
- Example data includes
  - exact make, model, type
  - engine size
  - power
  - max speed / acceleration
  - weight
  - number of doors
  - color
- Norwich Union Pay-as-you-Drive trial

Agenda

- External data
  - geodemographic data
  - geophysical data
  - vehicle data
  - banking data
- Geographical spatial analysis
- Product features of modeling interest
- Retention modeling / price optimization

Banking data

- Highly predictive of insurance claims experience
- Generally not used at point of sale in EU
- Many EU banks distribute insurance or own insurance companies
- If data collected with the correct box ticked, can be used for selective marketing

Example of predictive power of credit score (one-way of loss ratio)

Real GLM output cannot be disclosed in handouts

Graph shown in presentation showed strong effect of credit score on insurance claims experience

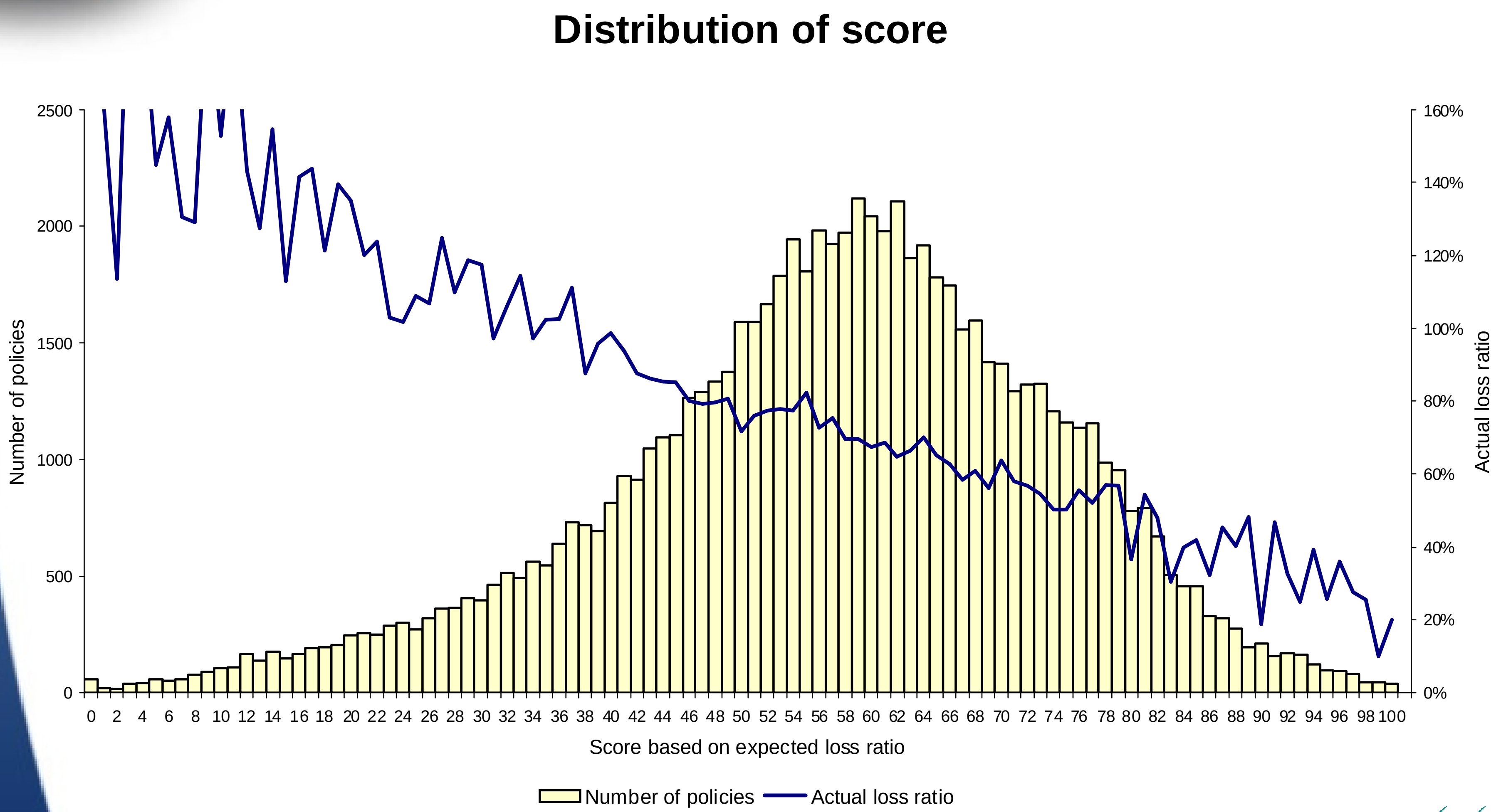
Banking data

- Credit score predictive, but so are lower level elements such as
  - average balance
  - whether or not in debt
- In one example over 12 banking factors were predictive in a GLM, some with effects of 1.5x or more
- Eg: strong increase in claims when policyholder in arrears on mortgage payments
  - suggests fraud element to the effect?

Case study: EU bank

- Distributed insurance products underwritten by a partner insurance company
- Bank could not easily change rates but shared in insurance profit
- Insurance penetration of banking customer base relatively small
- Bank derived profitability score based on banking factors which were not available at point of sale
- Score then used to market insurance selectively to banking customers

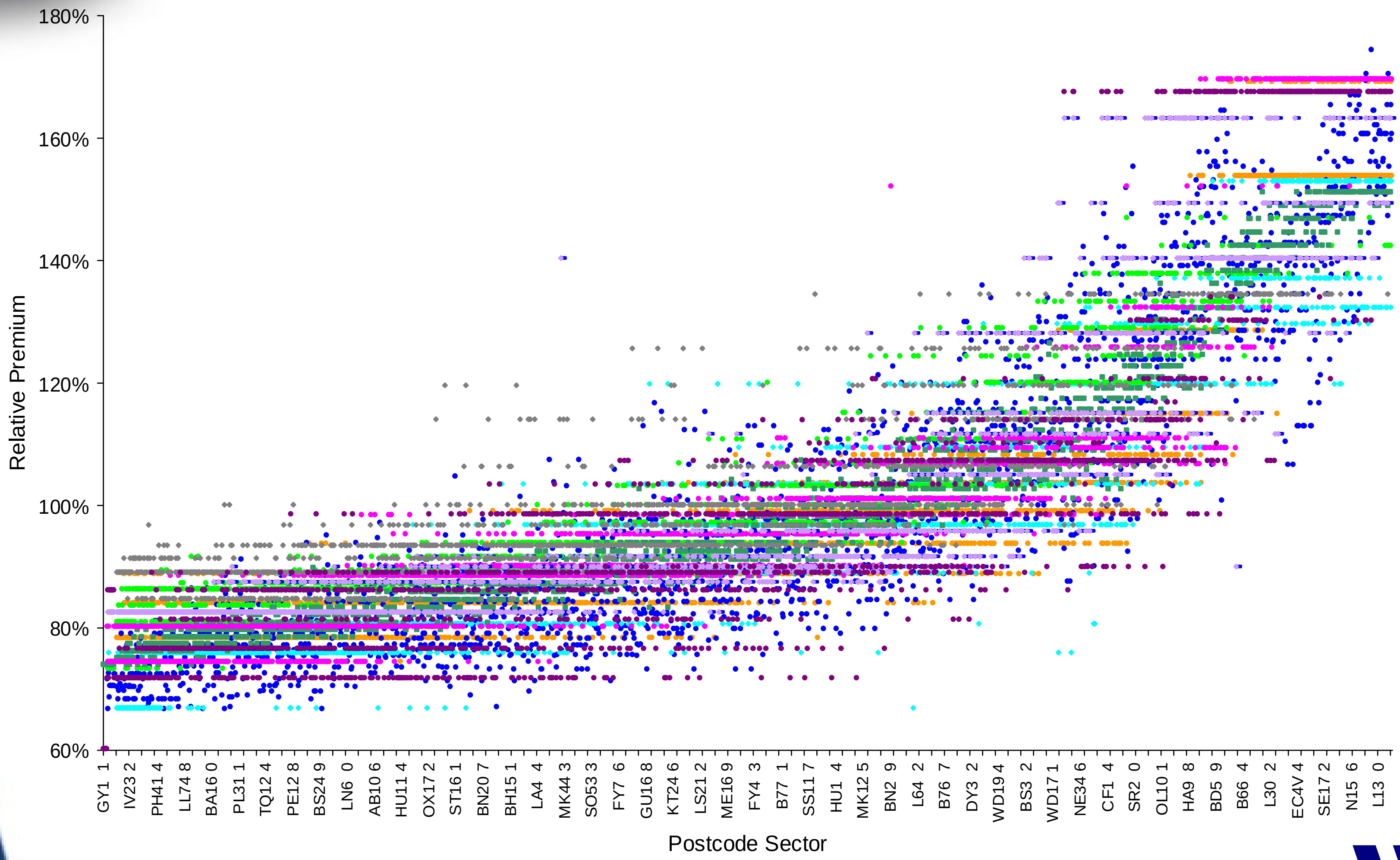
Profitability scoring



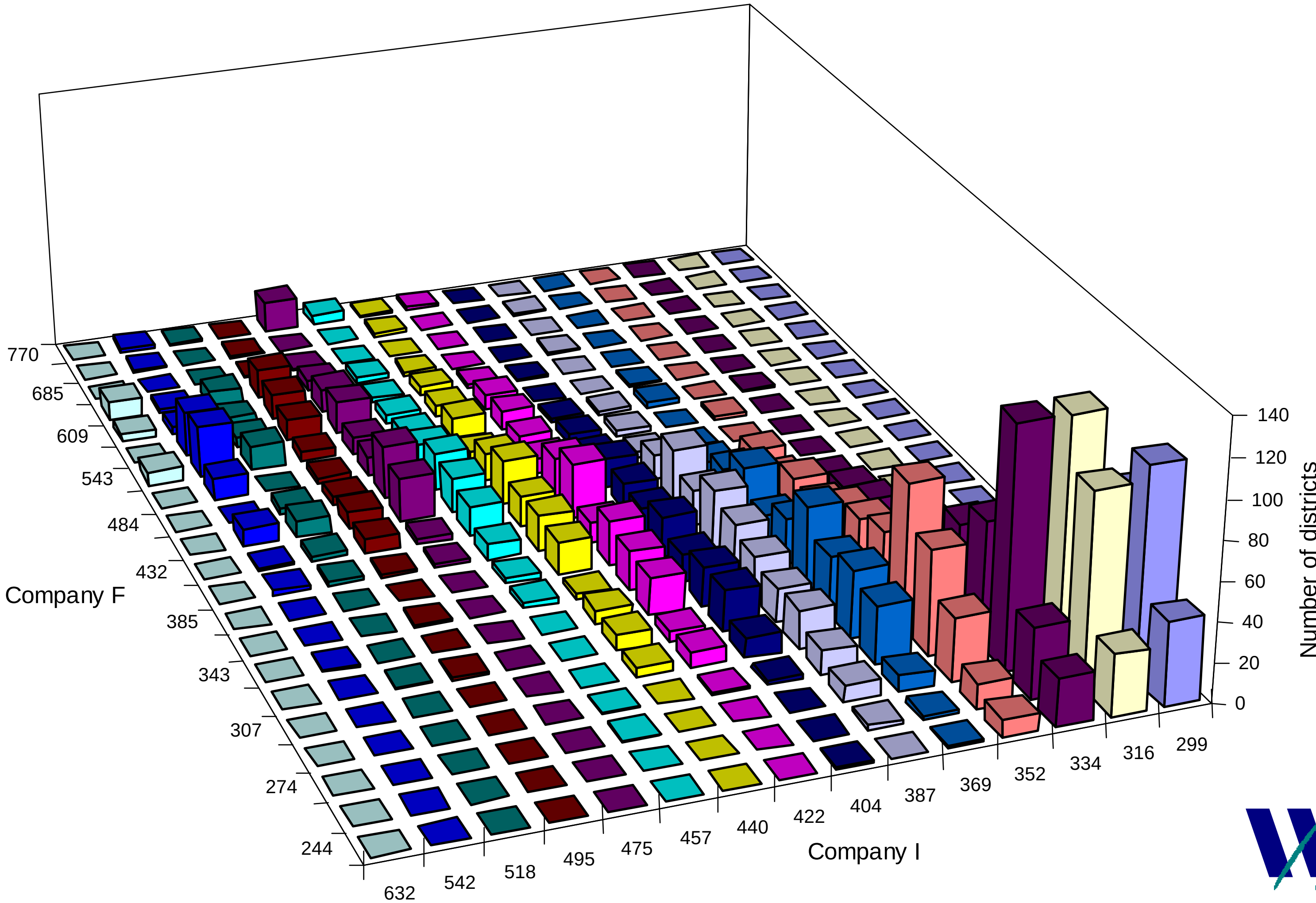
Agenda

- External data
- Geographical spatial analysis
- Product features of modeling interest
- Retention modeling / price optimization

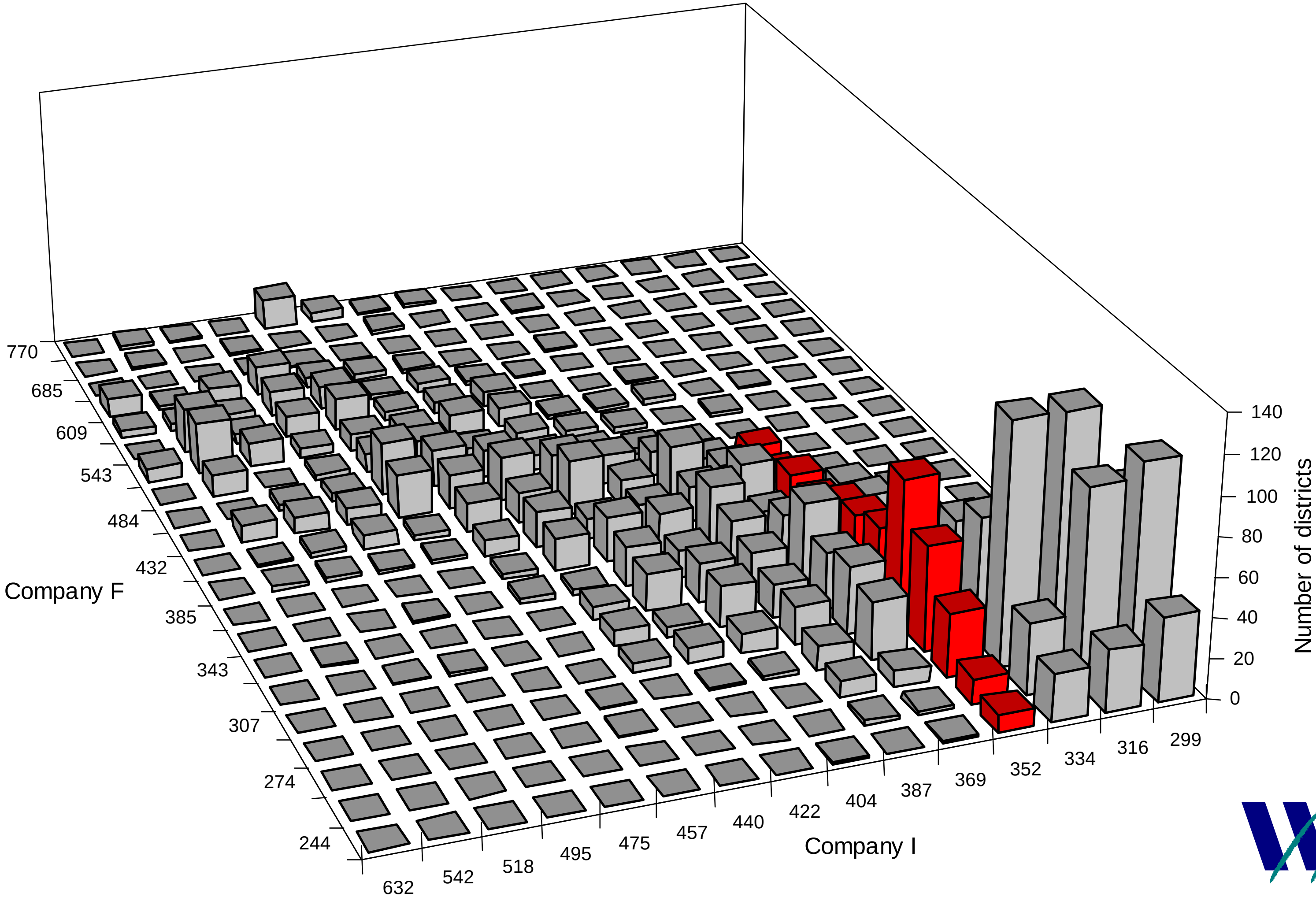
UK auto postcode loadings



Comparison of UK auto categorizations

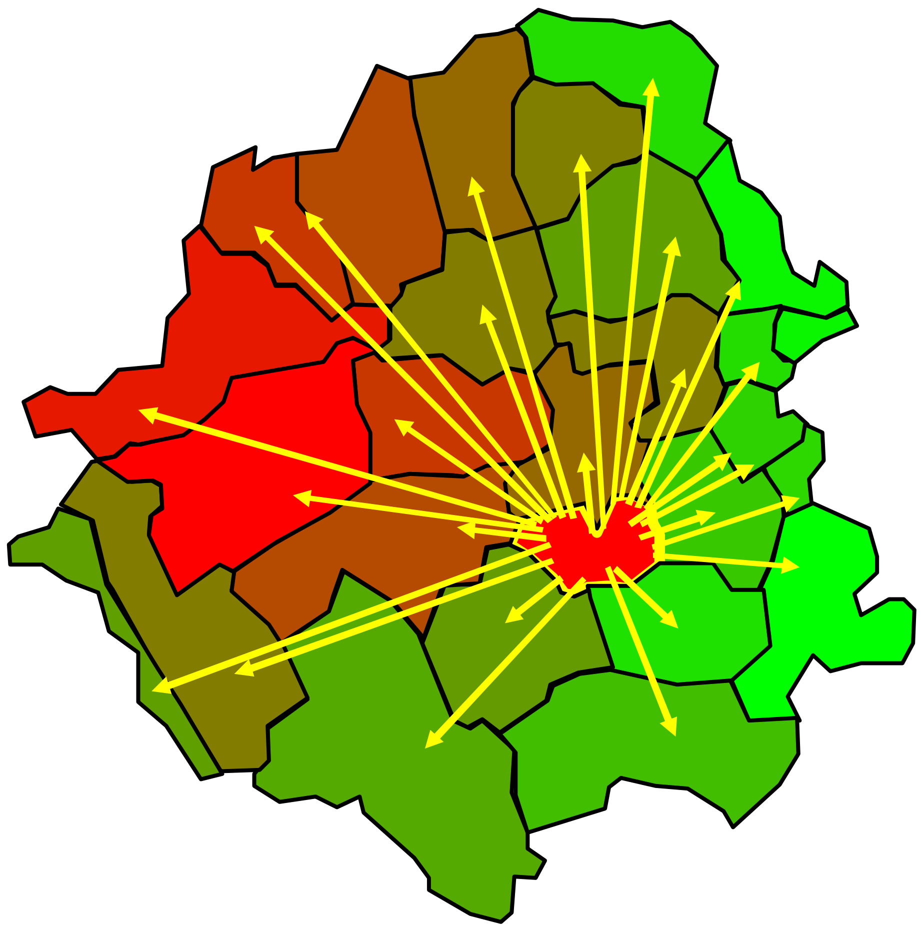


Comparison of UK auto categorizations



Spatial smoothing

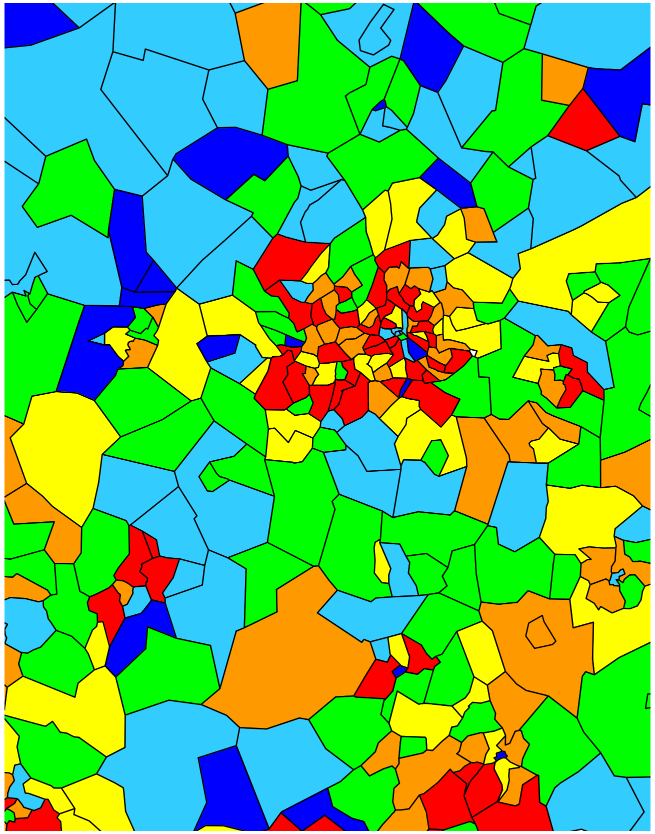
- Blends experience of one region with that of surrounding regions according to distance and credibility
- Credibility and spatial smoothing parameters are trained on actual data

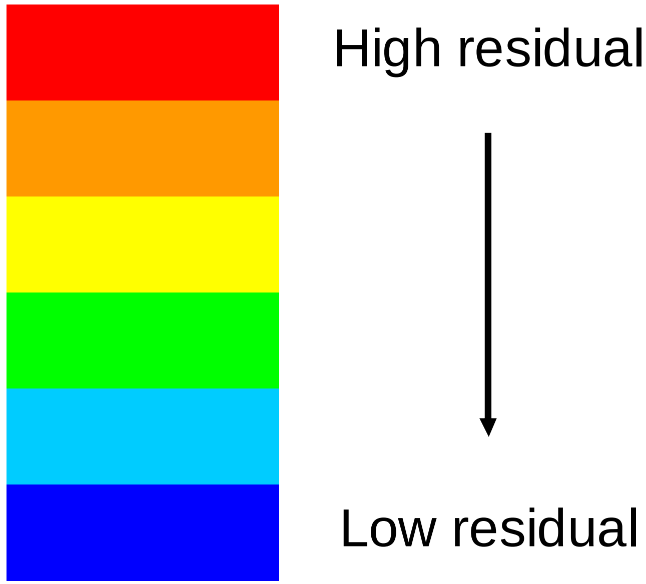


An approach

- More details of this approach available at
  www.casact.org/coneduc/ratesem/2004/handouts/anderson.pdf
- Standardize for other factors by fitting a GLM (excluding current zones)
- Consider "residual" risk by "region"
- Seek to make this residual risk more predictive
- Then categorize into zones to derive appropriate loadings

Residual risk





Model

$$r_i^* = Z(e_i).r_i + (1 - Z(e_i)) \sum_j e_j.r_j.f(d_{ij}) / \sum_j e_j.f(d_{ij})$$

where

r_i^*= smoothed residual r_i = unsmoothed residual

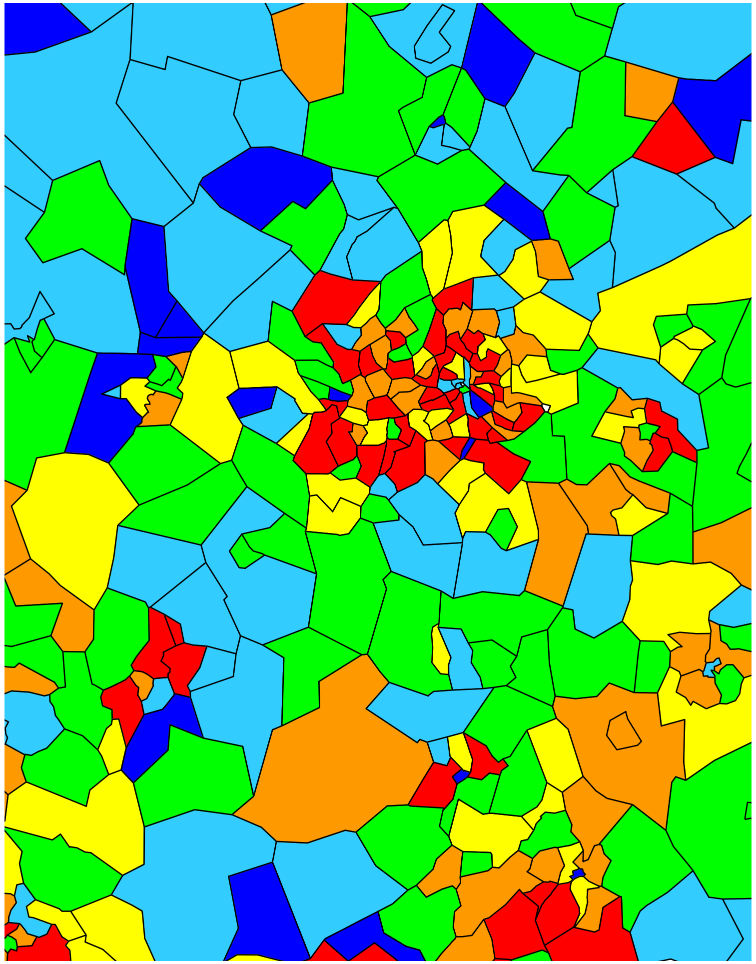
$Z(e_i) = \{ e_i / (e_i + a) \}^m$ e_i = exposure in region i

$$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 \}^{½}$$

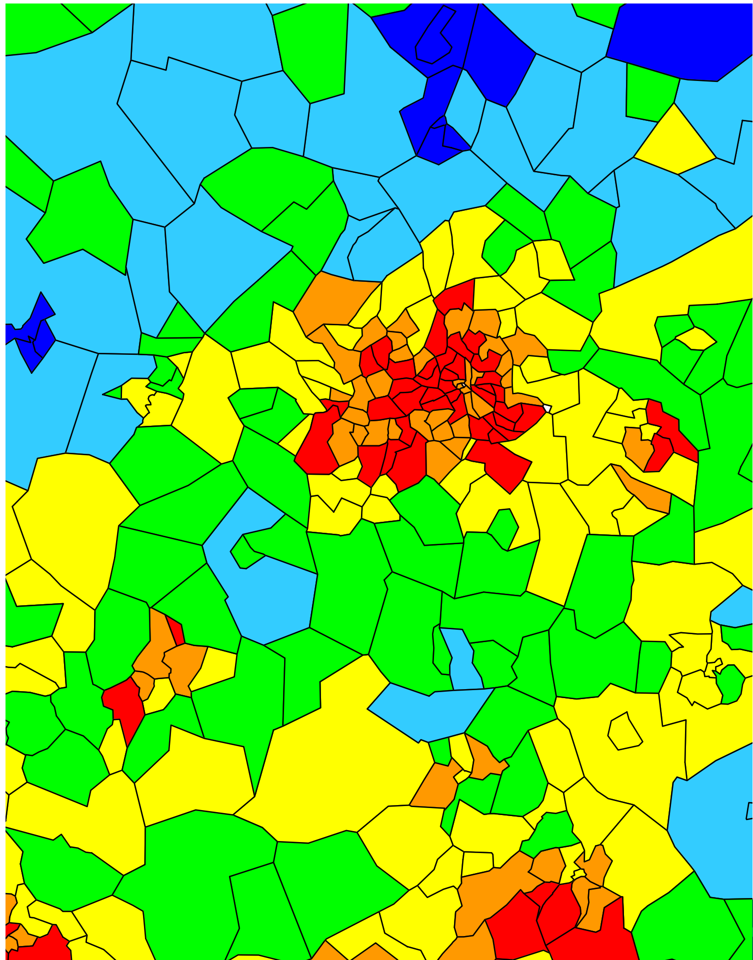
$f(d_{ij}) = 1/d_{ij}^n$ or $1/(d_{ij}^n + b^n)$ or $\exp(-n.d_{ij})$ etc

Example results

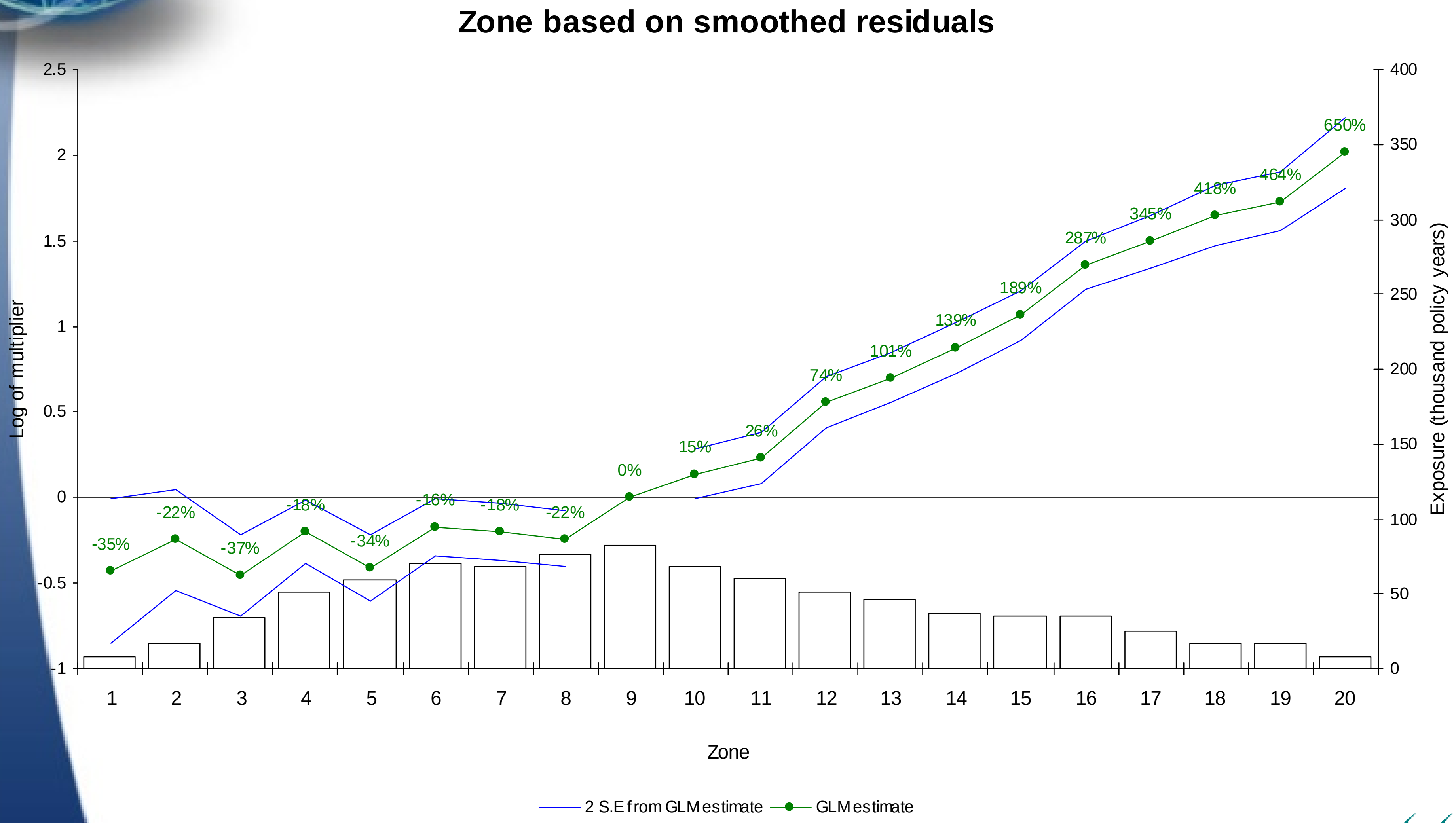
Unsmoothed residuals



Smoothed residuals

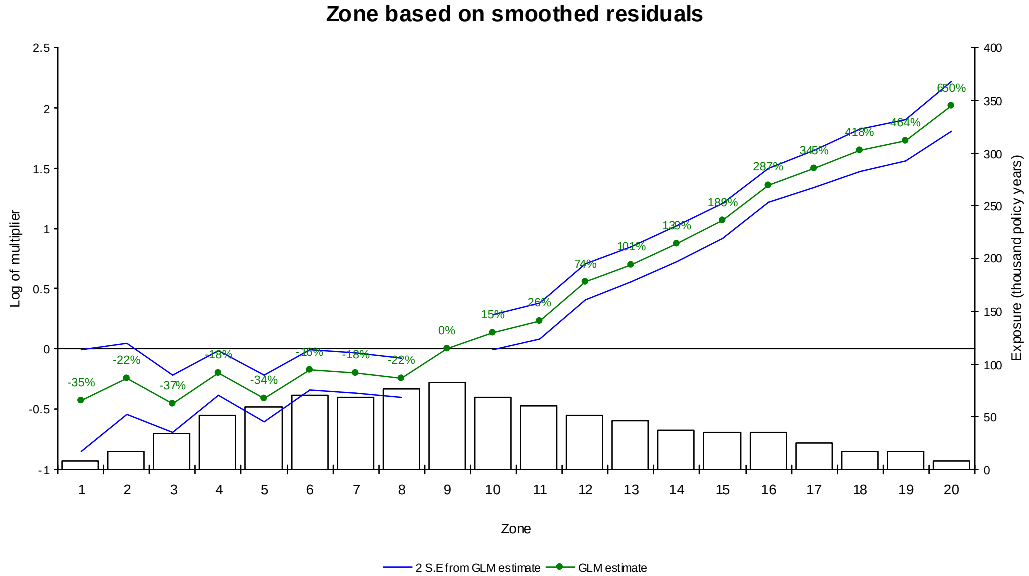


Predictive power of new zone on unseen data

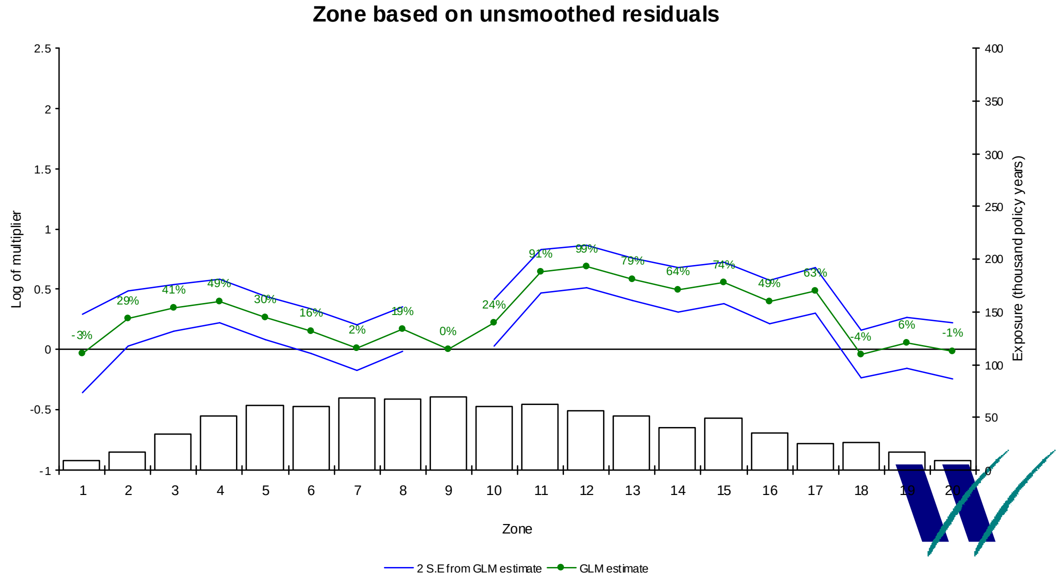


Predictive power of new zone on unseen data

Zone based on smoothed residuals



Zone based on unsmoothed residuals

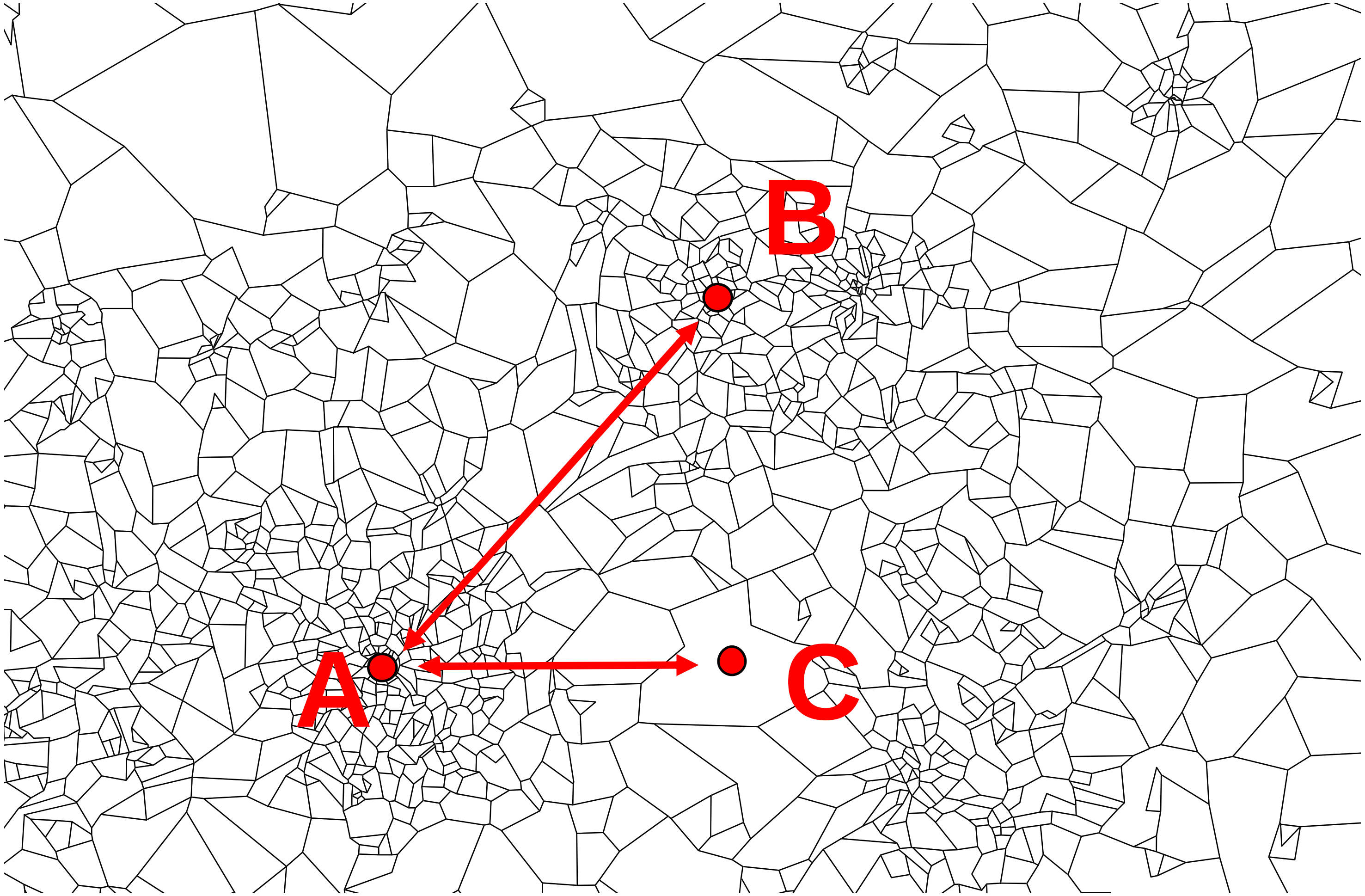


Parameters - international consistency $(f(d_{ij}) = 1/d_{ij}^n)$

		n	e for Z=20%
	UK	1.9	146
	UK	2.2	152
	UK	1.8	78
	France	2.0	104
	France	1.9	146
	Netherlands	1.8	61
	South Africa	2.2	106
	USA	2.5	127
	USA	1.9	106

Different metrics

Is A "closer" to B than C ?



Different metrics

$$r_i^* = Z(e_i).r_i + (1 - Z(e_i)) \sum_j e_j.r_j.f(d_{ij}) \,/\, \sum_j e_j.f(d_{ij})$$

where

r_i^*= smoothed residual r_i = unsmoothed residual

$Z(e_i) = \{ e_i / (e_i + a) \}^m$ e_i = exposure in region i

$$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 + (s.q_i - s.q_j)^2 \}^{½}$$

$f(d_{ij}) = 1/d_{ij}^n$ or $1/(d_{ij}^n + b^n)$ or $\exp(-n.d_{ij})$ etc

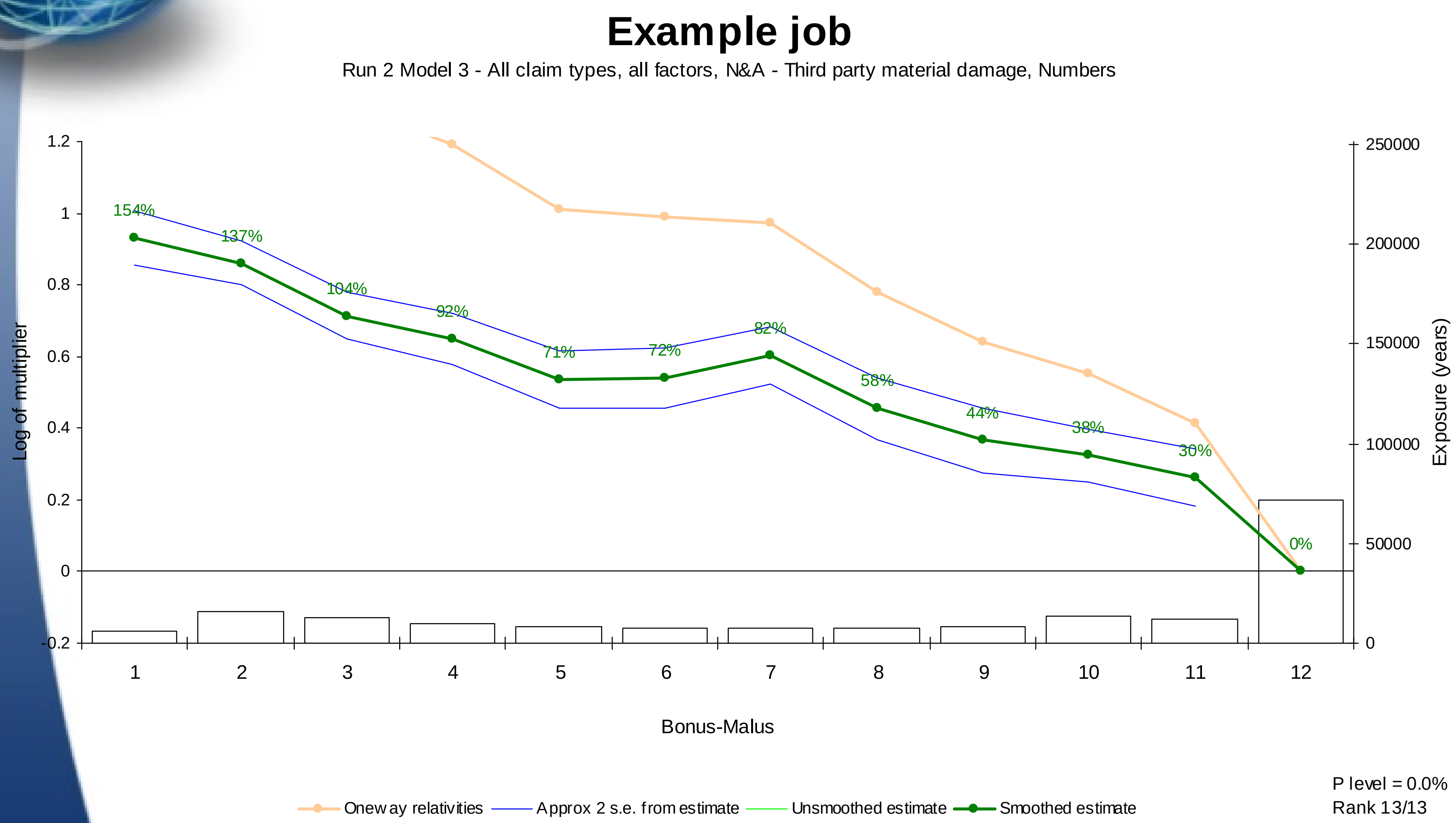
Agenda

- External data
- Geographical spatial analysis
- Product features of modeling interest
  - Bonus-Malus
  - Moderator / point of sale algorithms
- Retention modeling / price optimization

Bonus-Malus

- Scale of discounts based on number of years without a claim
- Enshrined into culture of many EU insurance markets
- Used to be mandatory in some countries
  - many EU insurers still scared to deviate from perceived norm
- Scales are of different lengths from one country to another, and have different transition rules
- Policyholder leaving one company will transfer an agreed measure of Bonus-Malus to another

Bonus-Malus - "realistic fictitious" example



Bonus-Malus

- Discounts generally wrong
- Other factors often moved around to compensate
- Nevertheless
  - discourages claims
  - can often be strongly predictive over and above other factors
  - indirectly picks up some element of risk factors not explained by rating factors
- In general mixed factors based on claim free years and tenure can be rather predictive

Agenda

- External data
- Geographical spatial analysis
- Product features of modeling interest
  - Bonus-Malus
  - Moderator / point of sale algorithms
- Retention modeling / price optimization

Types of rating structures - simple multiplicative (or additive/multiplicative)

$621.50 x

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
50-60	0.76
60+	0.78

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

Sex	Factor
Male	1.00
Female	1.25

Area	Factor
A	0.95
B	1.00
C	1.09
D	1.15
E	1.18
F	1.27
G	1.36
H	1.44

Types of rating structures - using a simple "moderator"

£621.50 x

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
50-60	0.76
60+	0.78

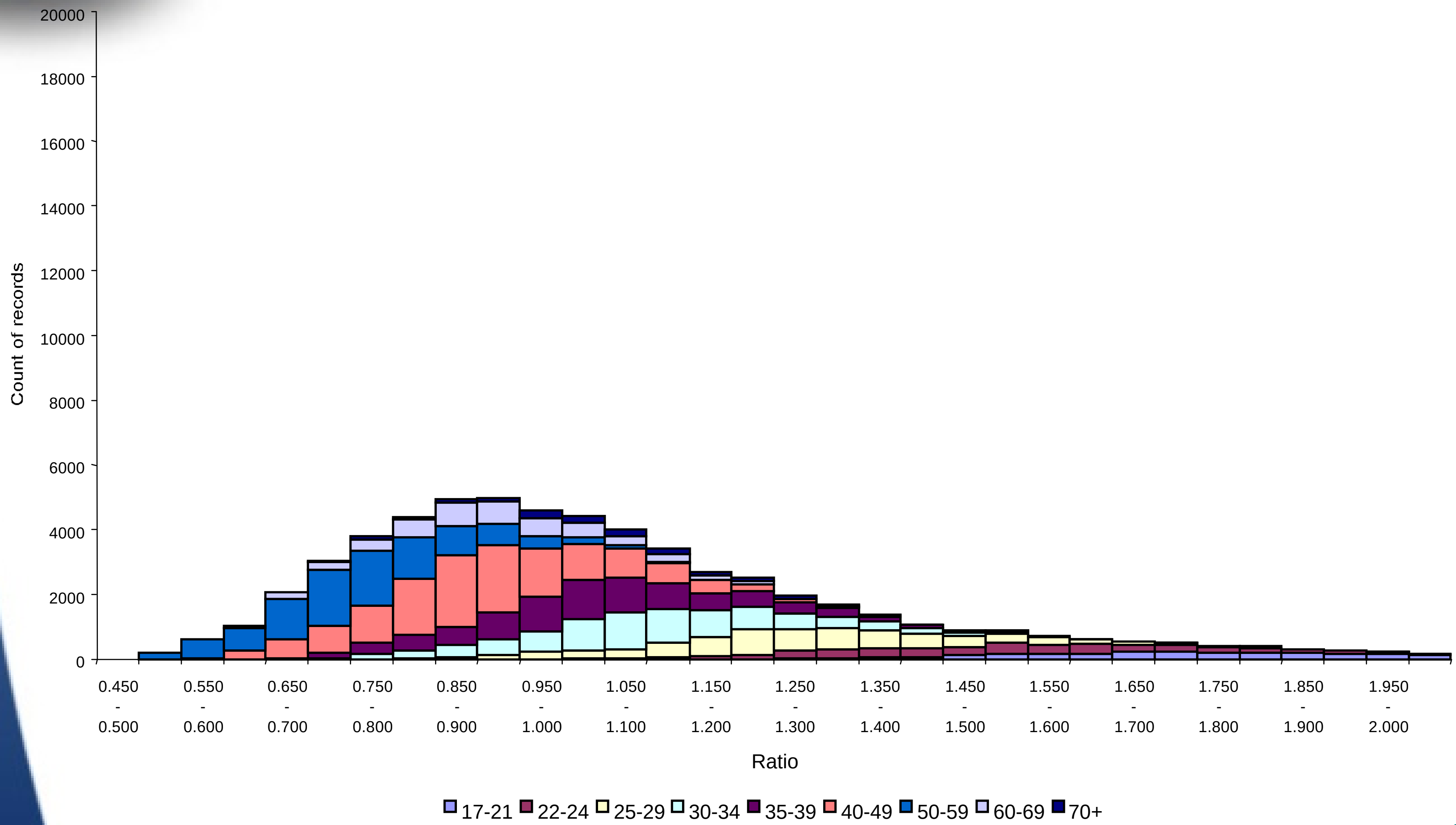
Group	Factor
1	0.54
2	0.65
3	0.73
4	0.85
5	0.92
6	0.96
7	1.00
8	1.08
9	1.19
10	1.26
11	1.36
12	1.43
13	1.56

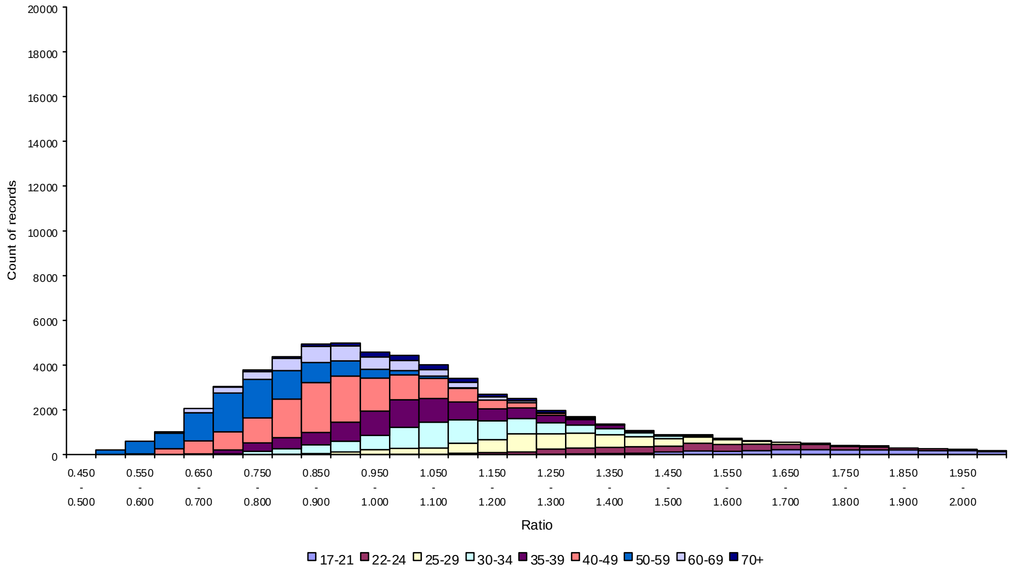
Sex	Factor
Male	1.00
Female	1.25

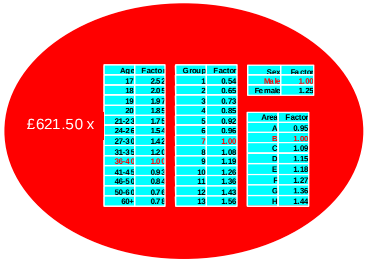
Area	Factor
A	0.95
B	1.00
C	1.09
D	1.15
E	1.18
F	1.27
G	1.36
H	1.44

Subject to
max +20%
min -10%

Example of use of moderator





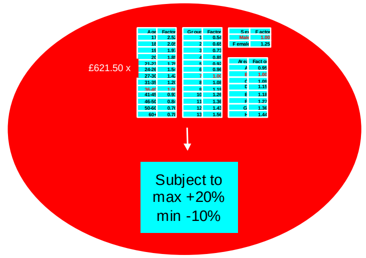


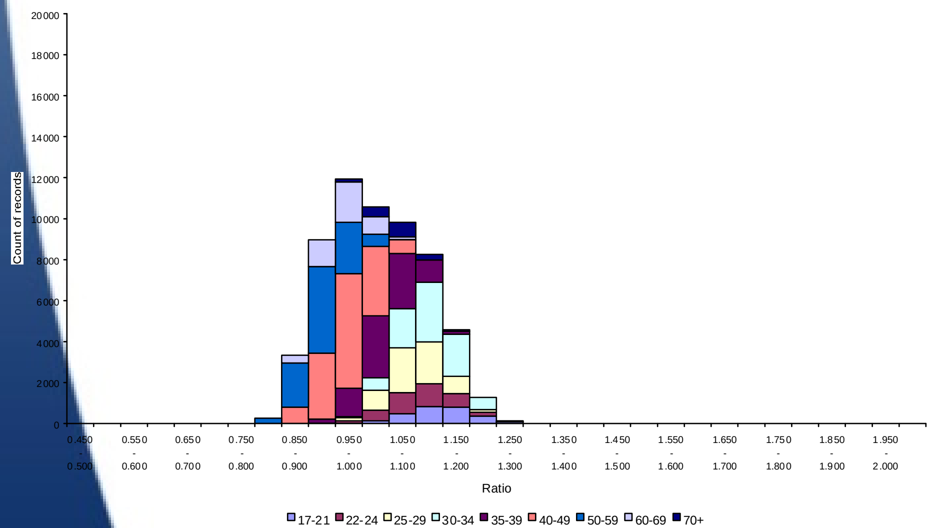
Age	Factor
17	2.52
18	2.05
19	1.97
20	1.85
21-23	1.75
24-26	1.54
27-30	1.42
31-35	1.20
36-40	1.00
41-45	0.93
46-50	0.84
50-60	0.76
60+	0.78

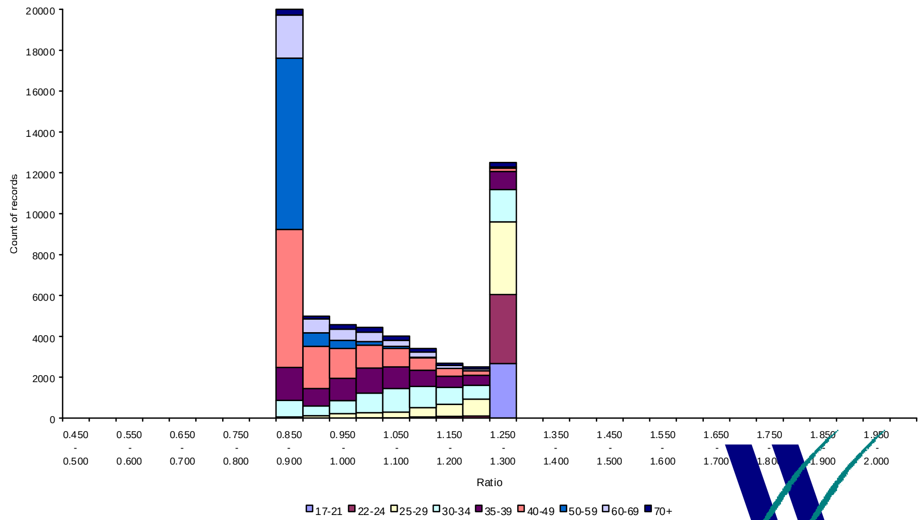
Group	Factor
1	0.54
2	0.65
3	0.73
4	0.85
5	0.92
6	0.96
7	1.00
8	1.08
9	1.19
10	1.26
11	1.36
12	1.43
13	1.56

Sex	Factor
Male	1.00
Female	1.25

Area	Factor
A	0.95
B	1.00
C	1.09
D	1.15
E	1.18
F	1.27
G	1.36
H	1.44



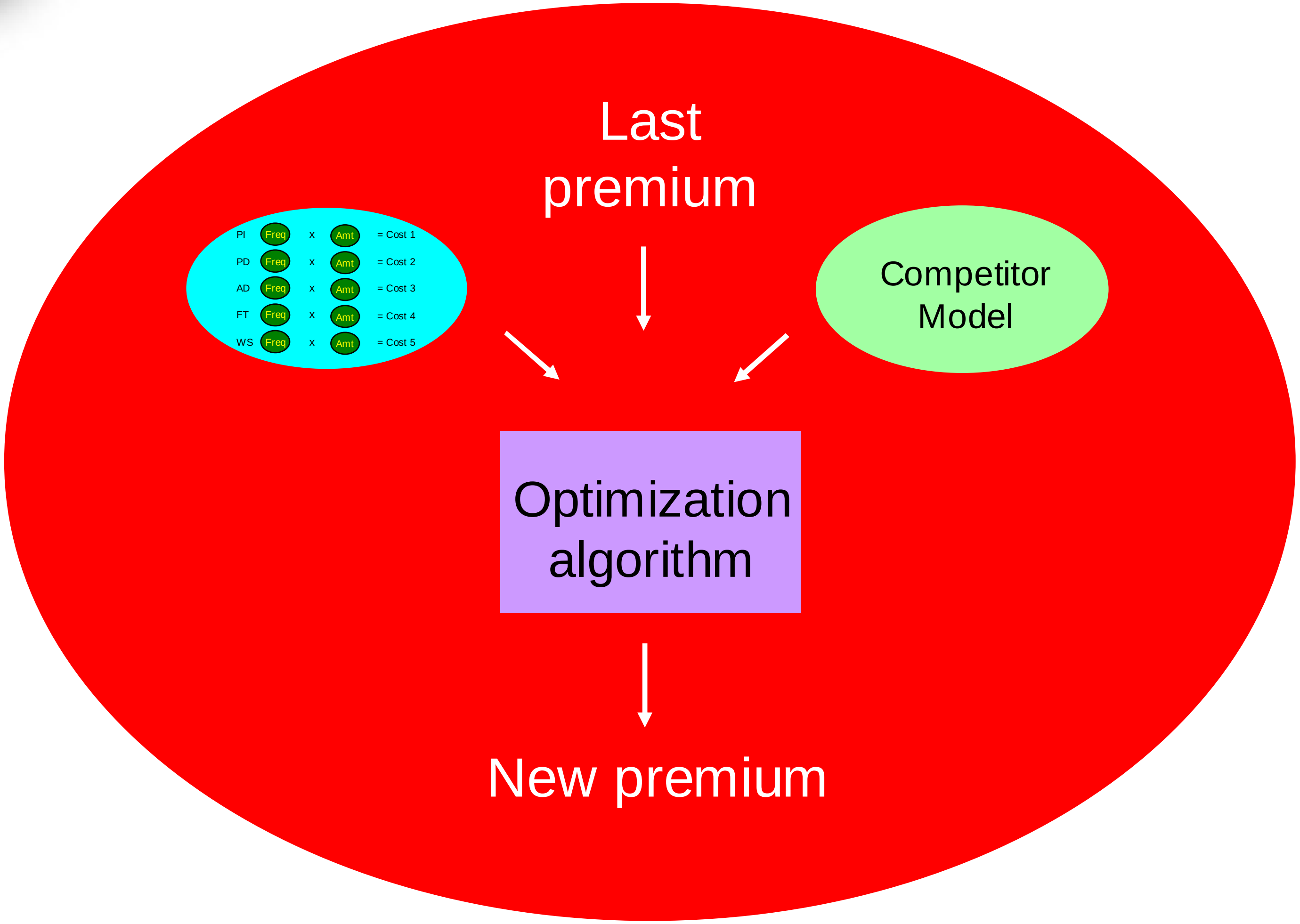




Moderator: pros/cons

- Advantages of moderators include:
  - moves everyone to optimal position (subject to acceptable premium increases) more quickly
  - can take into account elasticity for the type of person in question
  - can be less detailed work required regarding underlying parameterisation
  - less work required to parameterise in future
- Disadvantages
  - more onerous system requirements
  - harder to understand rating structure
  - likely to result in different quotes for renewals and new business for an identical risk
  - may not be too popular with some regulators?

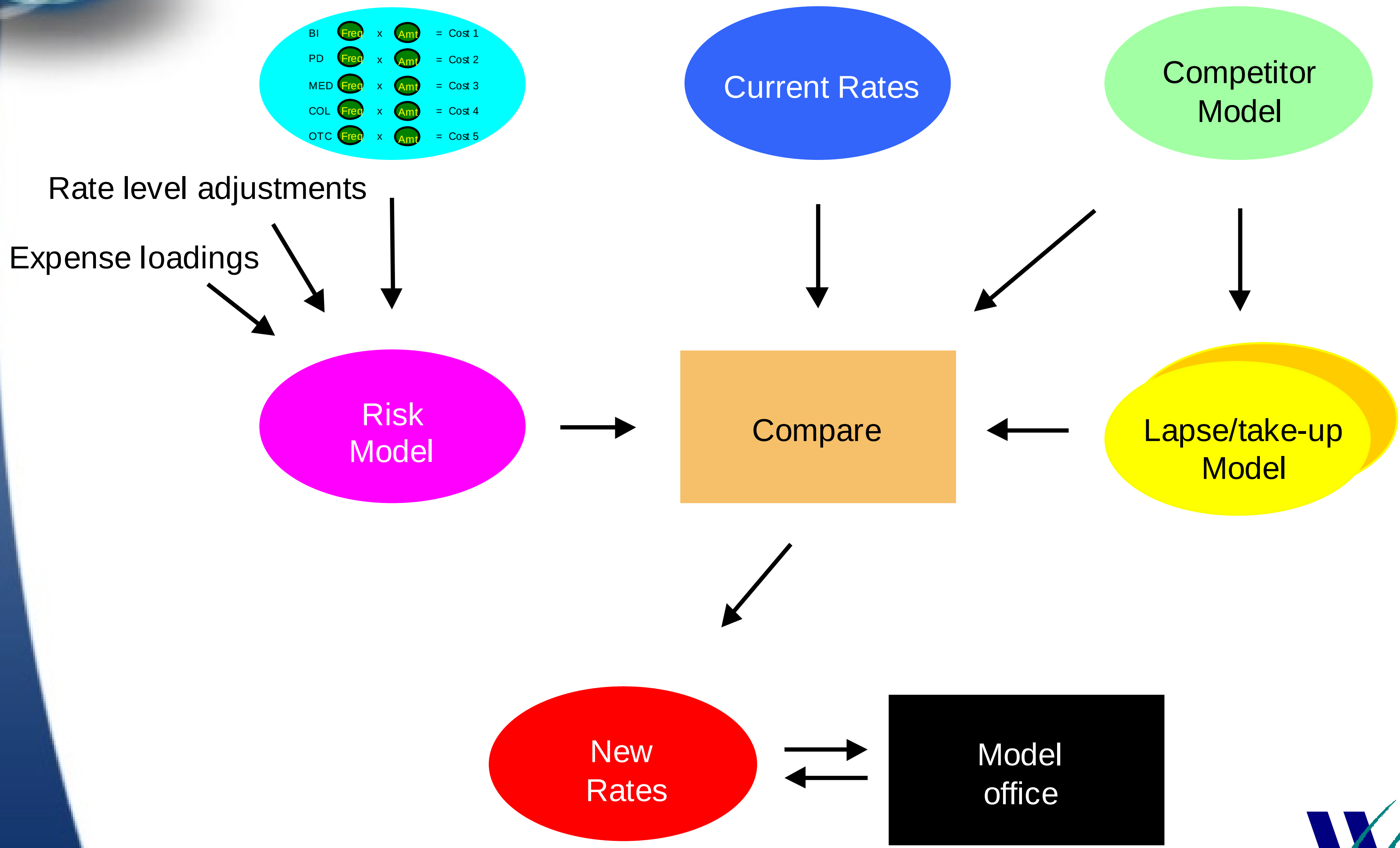
Rating structure - point of sale optimization algorithm



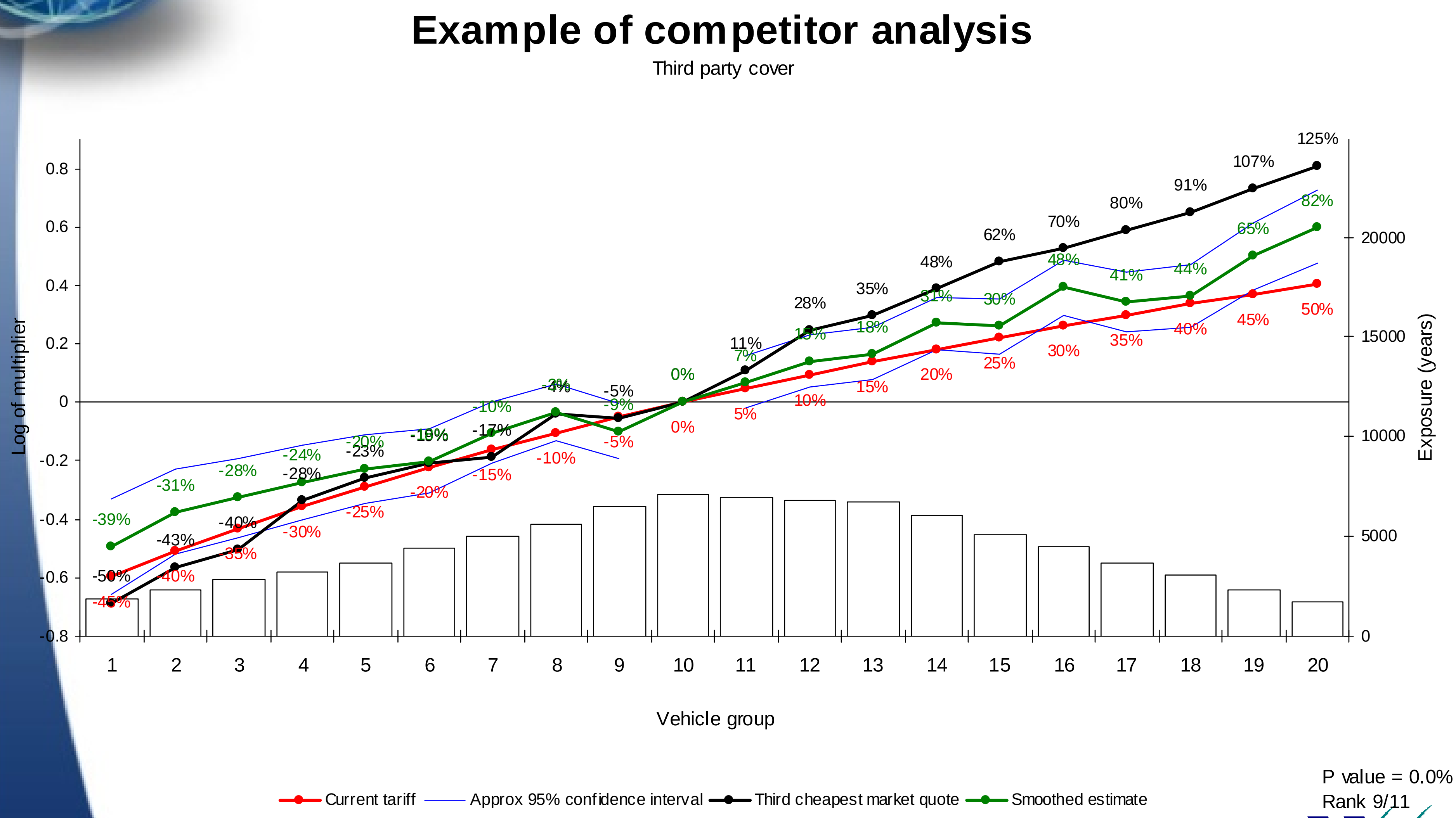
Agenda

- External data
- Geographical spatial analysis
- Product features of modeling interest
- Retention modeling / price optimization

The premium rating process

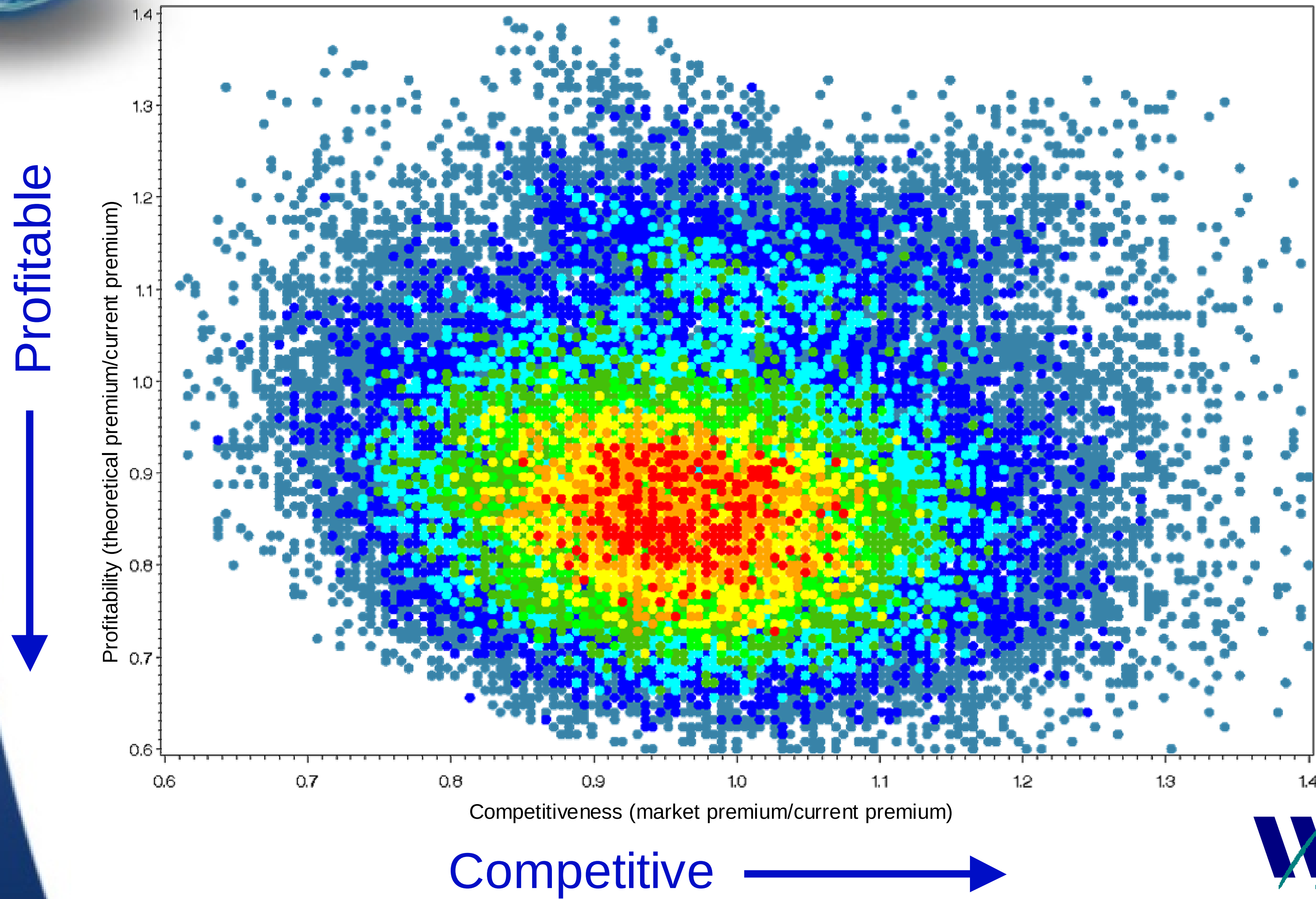


Considering the competitive position

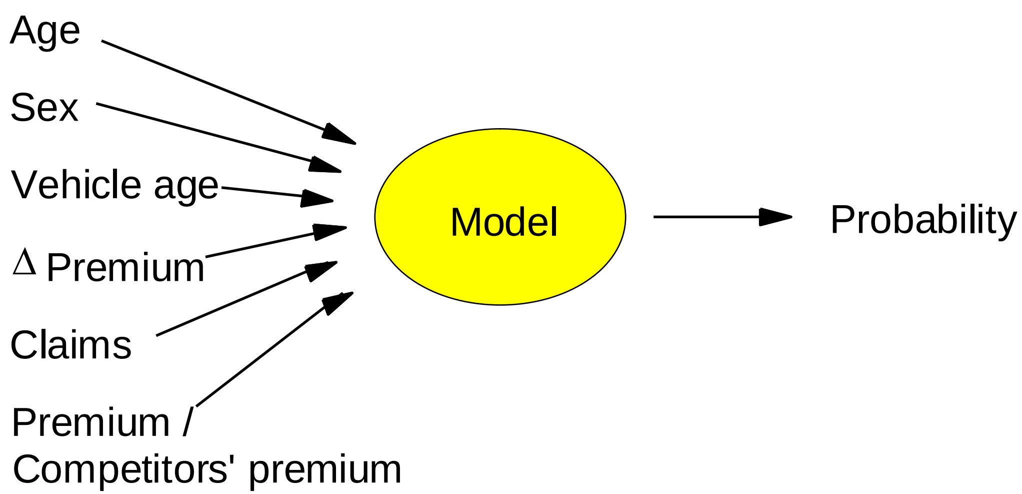


P value = 0.0%
Rank 9/11

Considering the competitive position



Modeling retention / conversion

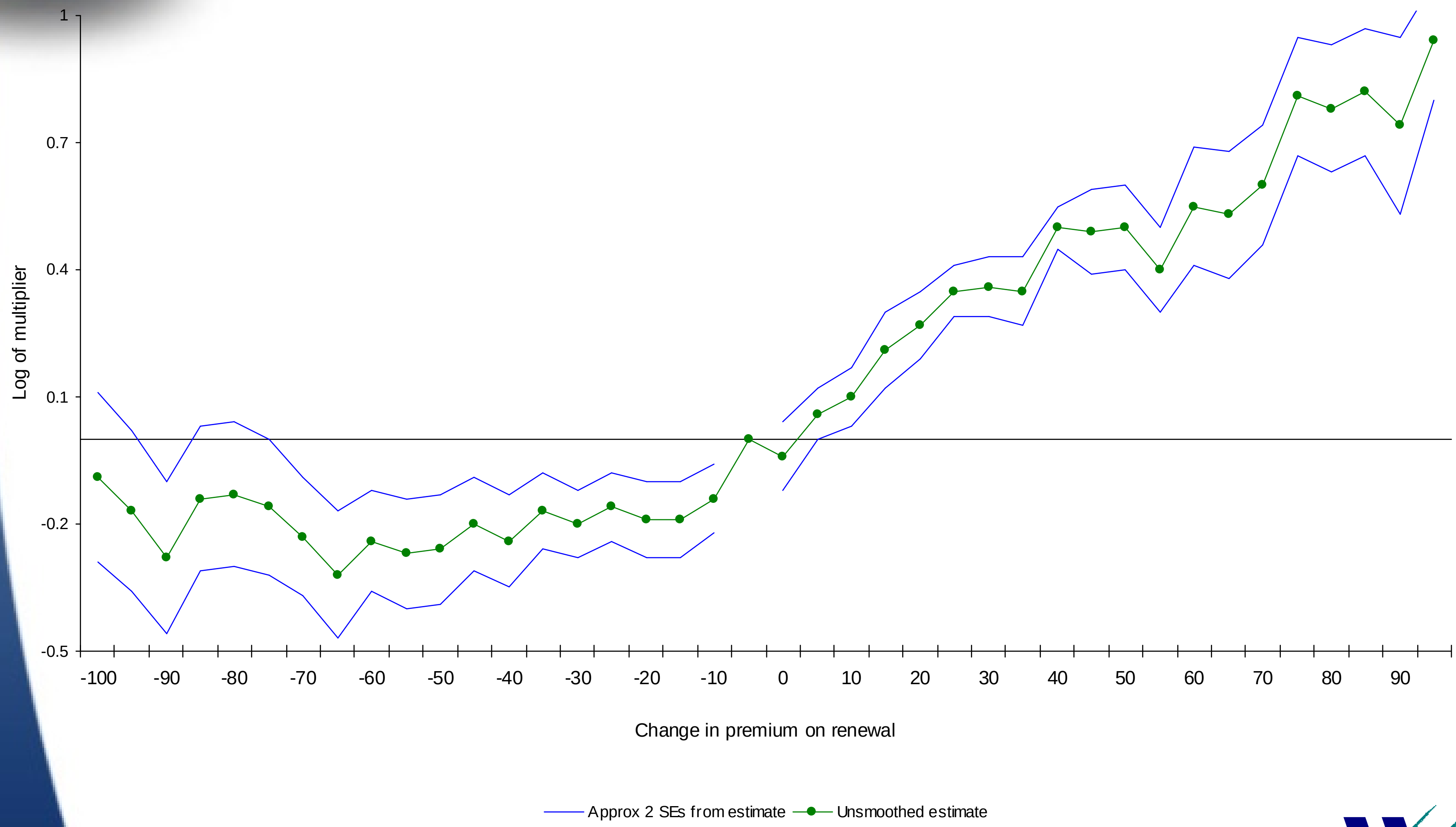


- Model
  - normal factors
  - softer factors
  - premium change /competitor factors
- Conversion models require data on failed quotes

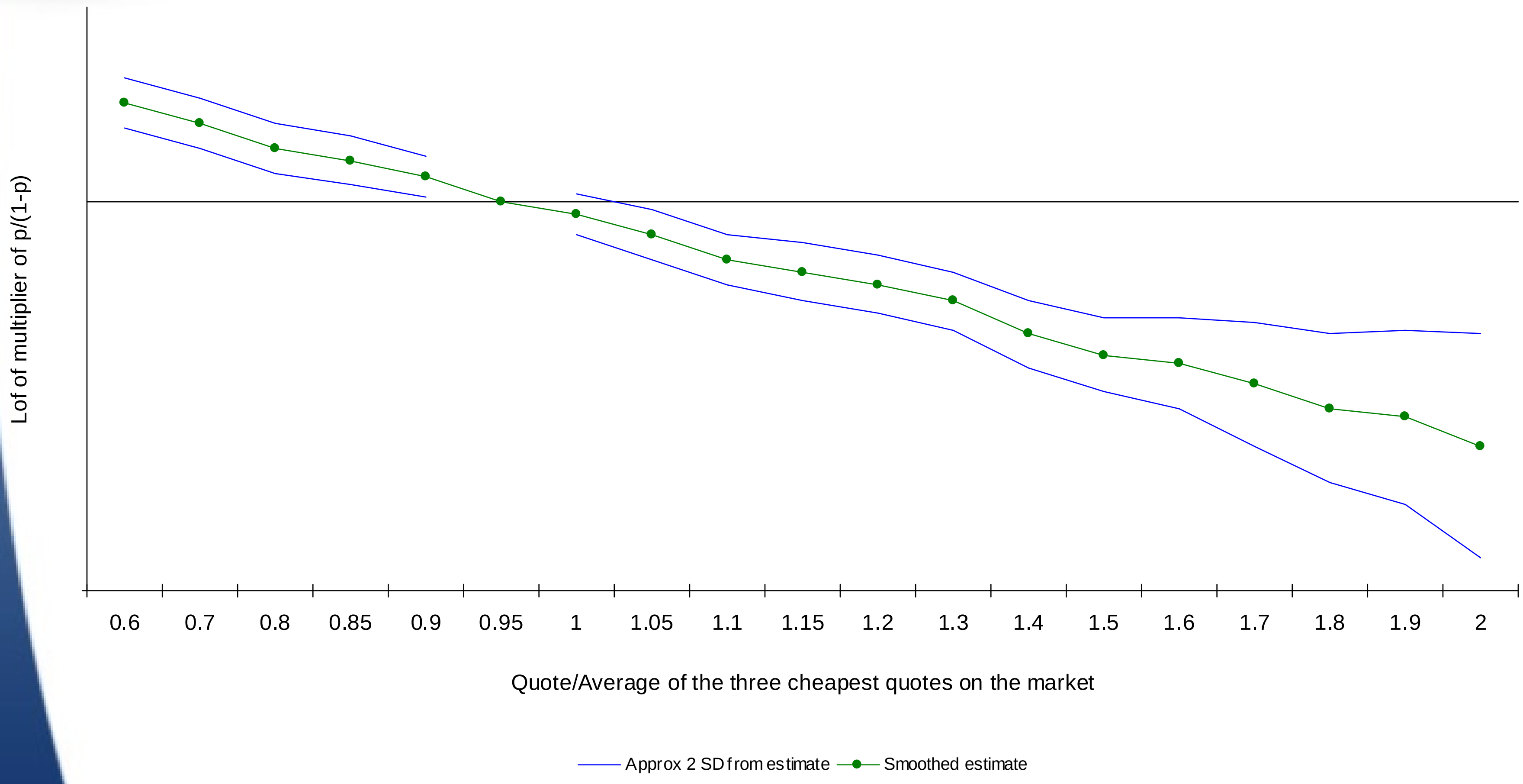
Modeling retention / conversion

- Factors
  - normal rating factors
  - endorsement activity
  - other products held
  - payment method
  - discount expectation
  - source
  - claims history
  - tenure
  - change in premium (this time and last)
  - competitiveness of premium
- Data
  - need healthy spread of historic rate changes to assess elasticity well
  - EU insurers can have more fun:
    - frequent changes
    - randomized price trials
    - generations of rates
- Uses
  - lifetime loadings
  - customer value calculations
  - price optimization

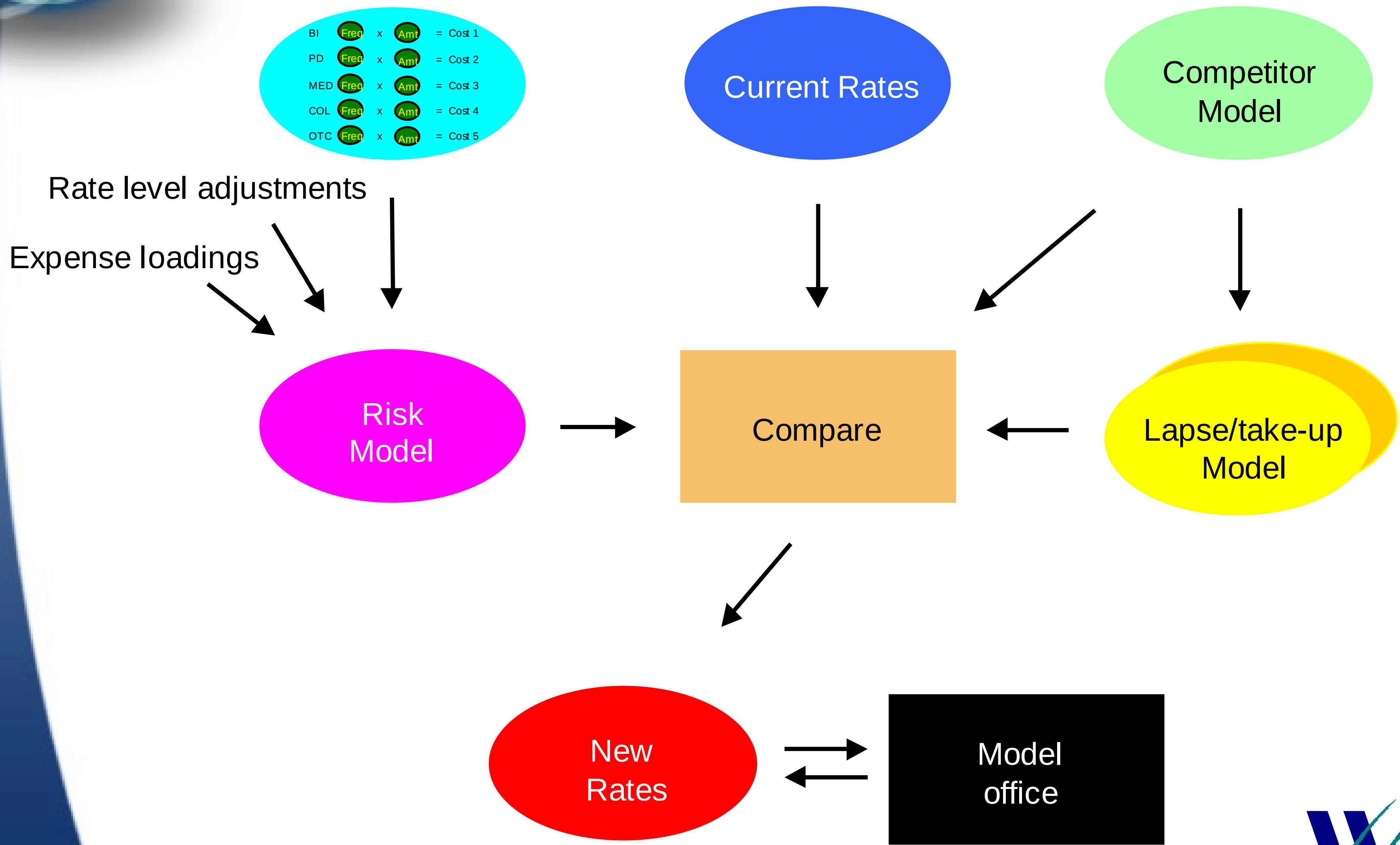
Effect of premium change on lapses



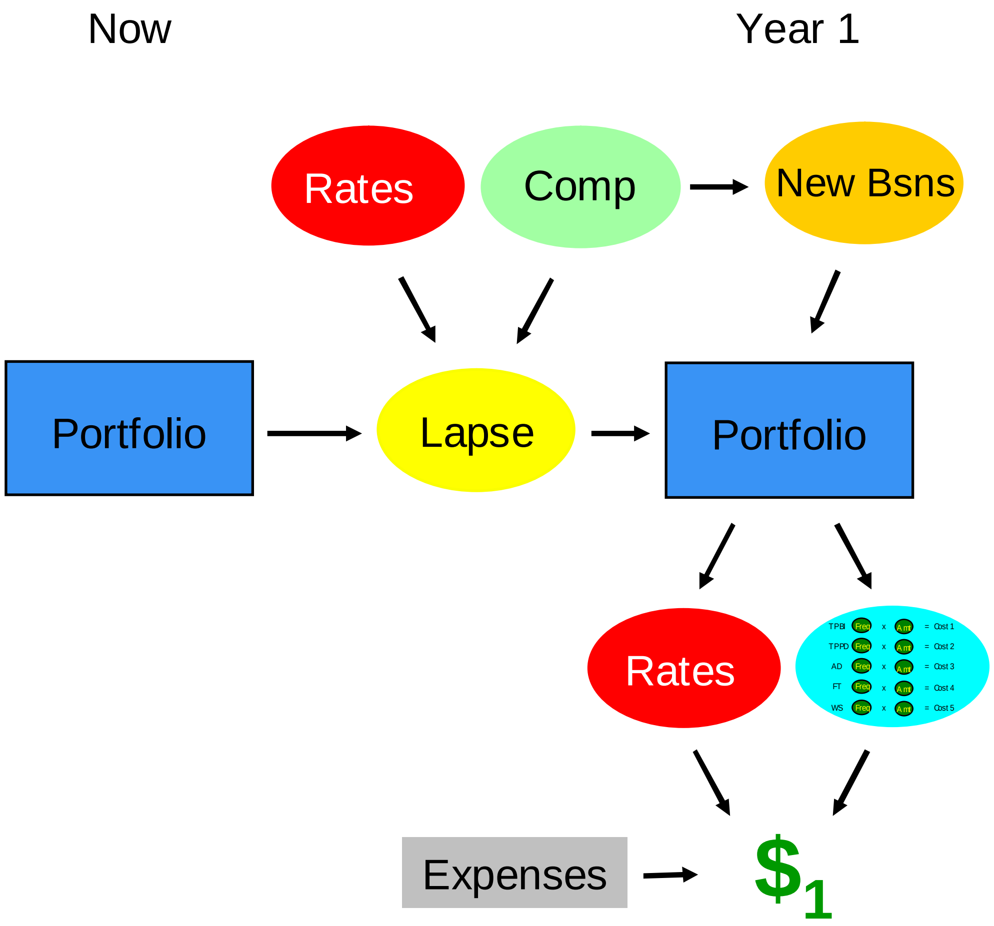
Effect of competitiveness on new business



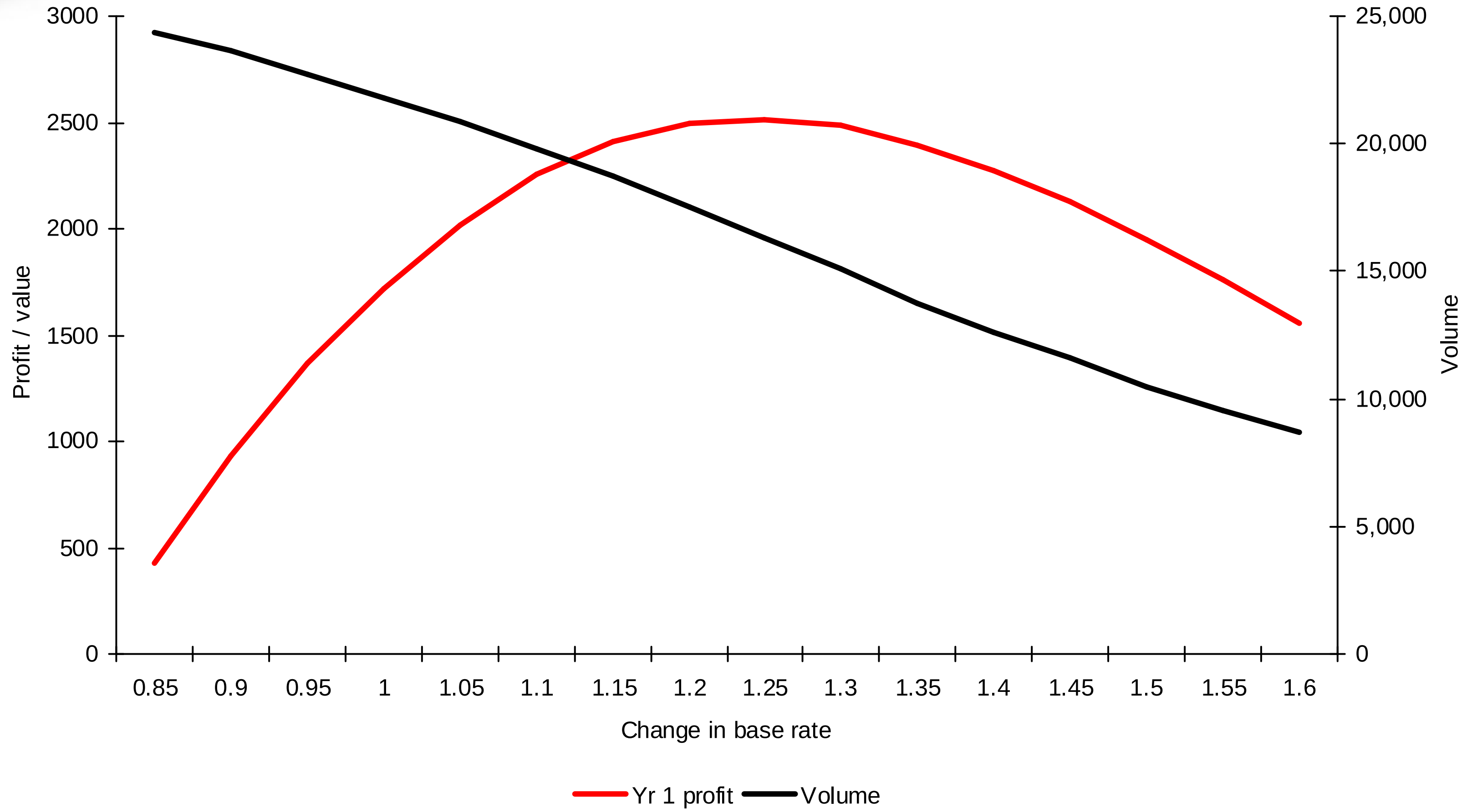
The premium rating process



Price optimization / model office scenario testing

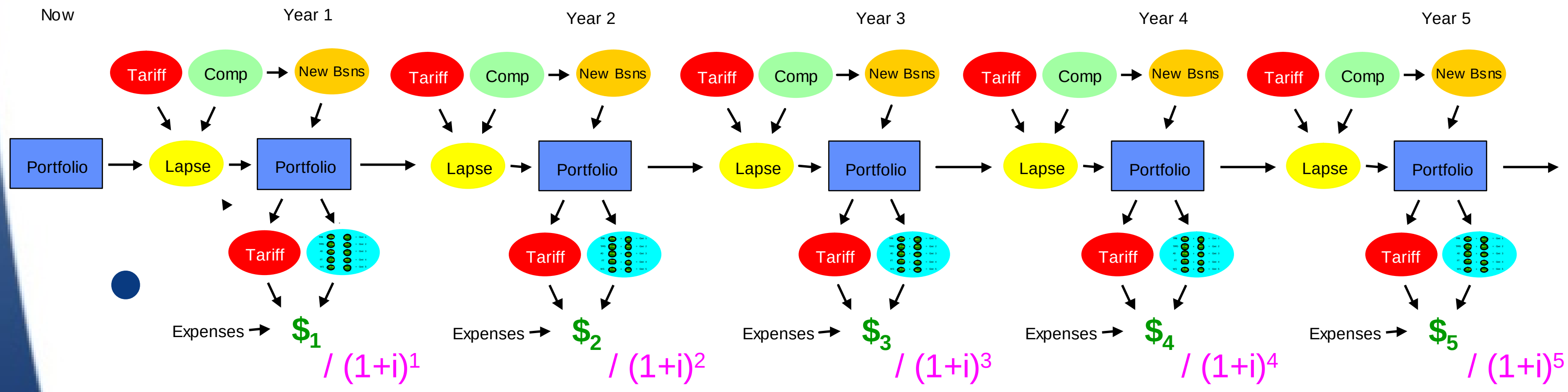


What do you optimize?



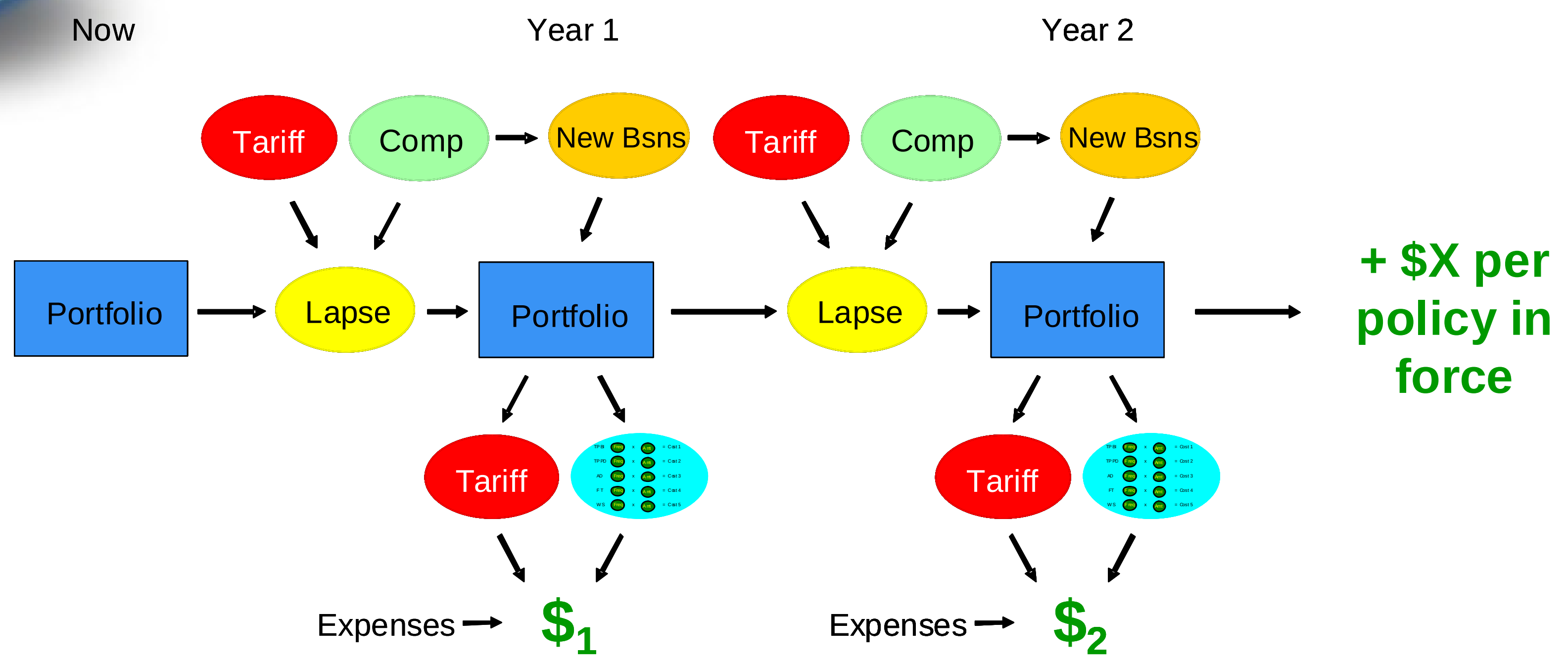
What do you optimize?

- Year 1 profit will not consider value of business in the future
- Could seek "a_x" for future profit - problematical
- Could calculate PVFP over a decade



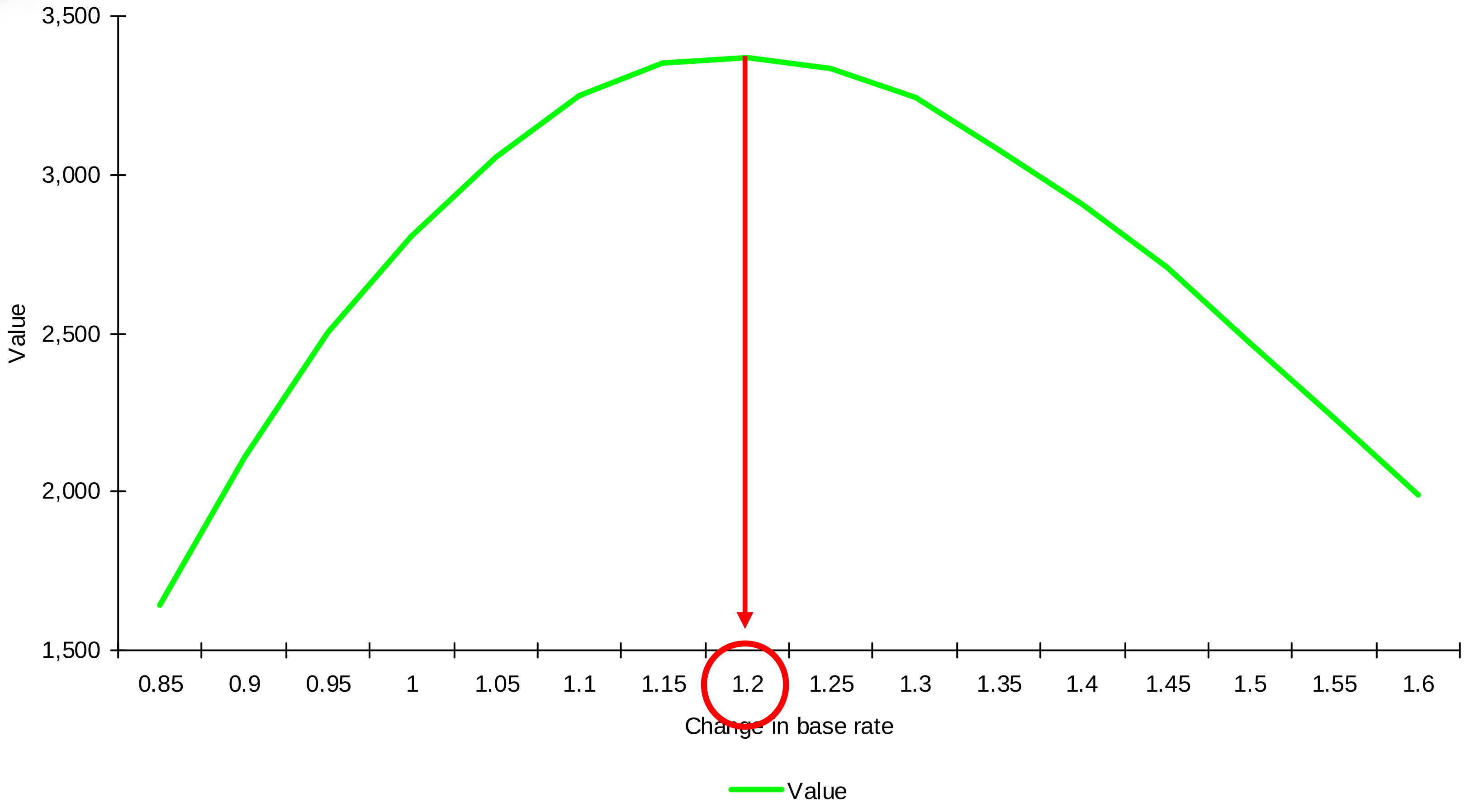
- but too many wobbly assumptions?

Pragmatic solutions?

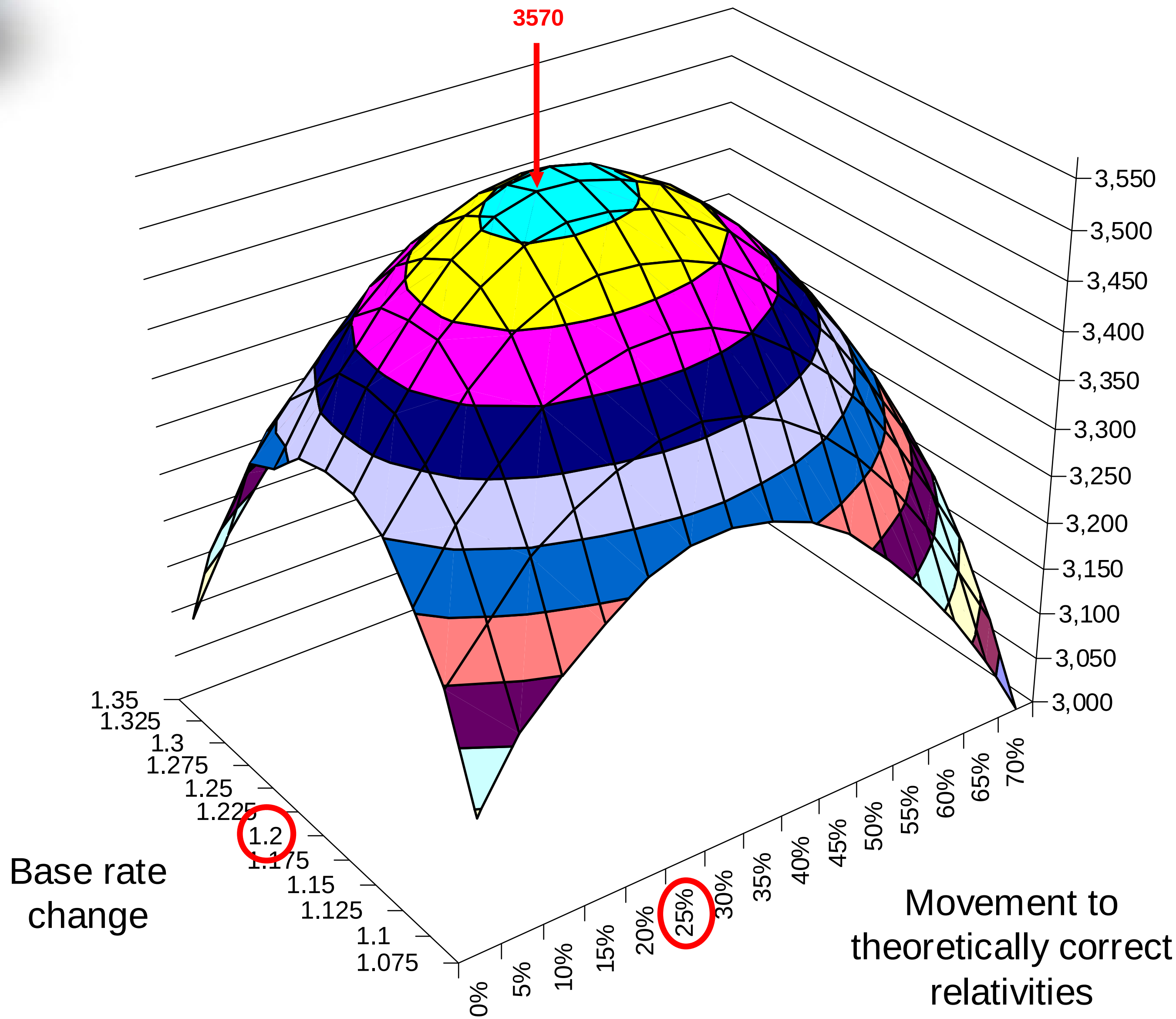


- Or optimize a profit measure subject to a minimum volume
- Or optimize volume subject to a minimum profit

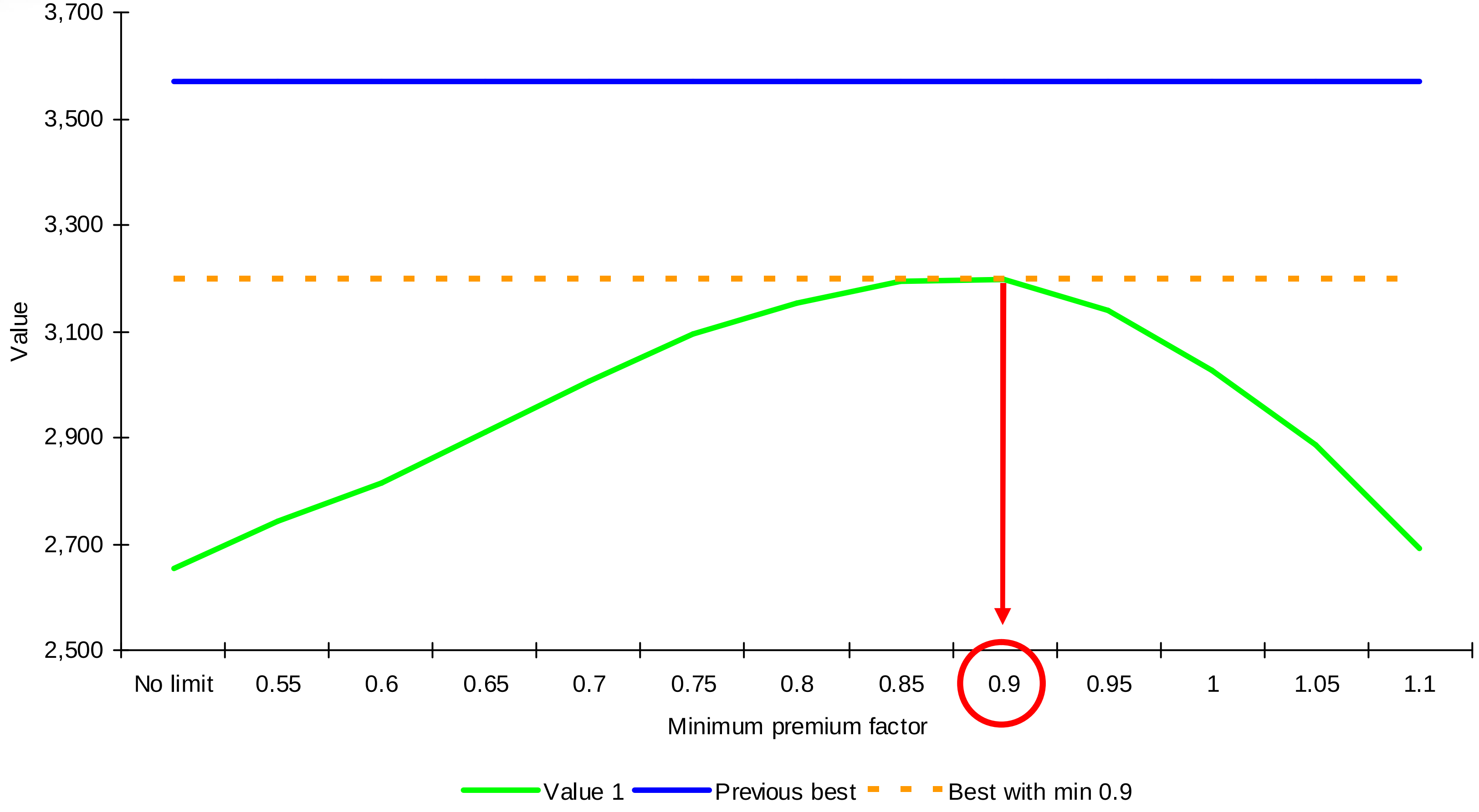
Base rate change



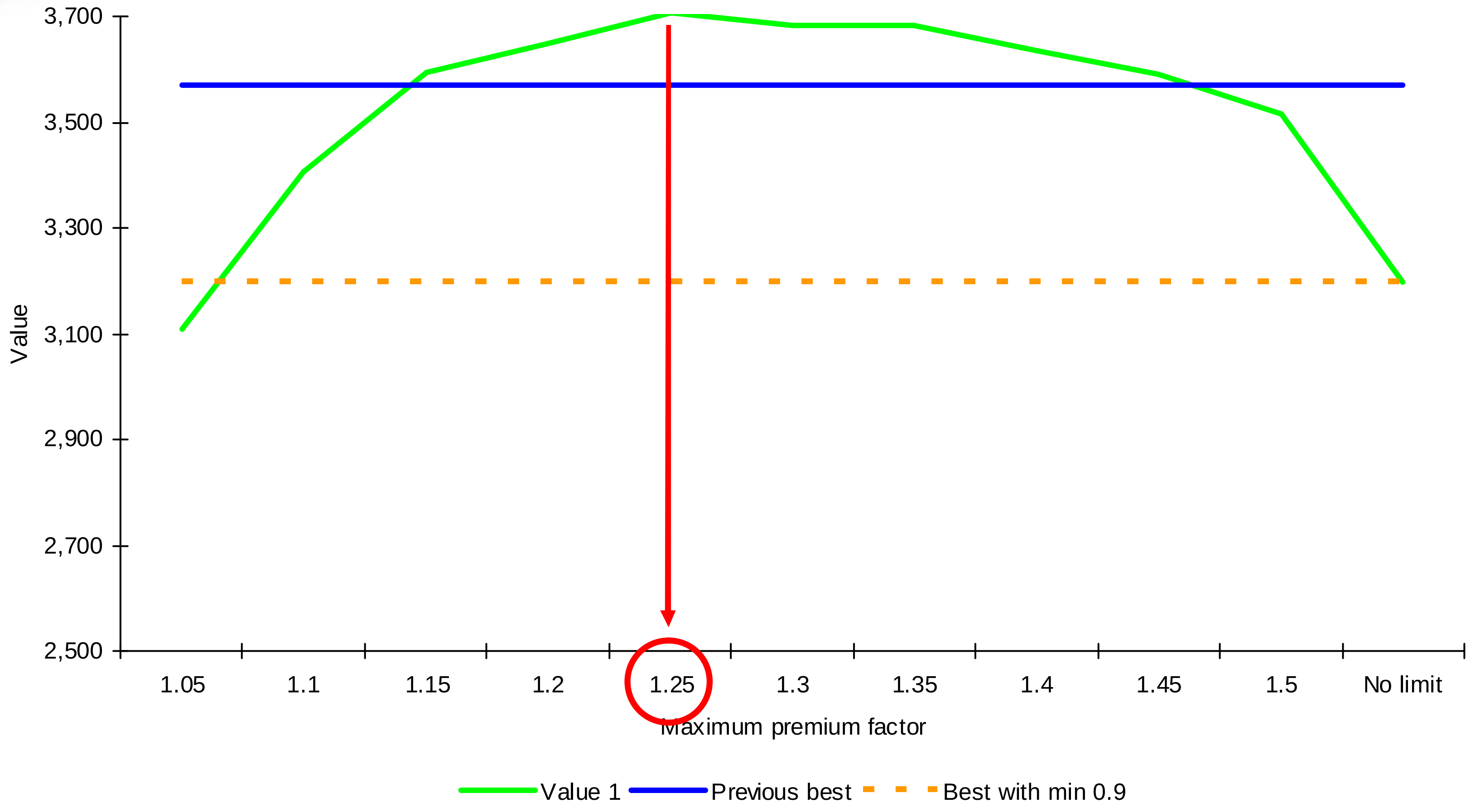
Base rate and relativity change



Calibration of a simple min/max moderator



Calibration of a simple min/max moderator



Other ideas

- Dynamic calibration of retention model at point of sale
- Multivariate models of expenses (eg young people amend cover more)

Modeling Lessons From Across the Pond - Insights from Decades of Deregulated Modeling

Duncan Anderson MA FIA
Watson Wyatt LLP

WWW.WATSONWYATT.COM