Institute
and Faculty
of Actuaries

Use of GLMs in a competitive market

Ji Yao and Simon Yeung, Advanced Pricing
Techniques (APT) GIRO Working Party

12 June 2013

About the presenters

Dr. Ji Yao is a manager with Ernst & Young's EMEIA insurance risk and actuarial services practice. He has extensive first-hand experience in various modelling for pricing, including risk models, demand models and price optimisation, with a solid background in mathematics and statistics. He is the chair of the Advanced Pricing Techniques (APT) GIRO working party.

JYao@uk.ey.com

Simon Yeung is currently a senior manager at Grant Thornton. Prior to joining Grant Thornton, Simon was the head of motor pricing at Saga for 3 years. Before that, he was a reserving manager at RBS Insurance for 3 years. Before joining RBS Insurance he worked for London market insurers, reinsurers and commercial insurers for four and half years. He is a member of the Advanced Pricing Techniques (APT) GIRO working party.

Simon.yeung@uk.gt.com

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Agenda

- Introduction
- Current market and uses of GLM
- Three overlooked facts of GLM and their implications
- Summary and Q&A

Introduction

- Advanced Pricing Techniques (APT) GIRO working party was created in 2012
- 22 members working in three work streams
  - GLM
  - Telematics pricing
  - Conversion/Elasticity modelling
- One workshop in GIRO 40 and one paper on GLM is being prepared
- We will focus on GLM in this presentation

Institute and Faculty of Actuaries

Current uses of GLM in the market

- Risk base pricing
- Cost plus approach
- Price optimisation

Institute and Faculty of Actuaries

Market Performance

Looking in more detail at:

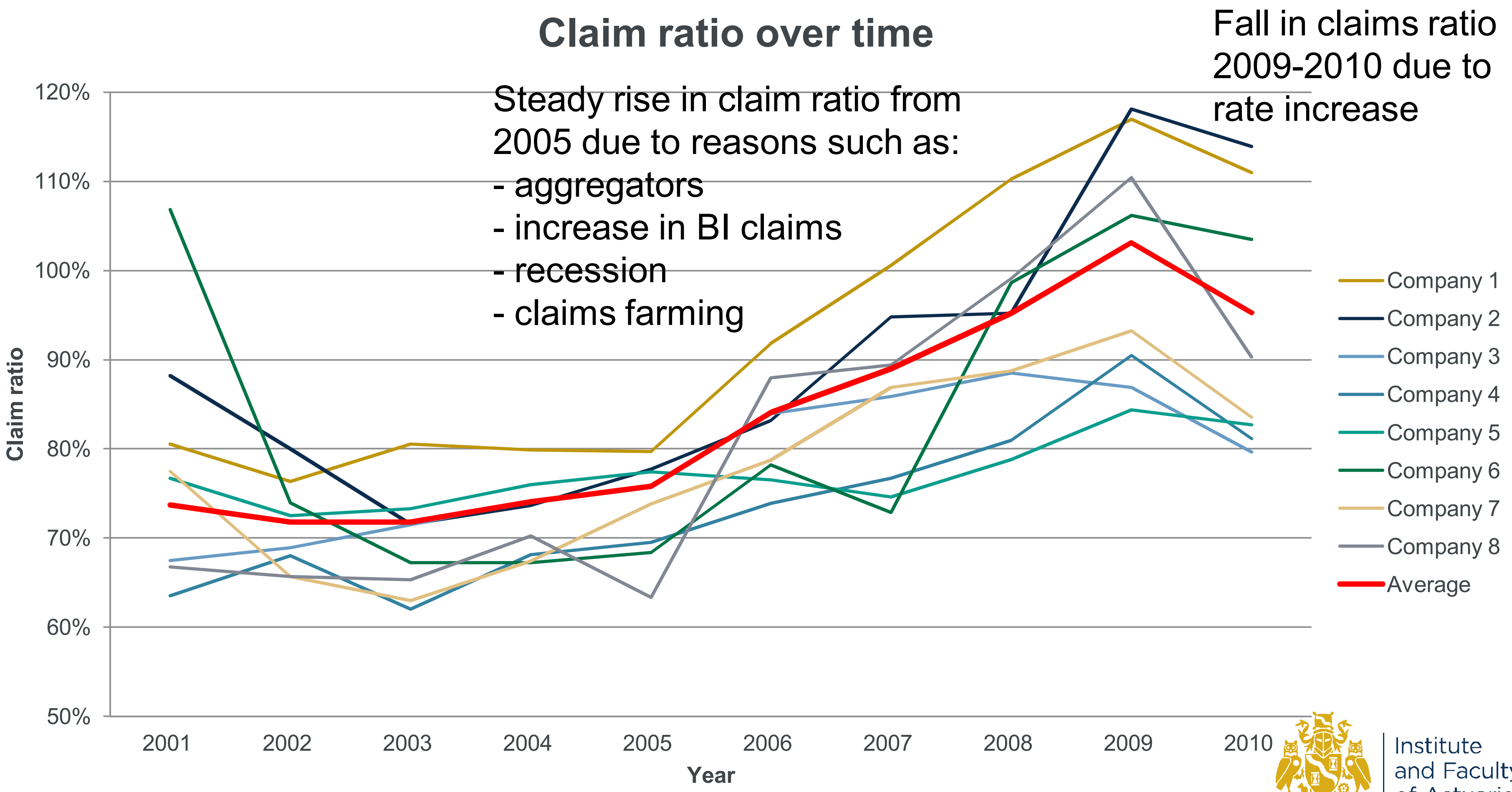
- Change in claim ratio and frequency over time
- What, if any, relationships can we derive between the two?
- How does this relate back to GLM modelling?

Data used:

- Cross section of market (8 companies)
- Totalling £4.7bn earned premium in 2010
- High level data taken from FSA returns

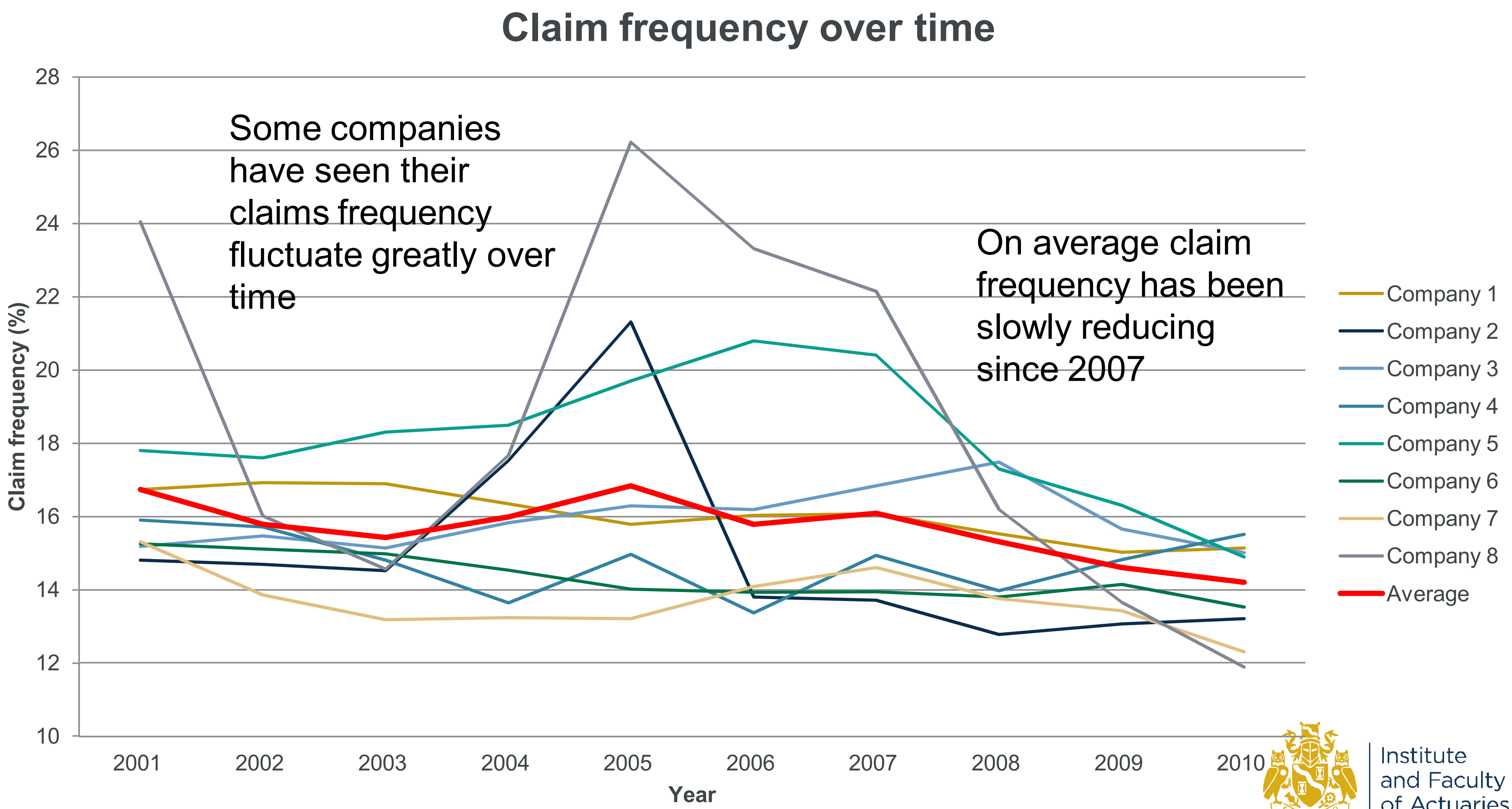
Institute and Faculty of Actuaries

Claim ratio over time



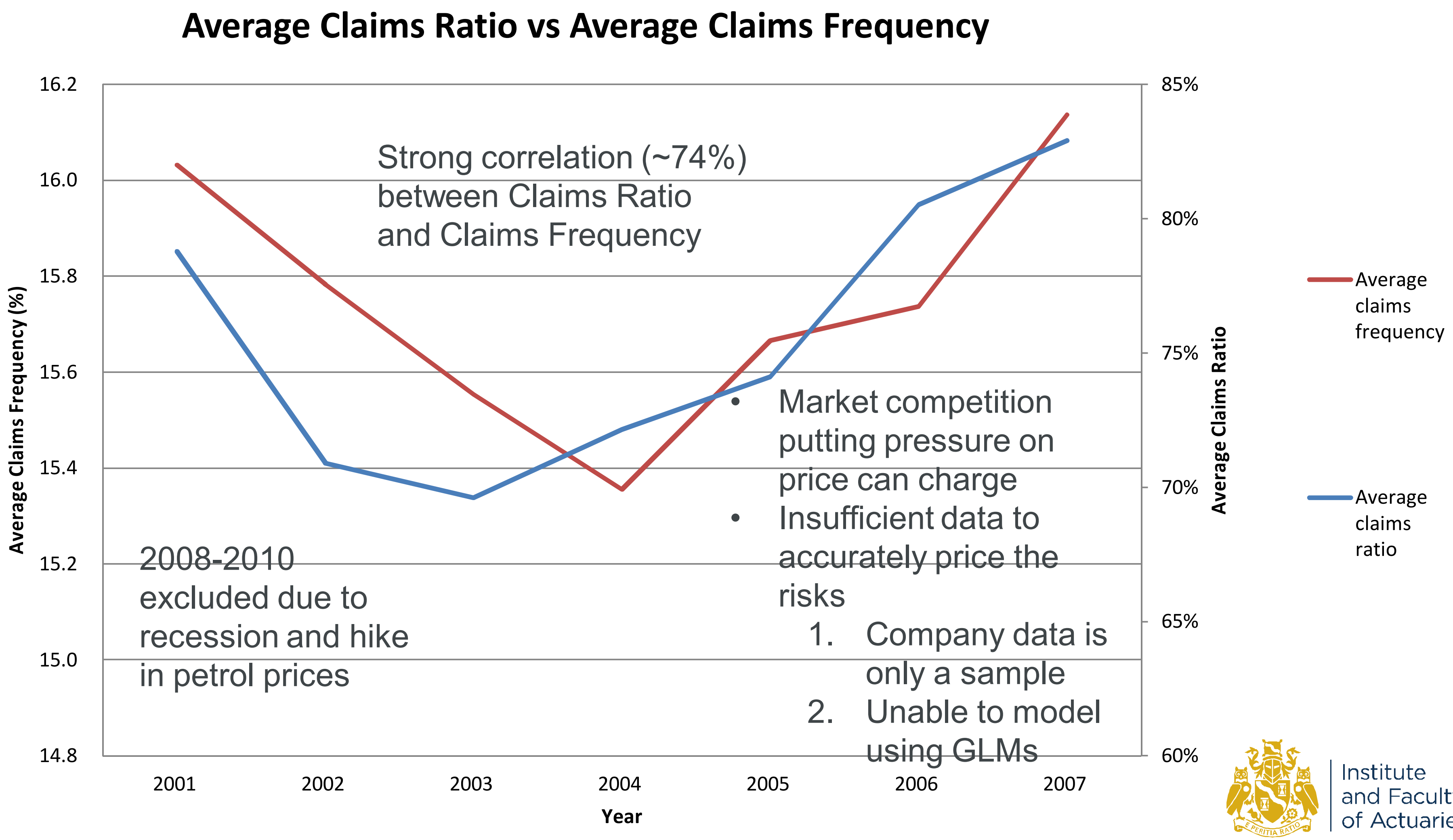
Institute and Faculty of Actuaries

Claim frequency over time

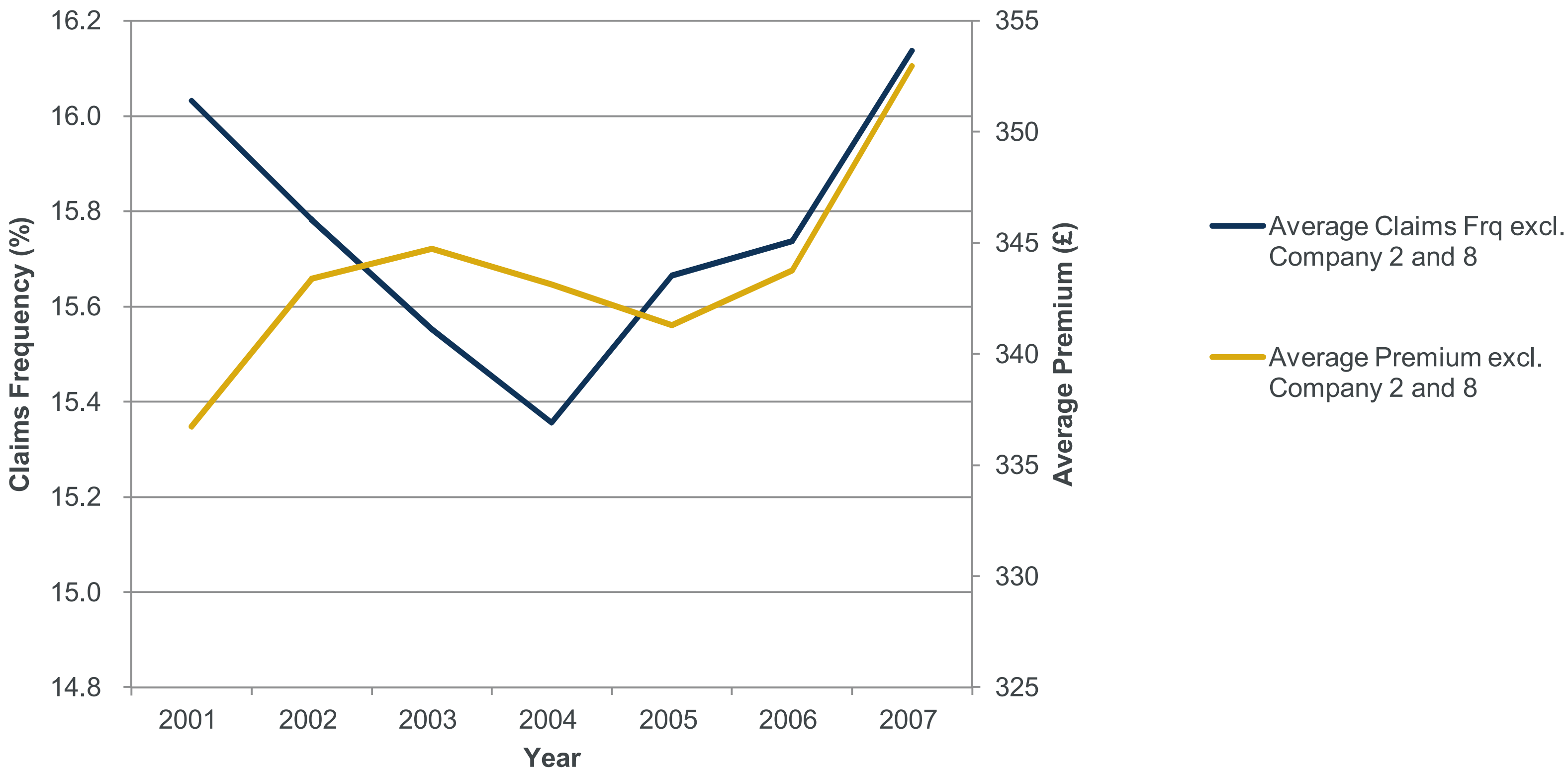


Institute and Faculty of Actuaries

Correlation between Frequency and ULR

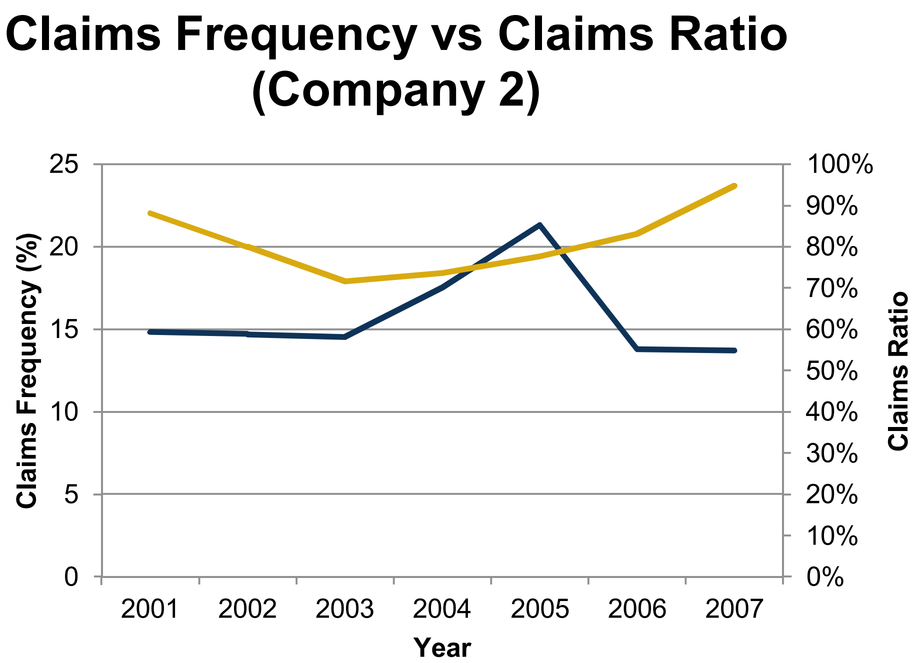


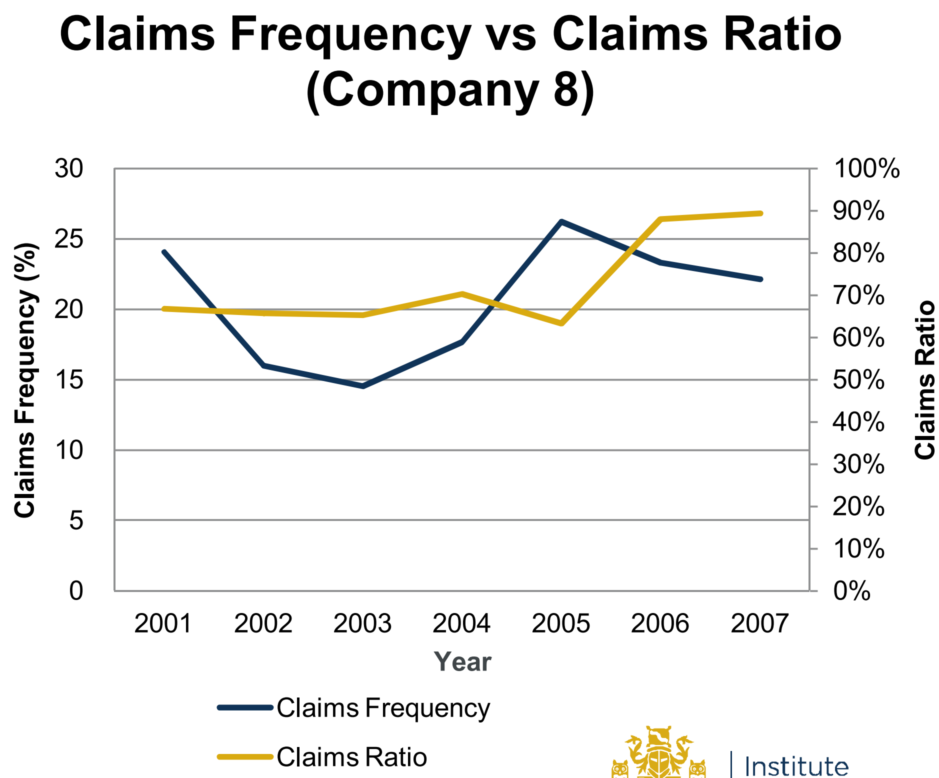
Frequency vs Average Premium



Change in mix of business

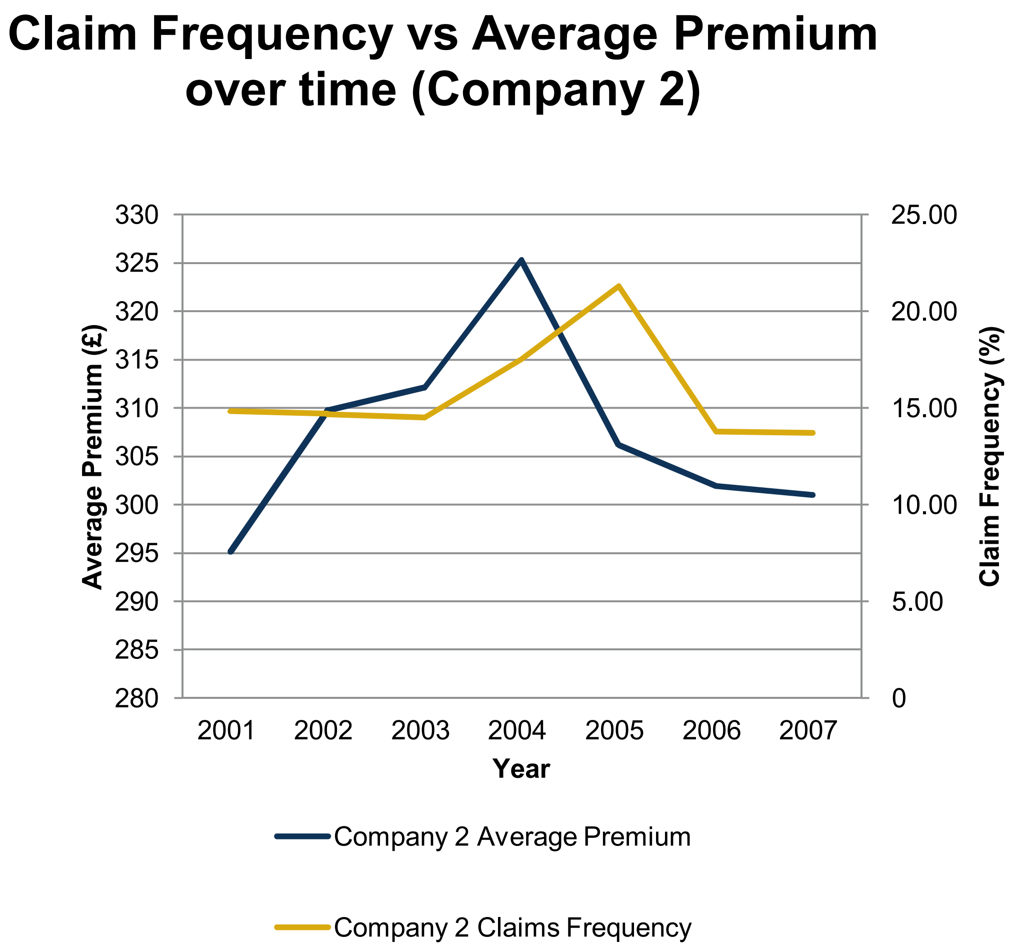
- Fluctuations in frequency due to changes in mix of business

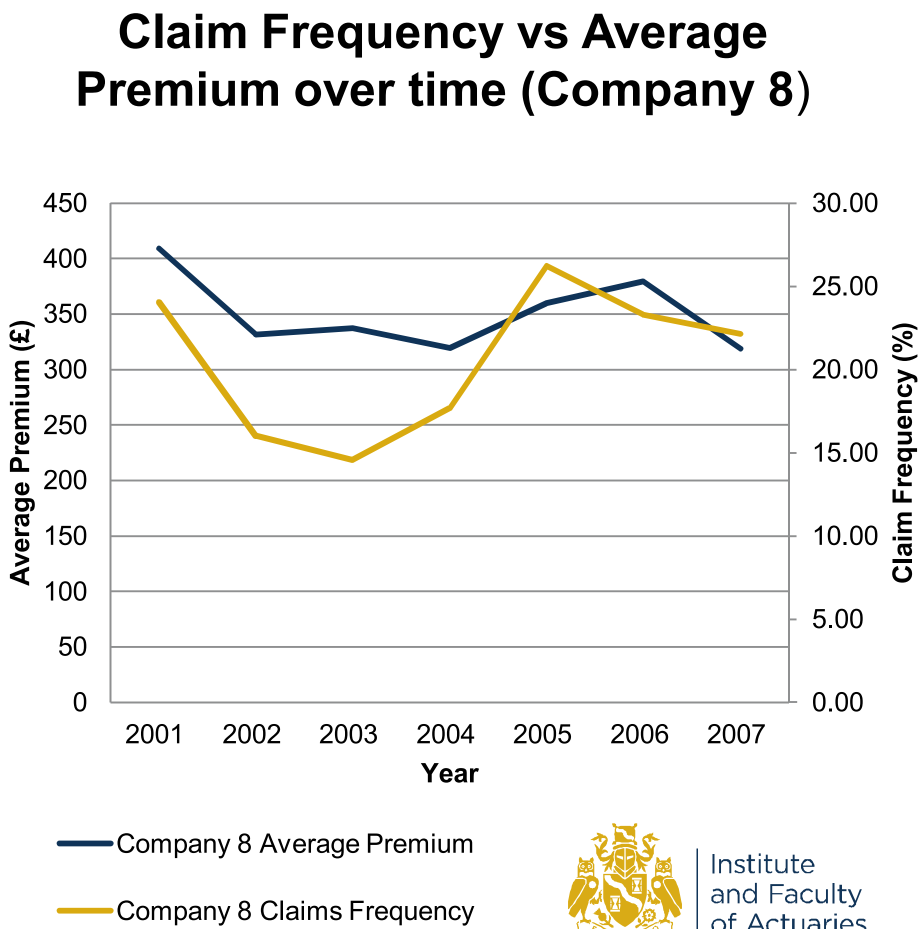




Institute and Faculty of Actuaries

Claim Frequency vs Average Premium





Institute and Faculty of Actuaries

There are a wide range of quoted premium on the market, while GLMs are used as a standard pricing technique throughout market.

	Annual Premium	Monthly Premium	Excess	Legal Cover	Courtesy Car	Breakdown Cover
esure	£598.89	1 x £104.80 10 x £54.75 £652.30	Vol: £300 Comp: £0 Excess: £300	£26.99 extra	✓	From £36.75
Sheilas' Wheels	£598.97	1 x £104.89 10 x £54.75 £652.39	Vol: £300 Comp: £0 Excess: £300	£26.99 extra	✓	From £36.75
Admiral	£806.66	1 x £102.96 11 x £102.85 £1234.31	Vol: £350 Comp: £0 Excess: £350	✓	✓	From £34.95 extra
Diamond Shaped for girls	£814.08	1 x £104.25 11 x £104.19 £1250.34	Vol: £350 Comp: £0 Excess: £350	✓	✓	From £34.95 extra
KWI	£1110.85	Check with provider	Vol: £300 Comp: £120 Excess: £420	£18.00 extra	✓	From £49.99
Nationwide On your side	£1206.82	Instalments not available	Vol: £300 Comp: £0 Excess: £300	£24.90 extra	✓	From £27.00

	Annual Premium	Monthly Premium	Excess	Legal Cover	Courtesy Car	Breakdown Cover
FLUXDIRECT	£1425.95	1 x £217.64 9 x £156.21 £1623.53	Vol: £300 Comp: £120 Excess: £420	✓	✓	Extra £59.00
TESCO Bank	£2658.99	1 x £443.19 10 x £248.17 £2924.89	Vol: £300 Comp: £150 Excess: £450	✓	✓	From £29.50
insure 2drive	£3113.05	1 x £622.61 10 x £275.37 £3376.31	Vol: £300 Comp: £150 Excess: £450	✓	✓	✗
Ladybird Insurance	£3782.07	1 x £761.41 9 x £385.78 £4233.43	Vol: £0 Comp: £3000 Excess: £3000	£25.00 extra	✓	£49.50 extra
Express Insurance	£4334.85	1 x £655.48 11 x £401.83 £5075.61	Vol: £300 Comp: £150 Excess: £450	£39.50 extra	✗	£70.75 extra
Drive	£7212.41	1 x £1081.86 9 x £777.80 £8082.06	Vol: £300 Comp: £120 Excess: £420	£15.00 extra	✓	From £60.00

Quotes for a 30 year old male with a clean license held for 10 years, for a 57 plate manual 1.6L ford focus style 5 door hatchback. Car is kept at home parked on the road, for social use only, approx 9000 annual mileage

Institute and Faculty of Actuaries

	Annual Premium	Monthly Premium	Excess	Legal Cover	Courtesy Car	Breakdown Cover
esure	£200.09	1 x £35.04 10 x £18.29 £217.94	Vol: £250 Comp: £150 Excess: £400	£30.00 extra	✓	From £36.75
M&S Bank	£259.73	1 x £38.96 11 x £22.27 £283.93	Vol: £250 Comp: £350 Excess: £600	From £24.99	✓	From £29.99
POST OFFICE	£294.23	1 x £58.94 11 x £24.28 £326.02	Vol: £250 Comp: £350 Excess: £600	✓	✓	£29.99 extra
HALIFAX	£322.72	1 x £48.49 11 x £27.88 £355.17	Vol: £250 Comp: £350 Excess: £600	£25.99 extra	✓	From £39.99
autonet CAR INSURANCE	£323.46	1 x £73.75 10 x £31.41 £387.85	Vol: £250 Comp: £350 Excess: £600	£39.95 extra	✓	✗
ASDA money	£376.59	1 x £102.82 11 x £31.36 £447.78	Vol: £250 Comp: £275 Excess: £525	✓	✓	Check with provider

	Annual Premium	Monthly Premium	Excess	Legal Cover	Courtesy Car	Breakdown Cover
AA	£377.43	1 x £37.74 11 x £35.41 £427.25	Excess: £370	£25.99 extra	✓	✗
ZURICH	£469.68	1 x £61.92 11 x £42.63 £530.85	Vol: £250 Comp: £150 Excess: £400	£28.99 extra	✓	£36.99 extra
bullseyeinsurance.co.uk Part of the Swinton Group	£505.64	1 x £48.25 11 x £48.25 £579.00	Vol: £250 Comp: £275 Excess: £525	From £30.00	✓	From £40.41
GA GENERAL ACCIDENT	£757.70	1 x £126.31 10 x £71.29 £839.21	Vol: £250 Comp: £50 Excess: £300	Optional (£27.00)	✓	Optional (from £33.18)
Drive	£1378.12	1 x £206.72 9 x £149.64 £1553.48	Vol: £250 Comp: £350 Excess: £600	£15.00 extra	✓	From £60.00
QUINN direct	£1702.49	1 x £363.25 10 x £145.28 £1816.05	Excess: £250	£22.00 extra	✗	✓

Quotes for a 40 year old married female with a clean license held for 15 years for a 59 diesel Golf GTD 2.0L 3 door hatchback. Car is kept at home and parked on a driveway for social use only, approx 7000 miles

Institute and Faculty of Actuaries

Current market and uses of GLM

- GLM is a standard approach for risk pricing and price optimisation
- Wide range of price for individual quote
- Wide range of performance for market player
- What causes the difference?

GLM technical details

A GLM consists of the following three components:

1. Random component

Each component of $\underline{Y}$ is independent and is from one of the exponential family of distributions.

2. Systematic component

A linear combination of the estimated parameters gives the linear predictor, $\underline{\eta}$:

$$\underline{\eta} = \mathbf{X}.\underline{\beta}$$

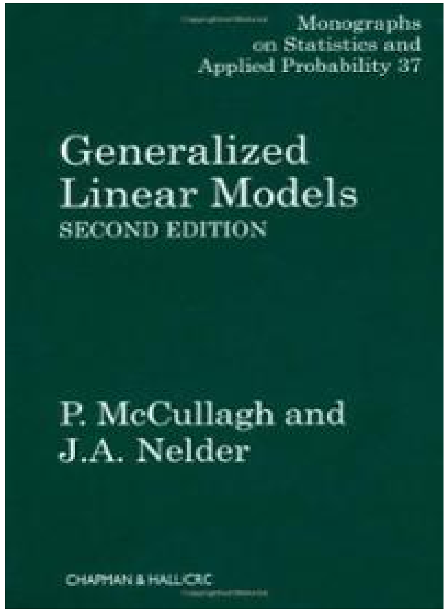
3. Link function

The relationships between the random and systematic components is specified via a link function, *g*, such that:

$$E[\underline{Y}] \equiv \underline{\mu} = g^{-1}(\underline{\eta})$$

4. Data

The dataset that GLM trained on.



Data

$$X$$
$$\downarrow$$
$$\eta = X \cdot \beta$$
$$\downarrow$$
$$E(Y) \equiv \mu = g^{-1}(\eta)$$

Institute and Faculty of Actuaries

Three overlooked facts of GLM

1. GLMs put either zero or full credibility into data
2. GLMs implicitly use median from the distribution of prediction
3. GLM results depend on the mixture of rating variables in the data

Institute and Faculty of Actuaries

Quiz 1: Average weight of yellow balls

- There is a bag of coloured balls. You sampled a few of them from the bag and obtained the following information:

Colour	Avg weight (kg)
Yellow	6
Red	4
Average All	5

- What is your estimation of the average weight of yellow balls?

A) Use average of yellow balls ONLY – 6kg

B) Use average of ALL balls – 5kg

C) Blended average weight of yellow balls and non-yellow balls

D) Other (with suggestions)

Institute and Faculty of Actuaries

GLM fact 1: GLMs put either zero or full credibility into data

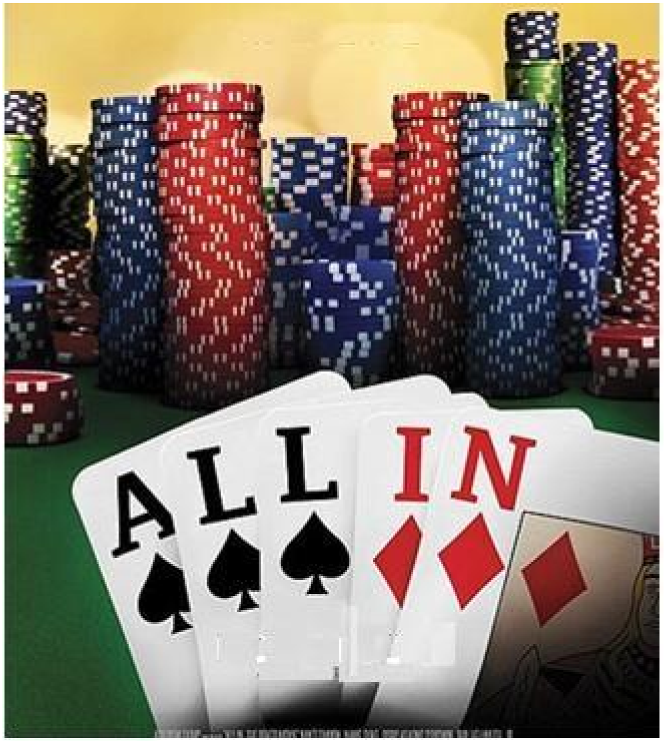
Data X

$$\downarrow$$

$$\eta = X \cdot \beta$$

$$\downarrow$$

$$E(Y) \equiv \mu = g^{-1}(\eta)$$



Institute and Faculty of Actuaries

A gradual approach to include data is needed in modelling

Data X

$$\downarrow$$
$$\eta = X \cdot \beta$$
$$\downarrow$$
$$E(Y) \equiv \mu = g^{-1}(\eta)$$

Sample 6 balls from the bag of yellow and red balls, and we obtained these weights:

Colour	Avg weight (kg)
Yellow	4
Yellow	6
Yellow	8
Red	2
Red	4
Red	6

Would you suddenly change your view because of the additional six balls?

Keep sampling and if we get 6 more identical balls as before:

Colour	Avg weight (kg)
Yellow	4
Yellow	4
Yellow	6
Yellow	6
Yellow	8
Yellow	8
Red	2
Red	2
Red	4
Red	4
Red	6
Red	6

Testing the ‘colour’ factor in a GLM shows that Yellow is not significantly different from Red at 95% confidence level (p-value=0.1336).
Avg weight of yellow balls = 5

Testing the ‘colour’ factor in a GLM shows Yellow is now significantly different from Red at 95% confidence level (p-value=0.0339).
Avg weight of yellow balls = 6

Institute and Faculty of Actuaries

An important implication is GLMs tend to push relativities and hence price towards extreme levels

Data

$$X$$
$$\downarrow$$
$$\eta = X \cdot \beta$$
$$\downarrow$$
$$E(Y) \equiv \mu = g^{-1}(\eta)$$

5kg

5.Xkg

6kg

- As the normal GLM practice is to calibrate the base rate after relativities are calculated, extreme relativities will result in more policies being priced at very low (or high) end
- Over-priced policies never get converted in a competitive market, so insurers are exposed to big under-pricing risk
- Linked to the observed diversified quoted premium on the market

Institute and Faculty of Actuaries

Generalised linear mixed models (GLMMs) provide a potential solution

- GLMMs are an extension to GLM, in which the linear predictor contains random effects to allow for correlation of the data in addition to the usual fixed effects.

- It provides a convenient way of applying credibility blending within GLM.

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \epsilon_i$$

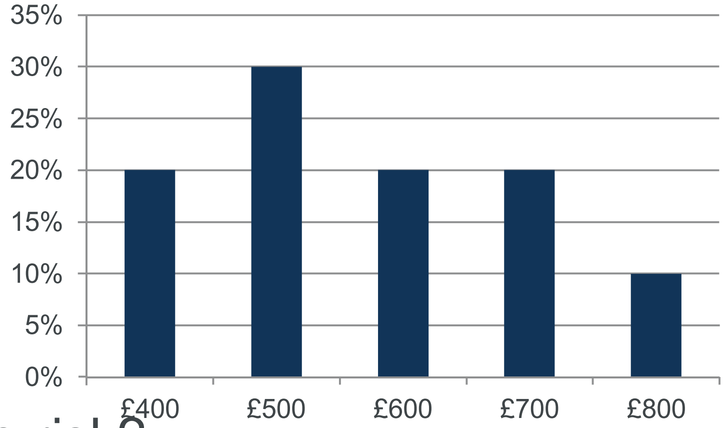
Random effect

Institute and Faculty of Actuaries

Quiz 2: Mean, median or mode? – a question not only relevant to reserving or capital

- A pricing analysis gives a range of possible prices for a risk as shown in the table below:

Price	Probability
£400	20%
£500	30%
£600	20%
£700	20%
£800	10%



- What is the price you will charge for the risk?

A) Mode - £500

B) Median - £550

C) Mean - £560

D) Other (with suggestions)

Institute and Faculty of Actuaries

GLM fact 2: GLMs implicitly use median from the distribution of prediction

Data X

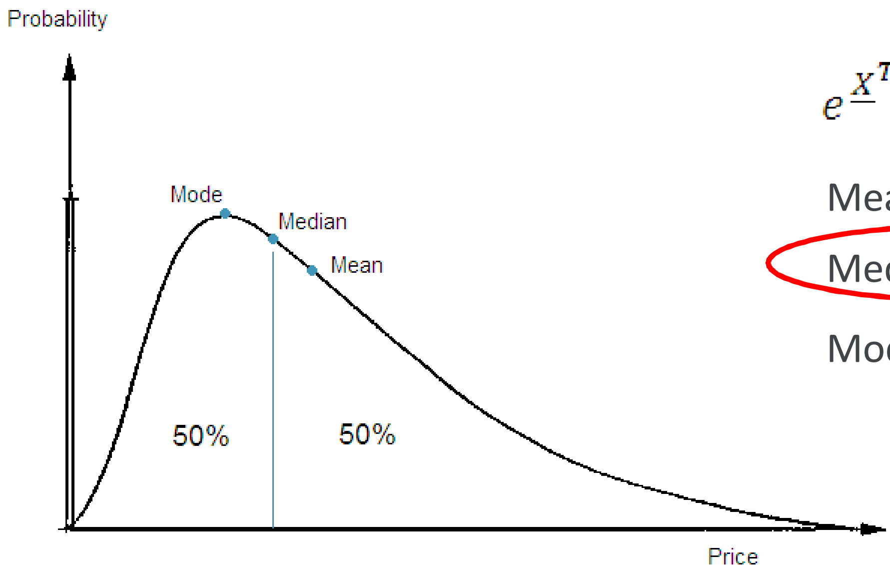
$\downarrow$

$\eta = X \cdot \beta$

$\downarrow$

$E(Y) \equiv \mu = g^{-1}(\eta)$

- The linear predictor $\Sigma X_i \beta_i$ is asymptotically normally distributed as all β_i are asymptotically multivariately normally distributed
- After the link function transformation, the prediction is no longer normally distributed.
- Take log link as an example:



$$e^{X^T\beta} = e^{\sum X_i \beta_i} \quad \text{where} \quad \beta_i \sim N(\widehat{\beta_i}, \sigma^2)$$

Mean: $E\left[e^{\sum X_i \beta_i}\right] = e^{\left[\sum X_i \widehat{\beta_i} + \frac{1}{2}\mathrm{var}(\sum X_i \beta_i)\right]}$

Median: $e^{\sum X_i \beta_i}$

Mode: $e^{\left[\sum X_i \widehat{\beta_i} - \mathrm{var}(\sum X_i \beta_i)\right]}$

Institute and Faculty of Actuaries

Link function is the dominant factor in shaping the distribution of prediction

Data X

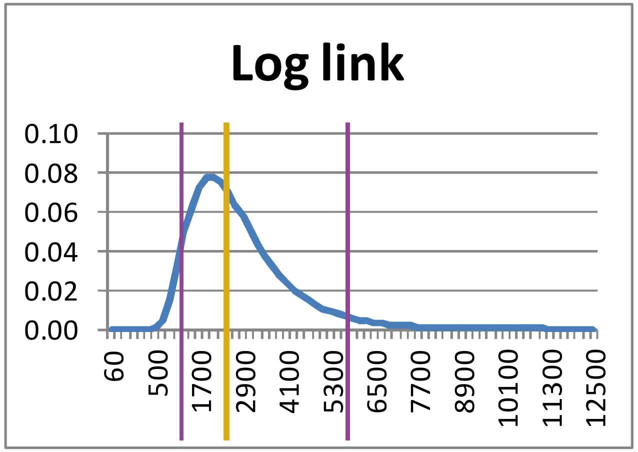
$\downarrow$

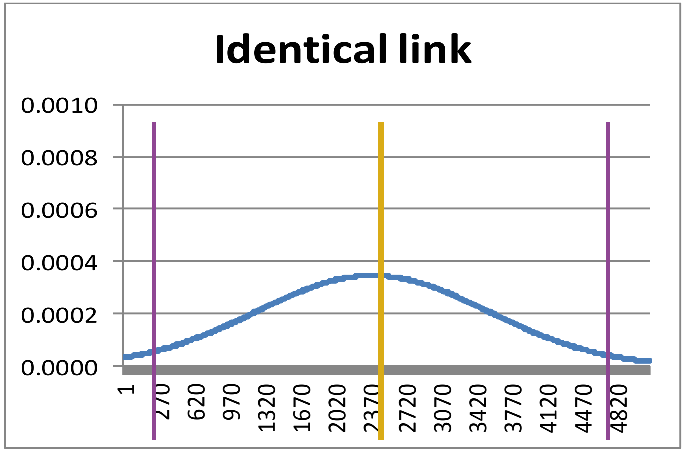
$\eta = X \cdot \beta$

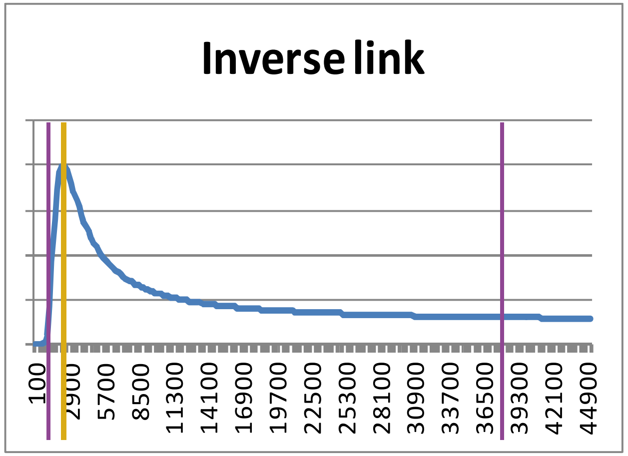
$\downarrow$

$E(Y) \equiv \mu = g^{-1}(\eta)$

Consider a severity model with Gamma error structure. Results for different link functions:







941 2400 6121

153 2400 4647

Lower bound Mean Upper bound

1240 2400 37656

These examples show that the upper and lower bounds could be very different, and the prediction is not always the mean of the distribution!

Institute and Faculty of Actuaries

Do GLMs systematically underestimate the cost?

Data

$$X$$
$$\downarrow$$
$$\eta = X \cdot \beta$$
$$\downarrow$$
$$E(Y) \equiv \mu = g^{-1}(\eta)$$

- For a distribution skewed towards the left, usually it is the case that Mode<Median<Mean, so the median used by GLMs is always lower than the mean
- To use mean, the term $\operatorname{var}\left(\sum X_i \beta_i\right)$ needs to be better understood and calculated. The key difficulty is the correlation matrix between β_i .

RowNa	Prm1	Prm2	Prm3	Prm4	Prm5	Prm6	Prm8	Prm9	Prm10	Prm11	Prm12	Prm14	Prm15	Prm16
Prm1	100%	-49%	-71%	-49%	-46%	-28%	-11%	-9%	-10%	-7%	-8%	-65%	-62%	-67%
Prm2	-49%	100%	66%	87%	89%	52%	-18%	-13%	-10%	-7%	1%	1%	3%	1%
Prm3	-71%	66%	100%	63%	66%	38%	-4%	-4%	-4%	-4%	2%	50%	48%	55%
Prm4	-49%	87%	63%	100%	84%	52%	1%	0%	0%	-1%	5%	3%	1%	2%
Prm5	-46%	89%	66%	84%	100%	50%	-6%	-5%	-4%	-8%	-1%	1%	1%	0%
Prm6	-28%	52%	38%	52%	50%	100%	-5%	-5%	-4%	-2%	1%	2%	2%	1%
Prm8	-11%	-18%	-4%	1%	-6%	-5%	100%	61%	62%	47%	31%	-5%	-6%	-3%
Prm9	-9%	-13%	-4%	0%	-5%	-5%	61%	100%	51%	40%	26%	-2%	-3%	-2%
Prm10	-10%	-10%	-4%	0%	-4%	-4%	62%	51%	100%	43%	29%	-1%	-3%	-1%
Prm11	-7%	-7%	-4%	-1%	-8%	-2%	47%	40%	43%	100%	25%	-1%	-2%	-2%
Prm12	-8%	1%	2%	5%	-1%	1%	31%	26%	29%	25%	100%	-1%	-1%	0%
Prm14	-65%	1%	50%	3%	1%	2%	-5%	-2%	-1%	-1%	-1%	100%	80%	88%
Prm15	-62%	3%	48%	1%	1%	2%	-6%	-3%	-3%	-2%	-1%	80%	100%	85%
Prm16	-67%	1%	55%	2%	0%	1%	-3%	-2%	-1%	-2%	0%	88%	85%	100%
Prm18	-22%	4%	5%	5%	4%	2%	-11%	-8%	-8%	-7%	-1%	1%	2%	6%
Prm19	-20%	1%	3%	3%	3%	0%	-9%	-6%	-8%	-5%	-2%	4%	3%	4%
Prm21	-16%	8%	2%	1%	3%	3%	4%	0%	0%	-1%	0%	-7%	0%	2%
Prm22	-11%	9%	5%	8%	7%	5%	2%	1%	1%	1%	1%	-4%	-1%	1%
Prm23	-4%	2%	-3%	-26%	0%	-8%	10%	7%	8%	5%	2%	-3%	0%	-3%
Prm24	-17%	12%	-6%	13%	-2%	4%	13%	8%	6%	5%	-1%	3%	3%	2%
Prm26	-42%	2%	2%	2%	-1%	1%	3%	0%	1%	1%	-1%	12%	9%	2%
Prm27	-49%	2%	4%	1%	0%	1%	4%	1%	0%	-1%	-1%	17%	13%	5%
Prm28	-50%	3%	3%	2%	0%	1%	3%	0%	0%	0%	-1%	15%	11%	3%
Prm29	-10%	0%	-2%	-1%	-1%	-1%	2%	0%	0%	-1%	0%	-8%	0%	-2%
Prm30	-10%	-1%	-2%	-1%	-1%	0%	3%	2%	1%	1%	1%	-6%	1%	-1%
Prm31	-19%	8%	6%	8%	8%	5%	4%	3%	3%	2%	2%	2%	3%	1%
Prm32	-33%	1%	0%	0%	-1%	0%	-2%	-2%	-2%	0%	0%	2%	-1%	1%
Scale	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

RowNa	Prm18	Prm19	Prm21	Prm22	Prm23	Prm24	Prm26	Prm27	Prm28	Prm29	Prm30	Prm31	Prm32	Scale
Prm1	-22%	-20%	-16%	-11%	-4%	-17%	-42%	-49%	-50%	-10%	-10%	-19%	-33%	0%
Prm2	4%	1%	8%	9%	2%	12%	2%	2%	3%	0%	-1%	8%	1%	0%
Prm3	5%	3%	2%	5%	-3%	-6%	2%	4%	3%	-2%	-2%	6%	0%	0%
Prm4	5%	3%	1%	8%	-26%	13%	2%	1%	2%	-1%	-1%	8%	0%	0%
Prm5	4%	3%	3%	7%	0%	-2%	-1%	0%	0%	-1%	-1%	8%	-1%	0%
Prm6	2%	0%	3%	5%	-8%	4%	1%	1%	1%	-1%	0%	5%	0%	0%
Prm8	-11%	-9%	4%	2%	10%	13%	3%	4%	3%	2%	3%	4%	-2%	0%
Prm9	-8%	-6%	0%	1%	7%	8%	0%	1%	0%	0%	2%	3%	-2%	0%
Prm10	-8%	-8%	0%	1%	8%	6%	1%	0%	0%	0%	1%	3%	-2%	0%
Prm11	-7%	-5%	-1%	1%	5%	5%	1%	-1%	0%	-1%	1%	2%	0%	0%
Prm12	-1%	-2%	0%	1%	2%	-1%	-1%	-1%	-1%	0%	1%	2%	0%	0%
Prm14	1%	4%	-7%	-4%	-3%	3%	12%	17%	15%	-8%	-6%	2%	2%	0%
Prm15	2%	3%	0%	-1%	0%	3%	9%	13%	11%	0%	1%	3%	-1%	0%
Prm16	6%	4%	2%	1%	-3%	2%	2%	5%	3%	-2%	-1%	1%	1%	0%
Prm18	100%	77%	-1%	2%	2%	2%	0%	-3%	0%	0%	-3%	0%	1%	0%
Prm19	77%	100%	1%	1%	2%	2%	-4%	-5%	-3%	-2%	-2%	-3%	-1%	0%
Prm21	-1%	1%	100%	19%	28%	30%	12%	10%	11%	4%	1%	-2%	7%	0%
Prm22	2%	1%	19%	100%	12%	17%	2%	3%	4%	4%	2%	2%	1%	0%
Prm23	2%	2%	28%	12%	100%	18%	1%	0%	1%	1%	2%	0%	2%	0%
Prm24	2%	2%	30%	17%	18%	100%	5%	4%	6%	1%	2%	3%	4%	0%
Prm26	0%	-4%	12%	2%	1%	5%	100%	81%	82%	22%	20%	24%	58%	0%
Prm27	-3%	-5%	10%	3%	0%	4%	81%	100%	94%	25%	23%	28%	66%	0%
Prm28	0%	-3%	11%	4%	1%	6%	82%	94%	100%	25%	24%	28%	67%	0%
Prm29	0%	-2%	4%	4%	1%	1%	22%	25%	25%	100%	10%	9%	21%	0%
Prm30	-3%	-2%	1%	2%	2%	2%	20%	23%	24%	10%	100%	8%	20%	0%
Prm31	0%	-3%	-2%	2%	0%	3%	24%	28%	28%	9%	8%	100%	21%	0%
Prm32	1%	-1%	7%	1%	2%	4%	58%	66%	67%	21%	20%	21%	100%	0%
Scale	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%

Institute and Faculty of Actuaries

GLM fact 3: GLM results depend on the mixture of rating variables in the data

Data

$$X$$
$$\downarrow$$
$$\eta = X \cdot \beta$$
$$\downarrow$$
$$E(Y) \equiv \mu = g^{-1}(\eta)$$

Driver Age	Car Age	Claim
Old	Old	0.2
Old	New	0.3
Young	Old	0.4
Young	New	0.6

Parameter	Level1	Estimate	StdErr
Intercept		0.4286	0.565
age	Old	-0.2411	0.6061
age	Young	0	0
carage	New	0.1339	0.5836
carage	Old	0	0
Scale		1	0

Driver Age	Car Age	Claim
Old	Old	0.2
Old	Old	0.2
Old	New	0.3
Young	Old	0.4
Young	New	0.6

Parameter	Level1	Estimate	StdErr
Intercept		0.4305	0.5594
age	Old	-0.2374	0.5827
age	Young	0	0
carage	New	0.1297	0.552
carage	Old	0	0
Scale		1	0

Institute and Faculty of Actuaries

GLMs results are dragged toward the segment where there is more data

Data

$$X$$
$$\downarrow$$
$$\eta = X \cdot \beta$$
$$\downarrow$$
$$E(Y) \equiv \mu = g^{-1}(\eta)$$

4 data points

Driver Age	Car Age	Claim	Prediction
Old	Old	0.2	0.1875
Old	New	0.3	0.32143
Young	Old	0.4	0.42857
Young	New	0.6	0.5625

5 data points

Driver Age	Car Age	Claim	Prediction
Old	Old	0.2	0.19315
Old	Old	0.2	0.19315
Old	New	0.3	0.3229
Young	Old	0.4	0.43053
Young	New	0.6	0.56027

- The dependency is not trivial. Some practical examples are:
  - Quote based premium model vs. sale based premium model
  - Modelled loss ratio for quotes vs. Sales
  - Time testing

Institute and Faculty of Actuaries

With a view to future is the key to mitigate this issue

Data X

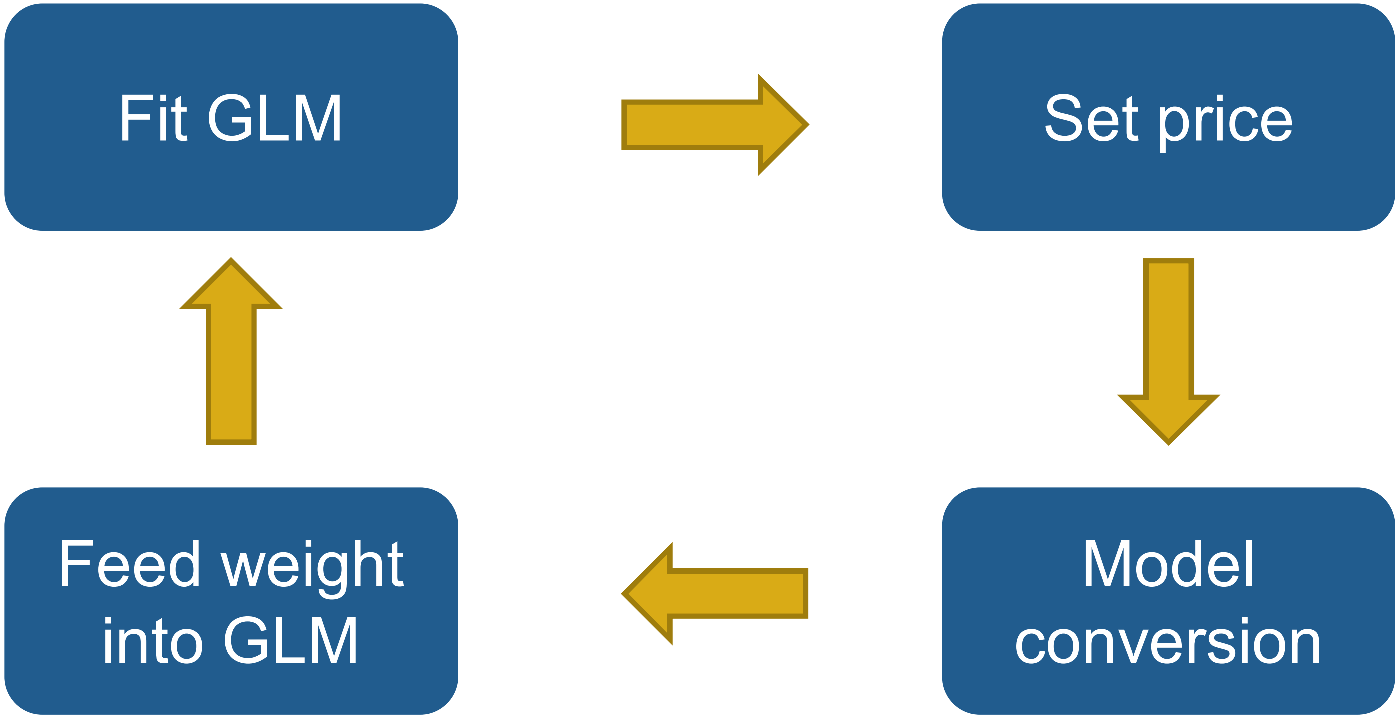
$$\downarrow$$

$$\eta = X \cdot \beta$$

$$\downarrow$$

$$E(Y) \equiv \mu = g^{-1}(\eta)$$

- GLM should be trained on expected future mixture of portfolio, rather than historical portfolio.
- Iterative modelling approach:



Institute and Faculty of Actuaries

Summary

- Significant variation in underwriting performance and quoted premiums in the current motor market pose challenges on the pricing techniques used in business.
- As the standard pricing technique, GLMs are coming cross new issues in a highly competitive market:
  - GLMs put either zero or full credibility into data
  - GLMs implicitly use median from the distribution of prediction
  - GLM results depend on the mixture of rating variables in the data
- Being able to understand and solve these issues could be one of the key ways to gain a competitive advantage in the market.

Institute and Faculty of Actuaries

Questions

Comments

Expressions of individual views by members of the Institute and Faculty of Actuaries and its staff are encouraged.

The views expressed in this presentation are those of the presenter.