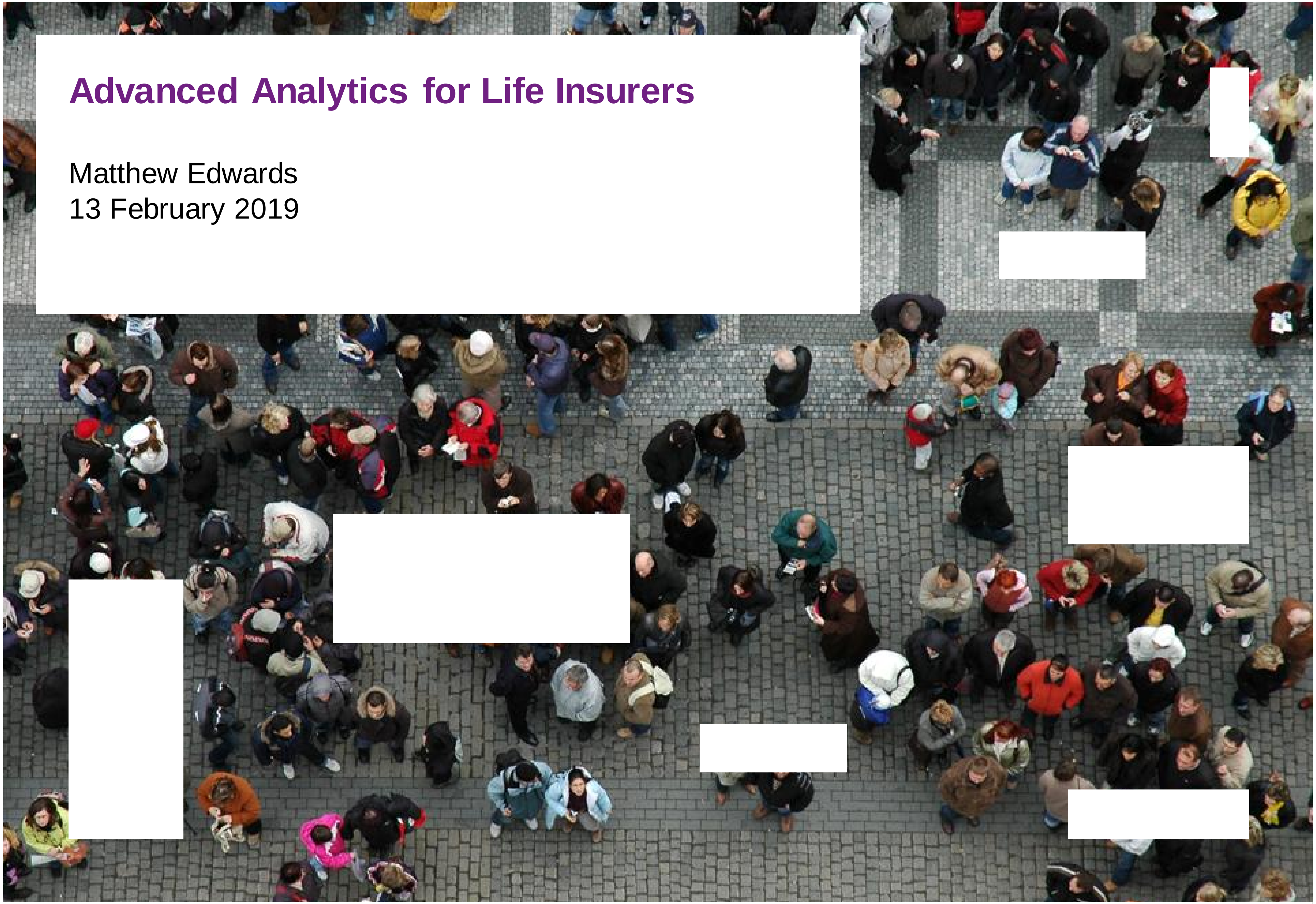
Advanced Analytics for Life Insurers

Matthew Edwards
13 February 2019



Agenda

Methodology

Mortality

Policyholder behaviour

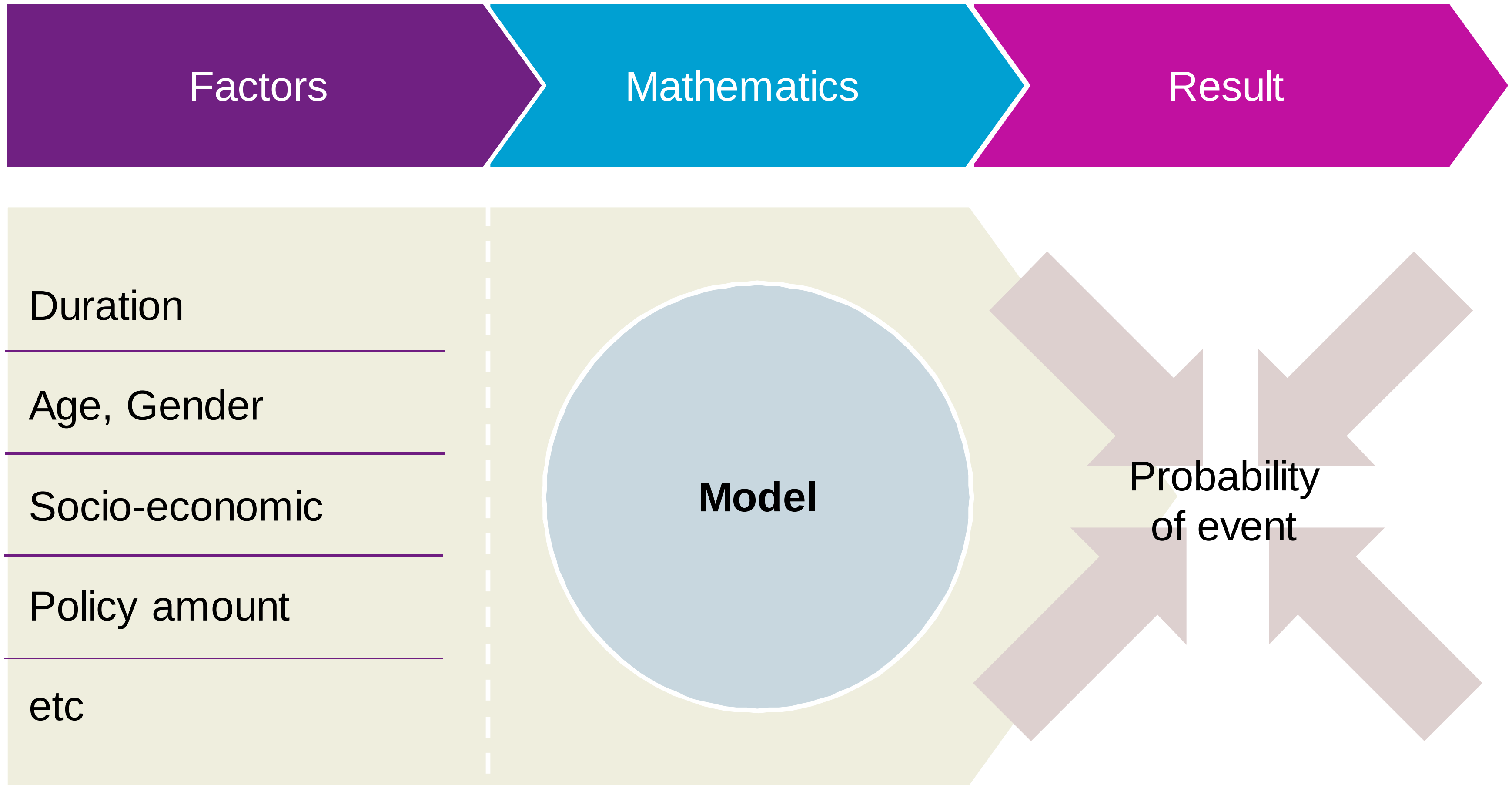
Proxy models

Other methods

Methodology

What is the basic methodology?

Multi-factor models (whether GLMs, pure or 'supervised' machine learning) provide a fast and powerful way to model the effect of many factors simultaneously, taking into account relationships between factors



Why do we need anything complicated?

Claims	Urban	Rural
Male	40	10
Female	10	10

Exposure	Urban	Rural
Male	200	100
Female	100	200

Risk	Urban	Rural
Male	20%	10%
Female	10%	5%

4x - true effect

One-way analysis of area

- Urban: 50 claims from 300 policies
- Rural: 20 claims from 300 policies
- Urban 2.5x worse than rural

One-way analysis of gender

- Male: 50 claims from 300 policies
- Female: 20 claims from 300 policies
- Males 2.5x worse than females

One way estimate

- Urban males have 2.5 x 2.5 = **6.25x** as many claims as rural females

One-way analysis fails to take account of correlations and result is (very) wrong

Generalised Linear Models

Basic mathematical structure

The link function – how the mean links to the systematic part

The impact of each factor. What we "solve" by maximising the likelihood

Error term – each observation of Y is assumed to come from some probability distribution

$$\vec{Y} = g^{-1}(\underline{X}.\vec{\beta}) + \vec{\varepsilon}$$

The observed data (eg deaths or lapses)

The design of our model (which factors we are going to investigate and how they are split up)

GLMs — typical structure

Postcode Group	Multiplier
Group A	0.86
Group B	0.92
Group C	1.00
Group D	1.14
Group E	1.26

- What the maths means in practice (example):

  Probability of event in year =

  Base level for observed population ×

  Factor 1 (based on age, sex) ×

  Factor 2 (based on postcode group) ×

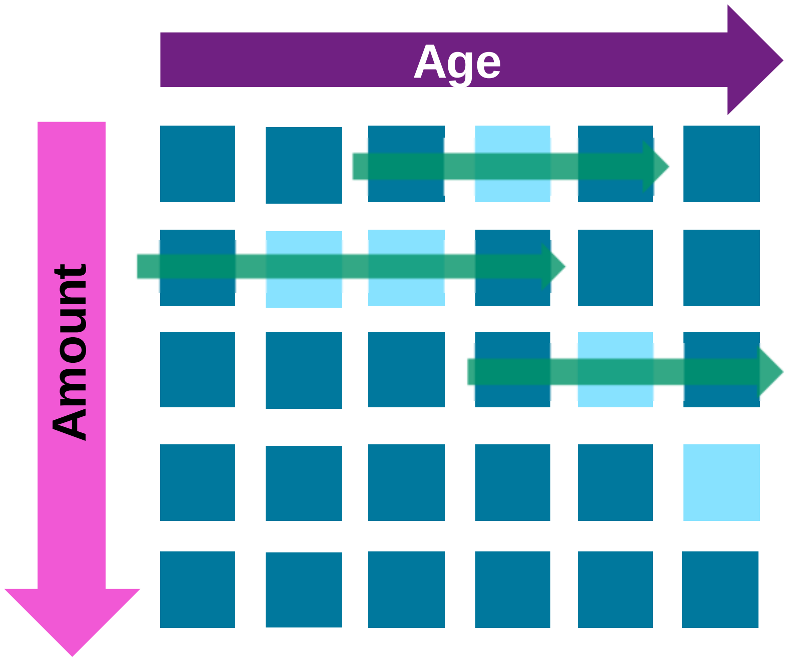
  Factor 3 (based on duration) ×

  Factor 4 (based on amount)…
- Each factor is a series of multiplicative coefficients
- All factors are considered simultaneously, allowing for correlations in the data automatically
- The GLM finds the factor coefficients that will best fit the data
- Method allows for the nature of the random process involved, and provides information about the (un)certainty of the result

Relatively simple way to anlayse effect of many factors properly – allowing for correlations and interactions between factors – simple 1-way techniques fail

Data Efficiency

- GLMs generally require > 1,000 of the ‘events’ in question (eg deaths or policy lapses)
- More is better
- This is the total over the period of the investigation – eg 200 deaths / year for 5 years
- GLMs can help improve data efficiency by ‘bridging’ over cells with no or little data (even more so if fitting a factor such as age with a parametric solution)



Light-shaded cells = no or little data

The GLM uses the relationship established in other rows / columns based on the ‘more data’ cells to ‘fill in’ for the ‘low data’ cells

Mortality

Mortality factors and their predictive power

Geography can be a powerful predictive factor in risk analysis

GLMs have become a standard tool to analyse mortality in major markets for different product types. Insurers can use them to improve pricing and financial reporting bases.

Typical mortality rating factors found to be significant

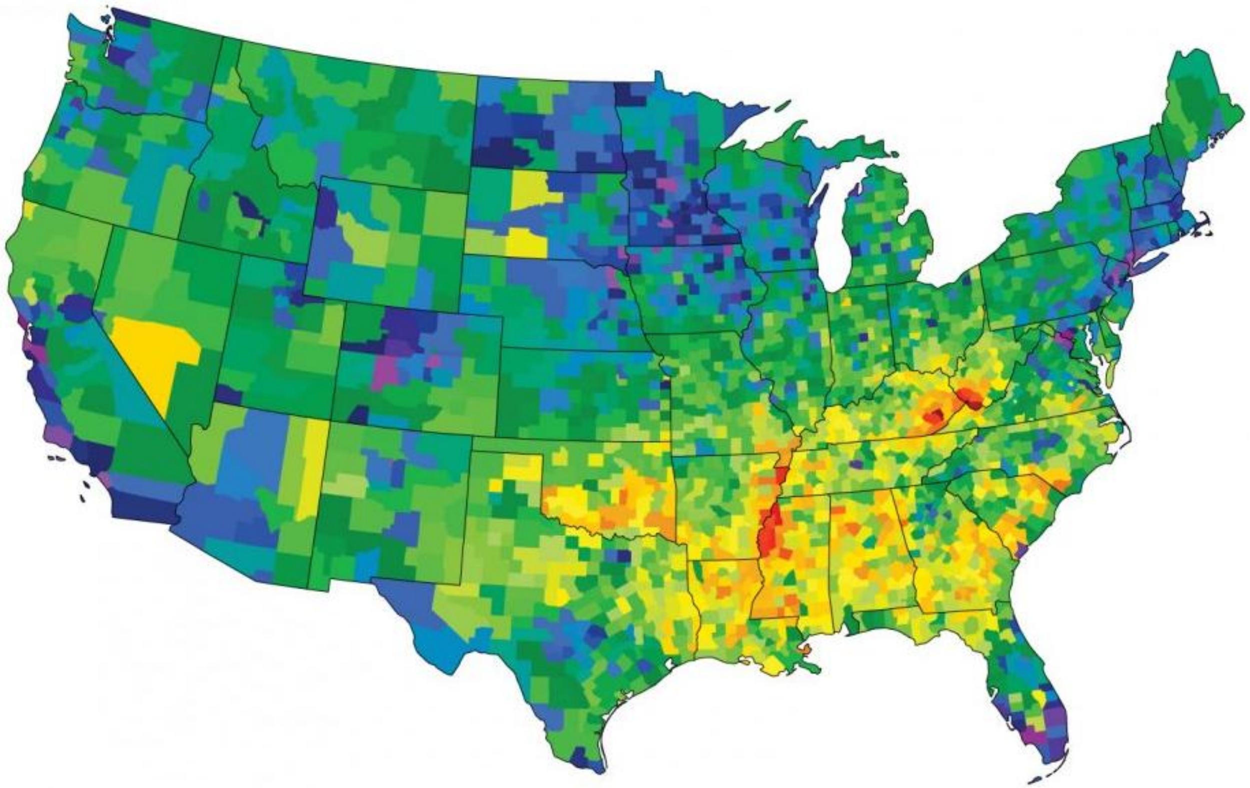
Time period (e.g., calendar year)

Personal data	Age Medical info — e.g., BMI Gender
Geographical data	Socio-geodemographic groups Portfolio-derived clusters Propensities by postcode
Policy data	Amount Duration Product type/options
Company data	Distribution

The map shows how we have derived postcode mortality effects using GLMs

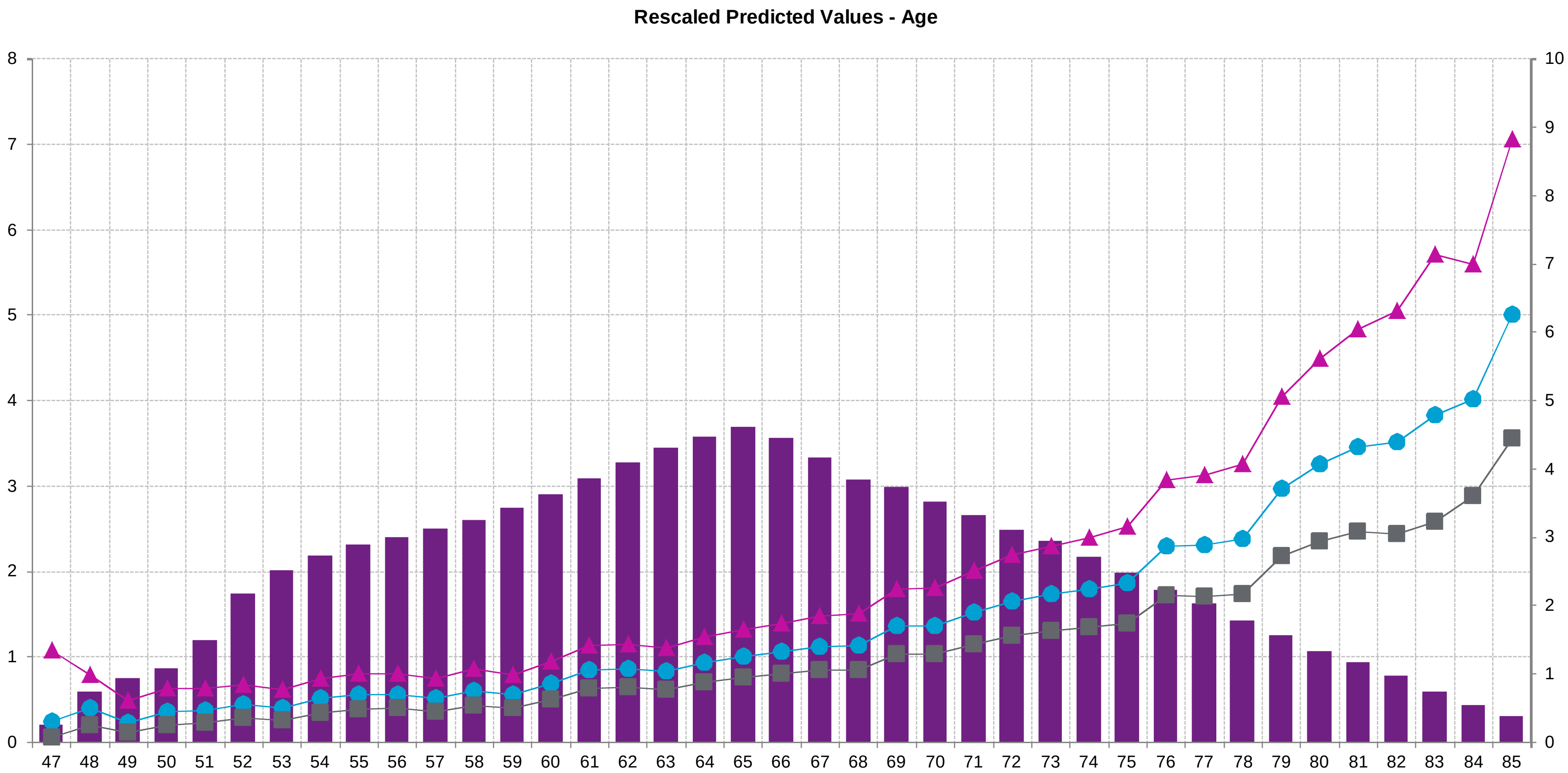
This type of analysis has worked in many countries

TOAMS 4 is Willis Towers Watson’s fourth study of U.S. life insurance mortality.
The study involves 23 participating companies, 124 million policy years, 1.5 million deaths.



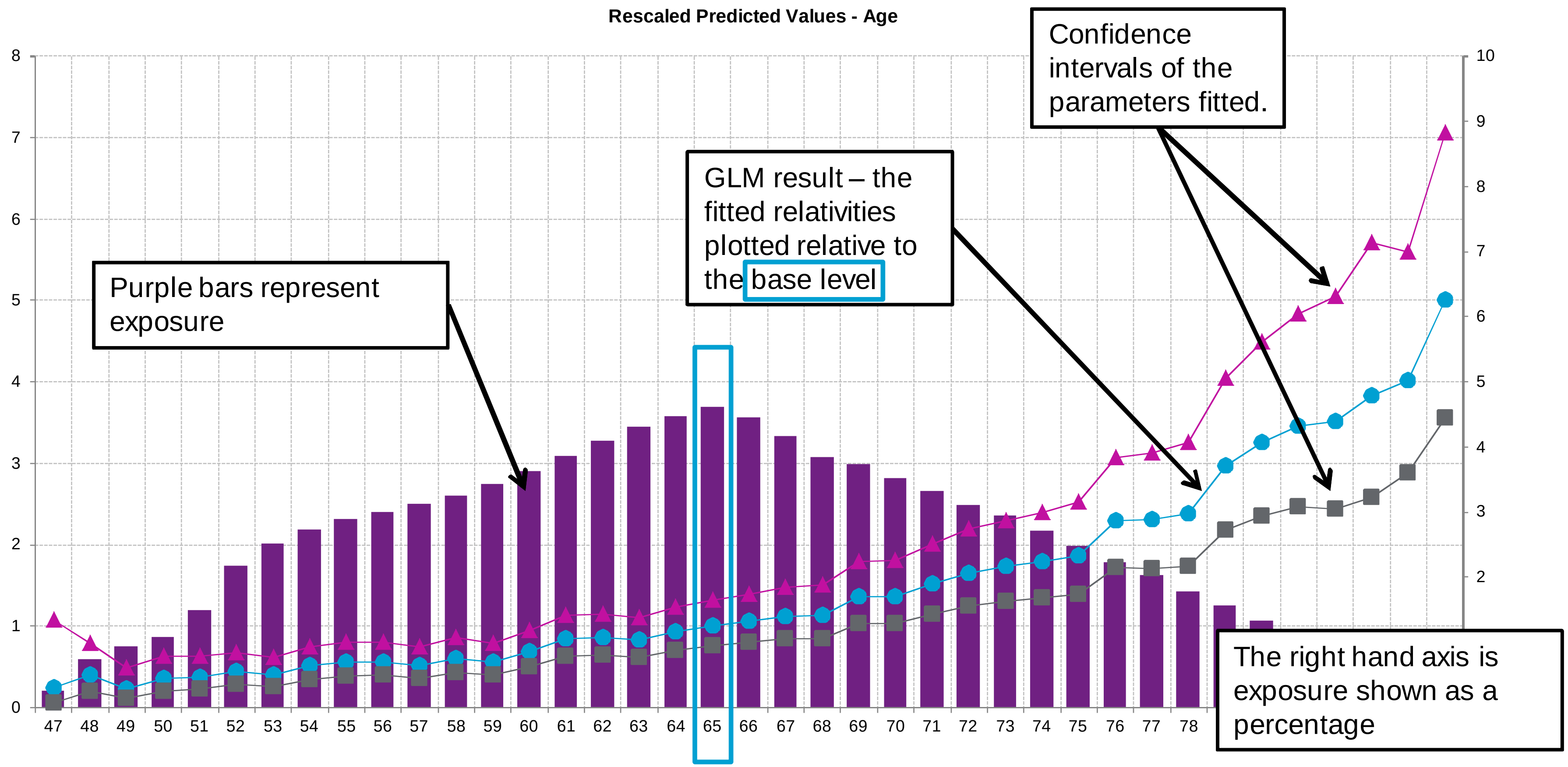
Effect of Age on mortality

(in conjunction with effects of all other factors)



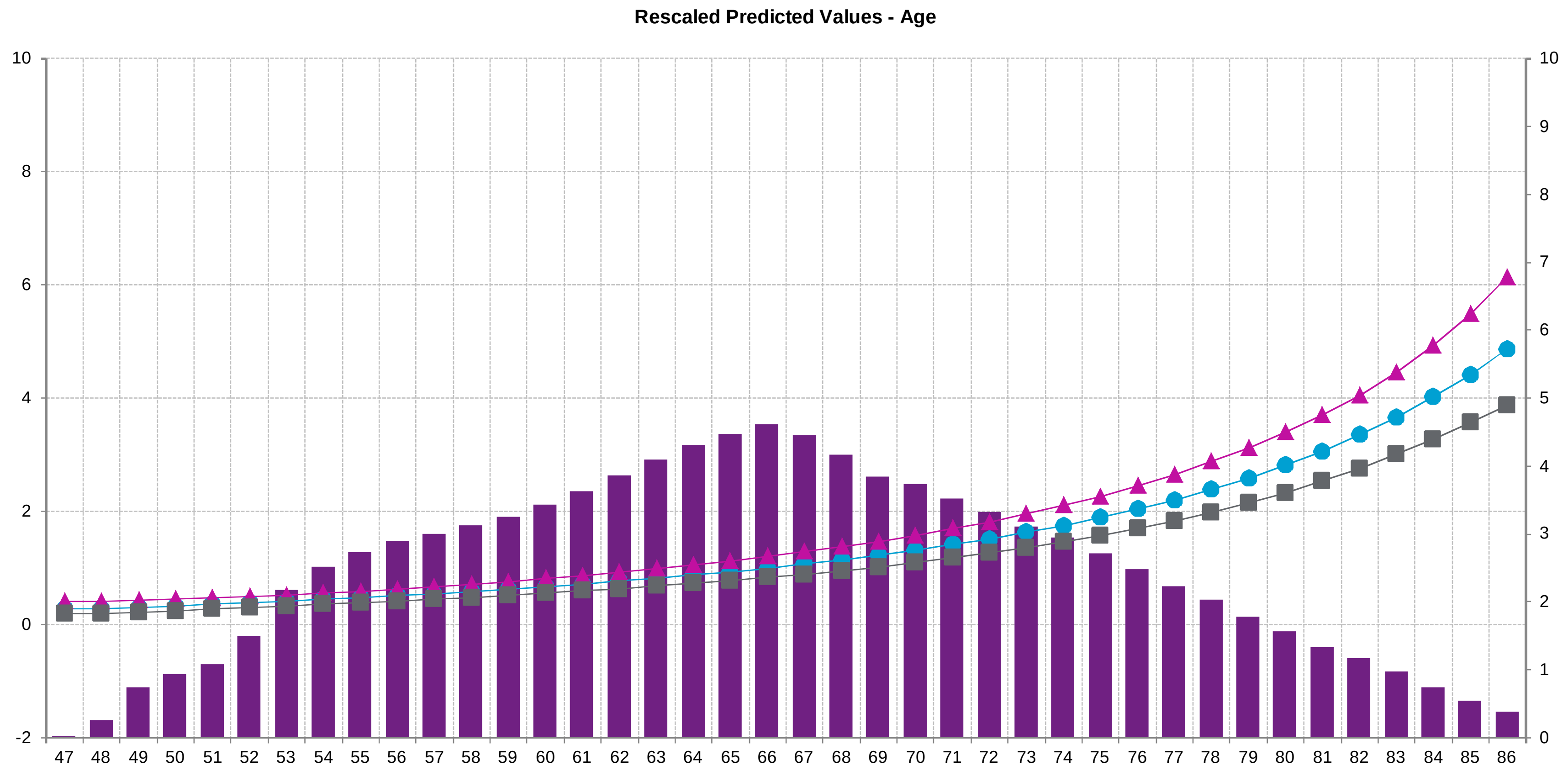
Effect of Age on mortality

(in conjunction with effects of all other factors)



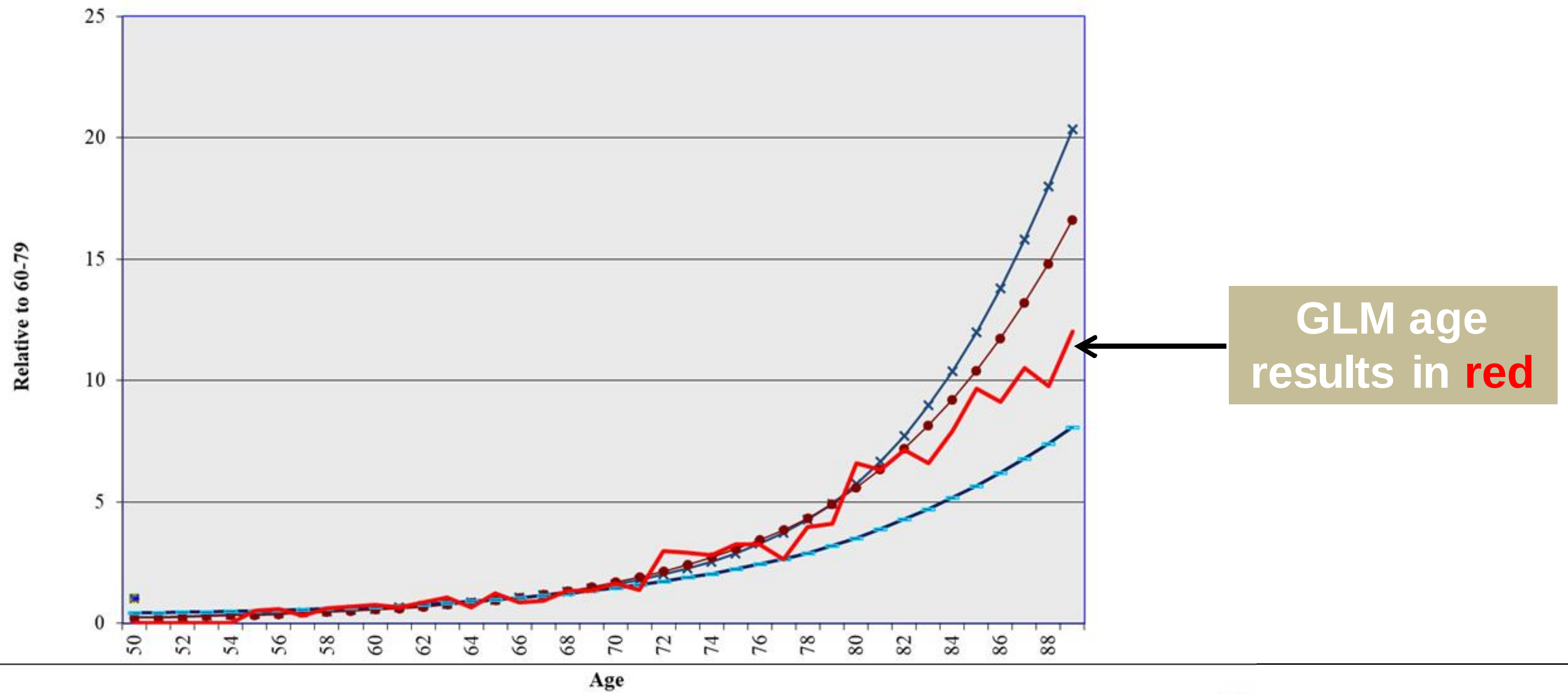
Effect of Age on mortality after smoothing

(in conjunction with effects of all other factors)



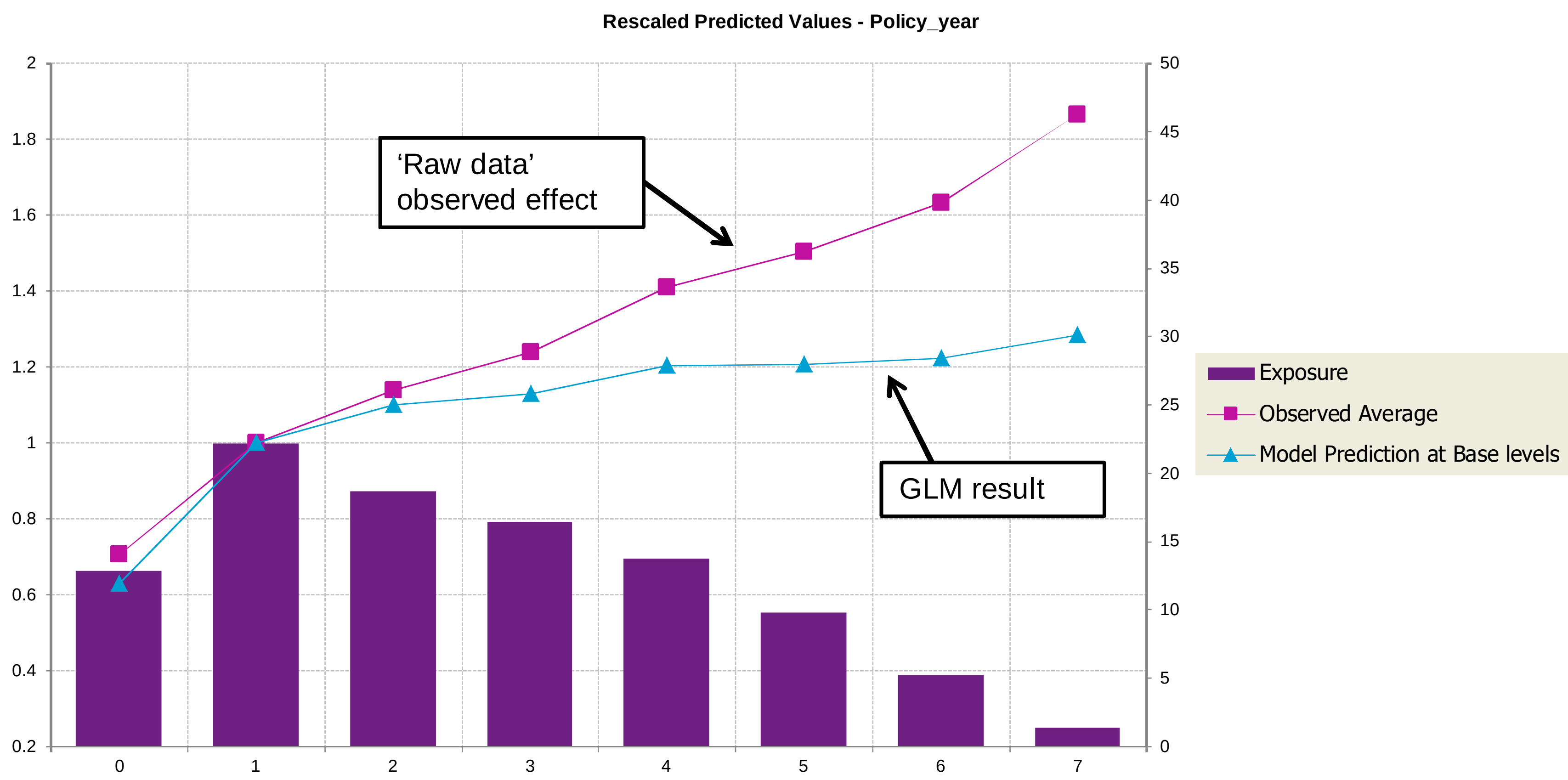
Allows us to define bespoke tables

- In the example below, the GLM age factor results clearly do not fit any of the three standard tables well – so we can create a bespoke table from the GLM.
- The smoothed curves are simple polynomials in log space, so aligns with Gompertz / Makeham but more flexible (for instance, we can use a spline to fit two curves)



Effect of duration on mortality

(in conjunction with effects of all other factors)



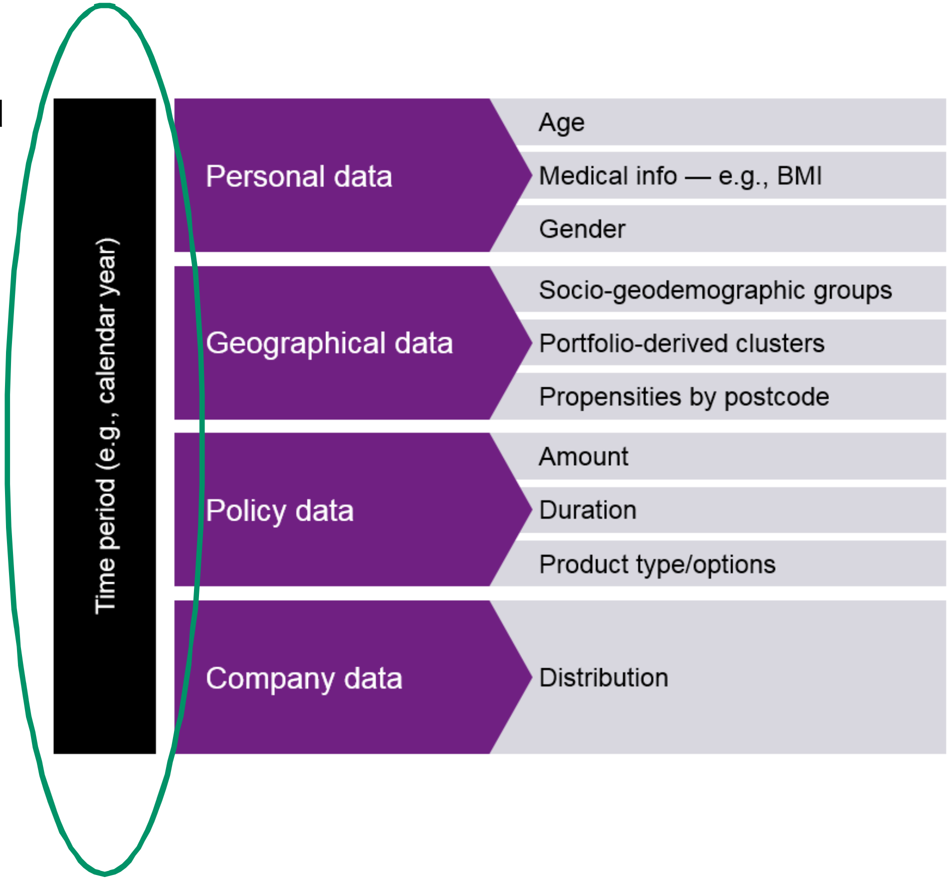
What has been the impact of this approach in UK market?

This approach has changed the way the UK market works

- In the UK at retirement (until 2014) everyone buys an annuity
- Three parts of the annuity market:
  - Normal individual annuities
  - Enhanced annuities
  - Bulk purchase annuities (insurer takes on liabilities of a pension scheme)
- Circa 2005 everyone got the same annuity rate.
- Since then all major firms introduced more segmentation and more individualised rates – postcode has been the main 'instrument'
- Enhanced and impaired lives annuities (reflecting health of the policyholder) also rated using such techniques – firms hold a great deal of information on the individuals
- *One of the largest UK annuity insurers has moved their reserving basis to a 'per policy' mortality curve approach – giving much greater accuracy in the future cash flow patterns*

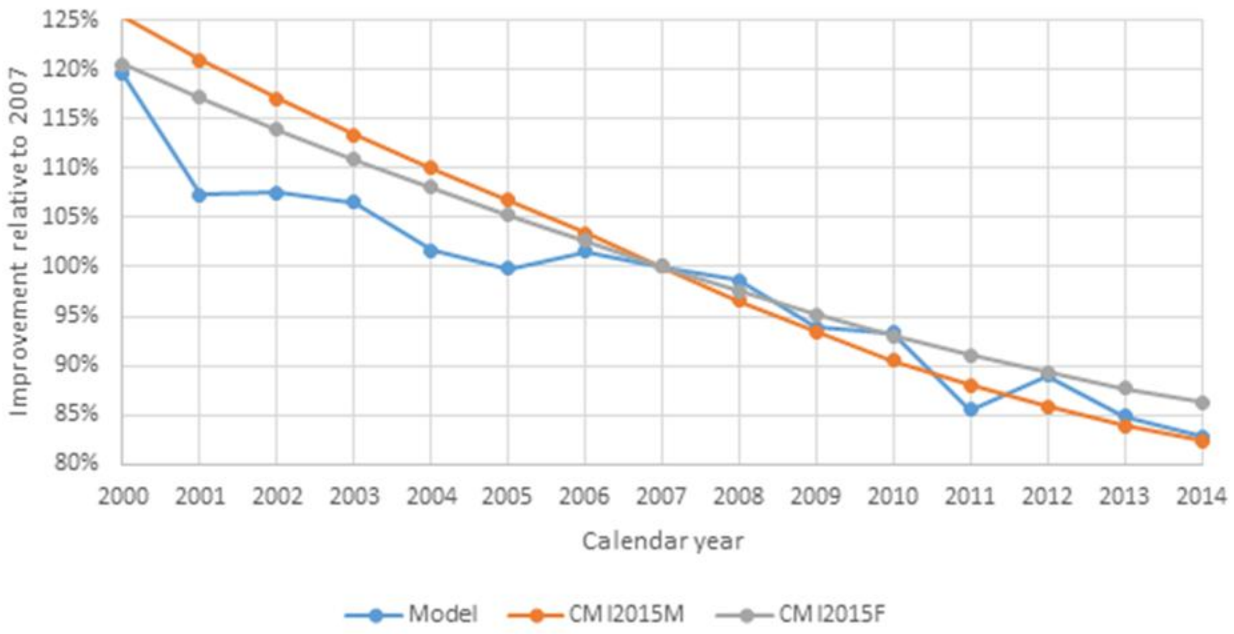
Mortality / longevity trend

- GLMs provide a multifactor perspective, and one of the most important factors in mortality analyses is the time period – typically using calendar year
- This allows us to quantify the mortality trend over time, taking account of all other factors in the model
- Simpler trend analyses are likely to provide misleading results because of the influence of other factors that may be changing over the period
- The GLM approach allows us to identify and quantify the trend as it relates to the specific portfolio, and this trend is often different from population mortality trends



Mortality / longevity trend (2)

The graph below shows a comparison of portfolio improvements against population improvements (as smoothed in CMI model)



We can extract the output and 'play with it' in Excel, or fit a parametric curve directly in the GLM

Policyholder behaviour

Factors typically found predictive in life GLM retention analyses

Time factor (calendar year) is also an important part of the model – picks up time trends

Personal data
- Age / Gender
- Other factors eg single/joint life?

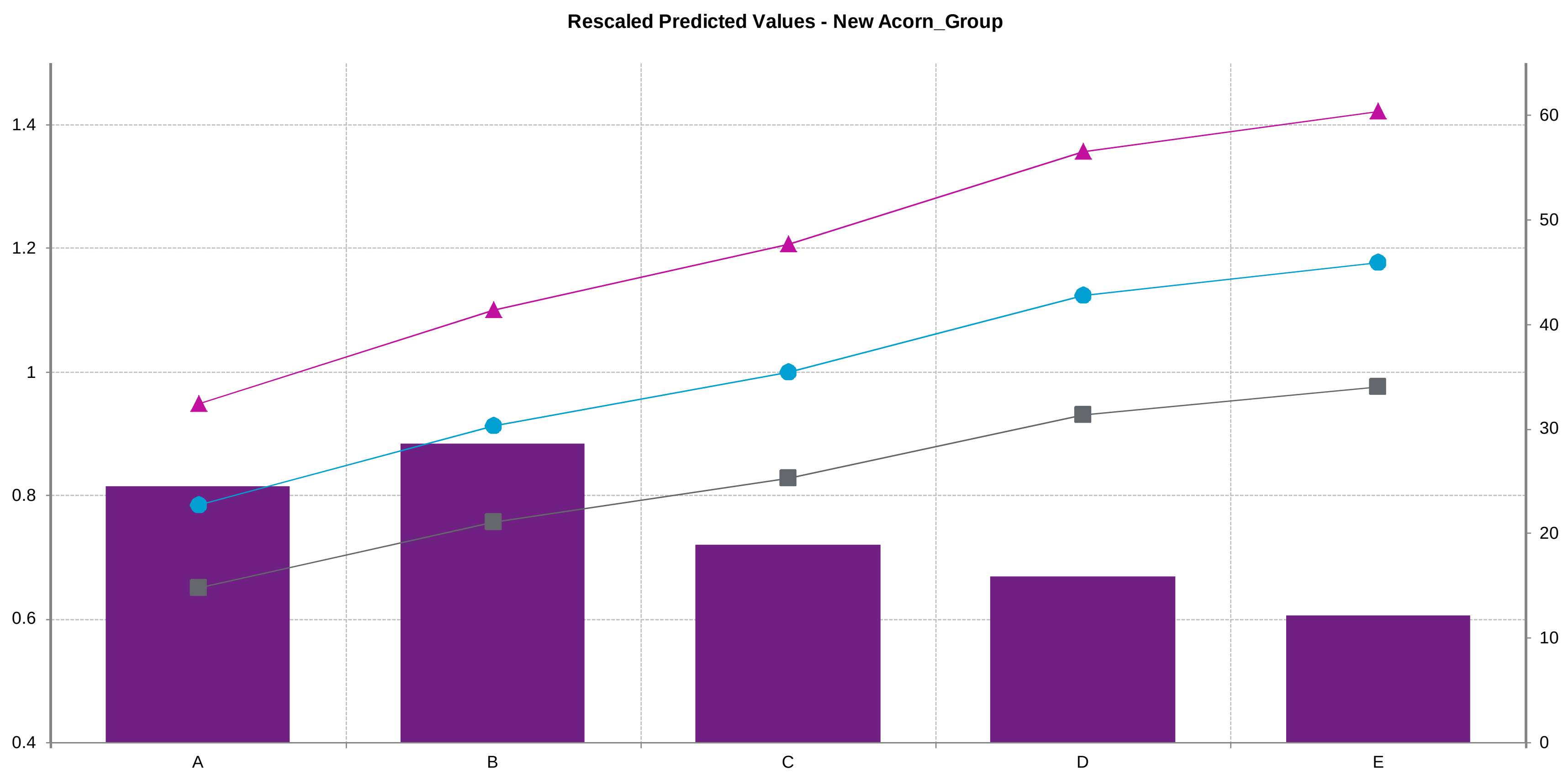
Geographical data
- Region
- Socio-demographic groups
- e.g. 3rd party postcode groupings

Policy data
- Policy duration
- Amount of benefit or premium
- Moneyness of guarantee/option

Company data
- Distribution
- Other policies/riders

Effect of Lifestyle (ie socio-economic) groups on surrender/lapse

(in conjunction with effects of all other factors)



Ways to use multi-factor persistency analyses to help the business

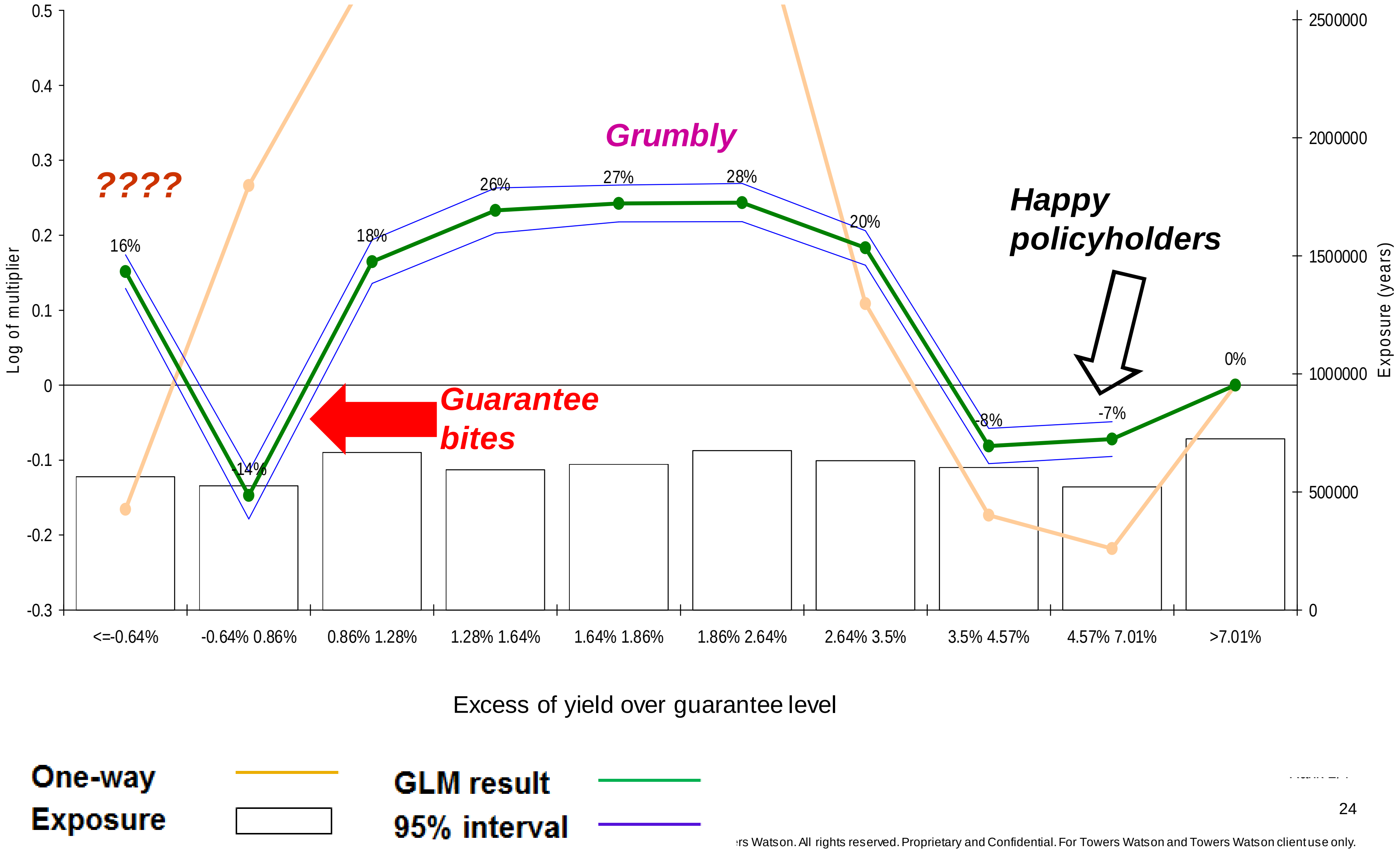
- More accurate and insightful forecasting of which policies are likely to surrender soon
- Cross-index high ‘about to surrender’ probabilities with policy value metric to identify which subsets to ‘save’ (eg offer instant maturity value increase on surrender request)

- Better assumptions for liquidity management (bespoke per-policy ‘off’ curves)
- Allows more accurate projections of cash requirements when joined with the per-policy £ amount

- Incorporating information on market indices into the analysis provides insight into dynamic relationships between markets and policyholder behaviour
- Improve consistency between economic assumptions and policyholder behaviour

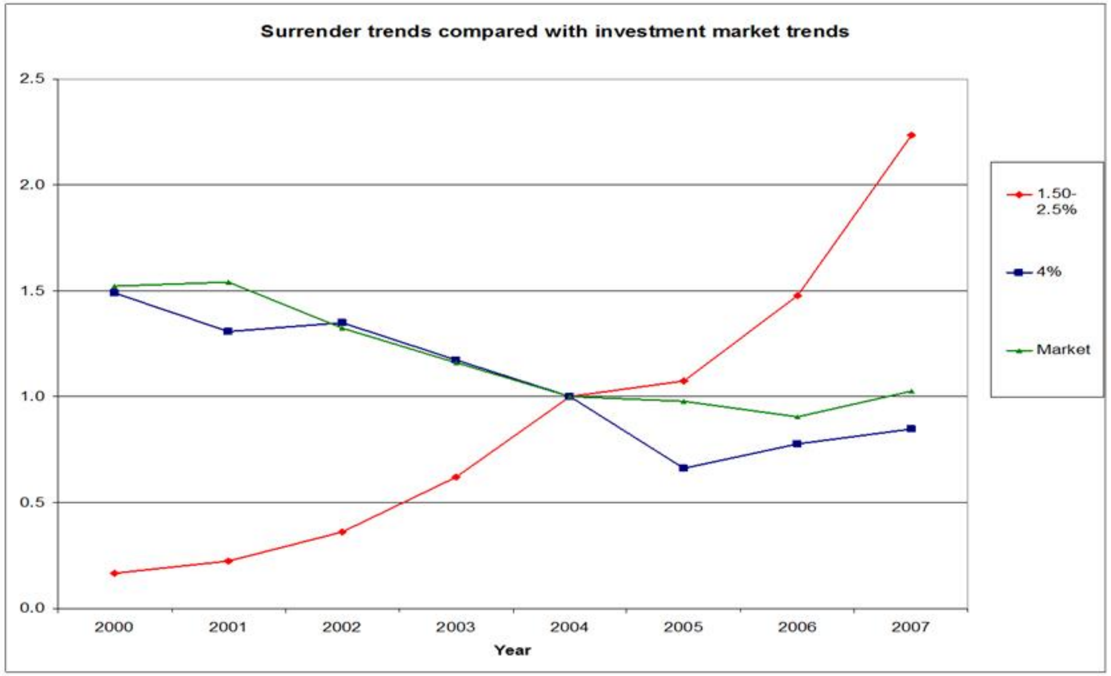
Note – these are in addition to enhancement of financial reporting

Life savings behaviour – market variations



Dynamic policyholder behaviour

- GLMs also allow us to investigate dynamic relationships between surrender rates and investment conditions
- The graph here shows the calendar year GLM result for high and low guarantee products compared with the yields in those years



- For high guarantees, market decreases lead to decreased surrenders – perhaps because policyholders value their guarantees more.
- For instance the results above could give us the following dynamic relationship to use in stochastic models:

For high guarantee products, multiply rates by {yield / 3.6%}

Proxy models

Using GLMs for new life applications

- GLMs provide a method to model
  - ***Some number***
- As a function of
  - ***Some factors***
- Previous examples related to frequency analysis
- We can also use GLMs to model amounts
  - Similar to claim amount analysis in P&C
- We can model amounts such as:
  - PVFP of a policy as a function of policy characteristics (age, duration …)
  - Economic capital of a portfolio as function of economic variables
  - Useful both as a check of the original ‘proper’ model and to create a simple proxy

Modelling year 1 capital as a function of ESG outputs

Simulation	Equity return	Property return	Pc1	Pc2	Pc3	Change in credit spread AAA	AA	A	Capital yr 1
1	-18.1%	-11.3%	1.1728	-0.0694	-0.0764	0.14%	0.19%	0.19%	351,956,232
2	37.5%	28.4%	-0.8093	0.1426	0.0376	0.15%	0.19%	0.20%	182,869,264
3	-16.0%	12.9%	-1.1597	-0.2165	0.0151	0.10%	0.08%	0.09%	295,234,182
4	34.0%	0.9%	1.5612	0.3284	-0.0514	0.40%	0.57%	0.60%	273,541,440
5	-7.6%	42.1%	5.7572	0.1840	-0.2618	0.00%	-0.06%	-0.08%	132,504,095
6	21.9%	-19.8%	-4.3497	-0.1075	0.1720	0.23%	0.32%	0.34%	401,335,715
7	-28.9%	0.8%	2.4245	0.0486	-0.1218	0.11%	0.14%	0.15%	310,364,028
8	58.4%	13.0%	-1.9035	0.0247	0.0747	0.08%	0.05%	0.01%	192,173,551
9	11.5%	45.0%	-2.9855	-0.6720	0.0549	0.00%	-0.07%	-0.10%	188,914,076
10	1.0%	-21.5%	3.8398	0.9810	-0.0792	0.22%	0.30%	0.32%	303,942,207
11	-8.5%	3.5%	3.5653	0.7616	-0.0941	0.02%	-0.05%	-0.03%	221,069,505
12	22.9%	10.0%	-2.7940	-0.5229	0.0594	0.07%	0.06%	0.05%	276,782,151
13	4.2%	-12.3%	-2.3709	0.1354	0.1092	0.17%	0.22%	0.24%	355,975,365
14	8.1%	29.9%	3.0075	-0.0947	-0.1628	0.35%	0.52%	0.57%	223,679,045
15	-9.1%	-9.2%	0.7996	-0.5391	-0.1028	0.02%	-0.04%	-0.06%	327,166,411
16	23.8%	25.5%	-0.4267	0.6100	0.0772	-0.02%	-0.12%	-0.14%	151,541,793
17	14.8%	12.6%	-4.6239	-0.1730	0.1771	0.36%	0.41%	0.51%	338,591,627
18	-2.0%	1.2%	6.1837	0.2795	-0.2719	0.20%	0.40%	0.45%	245,649,584
19	-28.8%	-4.4%	1.2037	0.2253	-0.0464	0.01%	-0.06%	-0.09%	303,185,900
20	58.2%	19.2%	-0.8391	-0.1285	0.0094	0.14%	0.32%	0.31%	188,498,682

Case study: assisting with economic capital project

- Client's EC team wanted a more scientific way to select 'equally bad' economic parameters to generate a 1/200 result
- Scope was restricted to ESG inputs, but we could have applied same approach to wider range of inputs e.g. ESG + mortality, lapses etc.
- Using GLMs, we were quickly able to:
  - Check which factor results seemed sensible, and which looked wrong (for follow-up investigation by client)
  - Check the correlations between factors (e.g. equity/property?)
  - Better understand the 'explanatory power' of factors
  - Decide which factors to adjust and which to drop from model
  - Generate a closed-form solution for the end year 1 capital
  - Advise on a set of 'equally bad' parameters for the 1/200 tail

Using GLMs to inform EC work

- Capital results (end year 1) for 30,000 sets of ESG output generated by usual life model
- What we modelled:

Capital amount =

Base level ×

Factor 1 (based on change in AAA credit spread) ×

… × (other credit spread factors) × ….

… × (yield curve change factors) × …

Factor 7 (based on equity return) ×

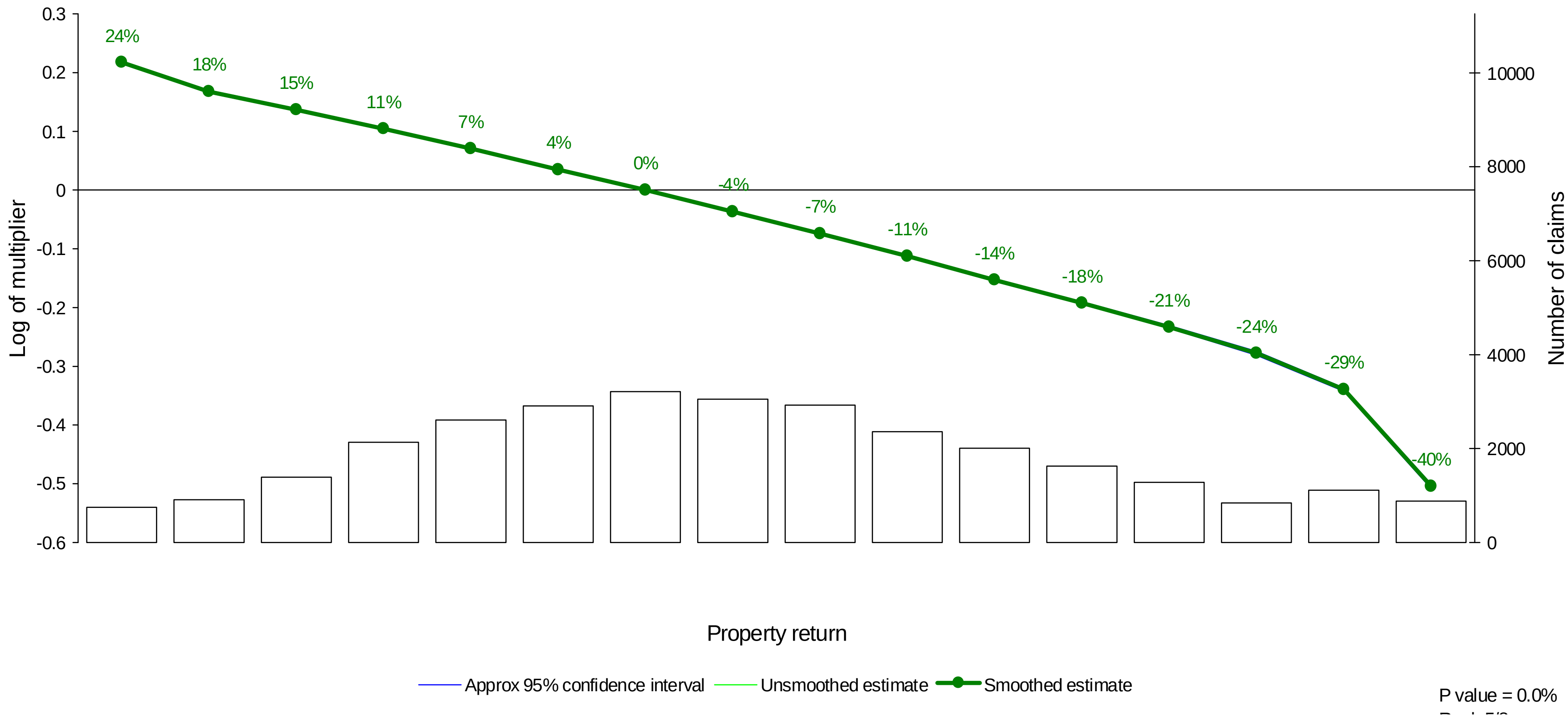
Factor 8 (based on property return) ×

GLMs for EC work: initial results (good)

Factor effect (property returns) has expected continuous and strong effect

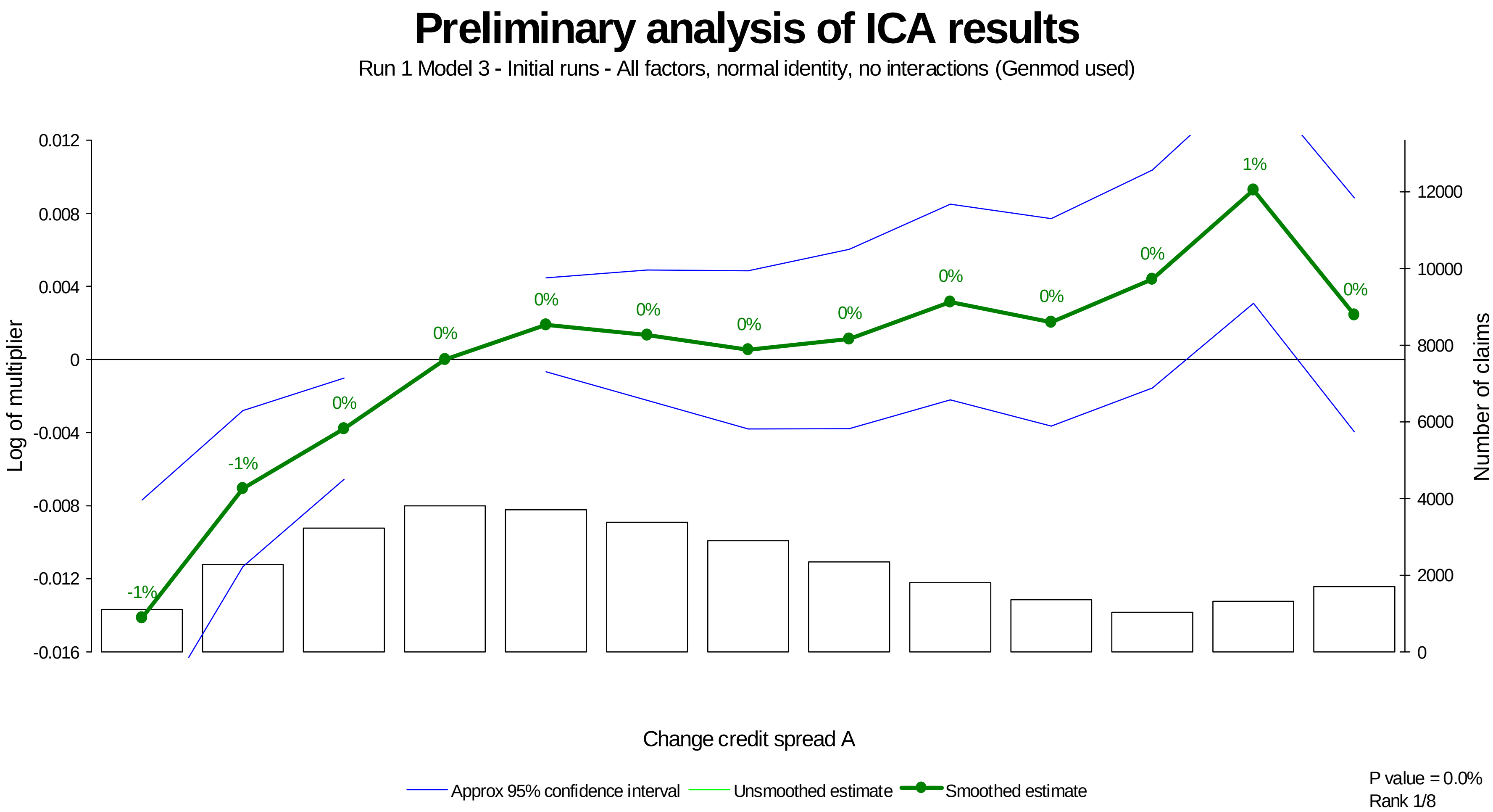
Preliminary analysis of ICA results

Run 1 Model 3 - Initial runs - All factors, normal identity, no interactions (Genmod used)



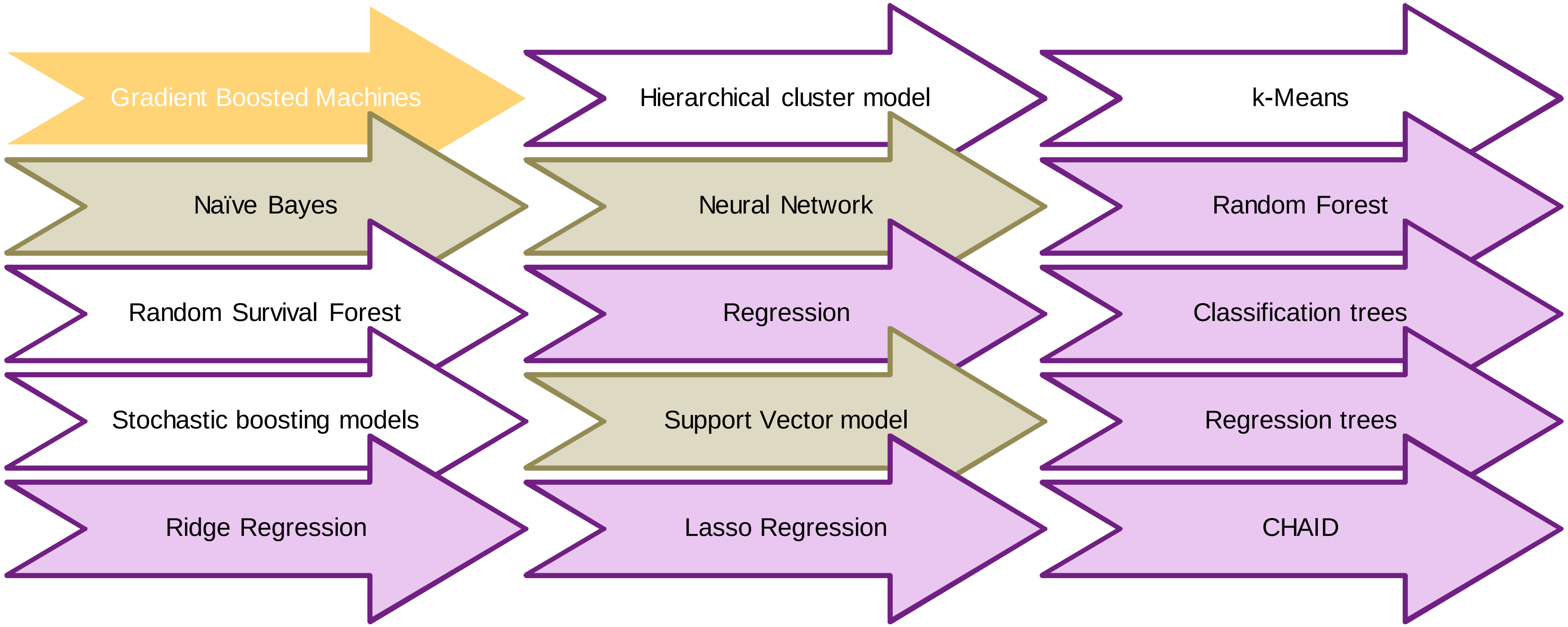
GLMs for EC work: initial results (bad)

GLM indicated a problem – checking in original model proved this was the case



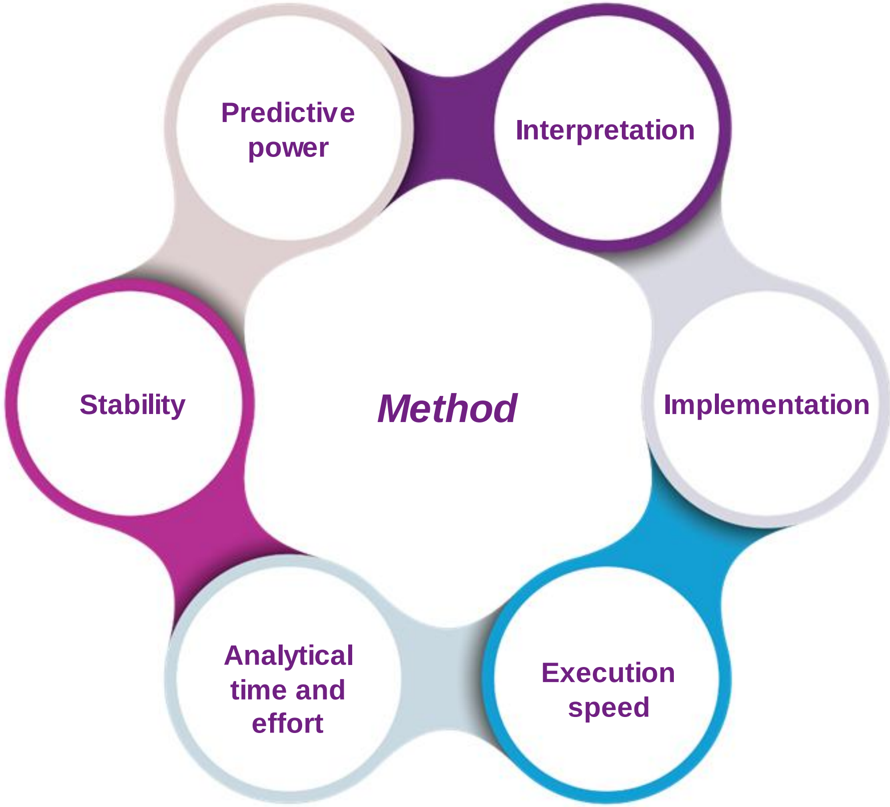
Other methods

A wide range of multifactor methods …

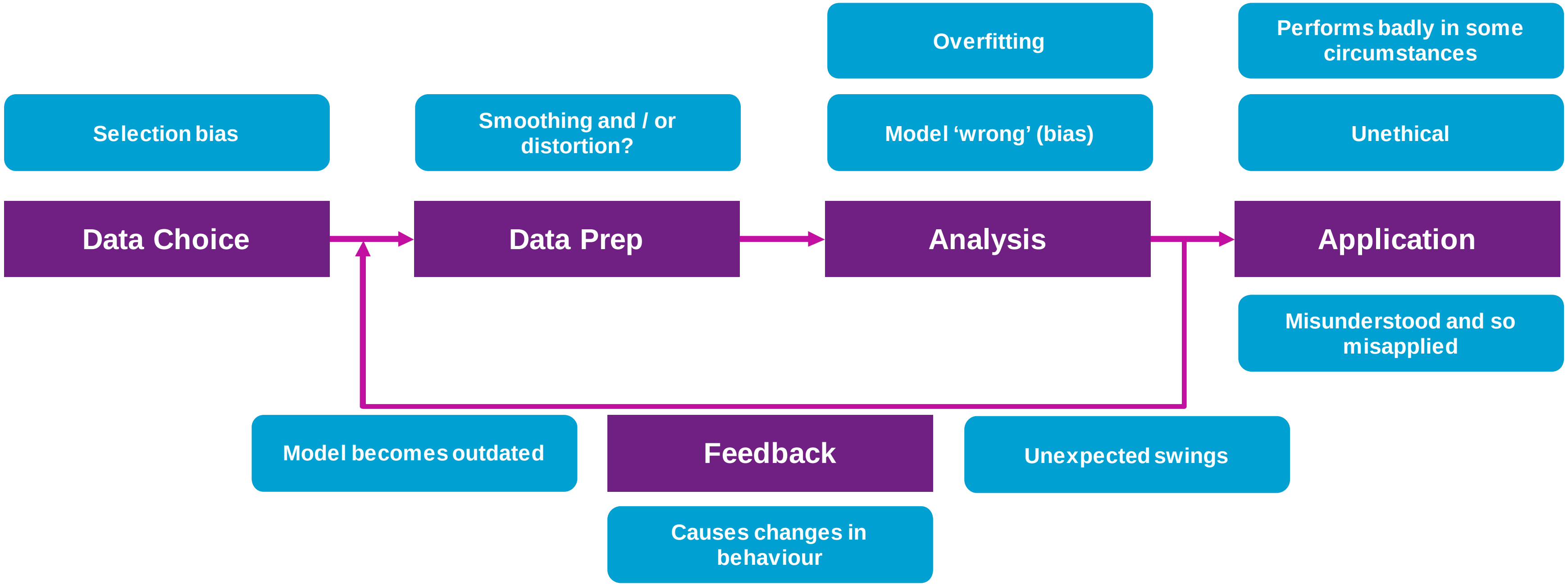


What technique is appropriate?

This depends on your criteria



What could go wrong?



https://sias.org.uk/events/2018-06-05-talk-analysing-analytics/

The end – Questions and Answers

Contact details

Matthew Edwards
Director, Insurance Consulting & Technology,
UK

T +44 1737 284771
M +44 7908 212317

Watson House
London Road
Reigate, UK

matthew.fj.edwards@willistowerswatson.com