GLM III - The Matrix Reloaded

Brent Petzoldt

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.

GLM III - The Matrix Reloaded

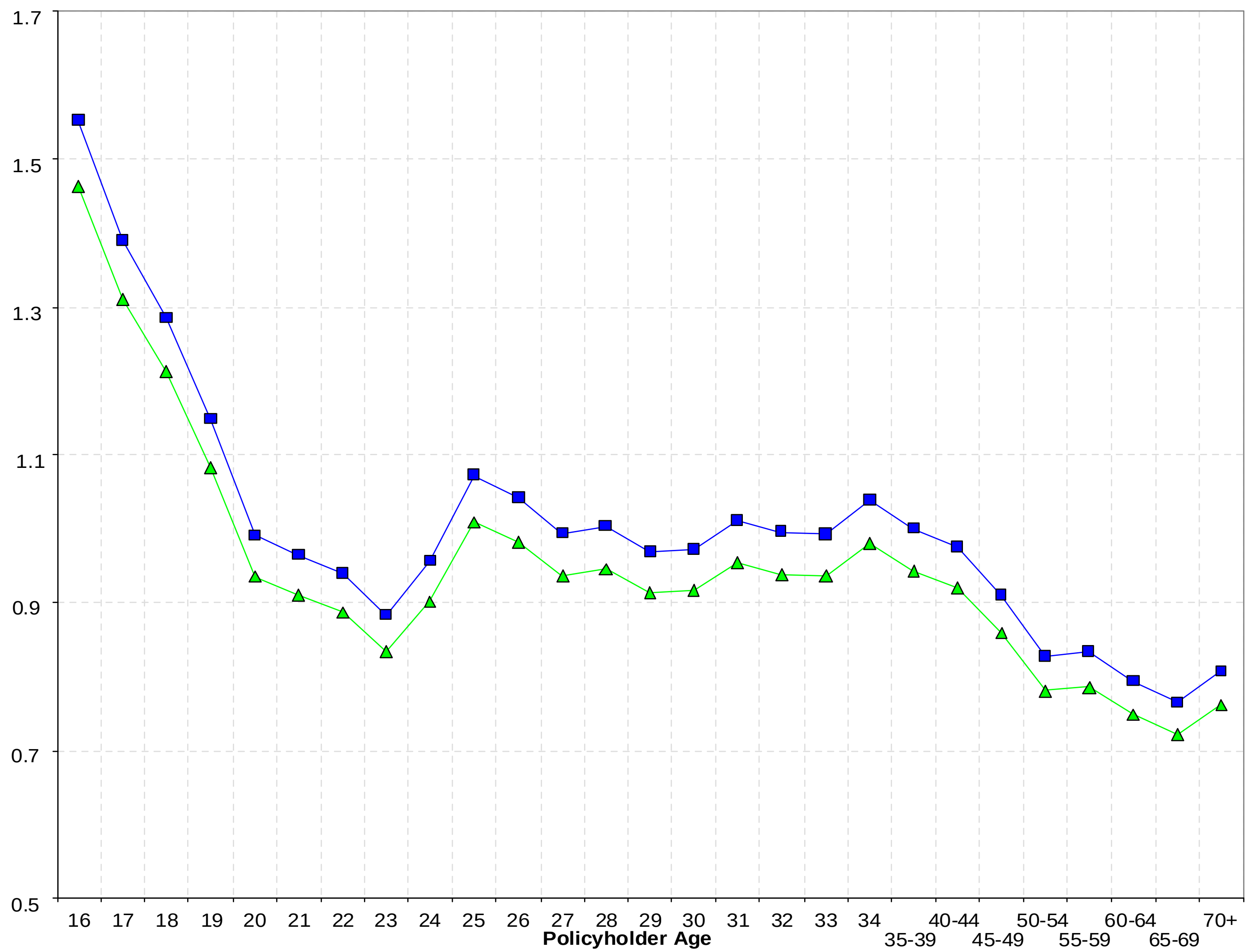
Brent Petzoldt

TOWERS WATSON

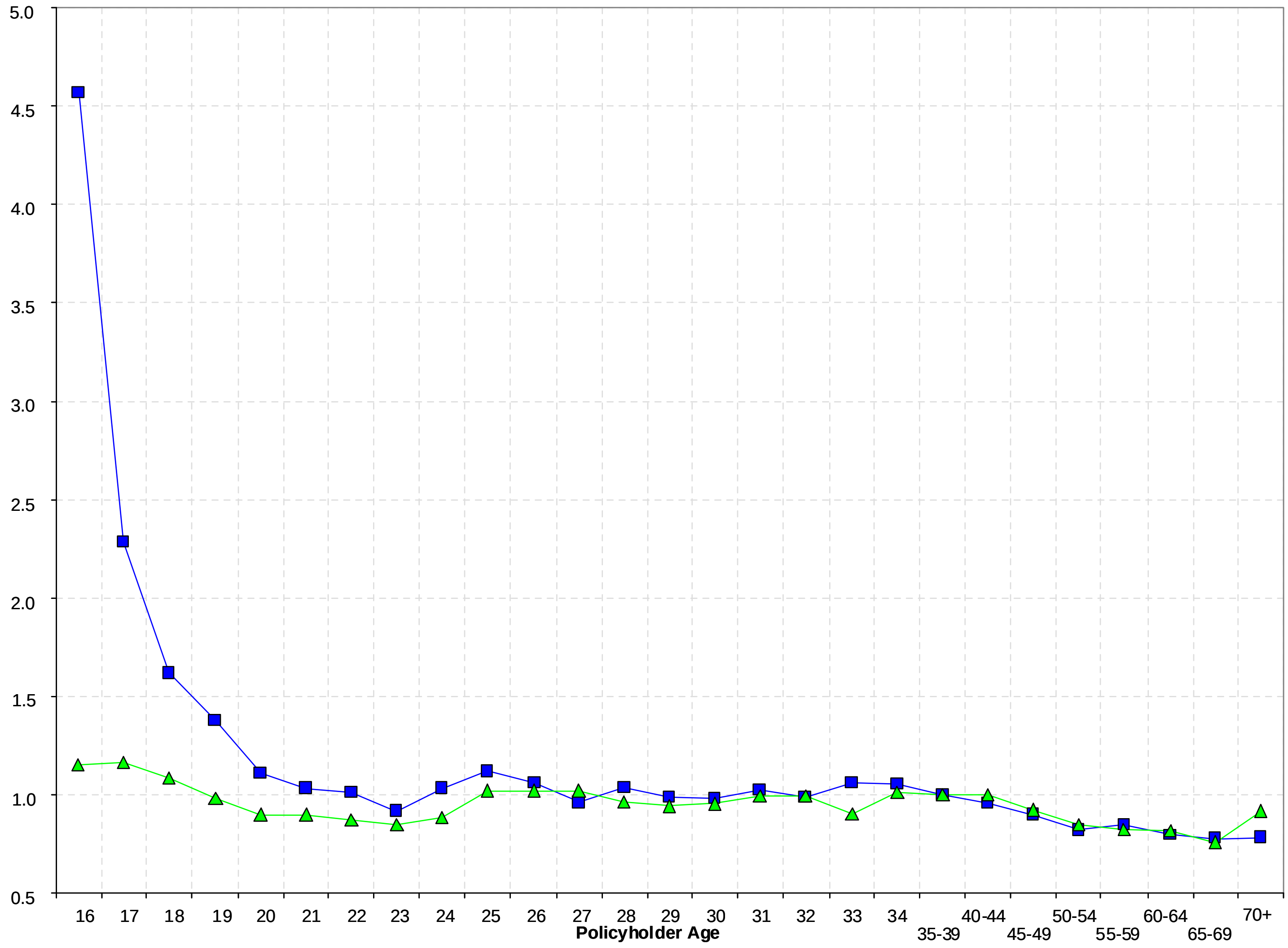
Agenda

- "Quadrant Saddles"
- The Tweedie Distribution
- "Emergent Interactions"
- Dispersion Modeling
- Modeling sparse claim types
- Driver Averaging
- Model Validation
- Man (with GLM) vs machine

Interactions



Interactions

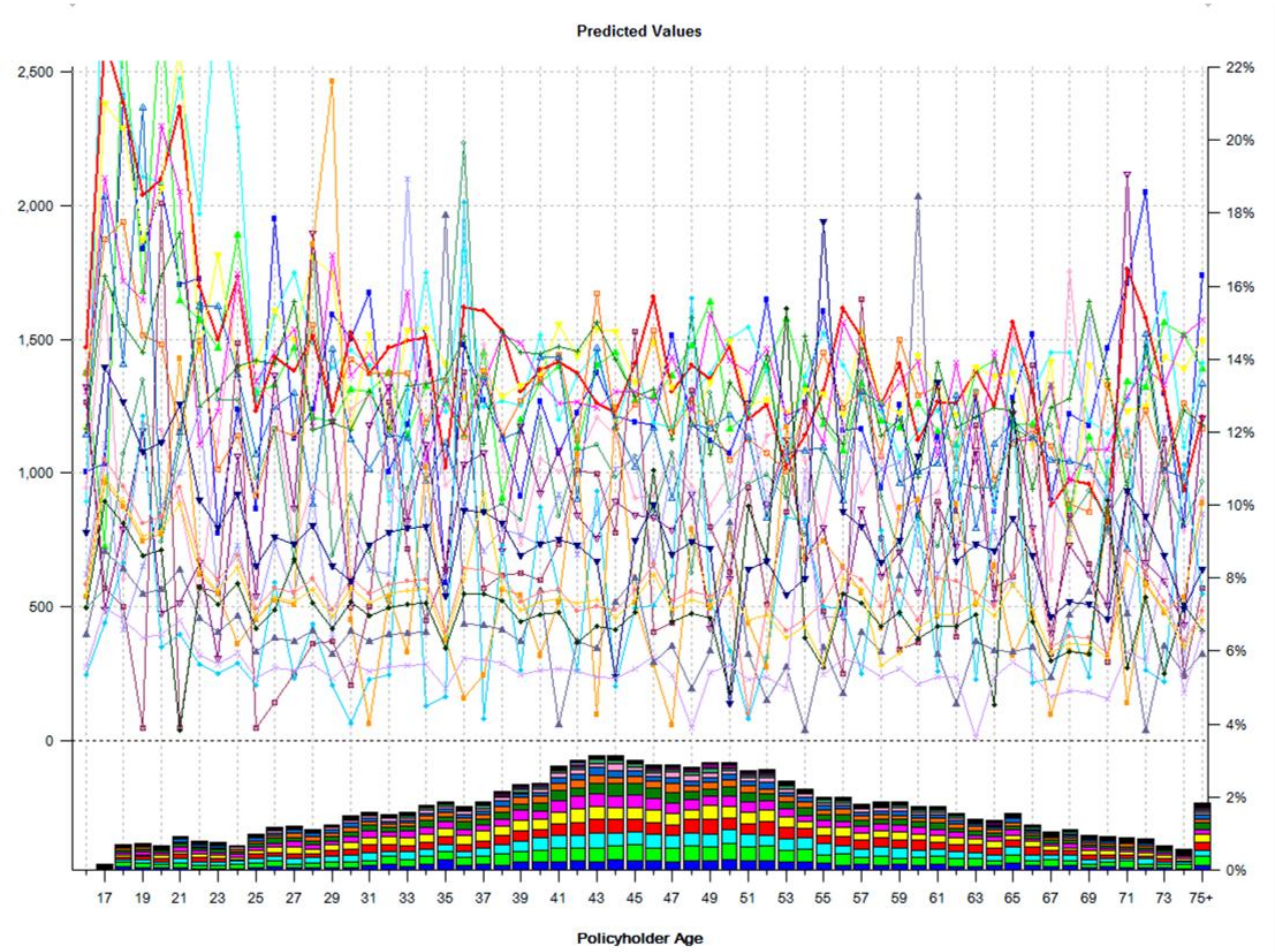


towerswatson.com

Why are interactions present?

- Because that's how the factors behave
- Because the multiplicative model can go wrong at the edges
  - 1.5 * 1.4 * 1.7 * 1.5 * 1.8 * 1.5 * 1.8 = 26!

Interactions



Interactions

Selected Interaction: Vehicle Age x Policyholder Age

Policyholder Age

Vehicle Age	N/A	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
0	0.0000	0.0861	1.3433	-0.2833	-0.1112	-0.6276	-0.4940	-0.9551	0.9797	-0.6060	-0.8082	-0.6622	-0.8114	-1.2990	-0.4380	1.0881	0.9502	1.19
1	0.0000	-0.3007	-0.3838	0.6784	-1.0827	1.5070	0.2896	-0.3146	0.4689	1.1002	0.0152	-1.2990	0.2485	-0.1511	-0.4321	-1.5809	0.2436	0.87
2	0.0000	0.5942	-0.5767	-1.3743	0.2555	1.5679	-0.3335	-0.3765	1.3385	1.3492	1.2615	-0.4747	0.9228	-0.1730	1.0359	-0.8213	-0.4516	1.09
3	0.0000	-0.0944	-1.6225	0.5895	-0.6255	0.4782	-1.4383	0.8664	-1.1652	0.3511	0.4272	-0.3256	-0.5740	-0.2573	-0.3694	0.3803	-0.6589	-0.7
4	0.0000	0.7341	-0.3873	-0.7813	0.1535	-1.0952	1.2768	0.0124	-0.2584	-0.2420	0.9662	0.8497	-0.9958	-1.4329	0.4942	-0.1446	1.6775	-0.3
5	0.0000	0.7834	-0.1512	0.6603	0.1302	-1.0948	-1.7072	-2.7967	-0.9688	-0.1618	-2.4681	-0.0443	-1.2045	0.6146	1.1614	1.8351	-1.5661	-0.2
6	0.0000	2.2230	-0.7901	-0.2768	-1.3420	-0.6245	1.1232	1.5019	0.5902	-0.0202	2.1912	0.1612	-1.7502	0.1939	-1.9662	1.5626	0.6414	-0.1
7	0.0000	-0.6514	0.4566	-1.4507	0.2114	-0.7806	-0.8530	-0.8567	0.0683	0.2886	-0.1552	1.3343	1.5336	-1.3740	1.5809	-1.1128	-0.8446	0.22
8	0.0000	-0.2580	0.2443	-1.0002	0.2211	0.6874	0.8802	1.4942	-1.7197	2.0866	-0.1492	0.5193	1.0116	2.4261	0.1893	1.9269	-1.1336	-0.7
9	0.0000	-1.4664	1.0500	-1.7925	-0.6770	0.6516	-1.4137	0.1120	-0.4953	-1.2055	-0.9852	-0.0994	0.4017	0.3569	-0.8758	-0.1560	0.6361	1.09
10	0.0000	-0.3617	-0.4029	1.1780	2.4451	-0.3206	0.3776	1.2449	2.3793	-1.0895	-0.5111	-1.7375	1.5678	0.9283	-0.4278	-1.9636	1.7165	-1.8
11	0.0000	-1.0350	1.2051	-0.2111	0.3438	-2.0220	1.4628	-0.7946	-0.1623	-0.4664	-0.0784	-0.2583	-1.3273	0.7165	-0.1358	-0.3915	0.2076	-0.6
12	0.0000	1.1877	-0.1965	1.8634	-0.2171	-0.4839	-0.7580	-0.2611	0.1789	-2.0832	0.1933	-0.9373	0.1420	-0.8524	1.2916	-1.8746	0.3296	1.12
13	0.0000	-0.7453	0.8332	0.7936	0.6080	-0.1837	-0.3785	-0.1706	-0.1664	-0.3504	-0.4945	2.2726	0.5549	-0.6972	0.0935	0.4392	-0.6115	-1.6
14	0.0000	-0.3047	0.8316	-0.8146	0.3477	0.3955	-0.2695	0.7418	-1.0084	-0.7375	-0.3714	1.5842	0.0045	0.6473	-0.9911	0.0465	-0.2720	0.33
15	0.0000	-0.1280	-1.0783	1.4620	-0.0174	1.8673	2.3668	0.1103	0.3009	0.8215	-0.7305	-0.6188	0.3240	-1.0383	-1.2315	-1.0858	-1.6775	0.17
16	0.0000	-0.7361	1.4342	2.1929	1.0053	1.2026	0.5048	0.0824	-0.1974	0.1701	1.3634	1.4209	1.2107	0.9176	0.2875	0.8263	-1.4325	1.85
17	0.0000	-0.6590	-0.6840	3.9395	-0.2883	1.5156	1.8105	4.0477	0.1852	-0.9165	2.6069	-0.0802	0.0289	0.1457	0.0409	0.2029	0.1843	-0.1
18	0.0000	-0.3412	-0.9500	0.6935	-0.5550	1.0272	1.5322	-0.6856	-0.6104	-0.4651	-0.4640	-0.3911	-0.5845	3.5499	-0.4948	-0.6204	-0.6273	-0.5
19	0.0000	-0.4783	-0.8589	-0.6467	-0.5014	-0.6032	0.7630	-0.5475	-0.5006	-0.5502	-0.5398	-0.4846	-0.3470	1.6578	-0.4972	-0.4311	1.8195	-0.6
20+	0.0000	-0.4626	-0.7337	0.9166	-0.4947	0.9693	-0.4324	-0.5751	-0.4146	-0.4206	-0.5885	-0.3000	1.7336	-0.2879	-0.4496	-0.2557	-0.4791	-0.4
Unknown	0.0000	2.4147	-0.7042	-0.6024	-0.5451	-0.4055	1.4308	-0.4851	-0.3577	1.5886	-0.4599	-0.4927	1.7167	2.6352	-0.4718	2.0103	1.5534	-0.4

Graph | Log | Un-Group | Group | Save | Hide Values | Minimum Weight 0

Back | Shortlist | Previous | Next

Interactions

Vehicle group

		β	β	–	β	β	β	β	β	β	β
	β	–	–	–	–	–	–	–	–	–	–
	β	–	–	–	–	–	–	–	–	–	–
	β	–	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–	–	–
Age	β	–	–	–	–	–	–	–	–	–	–
	β	–	–	–	–	–	–	–	–	–	–
	β	–	–	–	–	–	–	–	–	–	–
	β	–	–	–	–	–	–	–	–	–	–
	β	–	–	–	–	–	–	–	–	–	–
	β	–	–	–	–	–	–	–	–	–	–

Interactions

Vehicle group

Age

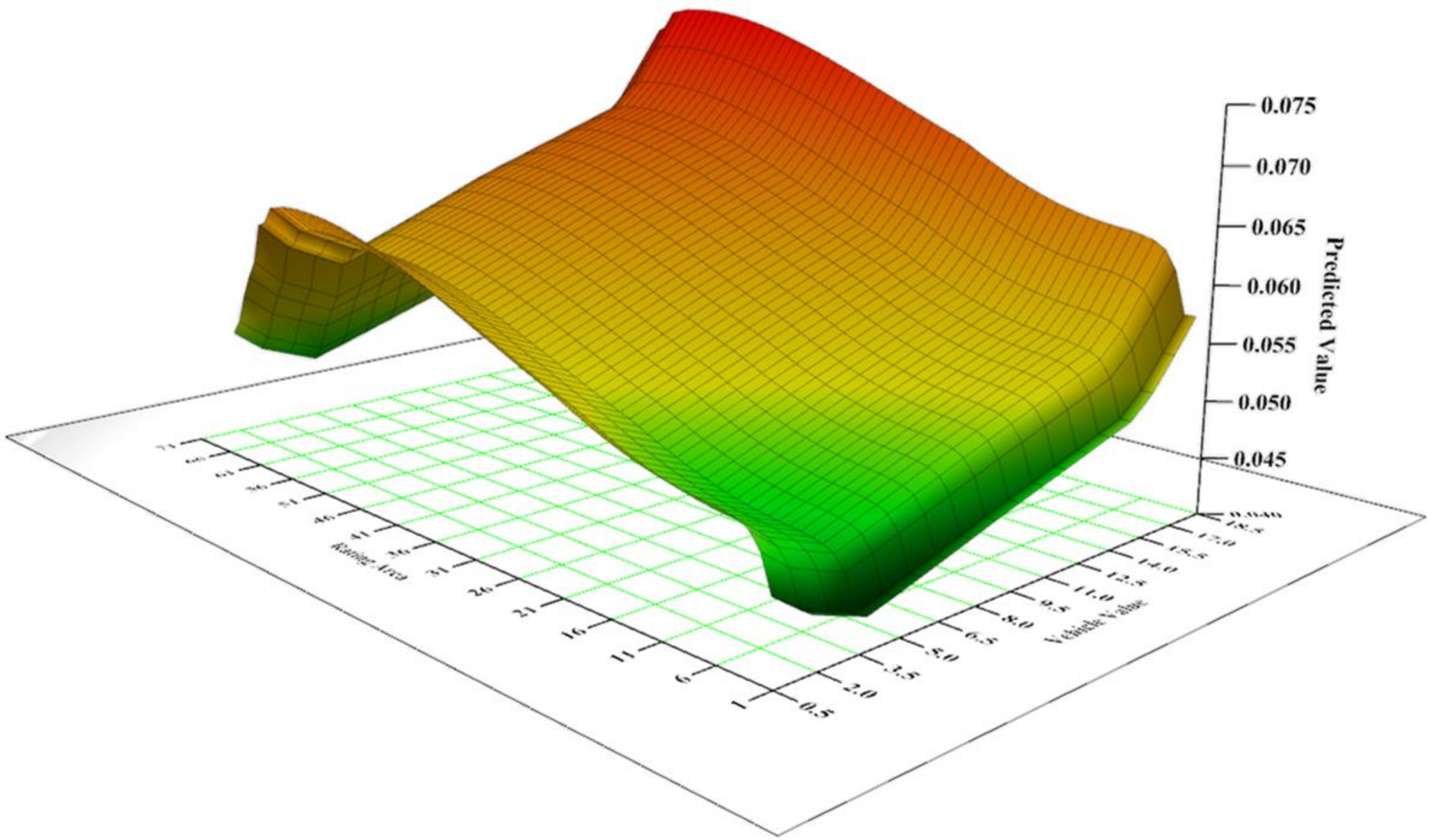
	β	β	–	β	β	β	β	β	β	β
β	–	–	–	–	–	–	–	–	–	–
β	–	–	–	–	–	–	–	–	–	–
β	–	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	–
β	–	–	–	–	–	–	–	–	–	–
β	–	–	–	–	–	–	–	–	–	–
β	–	–	–	–	–	–	–	–	–	–
β	–	–	–	–	–	–	–	–	–	–
β	–	–	–	–	–	–	–	–	–	–
β	–	–	–	–	–	–	–	–	–	–

Vehicle group

Age

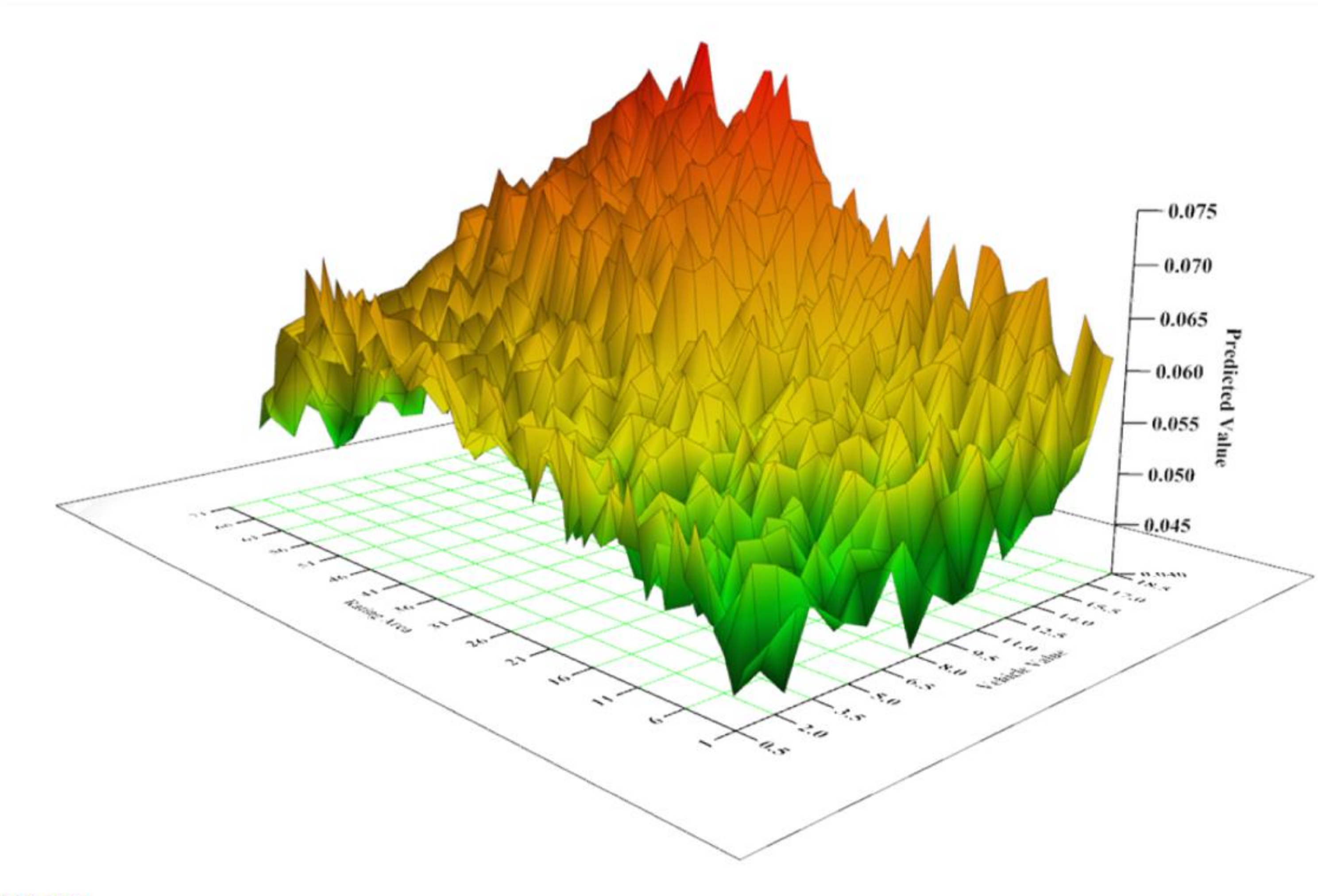
	β	β	–	β	β	β	β	β	β	β
β	β	β	–	β	β	β	β	β	β	β
β	β	β	–	β	β	β	β	β	β	β
β	β	β	–	β	β	β	β	β	β	β
–	–	–	–	–	–	–	–	–	–	–
β	β	β	–	β	β	β	β	β	β	β
β	β	β	–	β	β	β	β	β	β	β
β	β	β	–	β	β	β	β	β	β	β
β	β	β	–	β	β	β	β	β	β	β
β	β	β	–	β	β	β	β	β	β	β
β	β	β	–	β	β	β	β	β	β	β

Example



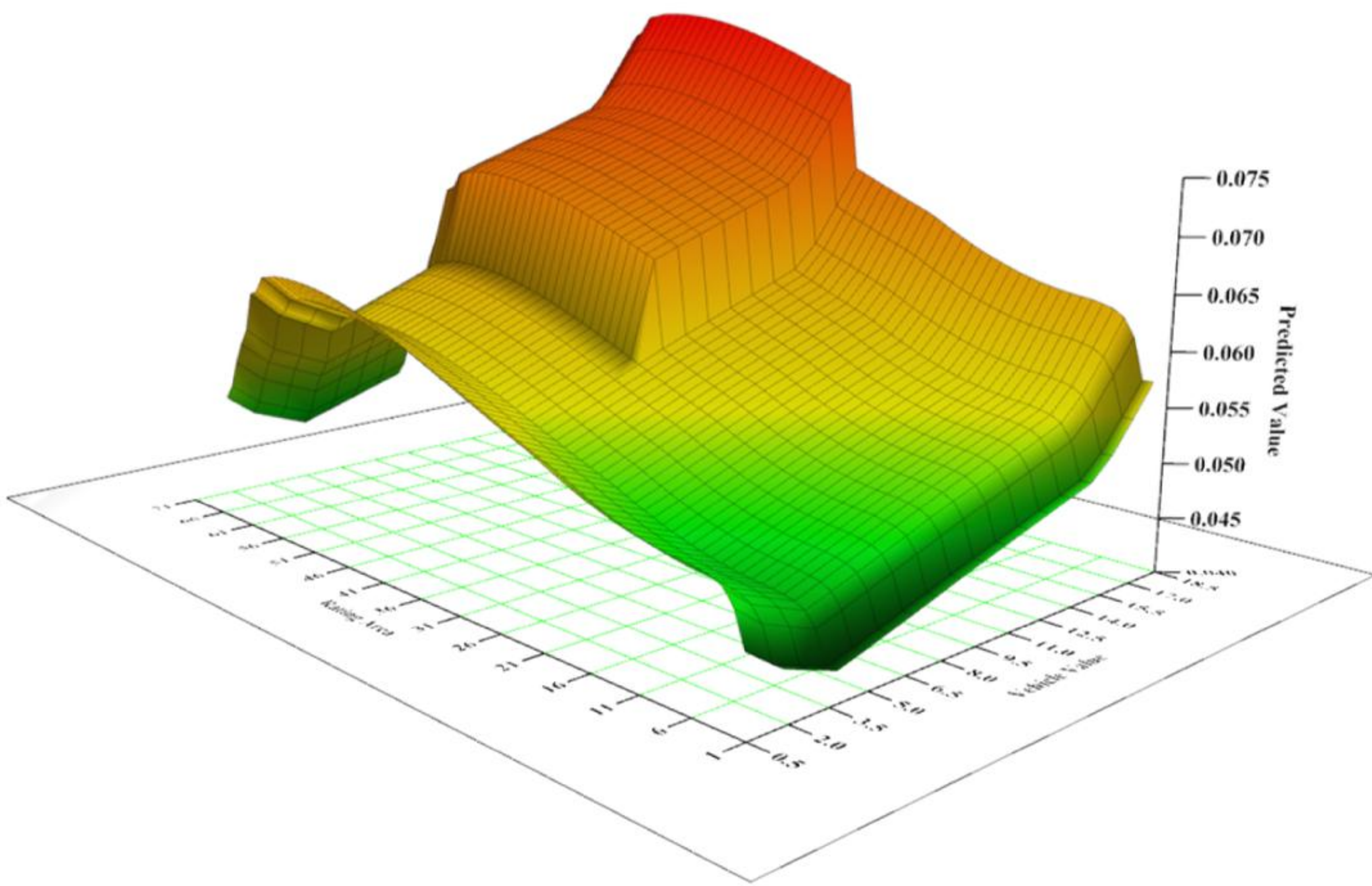
0.075
0.070
0.065
0.060
0.055
0.050
0.045
Predicted Value
Rating Area
Vehicle Value
1
6
11
16
21
26
31
36
41
46
0.5
2.0
3.5
5.0
6.5
8.0
9.5
11.0
12.5
14.0
15.5
17.0
18.5

Example

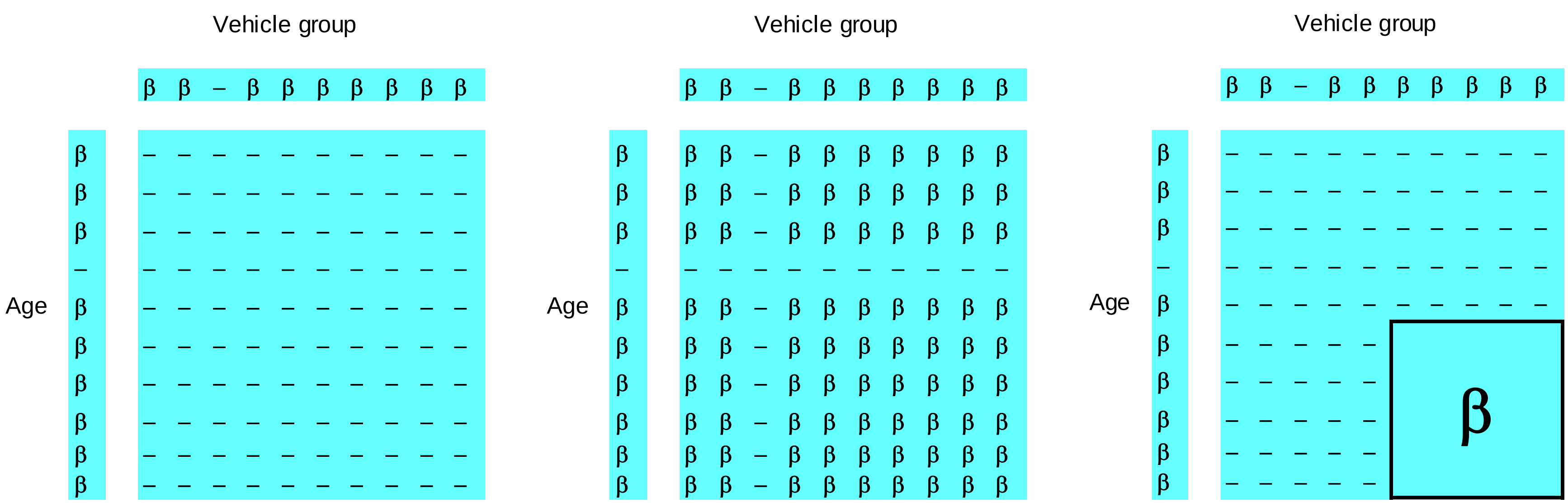


0.075
0.070
0.065
0.060
0.055
0.050
0.045
Predicted Value
Rating Area
Vehicle Value

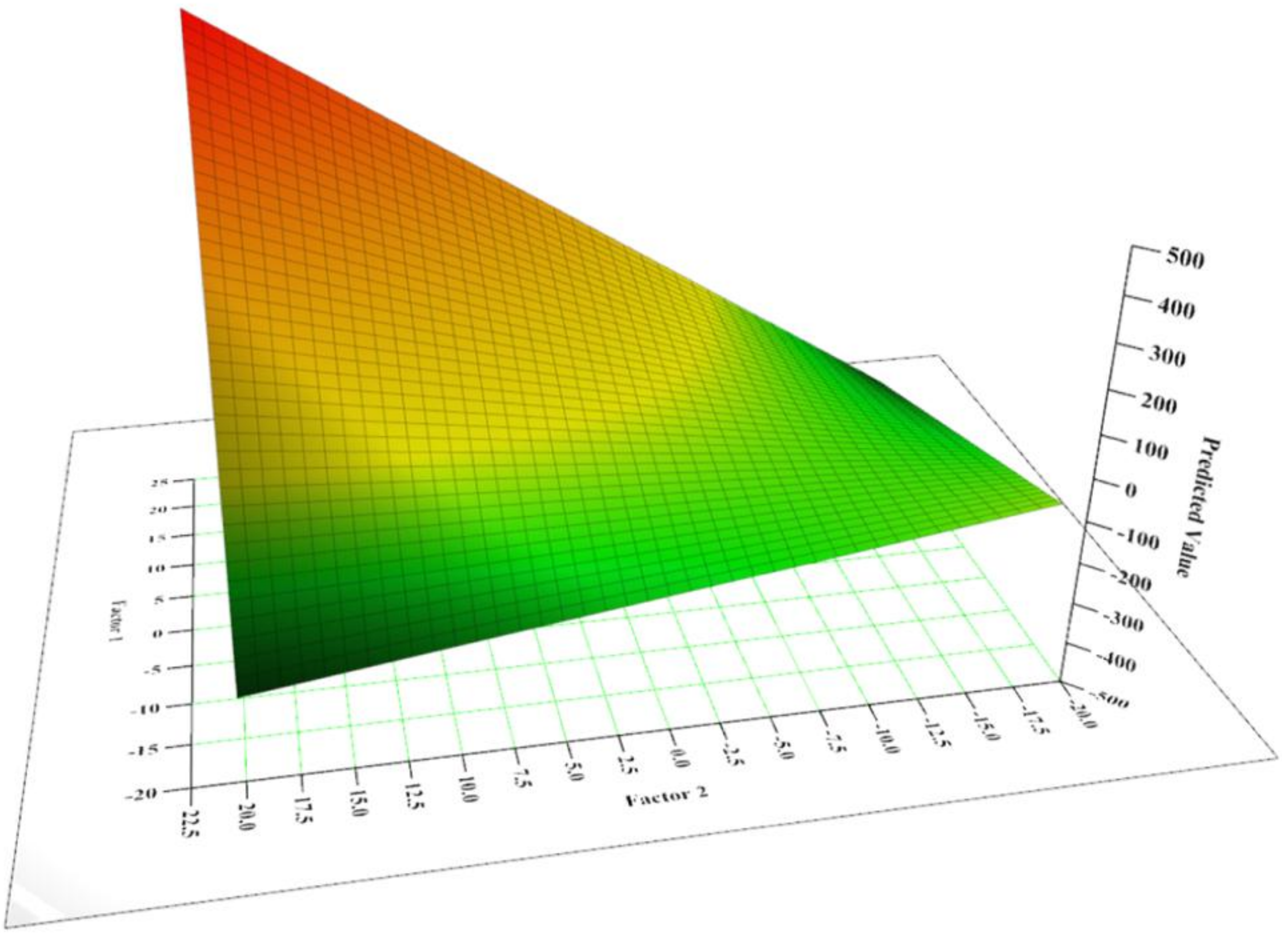
Example


Predicted Value
0.075
0.070
0.065
0.060
0.055
0.050
0.045
46
41
31
26
21
16
11
6
1
0.5
2.0
5.0
8.0
11.0
14.0
Vehicle Value

Interactions

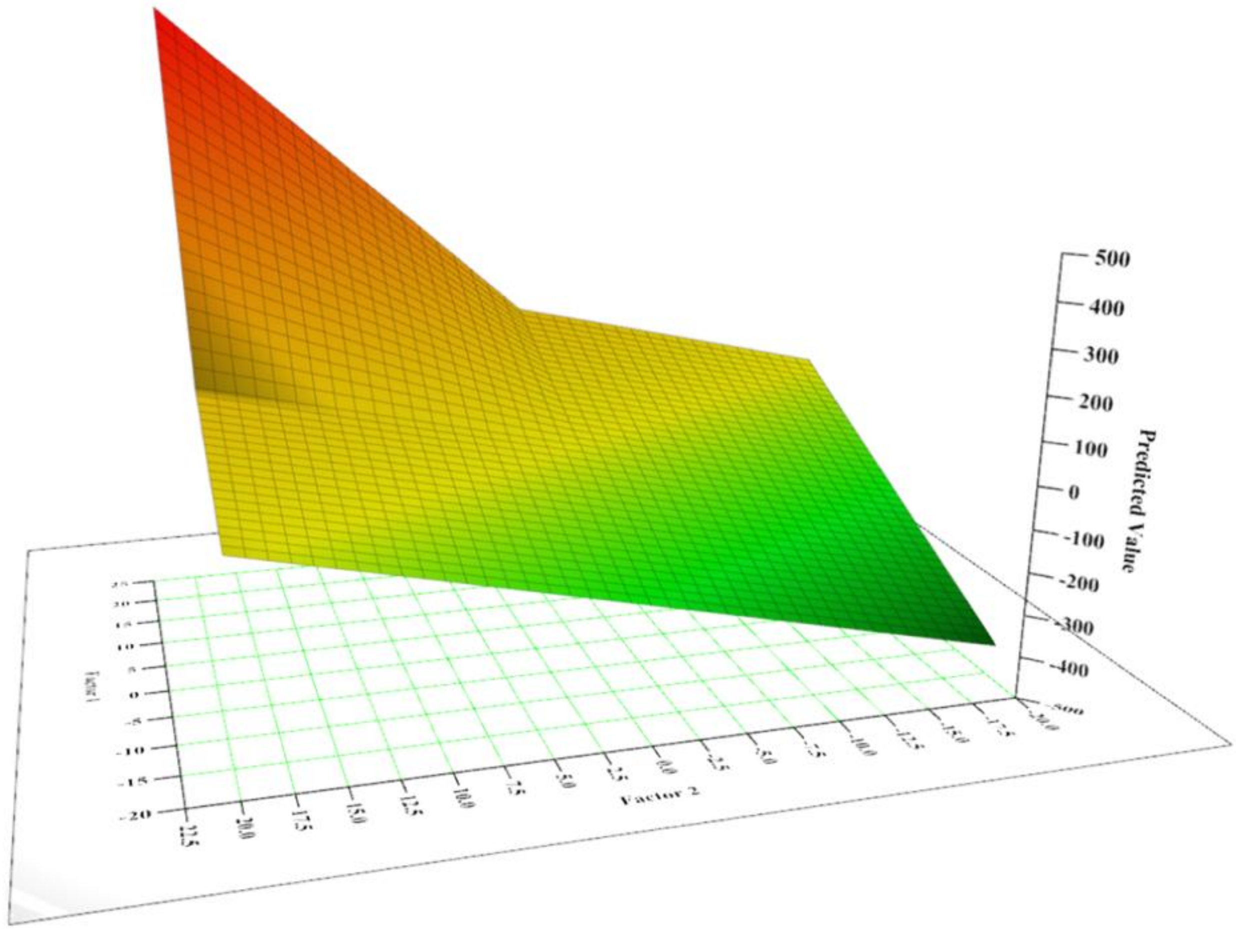


Saddles



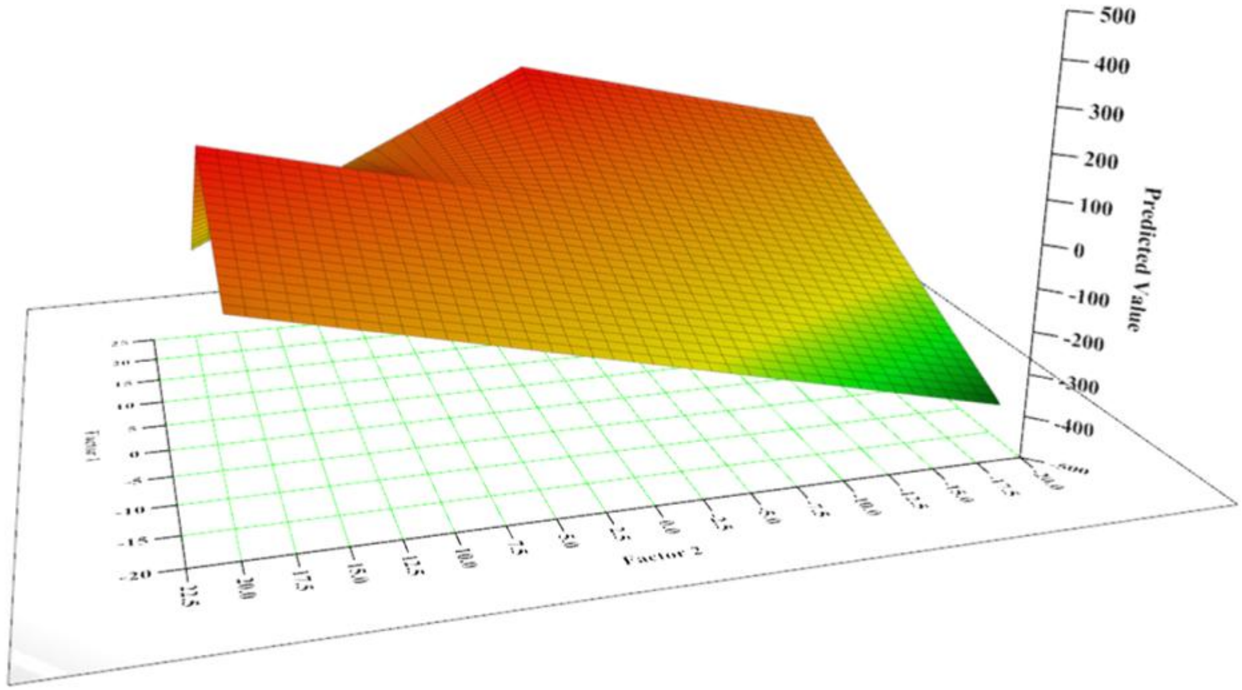
500
400
300
200
100
0
-100
-200
-300
-400
-500
Predicted Value
Factor 1
Factor 2

Saddles

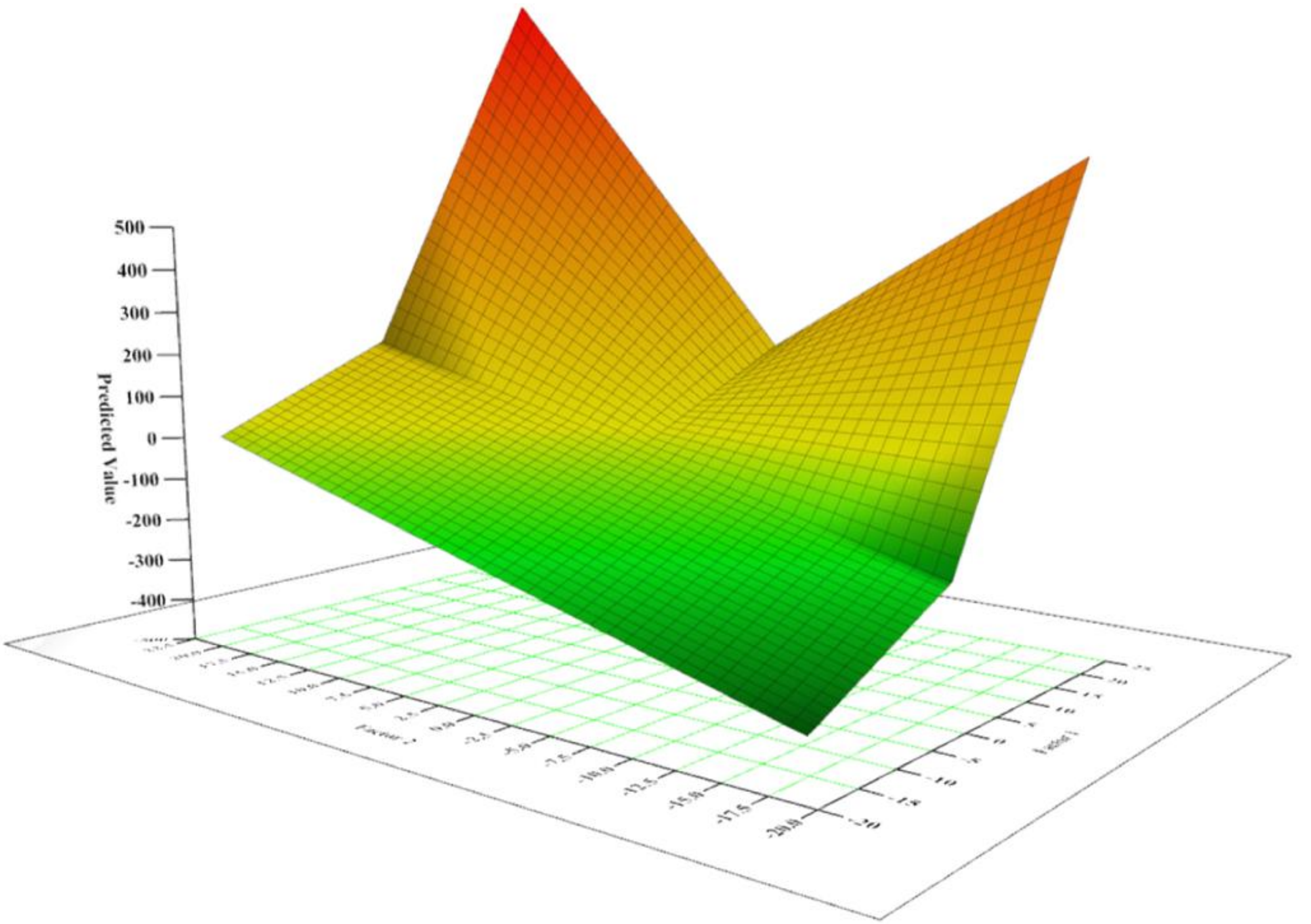


500
400
300
200
100
0
-100
-200
-300
-400
Predicted Value
Factor 2
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-20
-15
-10
-5
0
5
10
15
20

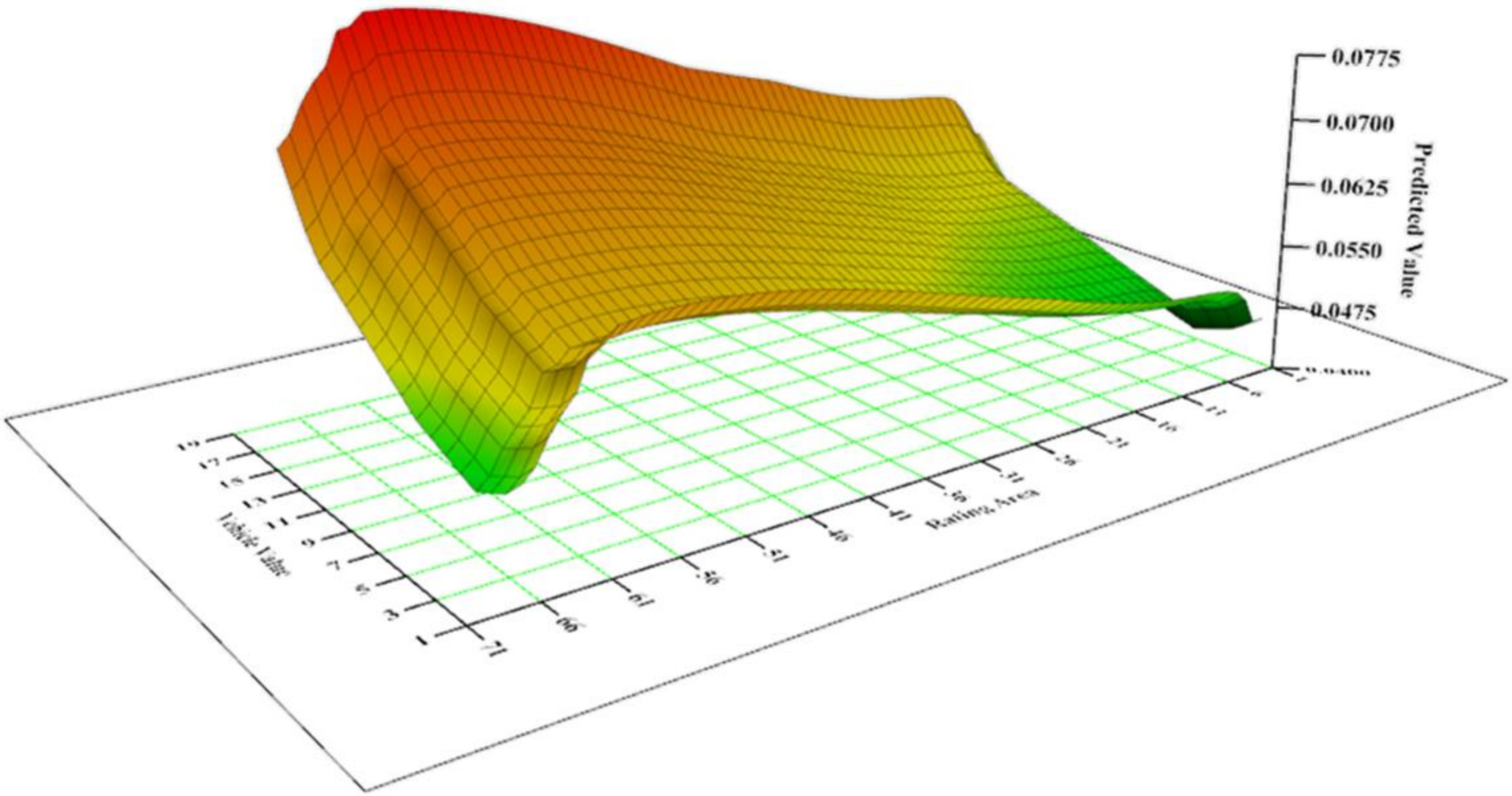
Saddles



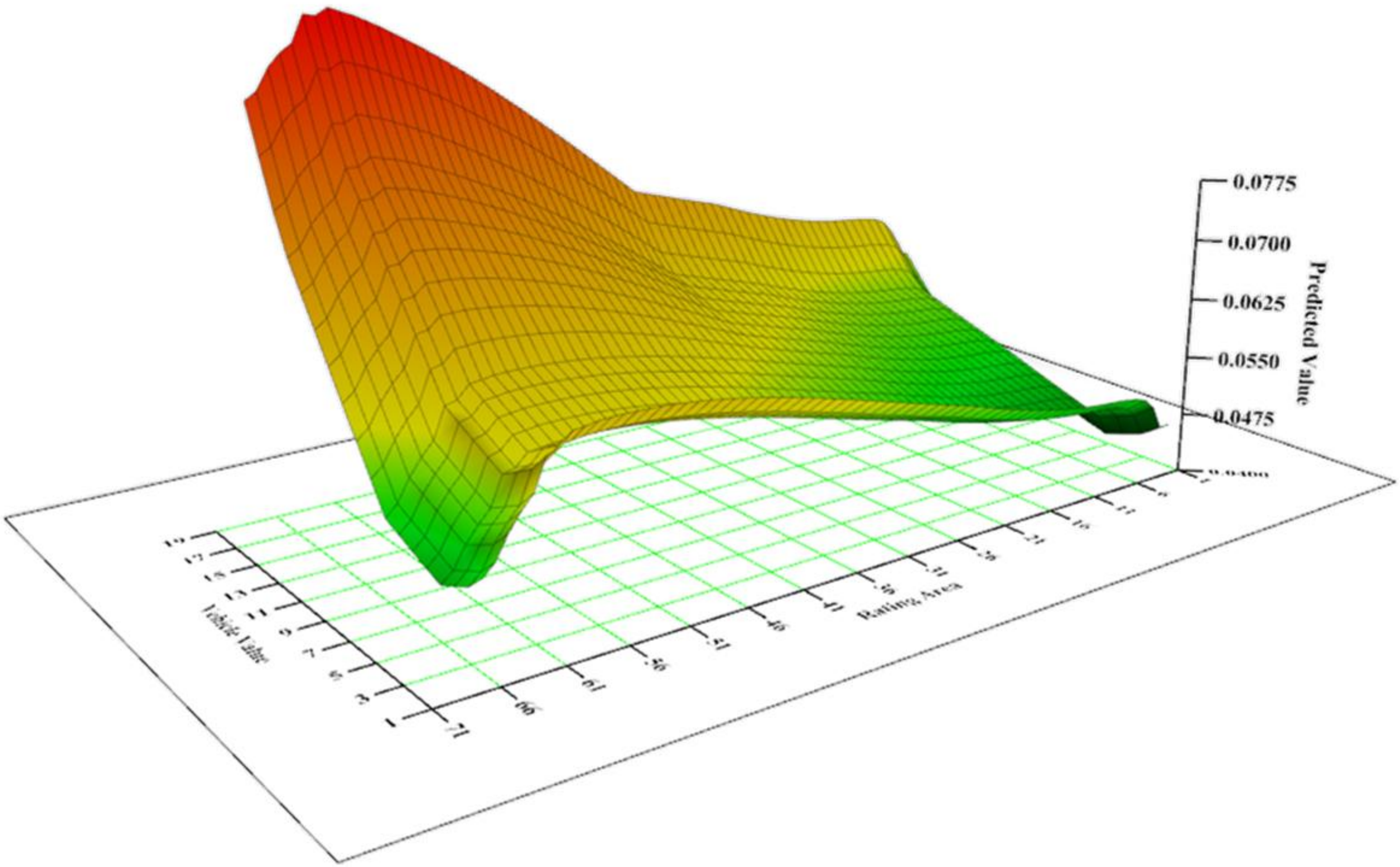
Saddles



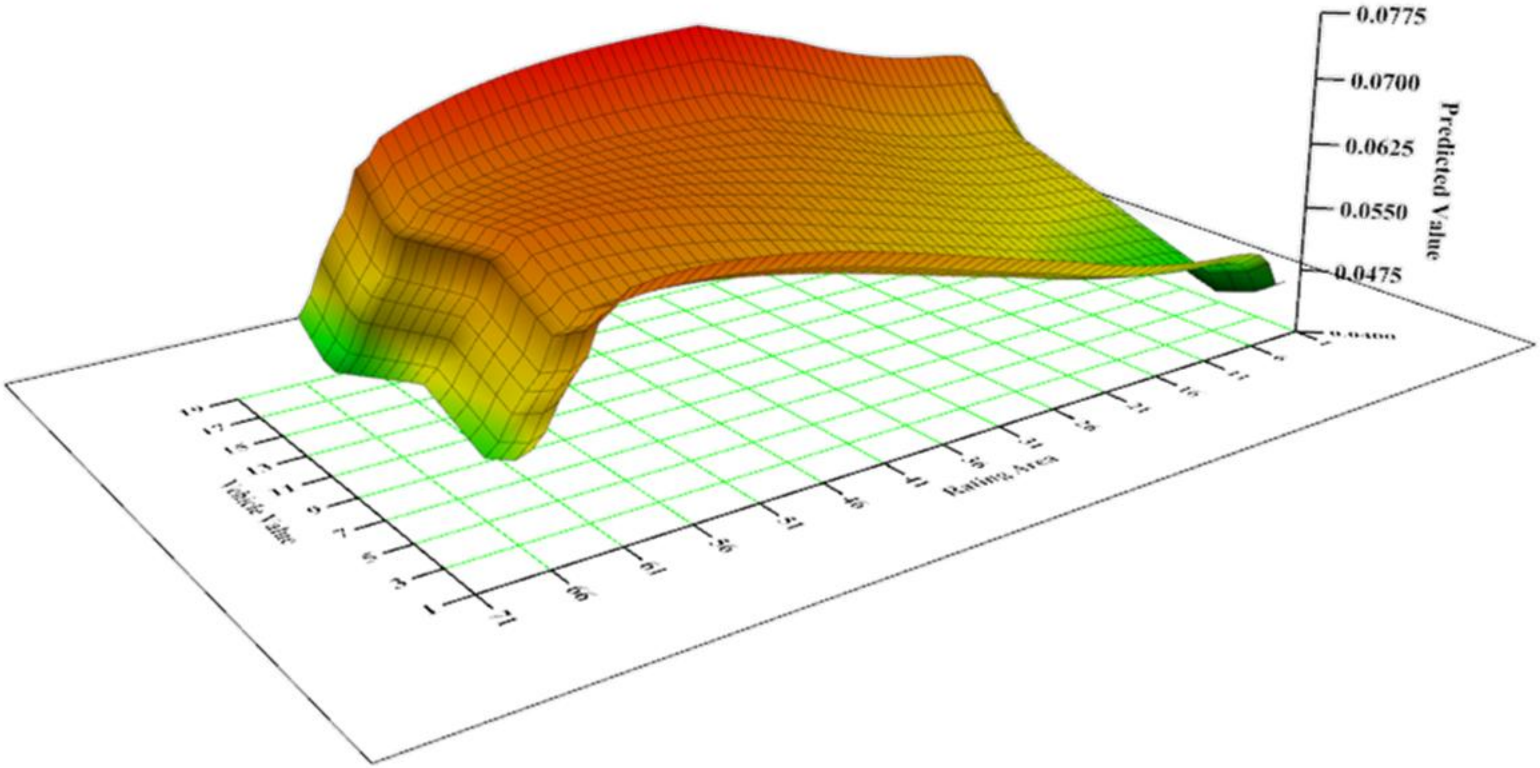
Saddles


0.0775
0.0700
0.0625
0.0550
0.0475
Predicted Value
Rating Area
Vehicle Value

Saddles

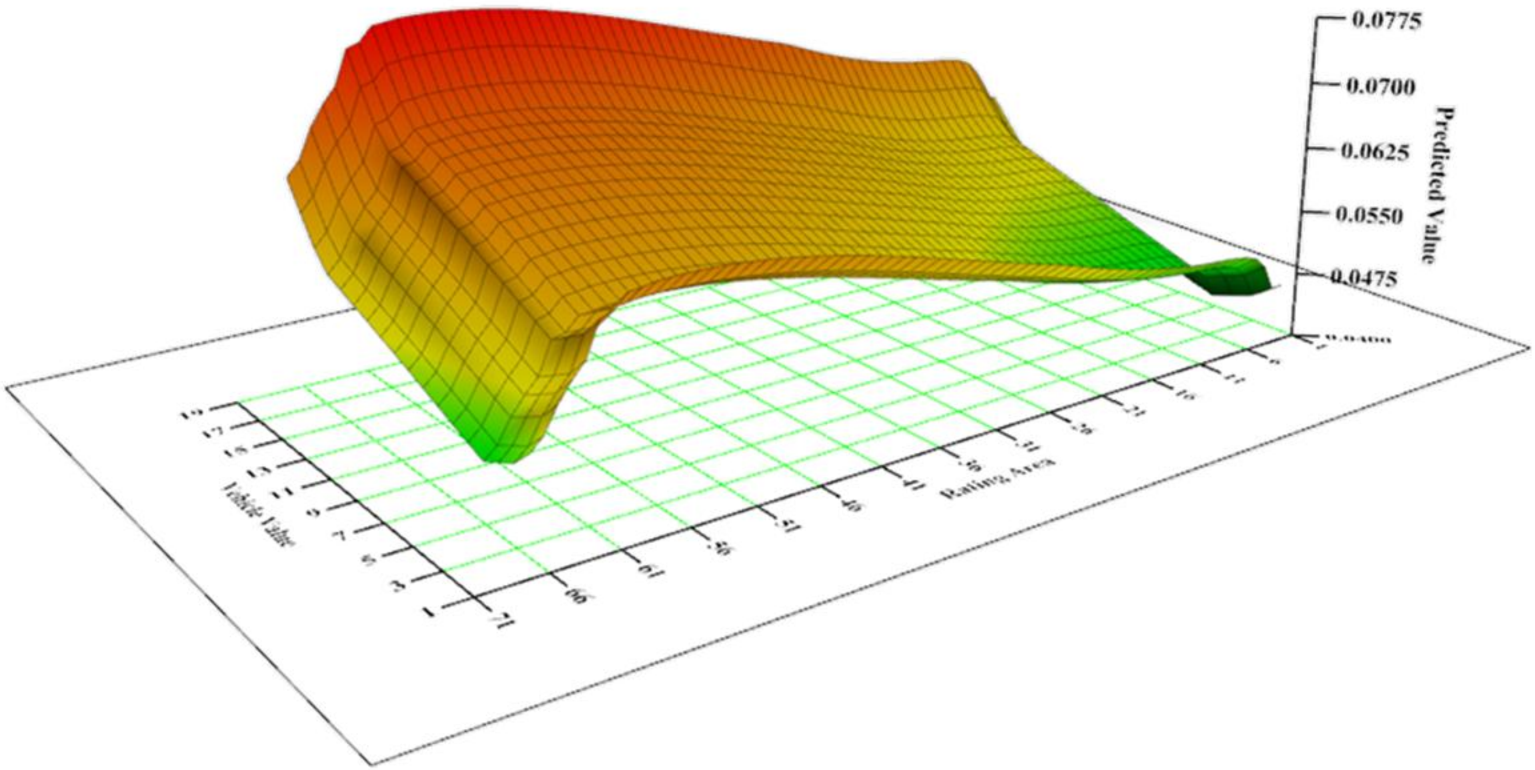

0.0775
0.0700
0.0625
0.0550
0.0475
Predicted Value
Rating Area
Vehicle Value

Saddles



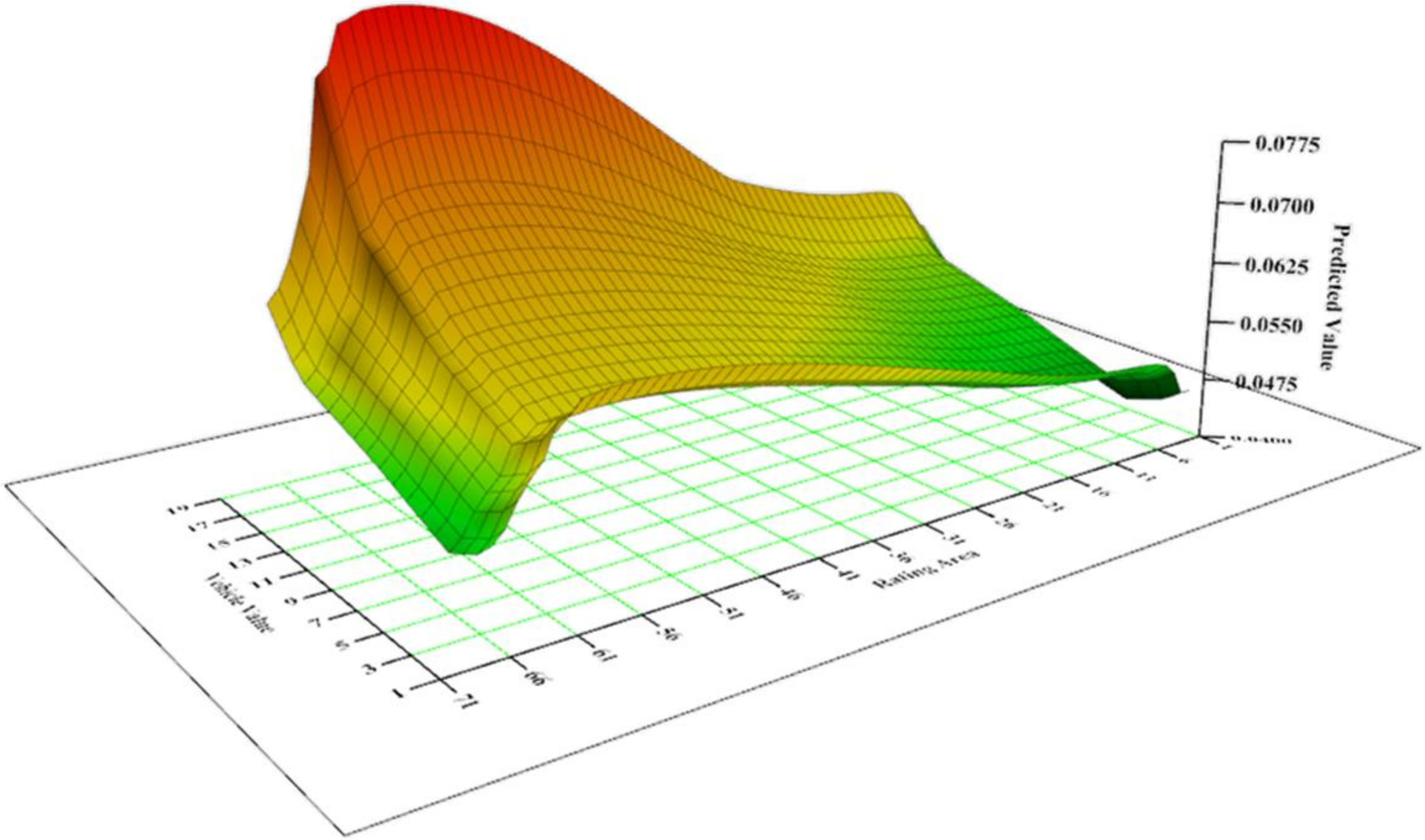
0.0775
0.0700
0.0625
0.0550
0.0475
Predicted Value
Rating Area
Vehicle Value

Saddles



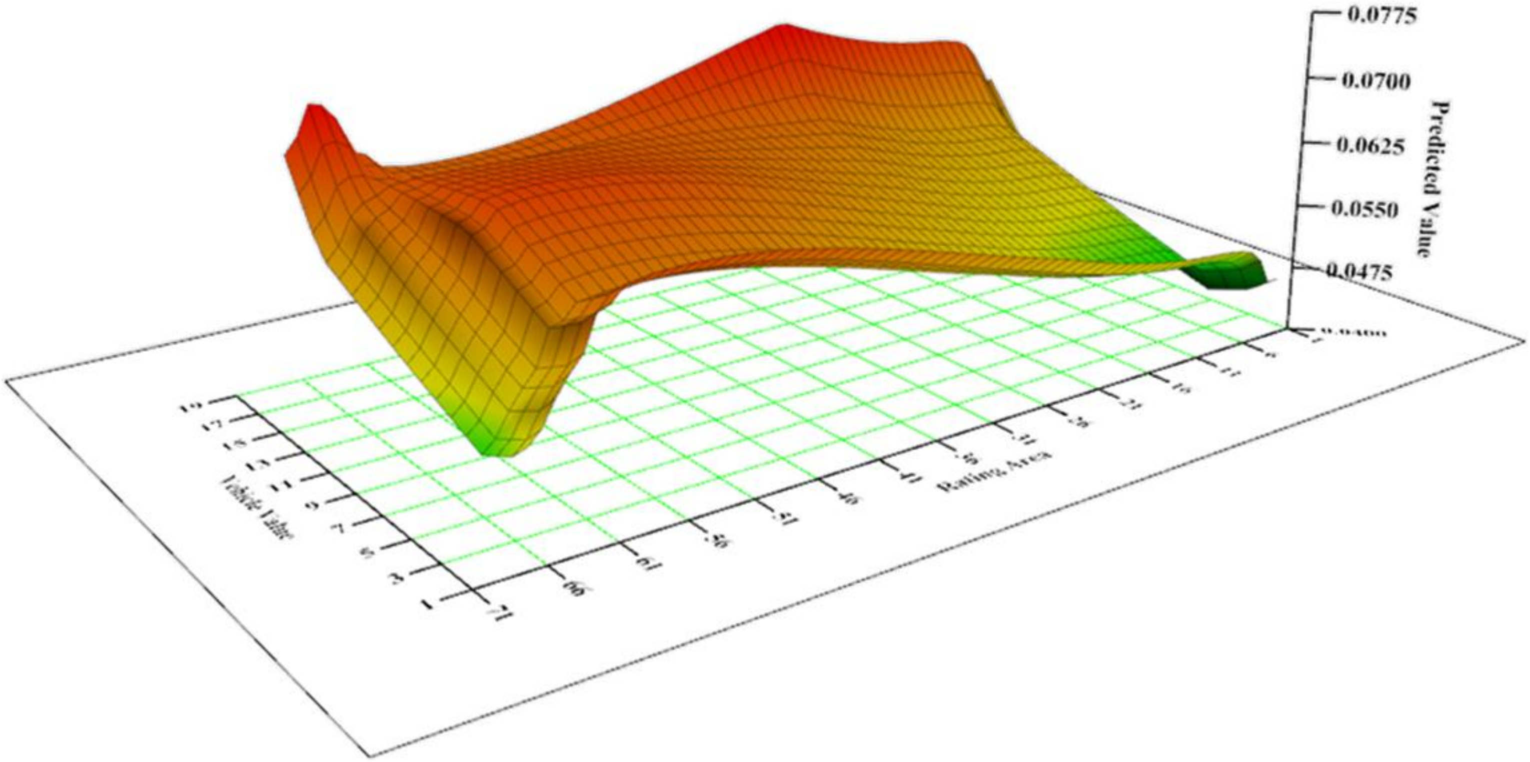
0.0775
0.0700
0.0625
0.0550
0.0475
Predicted Value
Rating Area
Vehicle Value

Saddles



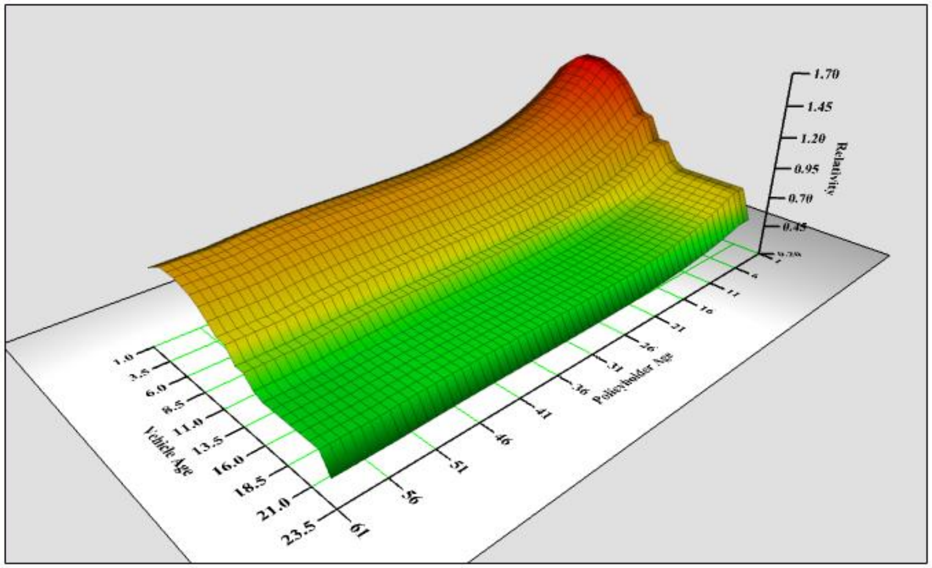
Predicted Value
0.0775
0.0700
0.0625
0.0550
0.0475
Rating Area
Vehicle Value

Saddles

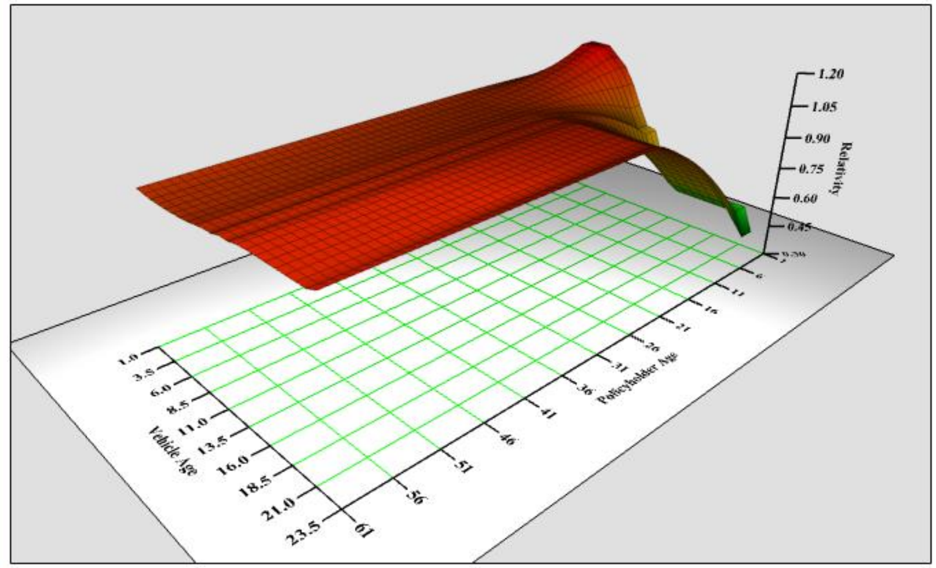


0.0775
0.0700
0.0625
0.0550
0.0475
Predicted Value
Rating Area
Vehicle Value

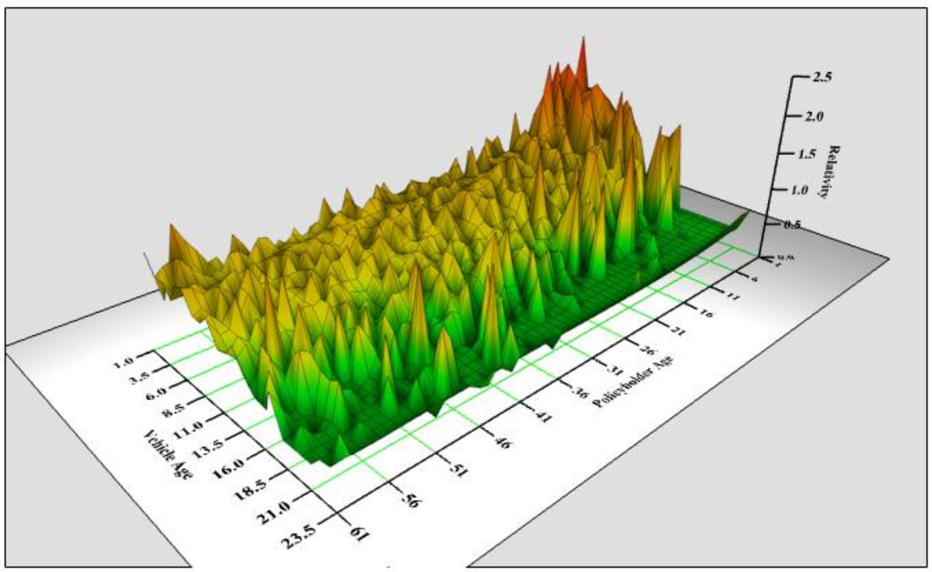
Original



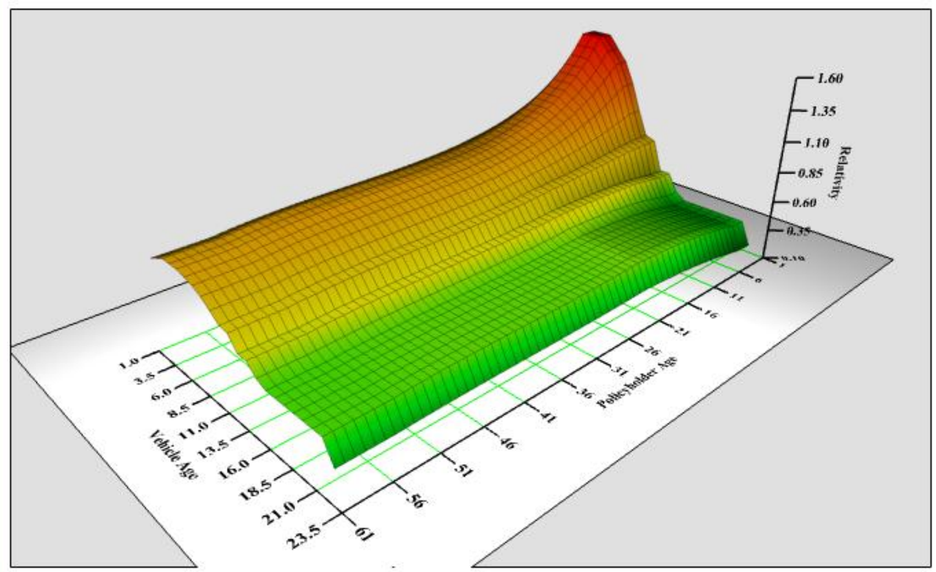
Saddle Parameter



Unsimplified

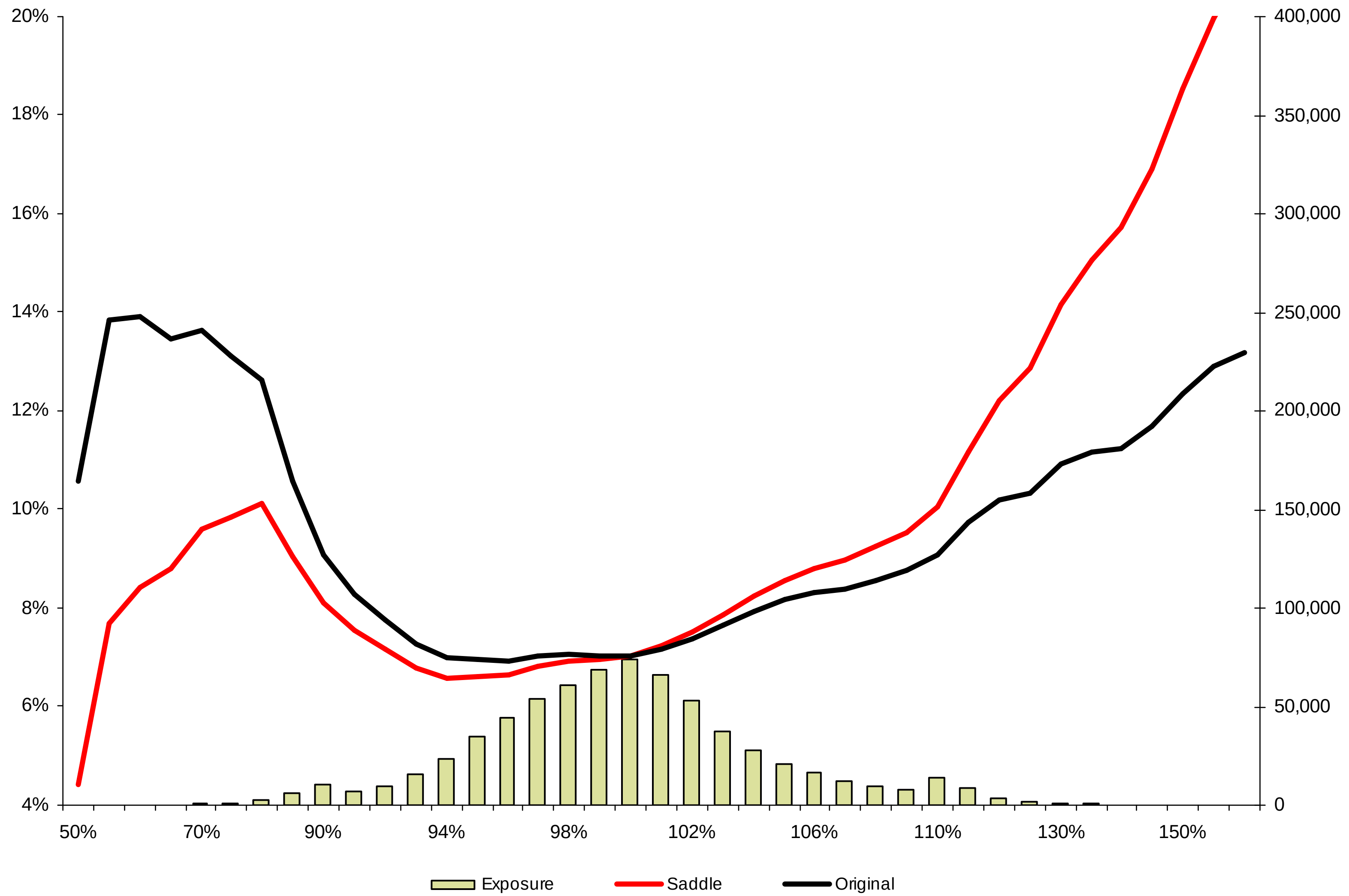


With Saddle



Saddles - model comparison

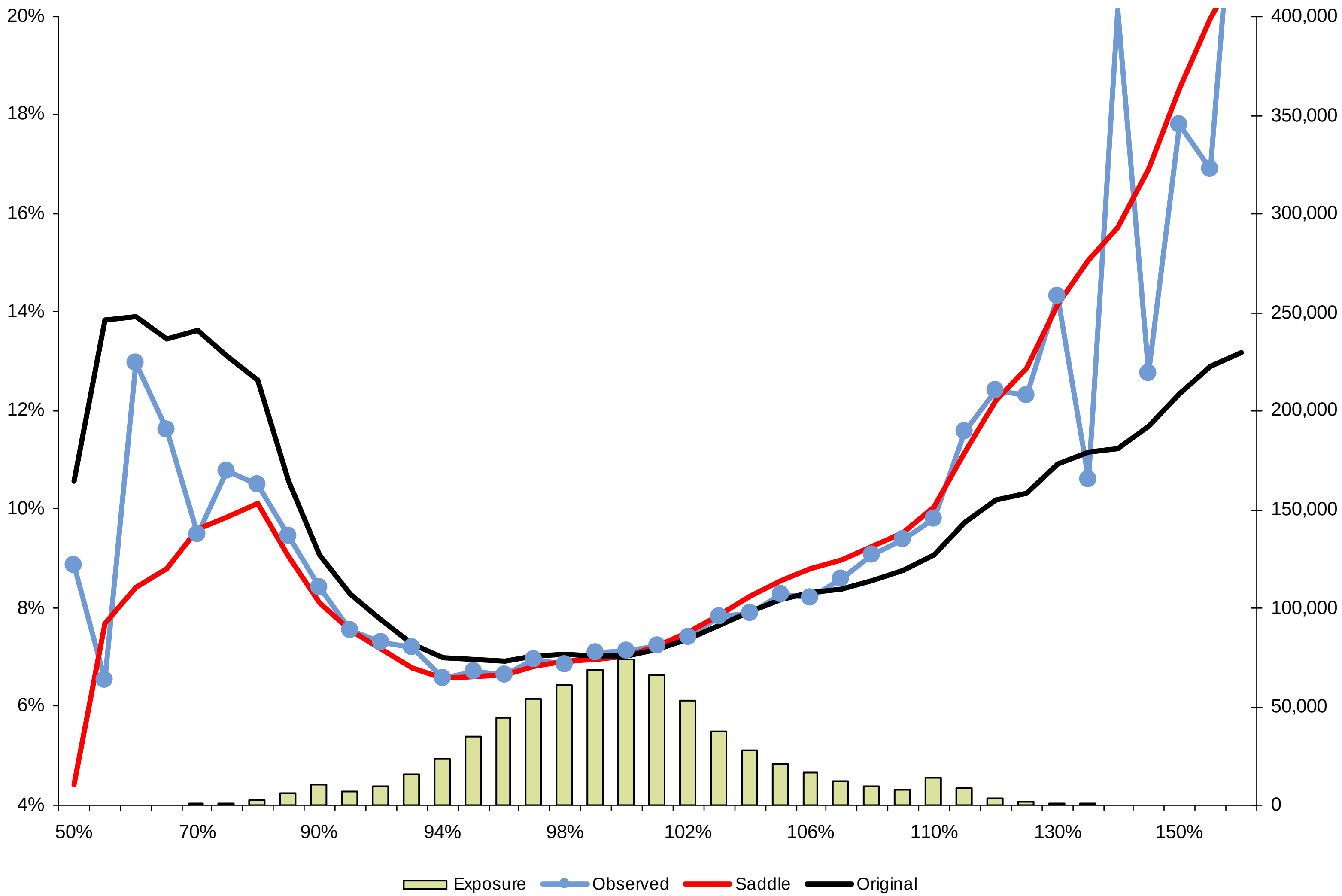
Motor frequency - out of sample



towerswatson.com

Saddles - model comparison

Motor frequency - out of sample



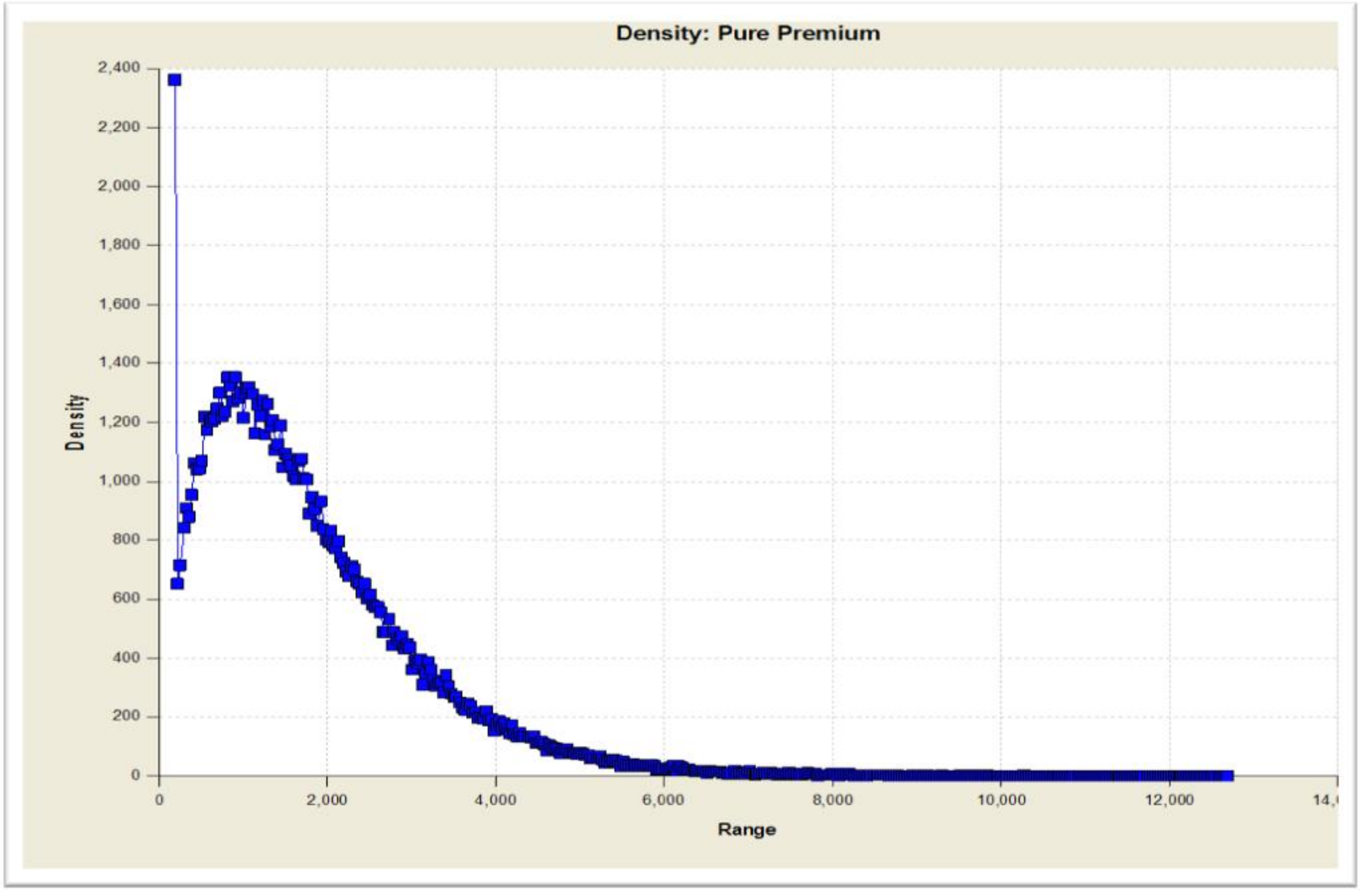
towerswatson.com

Agenda

- "Quadrant Saddles"
- The Tweedie Distribution
- "Emergent Interactions"
- Dispersion Modeling
- Modeling sparse claim types
- Driver Averaging
- Model Validation
- Man (with GLM) vs machine

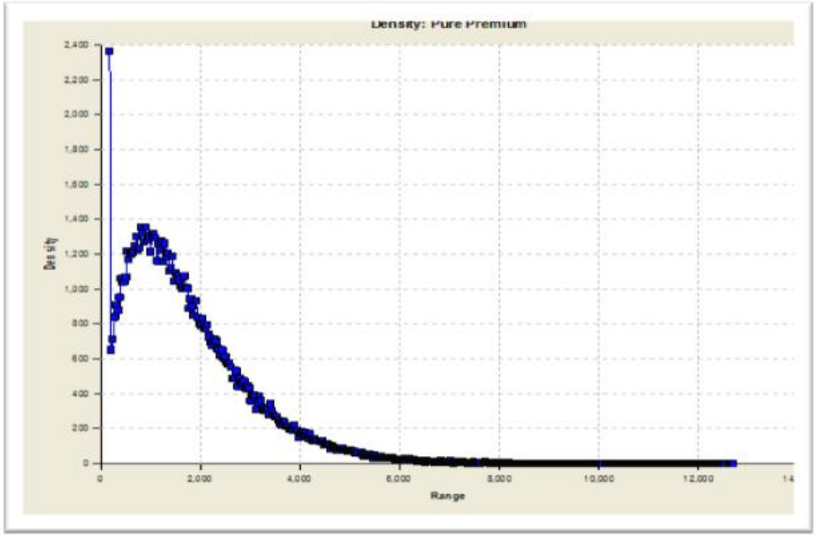
Tweedie GLMs

- Consider the following empirical probability distribution function



Tweedie GLMs

- Raw pure premiums
  - Incurred losses have a point mass at zero and then a continuous distribution
  - Poisson and gamma not suited to this
  - Tweedie distribution has
    - — point mass at zero
    - — a parameter which changes shape above zero



$$f_Y(y;\theta,\lambda,\alpha)=\sum_{n=1}^{\infty}\frac{\left\{(\lambda\omega)^{1-\alpha}\kappa_\alpha(-1/y)\right\}^n}{\Gamma(-n\alpha)n!\,y}\exp\{\lambda\omega[\theta_0 y-\kappa_\alpha(\theta_0)]\}\quad \text{for}\, y>0$$

$$p(Y=0)=\exp\{-\lambda\omega\kappa_\alpha(\theta_0)\}$$

Formulization of GLMs

- Generally accepted standards for link functions and error distribution

Observed Response	Most Appropriate Link Function	Most Appropriate Error Structure	Variance Function
--	--	Normal	μ^0
Claim Frequency	Log	Poisson	μ^1
Claim Severity	Log	Gamma	μ^2
Claim Severity	Log	Inverse Gaussian	μ^3
Raw Pure Premium	Log	Tweedie	μ^T
Retention Rate	Logit	Binomial	$\mu (1-\mu)$
Conversion Rate	Logit	Binomial	$\mu(1- \mu)$

Formulization of GLMs

- More formally:

$$\mathrm{Var}(Y) = \frac{\varphi V(\hat{\mu})}{\omega}$$

Variance Function

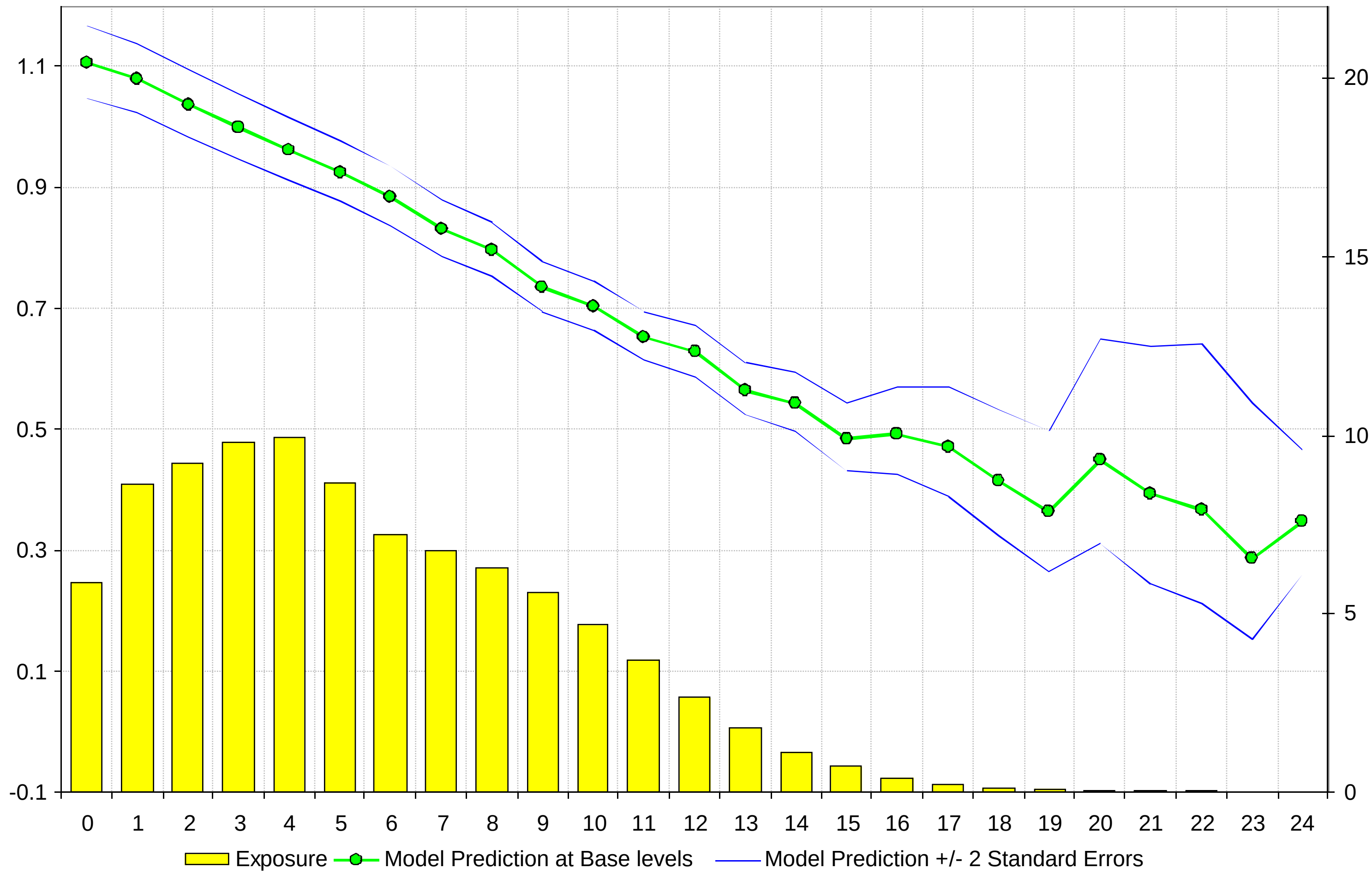
Scale Parameter

Prior Weights

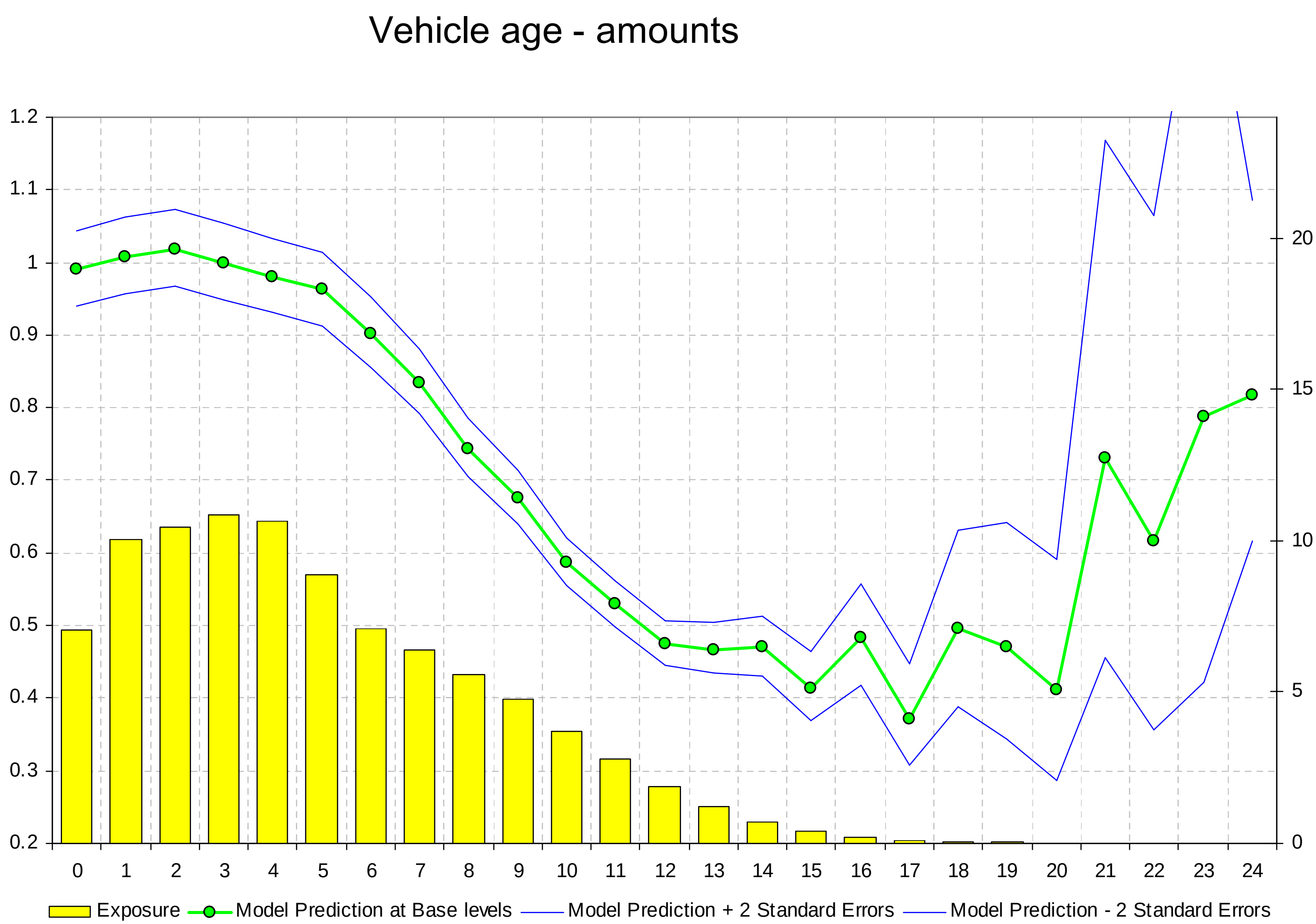
- Tweedie's Variance function: $V(\mu) = \mu^p$
  - $p=1$ Poisson
  - $p=2$ Gamma
  - $1<p<2$ Poisson/Gamma process
- Other concerns
  - Need to estimate both φ and p when fitting models
  - Typically $p \approx 1.5$ for incurred claims

Example 1

Vehicle age - frequency

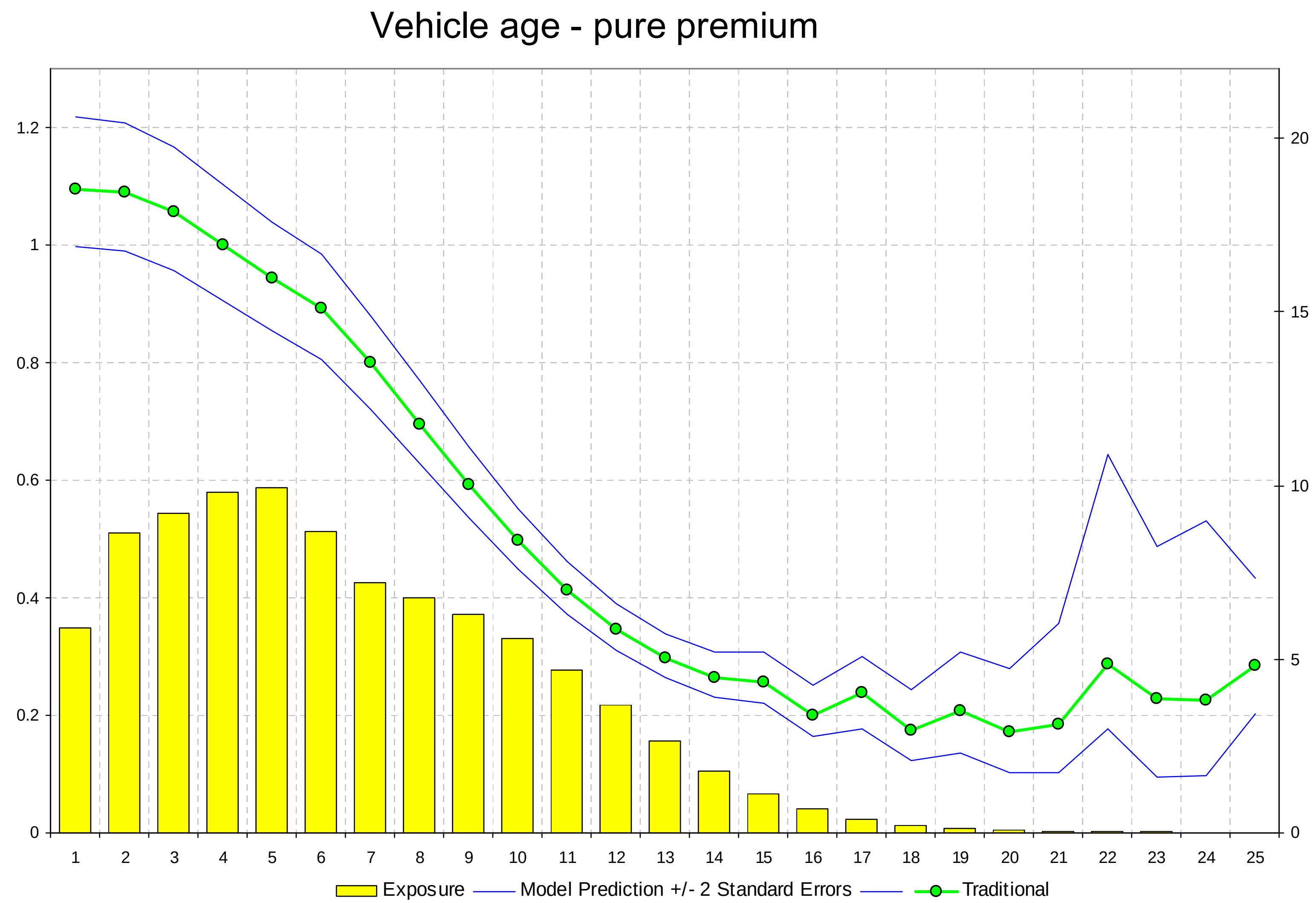


Example 1

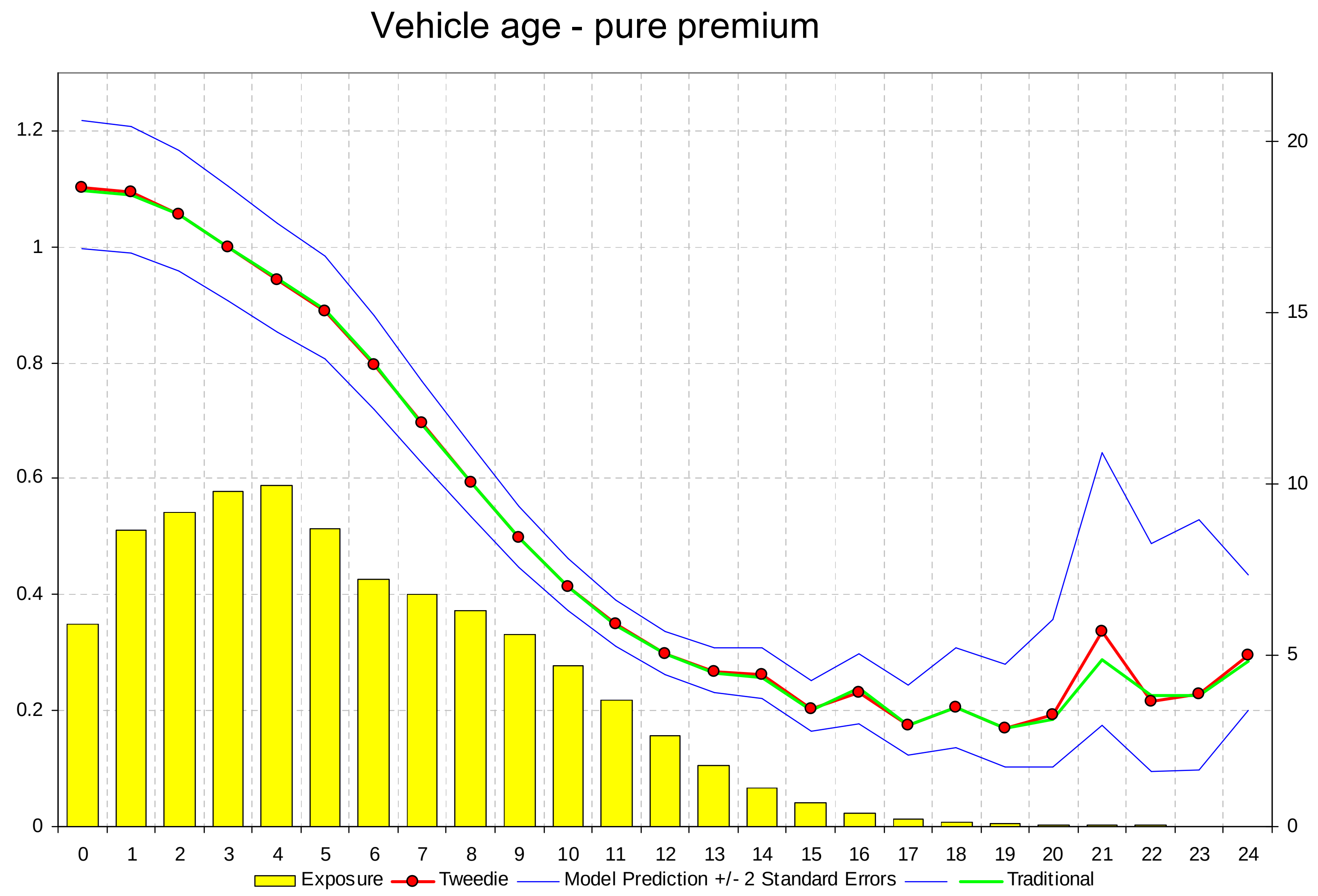


Vehicle age - amounts
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
20
15
10
5
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Exposure
Model Prediction at Base levels
Model Prediction + 2 Standard Errors
Model Prediction - 2 Standard Errors

Example 1

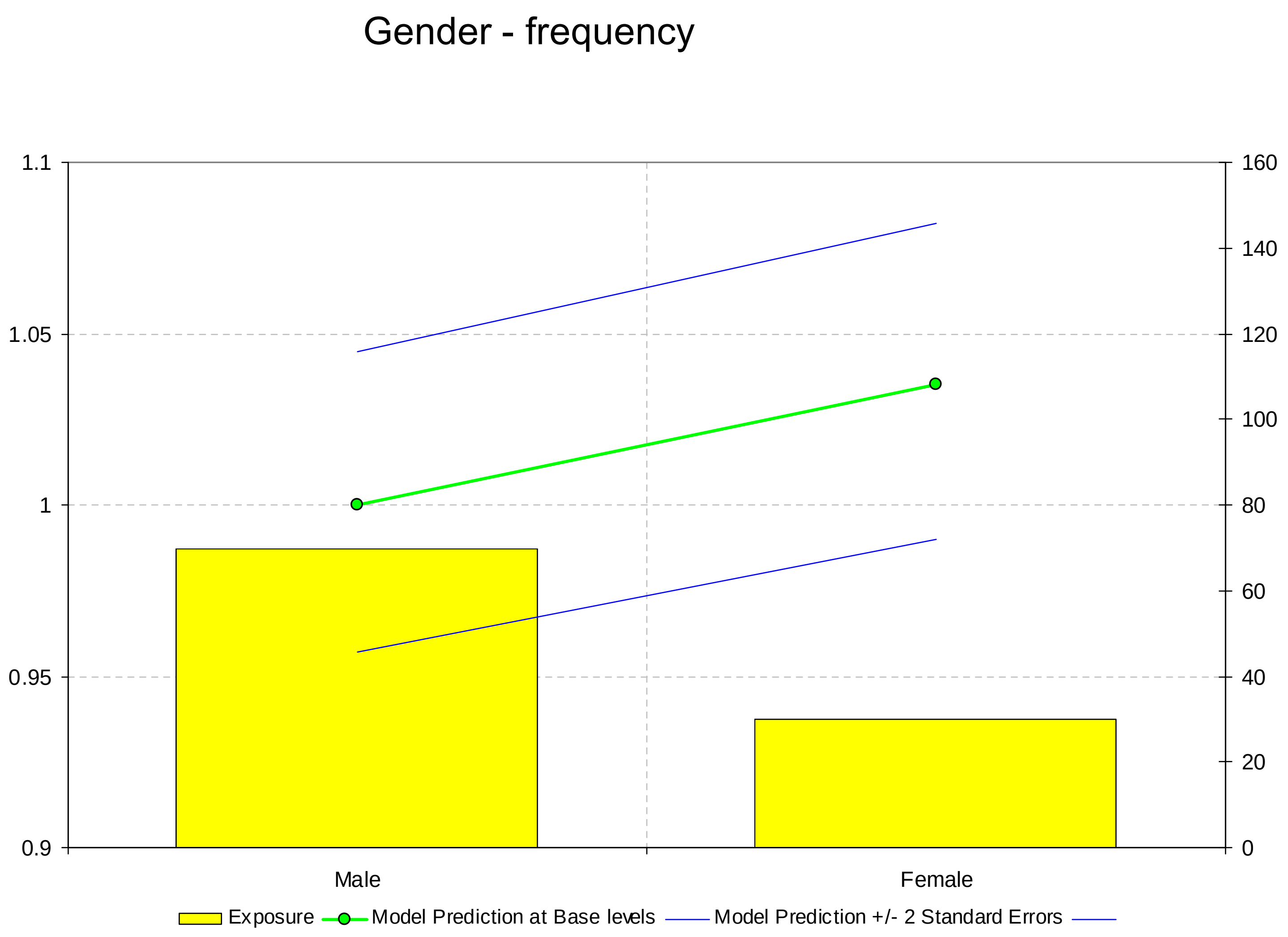

Vehicle age - pure premium
1.2
1
0.8
0.6
0.4
0.2
0
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Exposure
Model Prediction +/- 2 Standard Errors
Traditional

Example 1



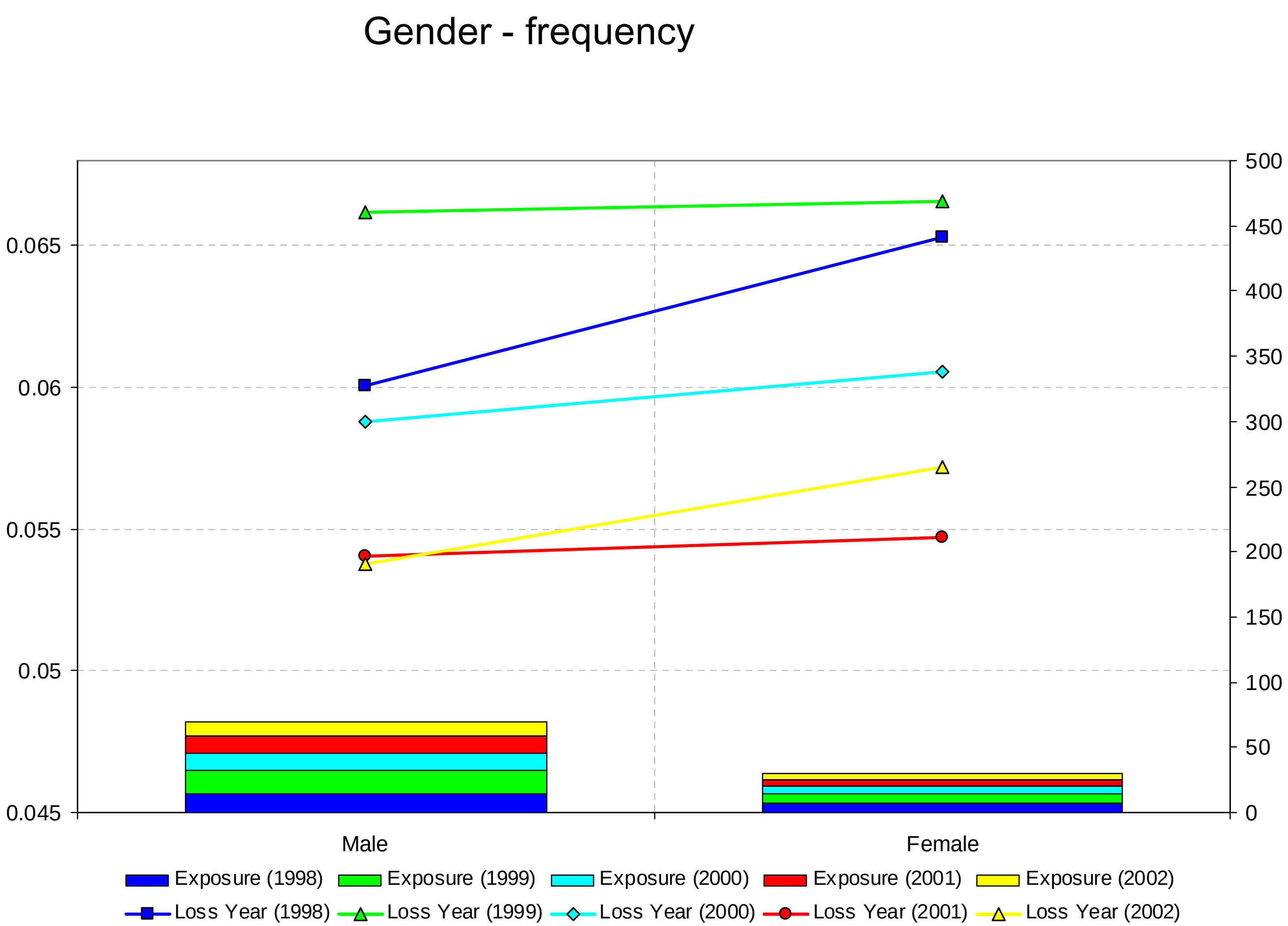
Vehicle age - pure premium
1.2
1
0.8
0.6
0.4
0.2
0
20
15
10
5
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Exposure
Tweedie
Model Prediction +/- 2 Standard Errors
Traditional

Example 2

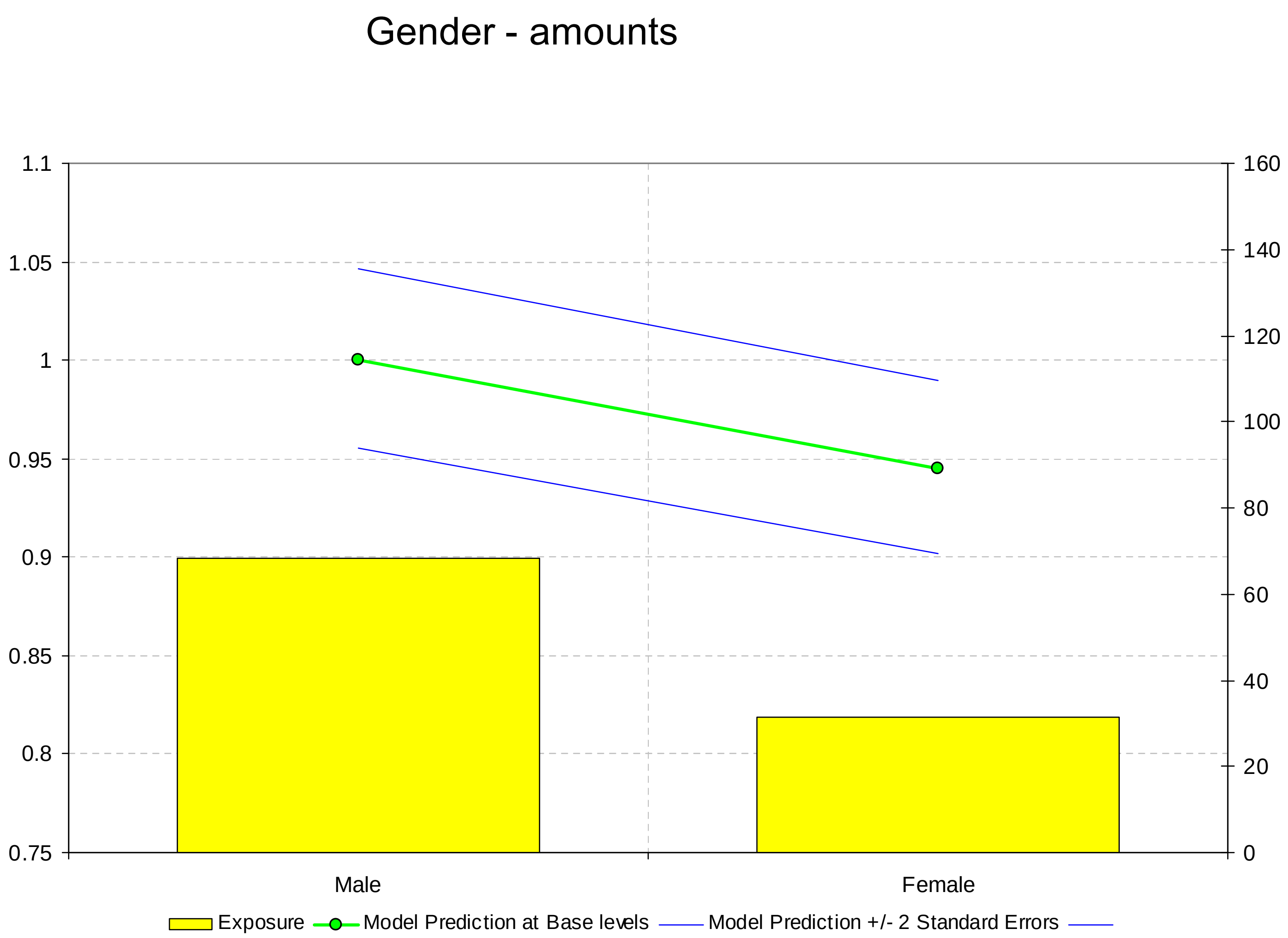


Gender - frequency
1.1
1.05
1
0.95
0.9
160
140
120
100
80
60
40
20
0
Male
Female
Exposure
Model Prediction at Base levels
Model Prediction +/- 2 Standard Errors

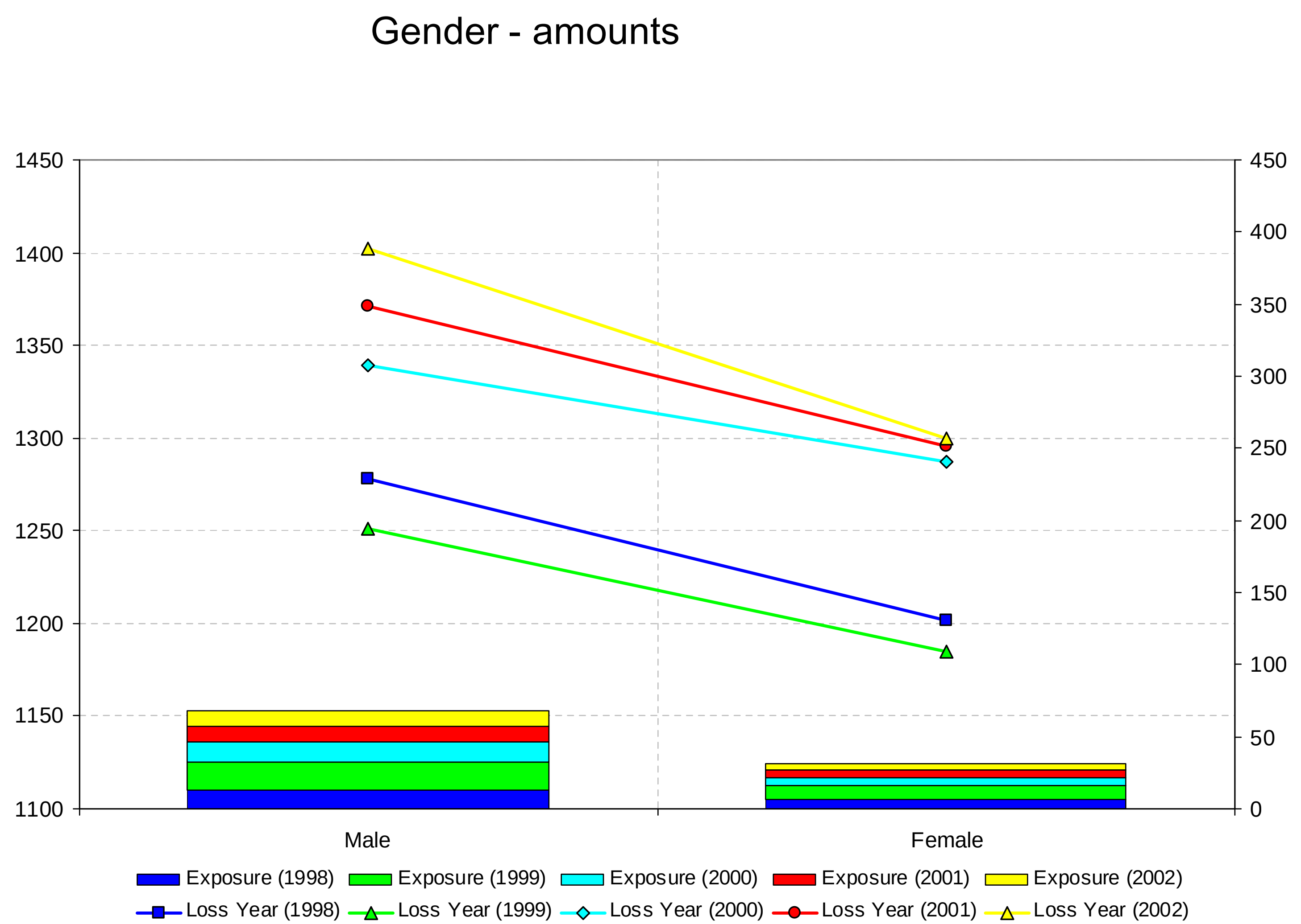
Example 2


Gender - frequency
0.065
0.06
0.055
0.05
0.045
500
450
400
350
300
250
200
150
100
50
0
Male
Female
Exposure (1998)
Exposure (1999)
Exposure (2000)
Exposure (2001)
Exposure (2002)
Loss Year (1998)
Loss Year (1999)
Loss Year (2000)
Loss Year (2001)
Loss Year (2002)

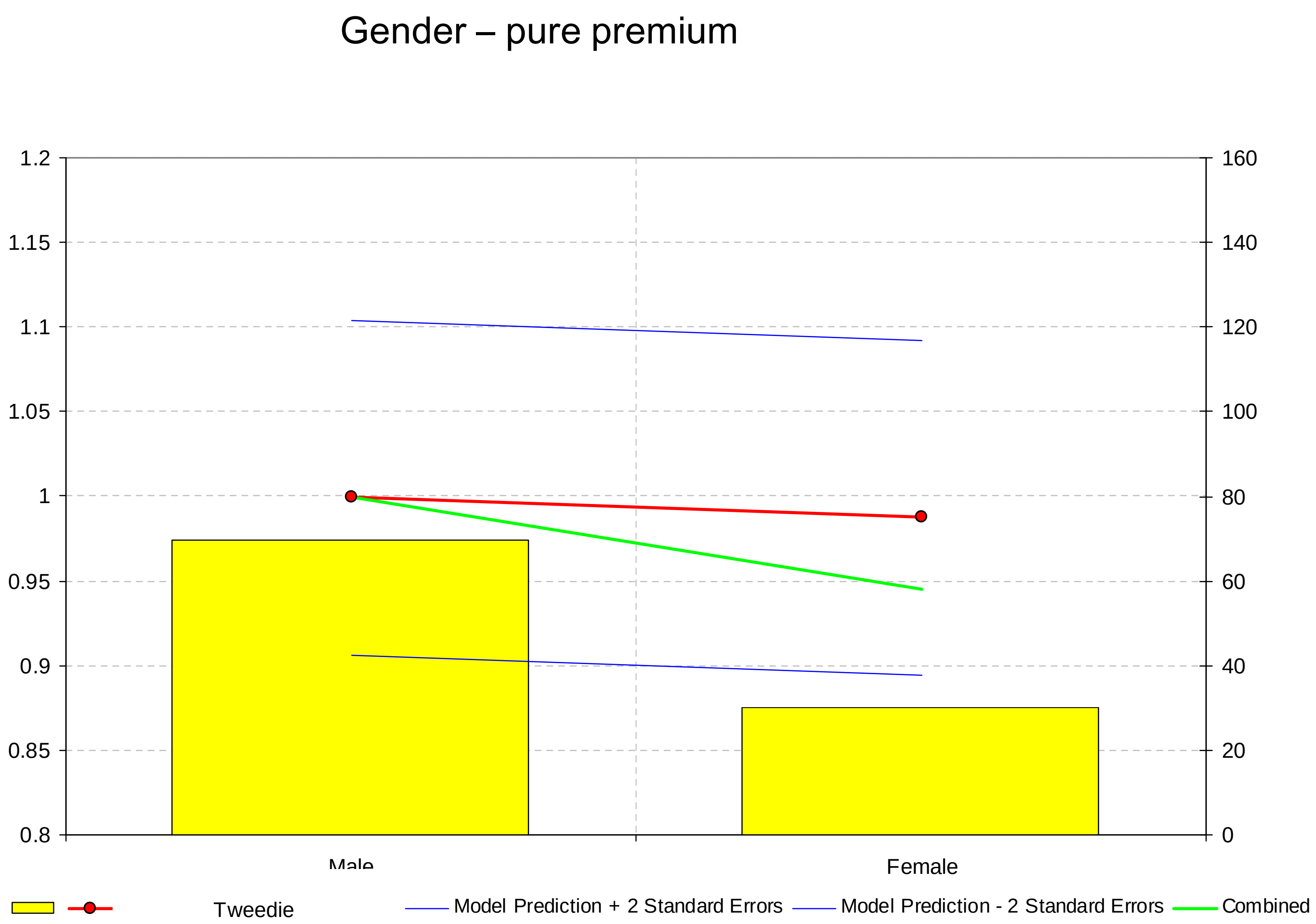
Example 2


Gender - amounts
1.1
1.05
1
0.95
0.9
0.85
0.8
0.75
160
140
120
100
80
60
40
20
0
Male
Female
Exposure
Model Prediction at Base levels
Model Prediction +/- 2 Standard Errors

Example 2


Gender - amounts
1450
1400
1350
1300
1250
1200
1150
1100
450
400
350
300
250
200
150
100
50
0
Male
Female
Exposure (1998)
Exposure (1999)
Exposure (2000)
Exposure (2001)
Exposure (2002)
Loss Year (1998)
Loss Year (1999)
Loss Year (2000)
Loss Year (2001)
Loss Year (2002)

Example 2



Gender – pure premium
1.2
1.15
1.1
1.05
1
0.95
0.9
0.85
0.8
160
140
120
100
80
60
40
20
0
Male
Female
Tweedie
Model Prediction + 2 Standard Errors
Model Prediction - 2 Standard Errors
Combined

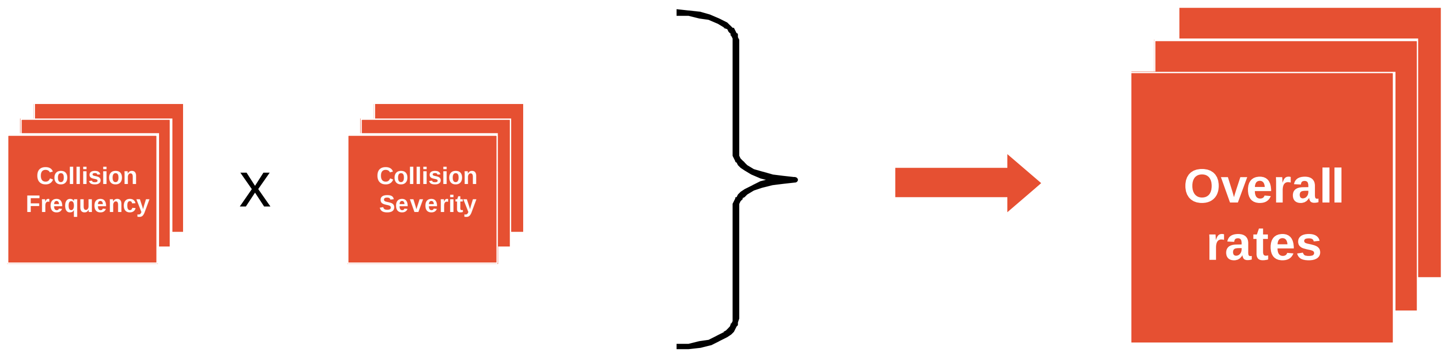
Tweedie GLMs

- Helpful when it's important to fit to incurred costs directly
- Similar results to frequency/severity traditional approach if frequency and amounts effects are clearly weak or clearly strong
- Distorted by large insignificant effects
- Removes understanding of what is driving results
- Smoothing harder

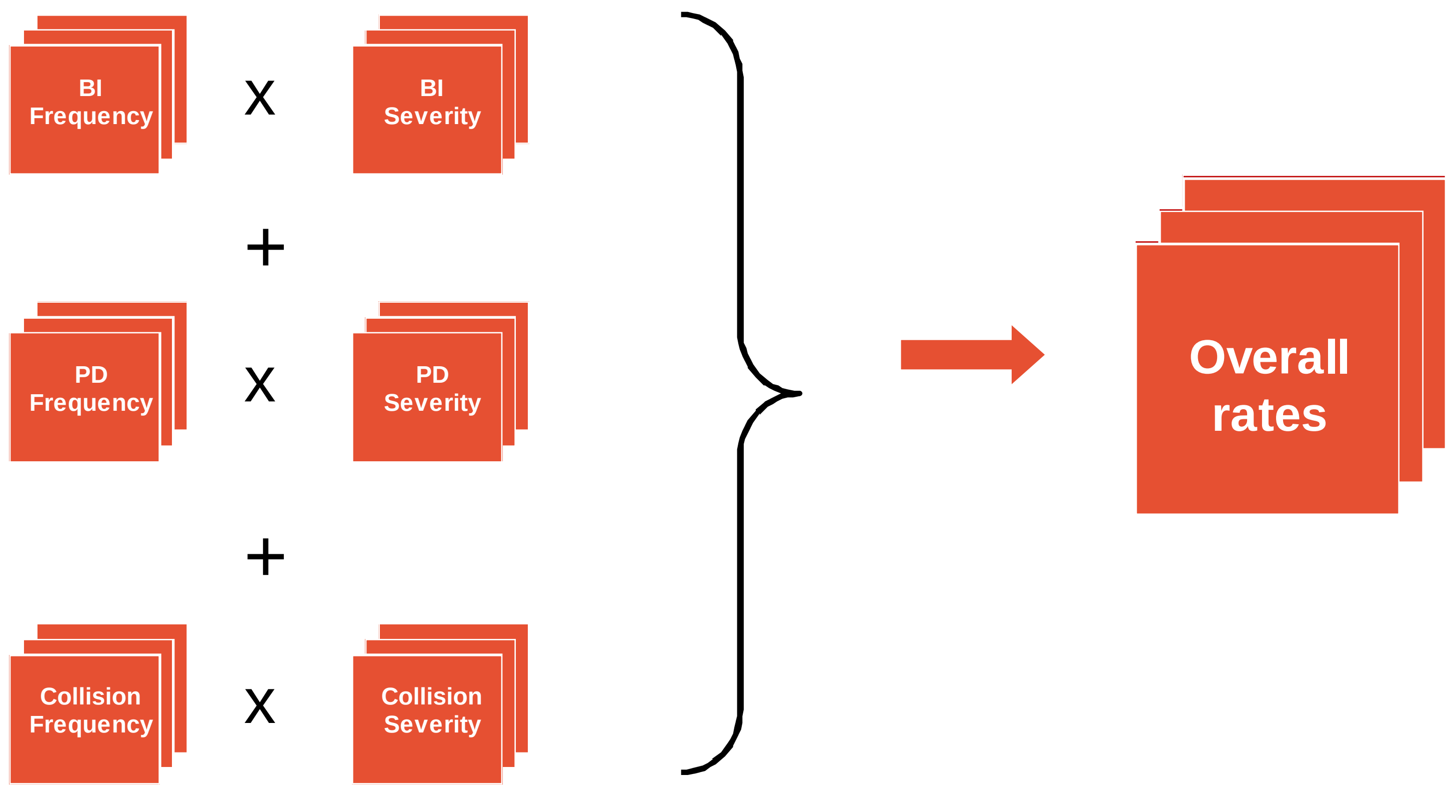
Agenda

- "Quadrant Saddles"
- The Tweedie Distribution
- "Emergent Interactions"
- Dispersion Modeling
- Modeling sparse claim types
- Driver Averaging
- Model Validation
- Man (with GLM) vs machine

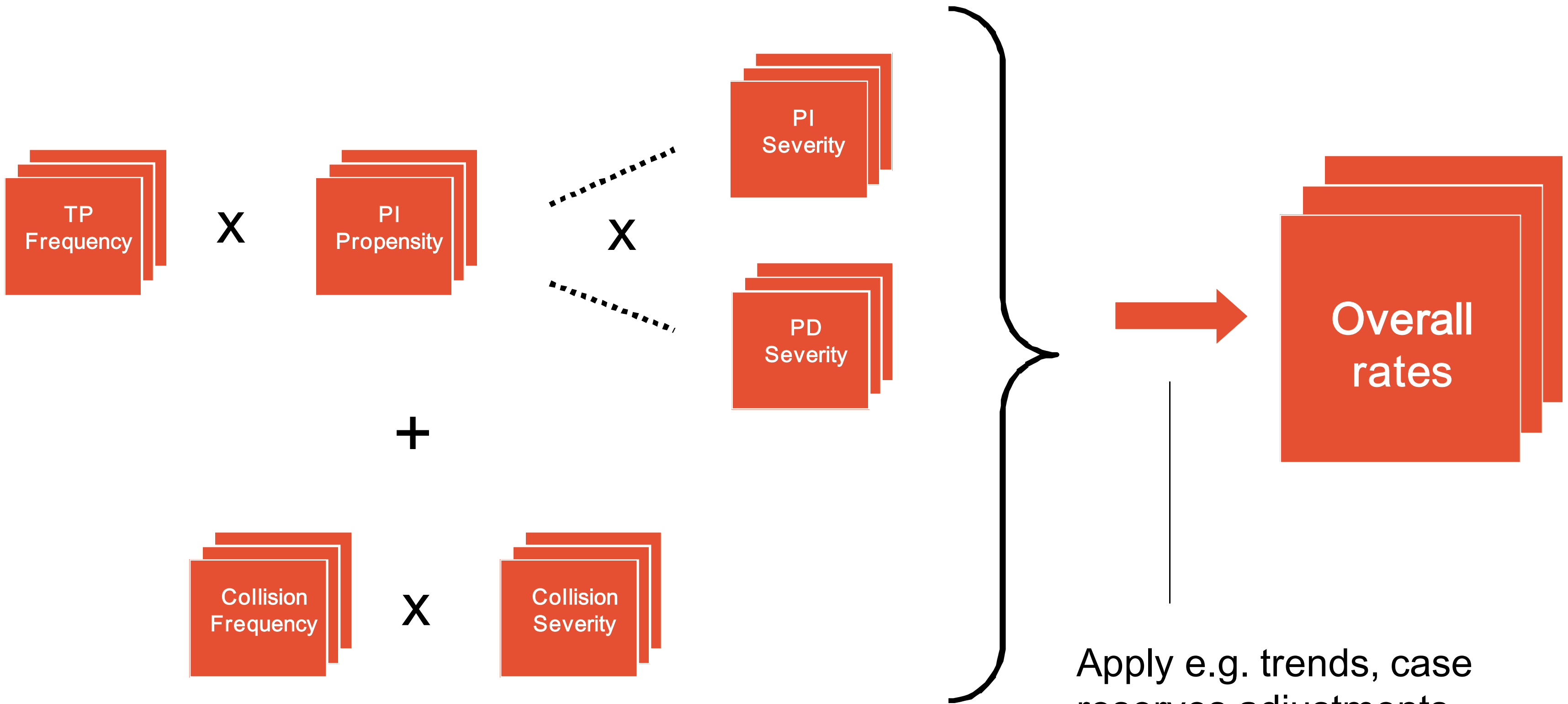
Combining models



Combining models

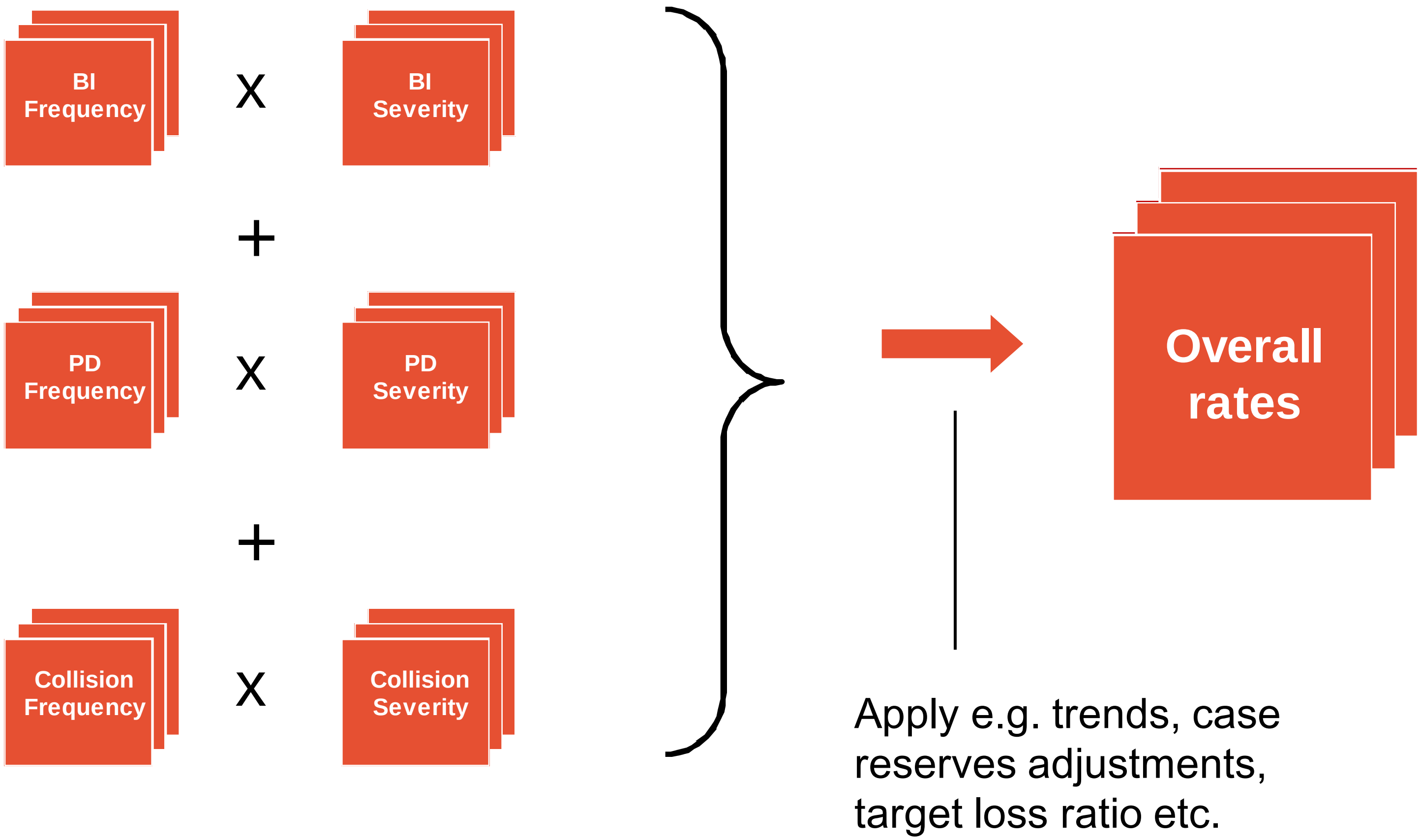


Combining models



Apply e.g. trends, case reserves adjustments, target loss ratio etc.

Combining models



Combining models

- Take models
- Take relevant mix of business
  - eg current in force policies
- For each record calculate expected frequencies and severities according to the models
- For each record, calculate expected total cost of claims "C"
- Fit a GLM to "C" using all available factors

Combining models

		PD Frequency	PD Severity	BI Frequency	BI Severity
Intercept		32%	$1000	12%	$4860
Gender	Male	1.000	1.000	1.000	1.000
	Female	0.750	1.200	0.667	0.900
Area	Town	1.000	1.000	1.000	1.000
	Country	1.250	0.700	0.750	0.833

Policy	**Gen**	**Area**	**Freq1**	**Sev1**	**Freq2**	**Sev2**	**RP1**	**RP2**	**RISKPREM**
…	…	…	…	…	…	…	…	…	…
82155654	M	T	32%	1000	12%	4860	320	583.20	903.20
82168746	F	T	24%	1200	8%	4374	288	349.92	637.92
82179481	M	C	40%	700	9%	4050	280	364.50	644.50
82186845	F	C	30%	840	6%	3645	252	218.70	470.70
…	…	…	…	…	…	…	…	…	…

Except…

Policy	Gen	Area	Freq1	Sev1	Freq2	Sev2	RP1	RP2	RISKPREM
…	…	…	…	…	…	…	…	…	…
82155654	M	T	32%	1000	12%	4860	320	583.20	903.20
82168746	F	T	24%	1200	8%	4374	288	349.92	637.92
82179481	M	C	40%	700	9%	4050	280	364.50	644.50
82186845	F	C	30%	840	6%	3645	252	218.70	470.70
…	…	…	…	…	…	…	…	…	…

- The global risk premium is not multiplicative
- In the town, women have a modeled claim cost 29% lower than men
  - 637.92/903.20=0.706
- In the country, women have a modeled claim cost 27% lower than men
  - 470.07/644.50=0.730

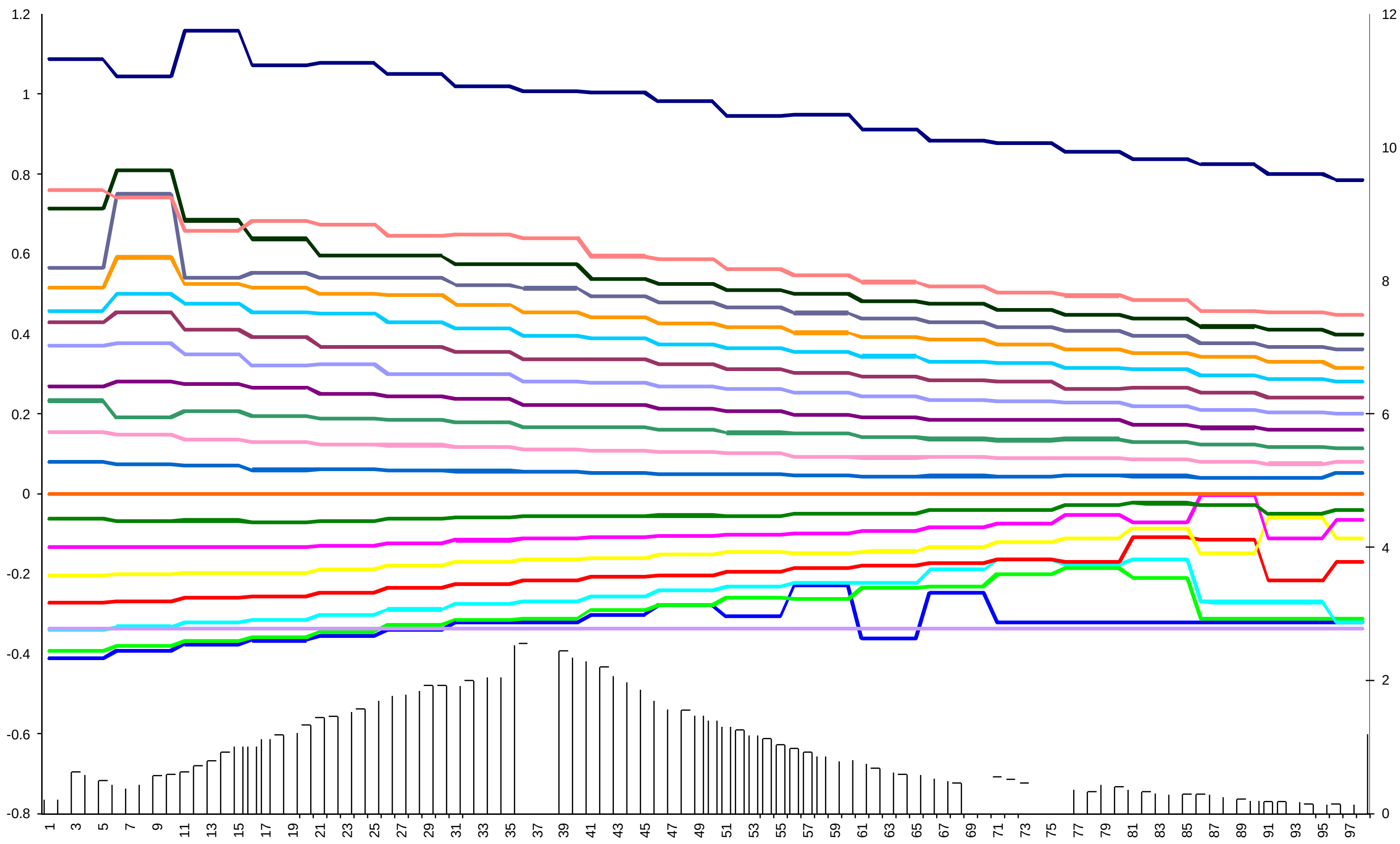
To solve…

Policy	Gen	Area	Freq1	Sev1	Freq2	Sev2	RP1	RP2	RISKPREM
…	…	…	…	…	…	…	…	…	…
82155654	M	T	32%	1000	12%	4860	320	583.20	903.20
82168746	F	T	24%	1200	8%	4374	288	349.92	637.92
82179481	M	C	40%	700	9%	4050	280	364.50	644.50
82186845	F	C	30%	840	6%	3645	252	218.70	470.70
…	…	…	…	…	…	…	…	…	…

- We can capture this result exactly with an interaction

Total risk premium			
Intercept		$903.20	
	Sex	Male	Female
Area	Town	1.000	0.706
	Country	0.714	0.521

Example "emergent" interaction



"Emergent" interactions

- In the above examples the interaction "emerged" from the risk premium step
- Emergent interactions are ***not*** risk insights, there is no subtle risk effect we have just discovered
  - The different behavior is by peril, and the rating factors are just bad proxies for the peril effects
- Emergent interactions are corrections to fix problems we have introduced
- Best solution is by peril pricing
  - Reflects true behavior
  - Underlying models simple to understand and implement
- If not, check for emergent interactions in the risk premium

Agenda

- "Quadrant Saddles"
- The Tweedie Distribution
- "Emergent Interactions"
- Dispersion Modeling
- Modeling sparse claim types
- Driver Averaging
- Model Validation
- Man (with GLM) vs machine

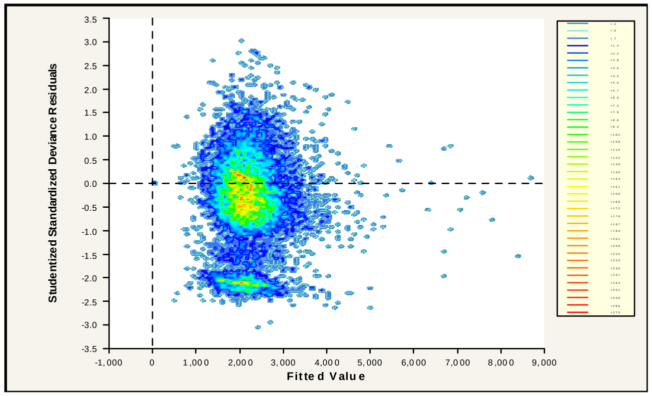
Modeling the Insurance Risk

ISSUE:

Heterogeneous exposure bases

- Different policies within the same line can cover entirely different structures (i.e. commercial property)

- Goal of a predictive model
  - Ideally would like to separate the heterogenous exposure bases
  - Joint-modeling techniques and quasi-likelihood functions allow for analysis of heterogeneous environment without separation



Two concentrations suggests two perils: split or use joint modeling

Heterogeneous Exposure Bases

- If possible should be modeled separately
  - If model together, exposures with high variability may mask patterns of less random risks
  - If loss trends vary by exposure class, the proportion each represents of the total will change and may mask important trends
  - Independent predictors can have different effects on different perils
- If cannot, use joint modeling techniques to improve overall fit

Generalized Linear Models

- Formulation of deviance – logarithm of a ratio of likelihoods

$$\frac{D}{a(\varphi)} = \ln\left(\frac{Act}{Exp}\right)^2$$

Where:

$$Act = f_Y\left(y;\tilde{\theta},\varphi\right) \ni E(Y) = y = b'\left(\tilde{\theta}\right)$$

$$Exp = f_Y\left(y;\hat{\theta},\varphi\right) \ni E(Y) = \hat{\mu} = b'\left(\hat{\theta}\right)$$

Then:

$$\frac{D}{a(\varphi)} = \ln\left(\frac{f_Y(y;\tilde{\theta},\varphi)}{f_Y(y;\hat{\theta},\varphi)}\right)^2 = 2\times\left[\frac{y\tilde{\theta} - y\hat{\theta} - b\left(\tilde{\theta}\right) + b\left(\hat{\theta}\right)}{a(\varphi)}\right]$$

Generalized Linear Models

- Analyzing the scale parameter
  - When modeling homogeneous data

$$a(\varphi) = \frac{\varphi}{\omega} \Rightarrow \varphi = \frac{D}{\text{dof}}$$

  - Heterogeneous data requires a more rigorous definition of the scale function
    - Scale parameter could vary in a systematic way with other predictors
    - Construct and fit formal models for the dependence of both the mean and the scale

Dispersion Model Form

- Double generalized linear models
  - Response model

$$Y \sim f_Y(y;\theta;\varphi)$$
$$E(Y) = b'(\theta)$$
$$Var(Y) = \frac{\varphi b''(\theta)}{\omega}$$

  - Dispersion model

$$D \sim f_D(d;\xi,\tau)$$
$$E(D) = b'(\xi)$$
$$Var(D) = \frac{\tau b''(\xi)}{\omega}$$

Where

$$d = \frac{(Y-\mu)^2}{V(\mu)}$$

Dispersion Model Form

- Dispersion adjustments
  - Pearson residual has excess variability (deviance residual has bias)

Distribution	Adjustment
Normal	0
Poisson	f/(2m)
Gamma	3f

  - Parameter in the adjustment term is the scale parameter from the original response model

Dispersion Model Results

- Dispersion model is integrated with original response model

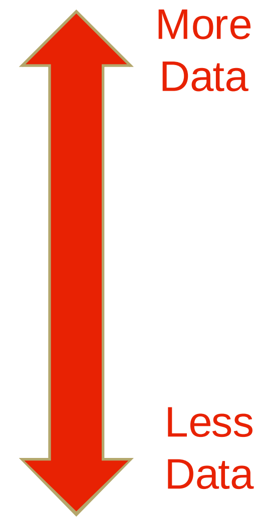
	Response	Weight
Initial Response Model	Loss / Exposure	Exposure
Dispersion Model	Squared Pearson Residual	Exposure/ (Exposure + Adjustment)
Final Response Model	Loss / Exposure	Exposure/ Squared Pearson Residual

Agenda

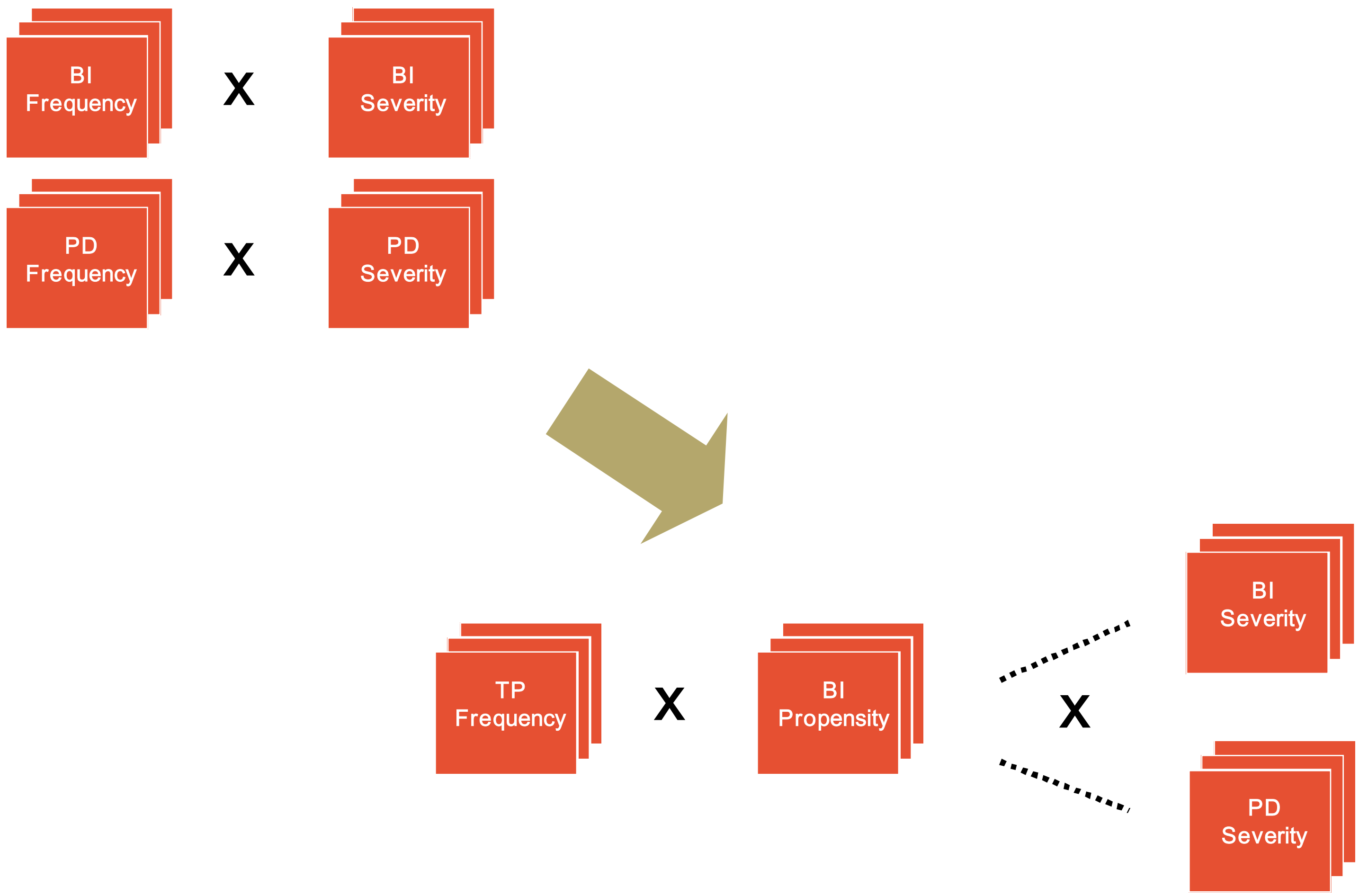
- "Quadrant Saddles"
- The Tweedie Distribution
- "Emergent Interactions"
- Dispersion Modeling
- Modeling sparse claim types
- Driver Averaging
- Model Validation
- Man (with GLM) vs machine

Amplification of the BI signal using PD experience

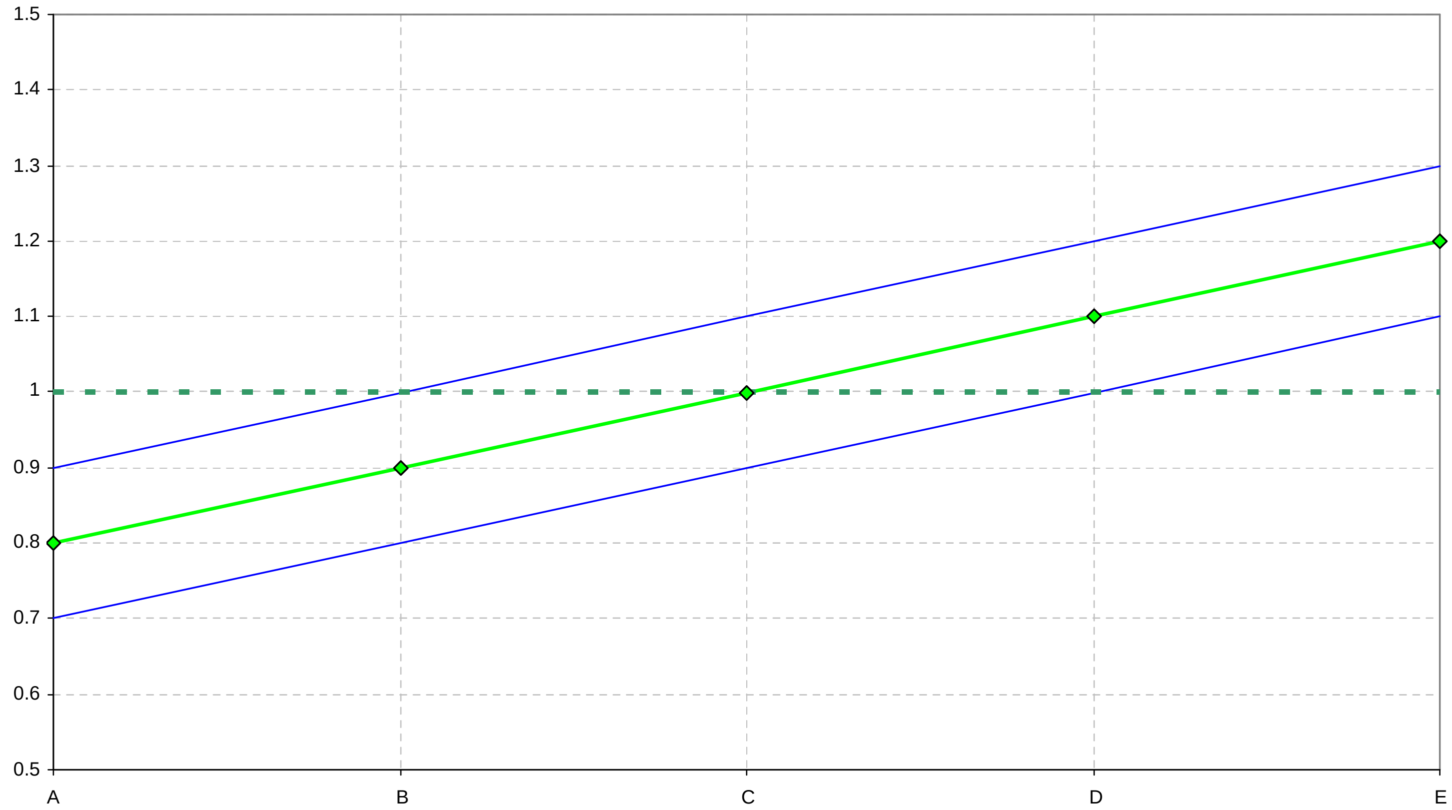
- Fit straight to BI
- Use PD model as a guide in free fitting BI
- Use PD model structure
- Offset PD relativities onto BI data as starting point
- BI/PD proportion model:
  - BI frequency = BI/PD proportion * PD frequency



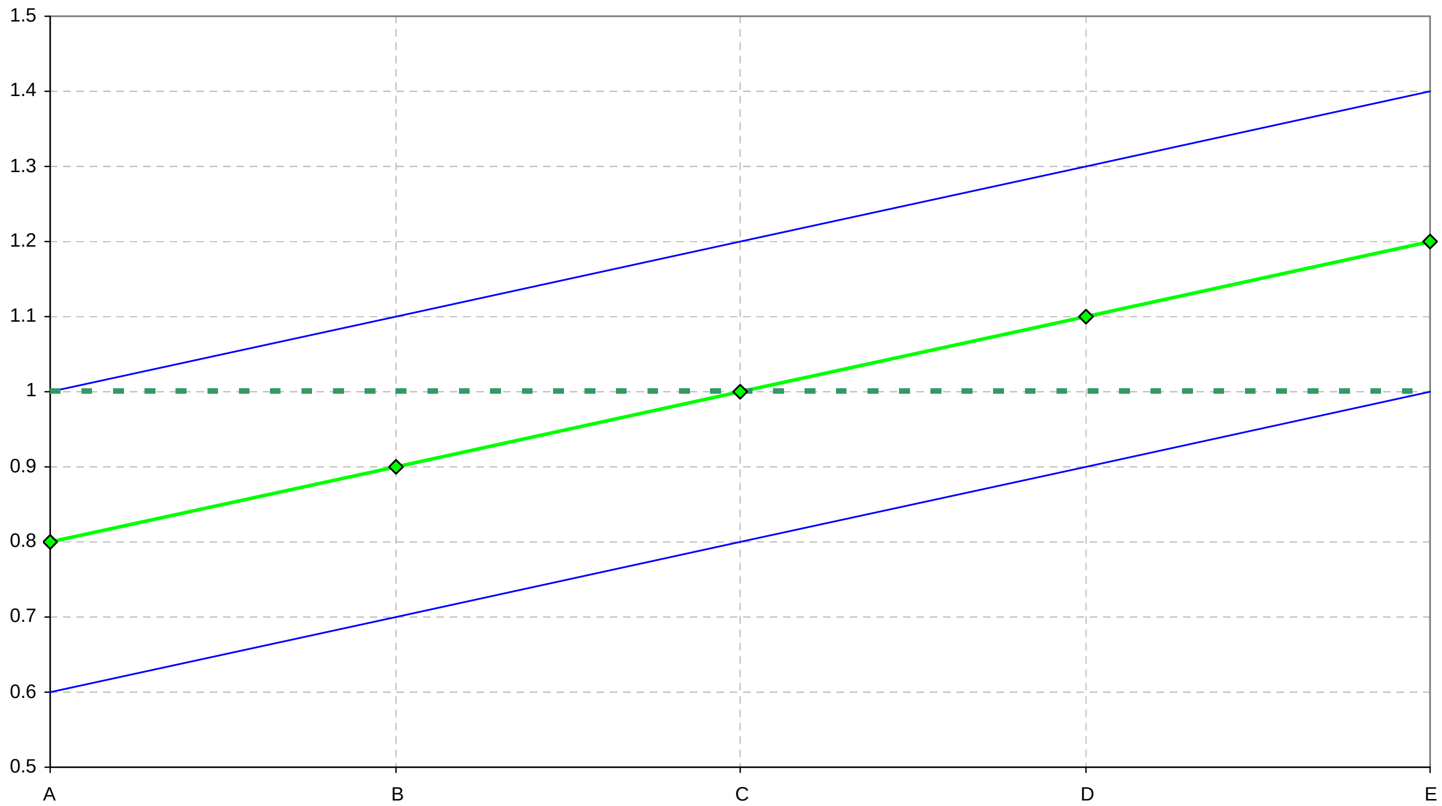
Reference models



Reference models

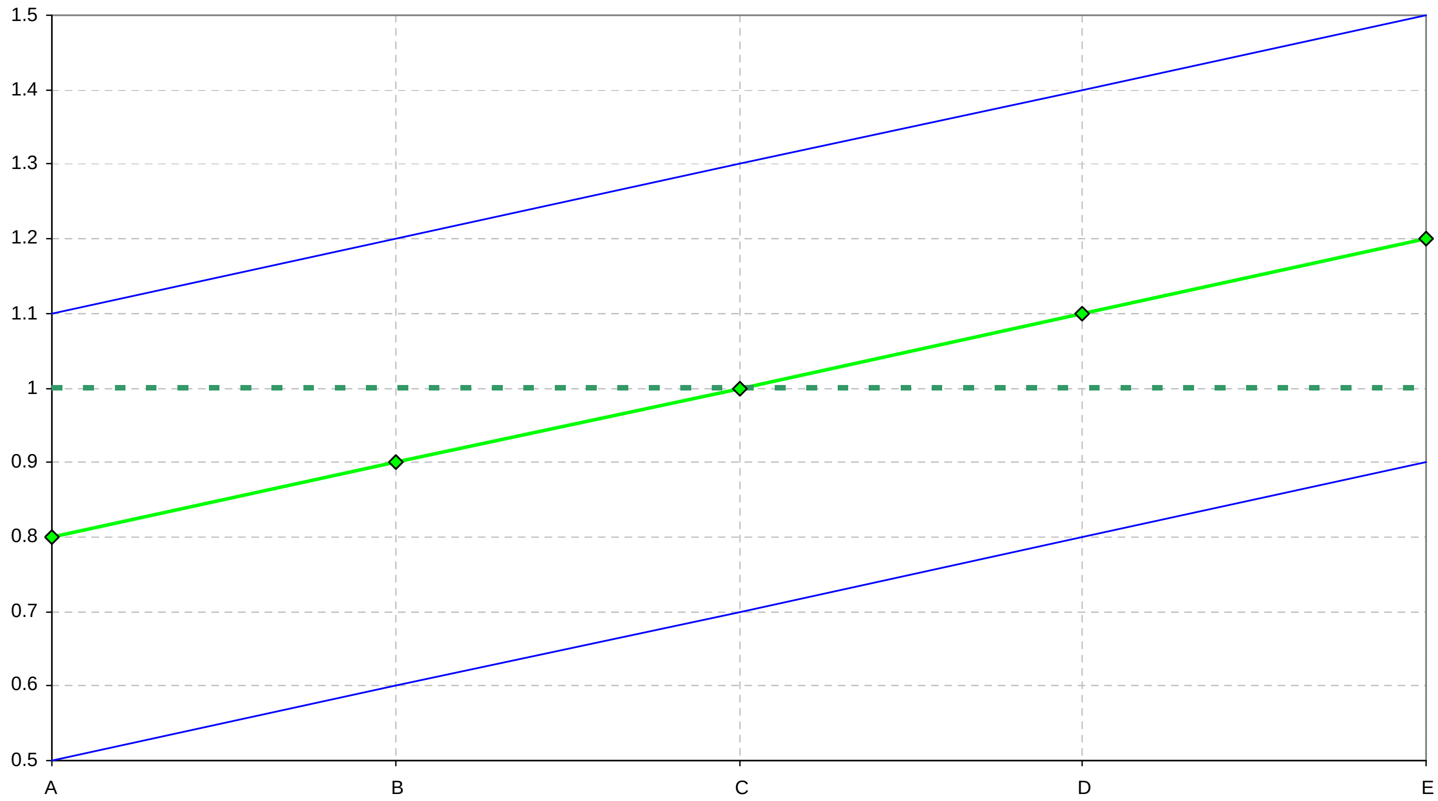


Reference models



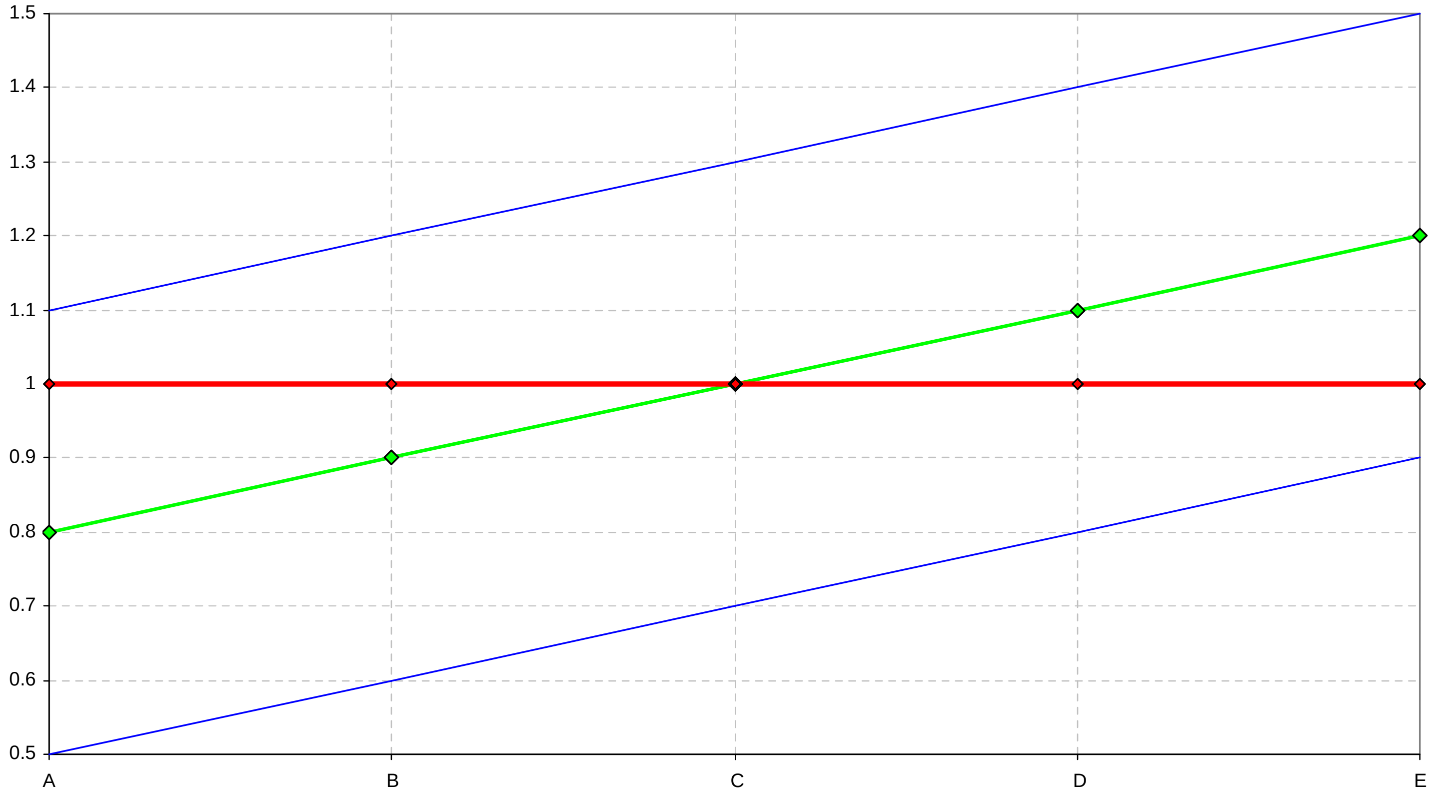
towerswatson.com

Reference models

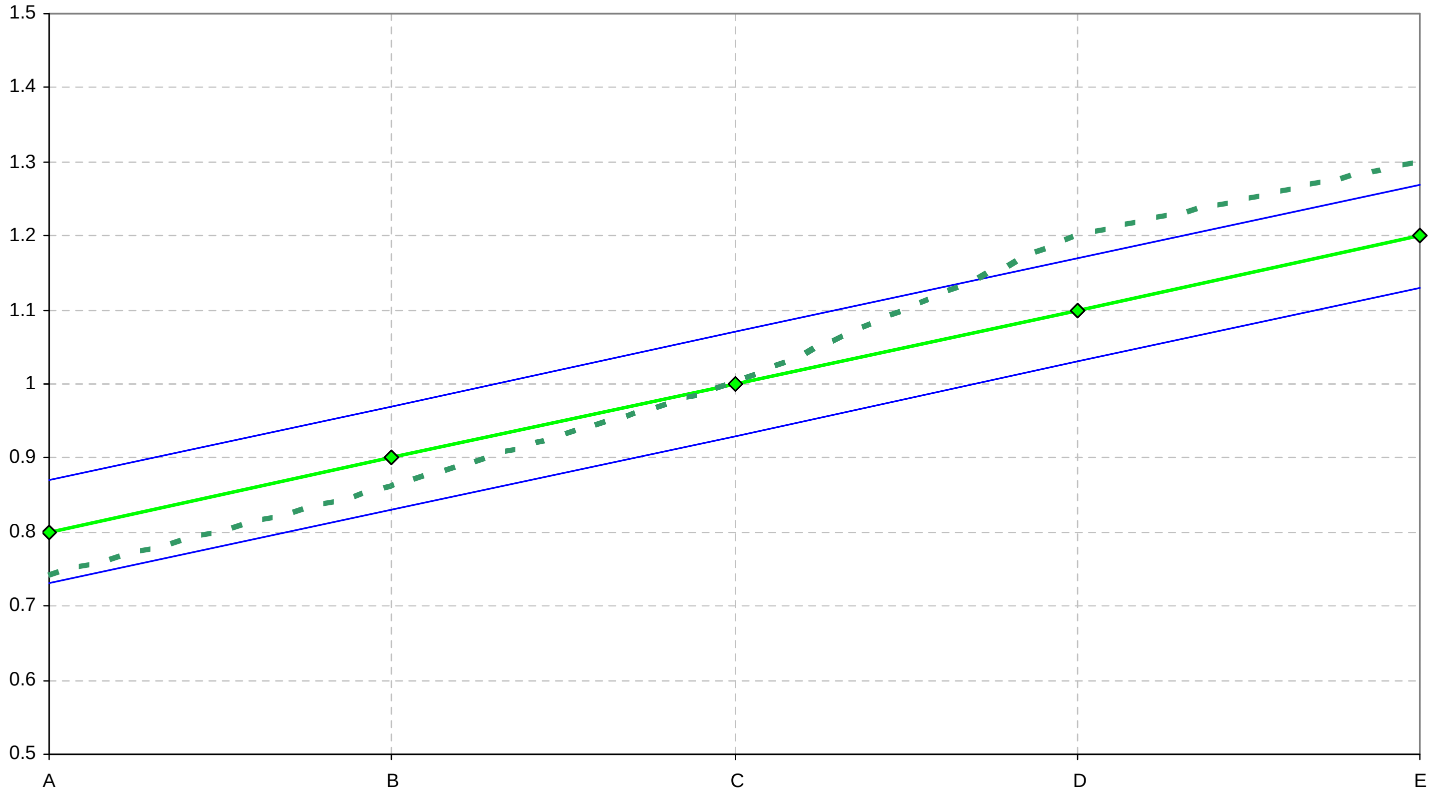


towerswatson.com

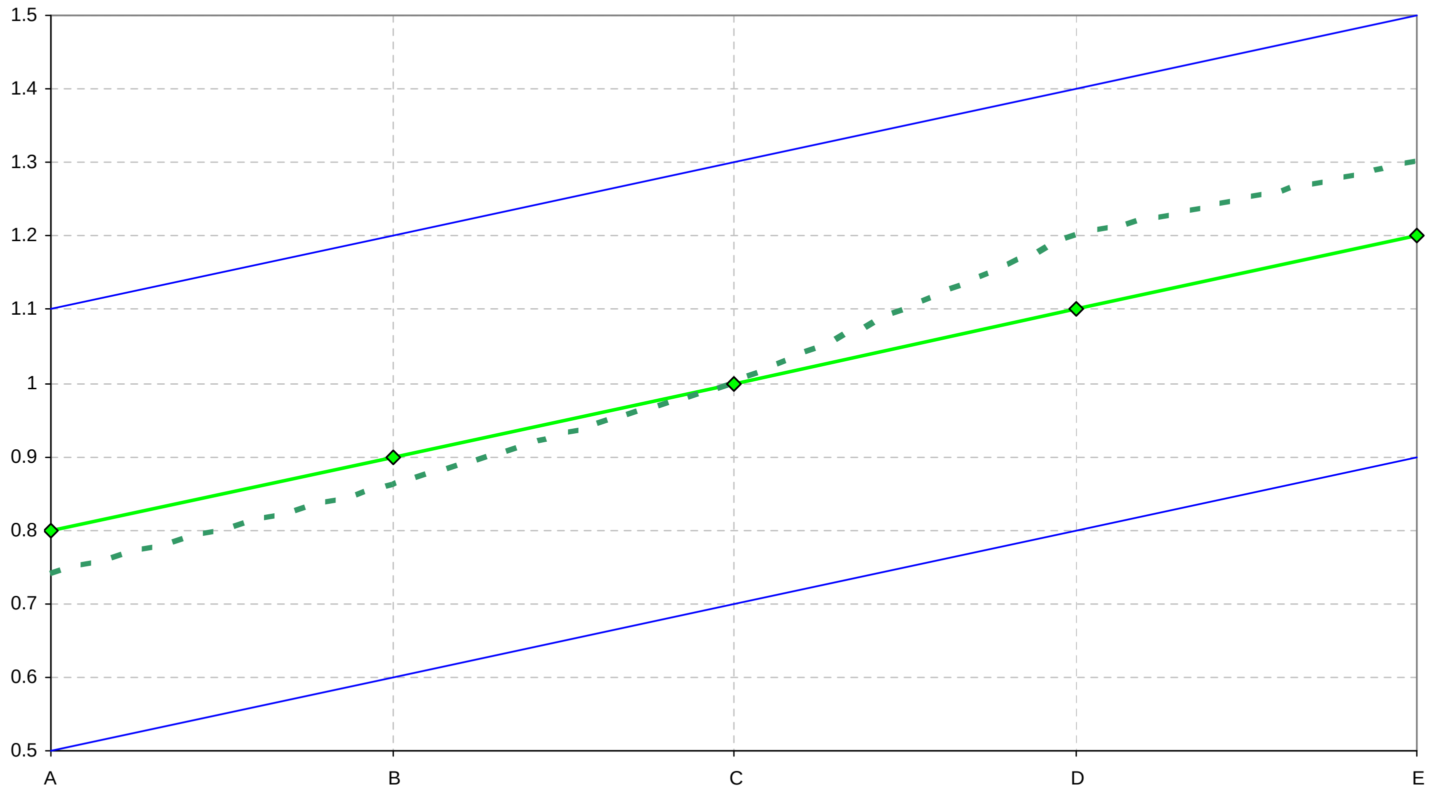
Reference models



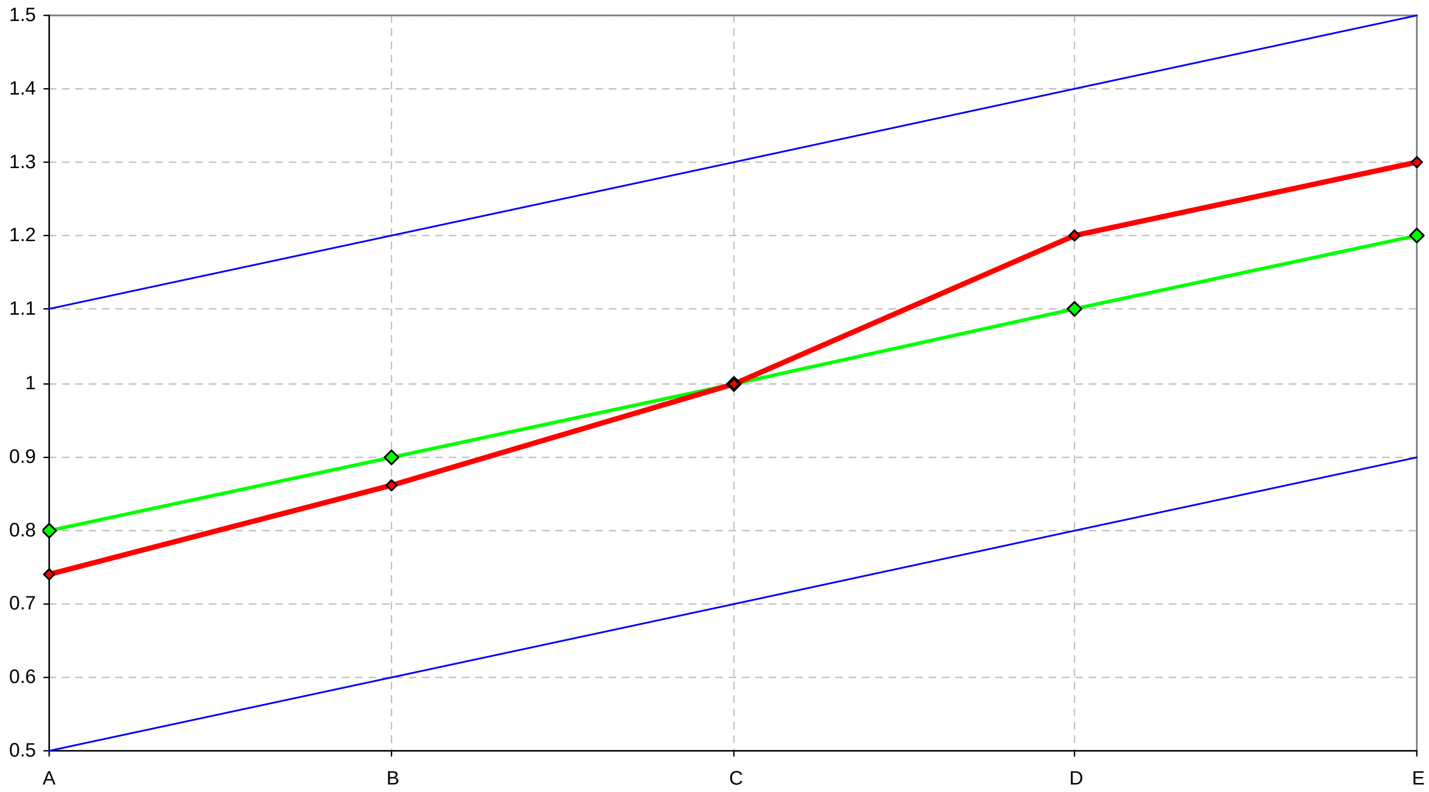
Reference models



Reference models



Reference models



towerswatson.com

Reference models

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

Offset term

- When modeling BI set PD fitted values to be offset term
- GLM will seek effects over and above assumed PD effect

Experiment

(1) GLM on BI claims on all the data - the "correct" answer

Real large company

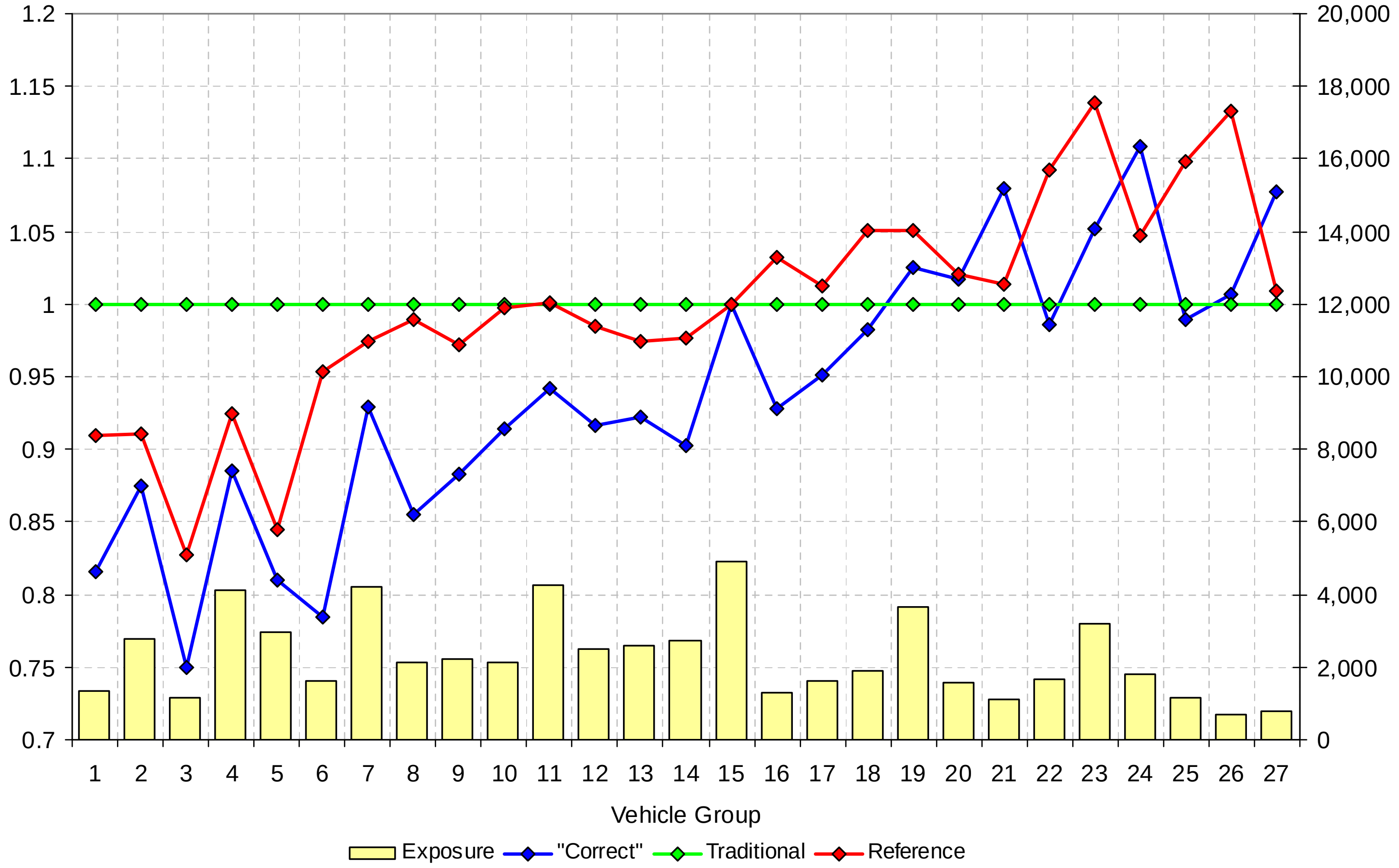
10% random sample

Pretend small company

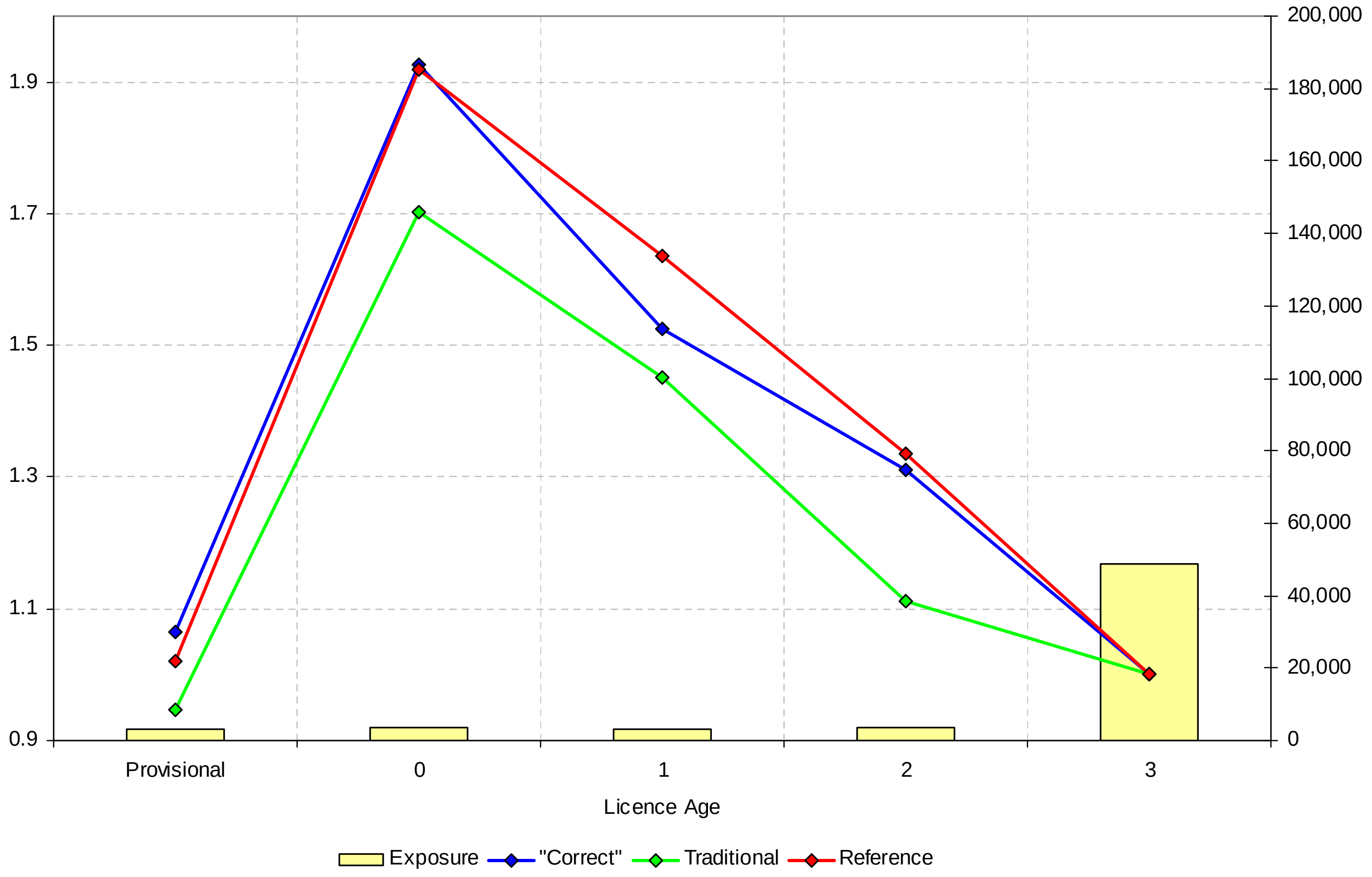
(2) Traditional GLM on BI claims on the "small company"

(3) Propensity reference model on BI claims cf PD claims

Example result



Example result



Agenda

- "Quadrant Saddles"
- The Tweedie Distribution
- "Emergent Interactions"
- Dispersion Modeling
- Modeling sparse claim types
- Driver Averaging
- Model Validation
- Man (with GLM) vs machine

Household Averaging

- Historically companies assigned operators to vehicles for the purpose of rating
- More recently driver averaging strategies are deployed to capture household
- Average may consider all drivers or a subset
  - This choice may affect other household composition factors
- Types of averages
  - Straight vs. geometric average
  - Weighted average
  - Modified
  - Average/assignment hybrid
- Modeling data needs to mimic the transaction

Model Design

- In all modeling projects, it is imperative that the data set up mimic the rating
- Consider the following example…

Vehicle	Operator	Vehicle Rate
V1	Dad	$500
V2	Mom	$450

Operator	Class Factor
Dad	0.80
Mom	0.85
Junior	2.80

  - Assume Mom had a $1000 claim in Dad's car

Assignment

- <u>Driver assignment methodology</u> each record represents a single vehicle with one assigned operator

Veh	Op	Sym	MYR	Age	Sex	Type	Yths	Drvrs	Vehs	Exp	Clm	Losses	Prem
V1	Junior	17	2006	16	M	OO	1	3	3	1	1	1,000	1,400.00
V2	Mom	17	2005	43	F	PO	1	3	3	1	0	0	382.50

  - Operator characteristics based on <u>assigned operator</u>
  - Vehicle characteristics based on vehicle
  - Policy characteristics “catch” other drivers
  - Losses assigned to vehicle

Straight Average

- Straight average methodology:

$$VehicleFactor \times \frac{(Op1Factor + Op2Factor + Op3Factor)}{3}$$

- Which can be deconstructed:

$$VehicleFactor \times \frac{(Op1Factor)}{3}$$
$$VehicleFactor \times \frac{(Op2Factor)}{3}$$
$$VehicleFactor \times \frac{(Op3Factor)}{3}$$

Straight Average

- Straight average methodology each record represents a single vehicle and operator combination

Veh	Op	Sym	MYR	Age	Sex	Yths	Drvrs	Vehs	Exp	Clm	Losses	Prem
V1	Dad	17	2006	45	M	1	3	3	1/3	0	0	133.33
V1	Mom	17	2006	43	F	1	3	3	1/3	1	1,000	141.67
V1	Junior	17	2006	16	M	1	3	3	1/3	0	0	466.67
V2	Dad	17	2005	45	M	1	3	3	1/3	0	0	120.00
V2	Mom	17	2005	43	F	1	3	3	1/3	0	0	127.50
V2	Junior	17	2005	16	M	1	3	3	1/3	0	0	420.00

- Policy characteristics are same, but less predictive
- Exposure split amongst the vehicle
- Losses assigned to vehicle/operator combination
- iid is a major concern
- What about Comprehensive?

Geometric Average

- Geometric average methodology:

$$VehicleFactor \times (Op1Factor + Op2Factor + Op3Factor)^{1/3}$$

  - No direct decomposition

Geometric Average

- <u>Geometric methodology</u> each record represents a single vehicle

Veh	Sym	MYR	# of Dads	# of Moms	# of Juniors	Exp	Clm	Losses	Prem
V1	17	2006	1/3	1/3	1/3	1	1	1,000	619.72
V2	17	2005	1/3	1/3	1/3	1	0	0	557.74

  - Policy characteristics are same, but less predictive
  - Predictors are translated to counts
  - Losses assigned to vehicle
  - More challenging to add operator interactions or variates

Weighted Average

- <u>Weighted average methodology</u> for a straight average approach

Veh	Op	Sym	MYR	Age	Sex	Type	Yths	Drvrs	Vehs	Exp	Clm	Losses	Prem
V1	Dad	17	2006	45	M	PO	1	3	3	1/3	0	0	133.33
V1	Mom	17	2006	43	F	OC	1	3	3	1/3	1	1,000	141.67
V1	Junior	17	2006	16	M	OC	1	3	3	1/3	0	0	466.67
V2	Dad	17	2005	45	M	OC	1	3	3	1/3	0	0	120.00
V2	Mom	17	2005	43	F	PO	1	3	3	1/3	0	0	127.50
V2	Junior	17	2005	16	M	OC	1	3	3	1/3	0	0	420.00

  - Creates a relationship between the vehicle and the operator
  - Uses the model to determine the weights
  - More *accurate* as it requires more information

$$VehicleFactor1 \times \frac{(Op1Factor * PO + Op2Factor * OC + Op3Factor * OC)}{3}$$

Agenda

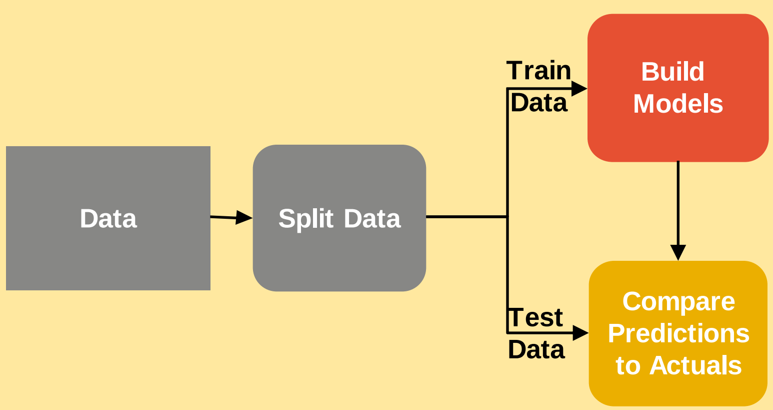
- "Quadrant Saddles"
- The Tweedie Distribution
- "Emergent Interactions"
- Dispersion Modeling
- Modeling sparse claim types
- Driver Averaging
- Model Validation
- Man (with GLM) vs machine

Validate Models

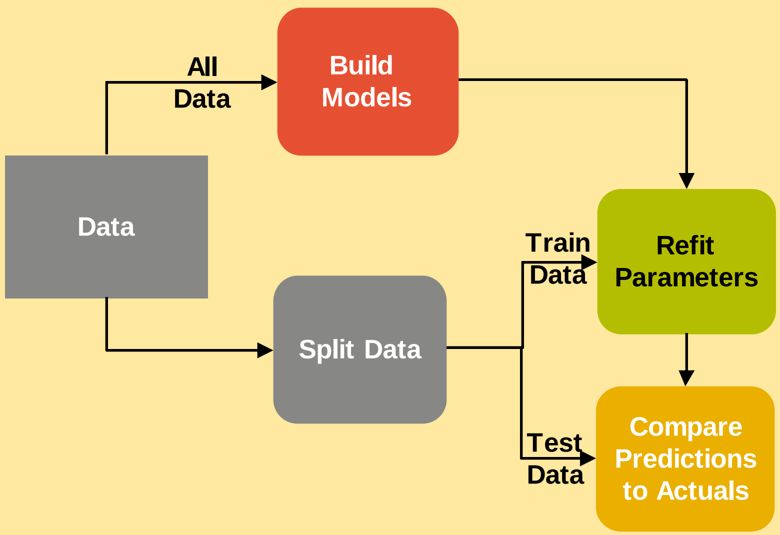
Holdout samples

- Holdout samples are effective at validating model
  - Determine estimates based on part of data set
  - Uses estimates to predict other part of data set

Full Test/Training for Large Data Sets



Partial Test/Training for Smaller Data Sets

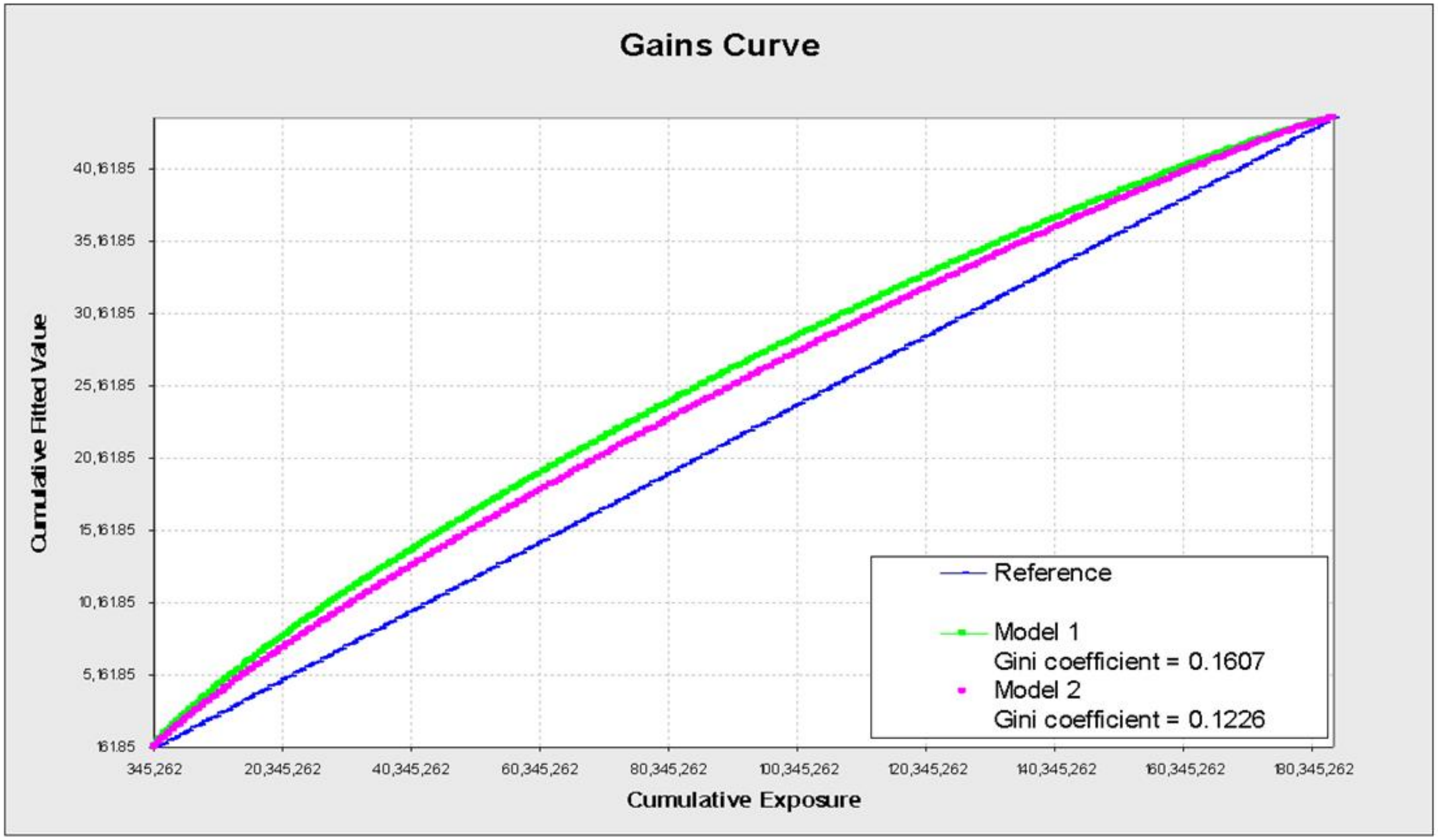


Predictions should be close to actuals for heavily populated cells

Validate Models

Gains curves

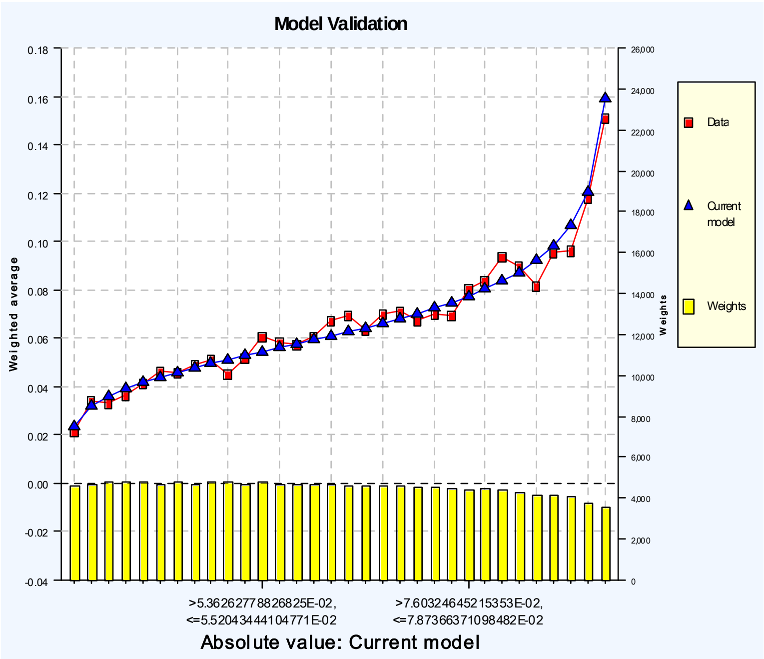
- Compare predictiveness of models

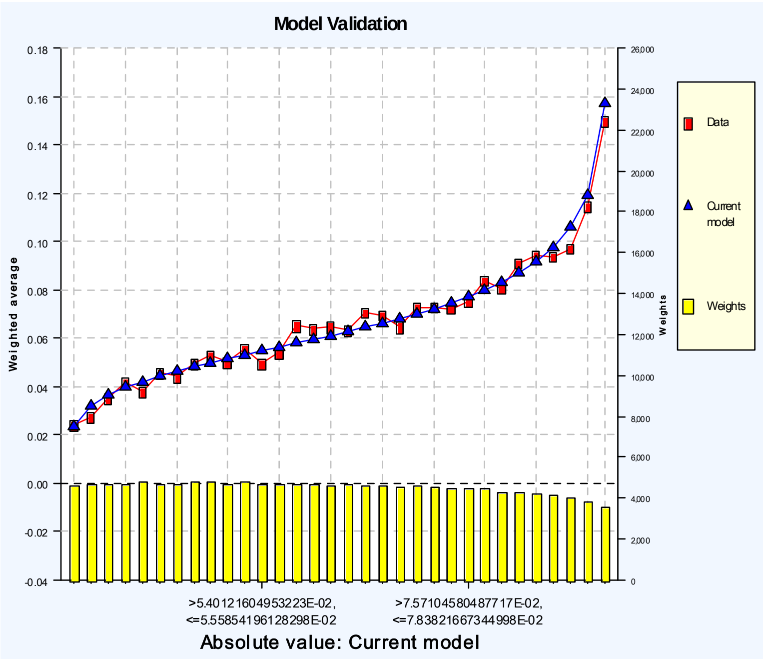


Validate Models

Lift curves

- Compare predictiveness of models

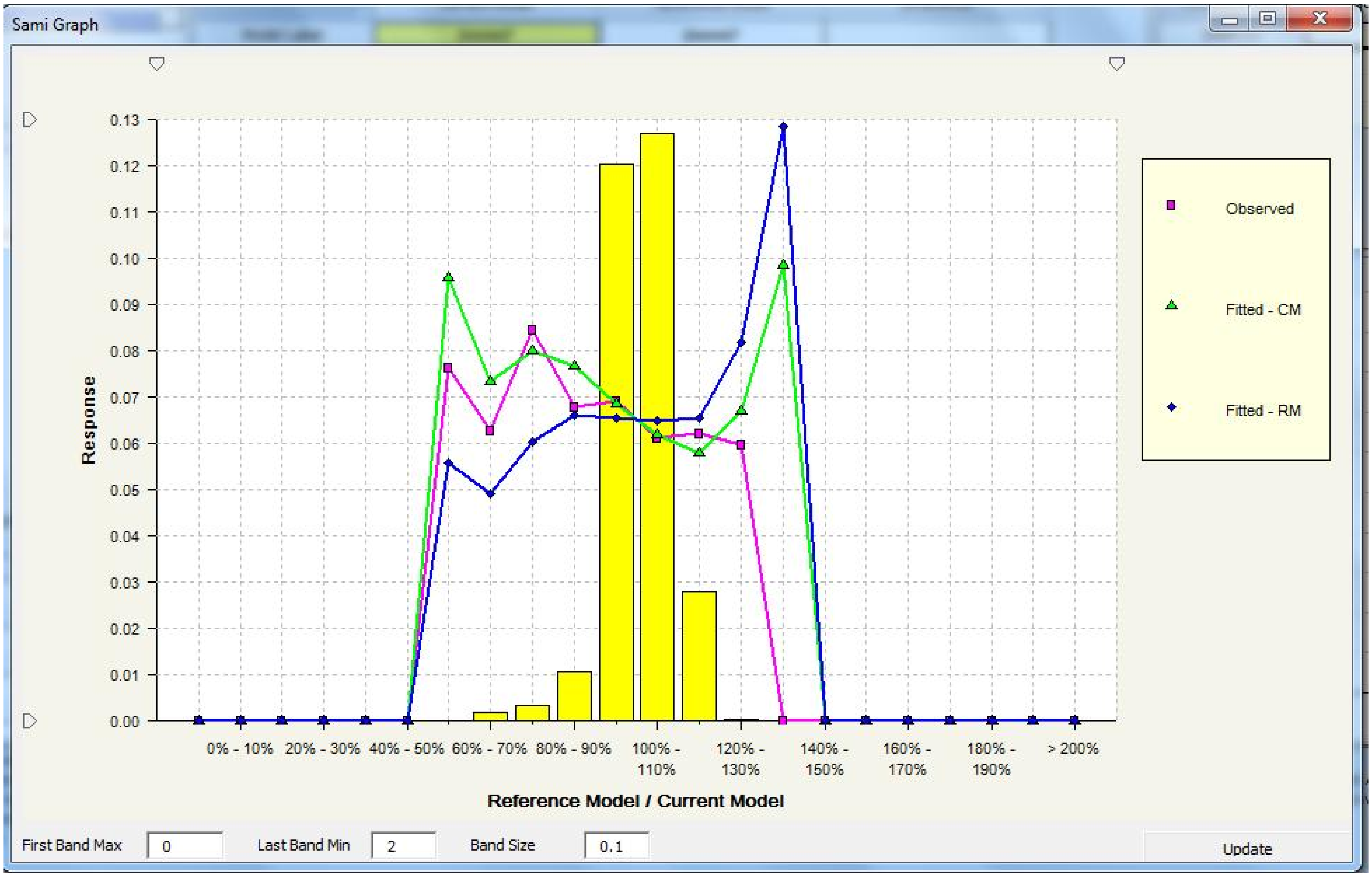




- More intuitive but difficult to assess performance

Validate Models

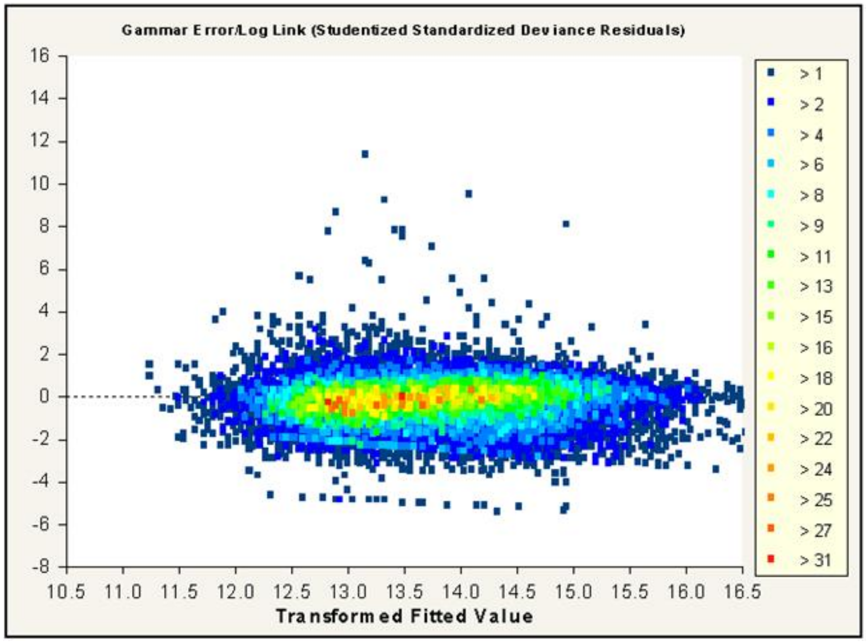
X-Graphs

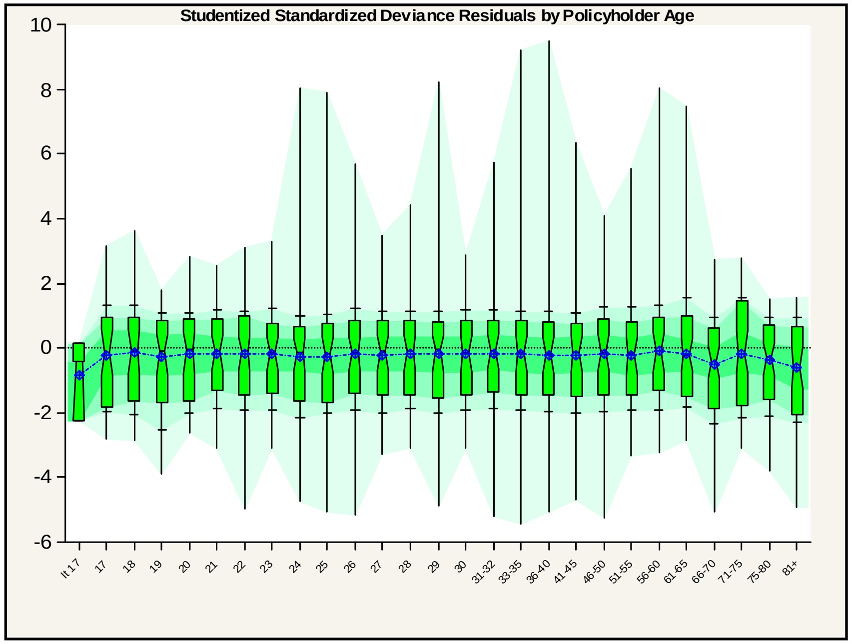


Validate Models

Residual analysis

- Recheck residuals to ensure appropriate shape





Is the contour plot symmetric?

Does the Box-Whisker show symmetry across levels?

Agenda

- "Quadrant Saddles"
- The Tweedie Distribution
- "Emergent Interactions"
- Dispersion Modeling
- Modeling sparse claim types
- Driver Averaging
- Model Validation
- Man (with GLM) vs machine

Machine vs man

VS

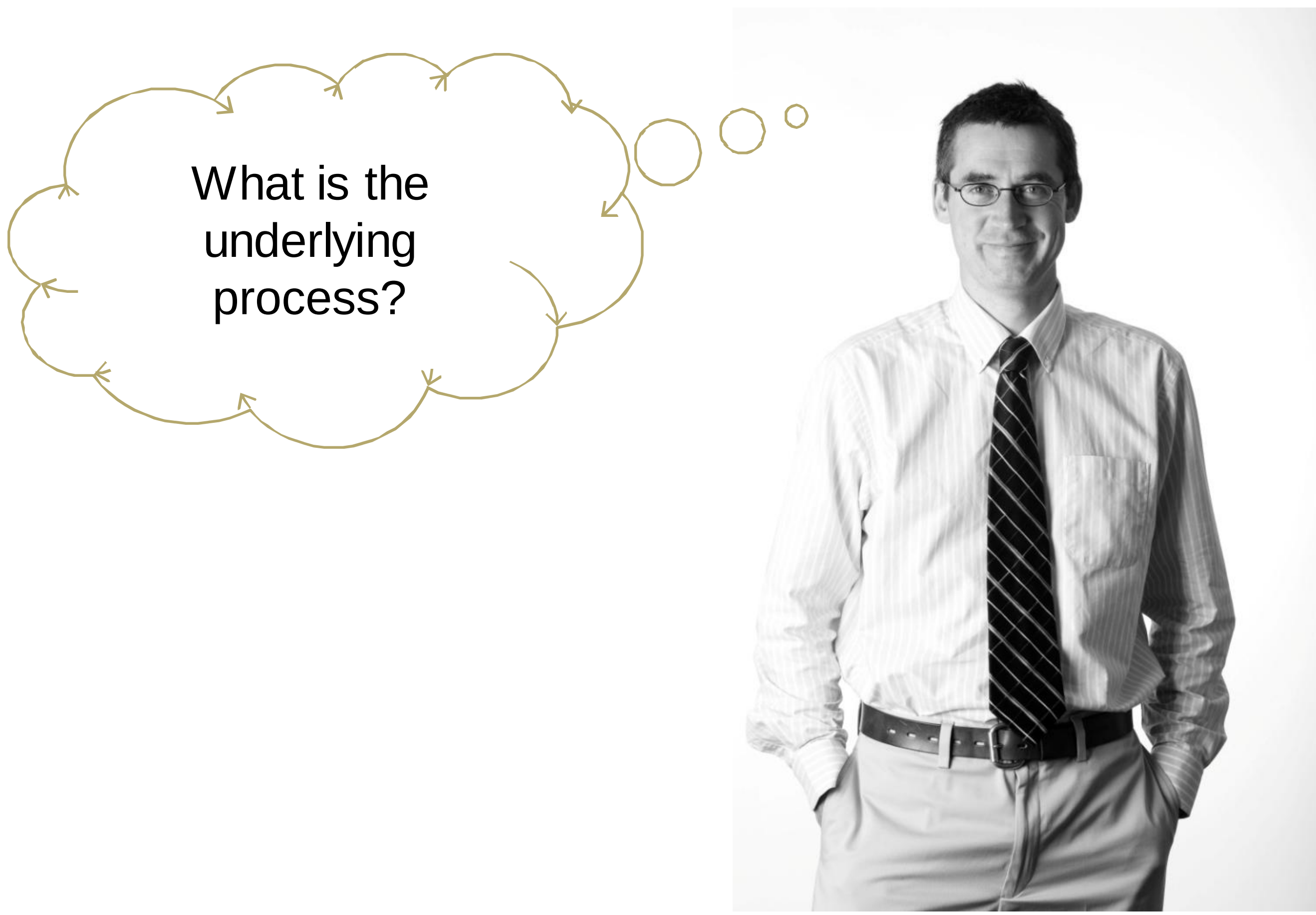
Machine vs man

VS

Machine vs man

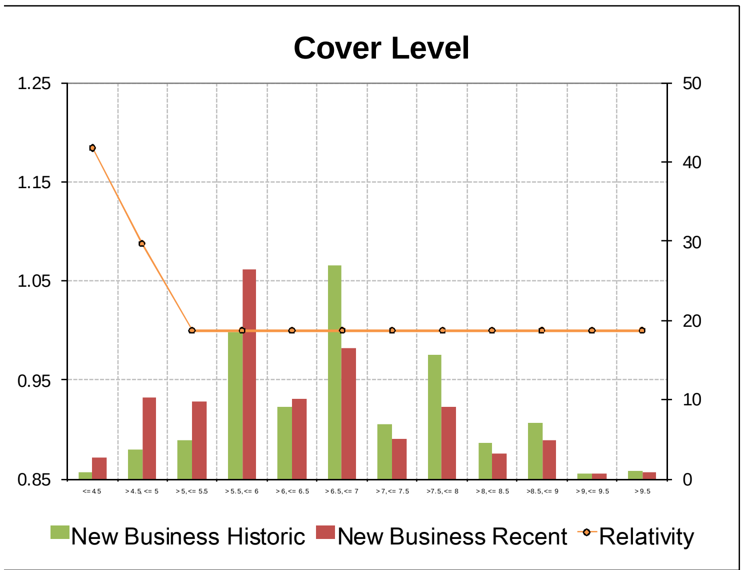
VS

Machine vs man

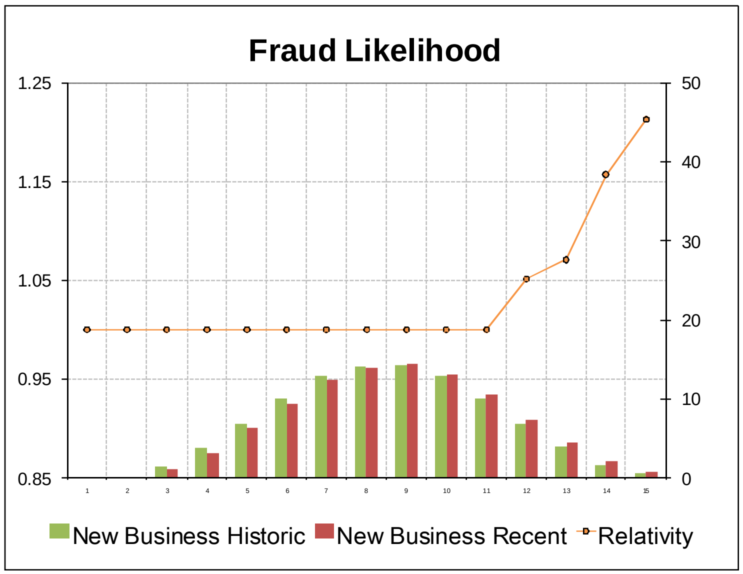


Machine vs man

Underwriting

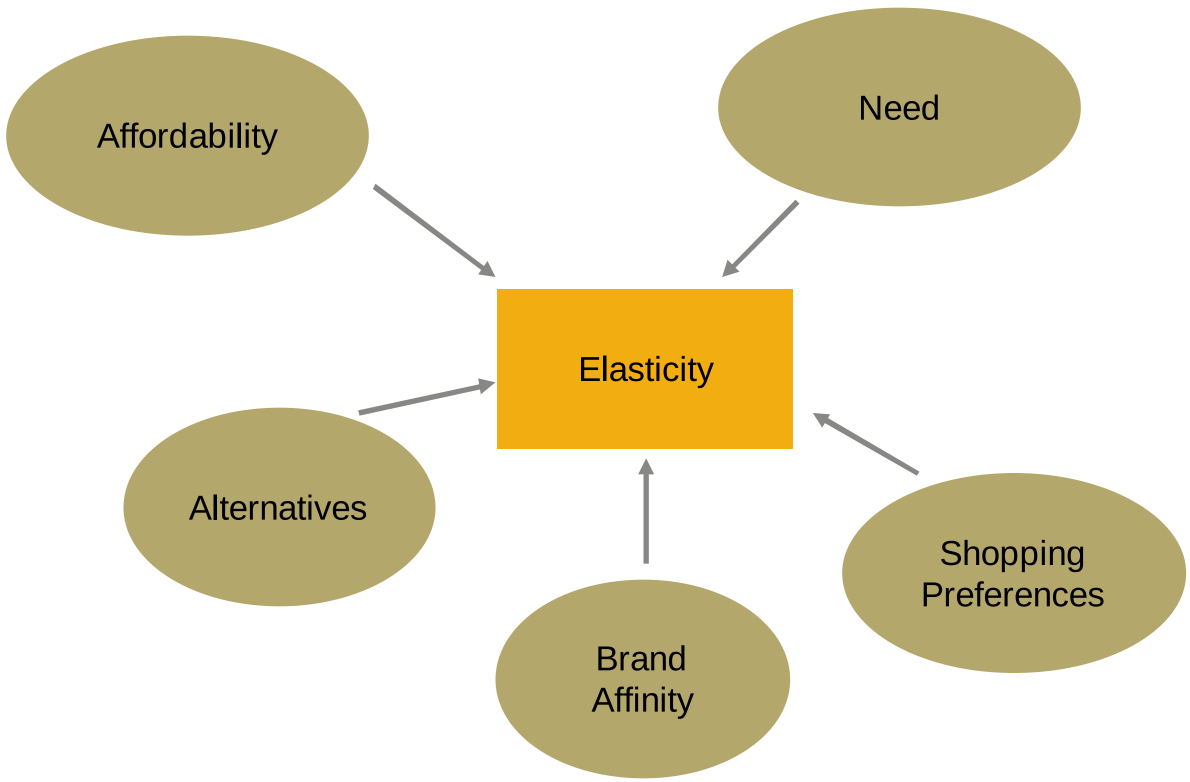


Claims



Machine vs man

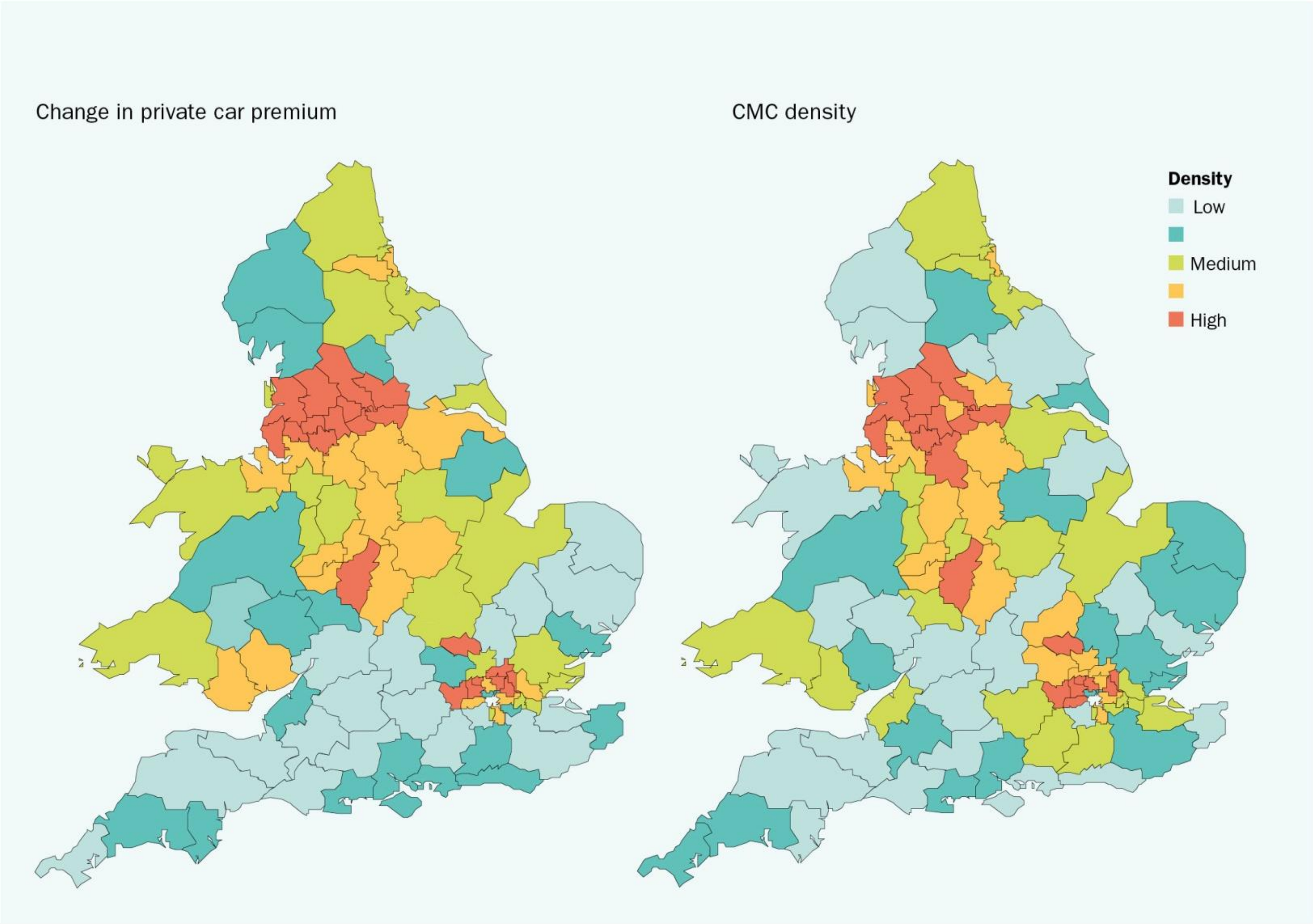
Drivers of elasticity



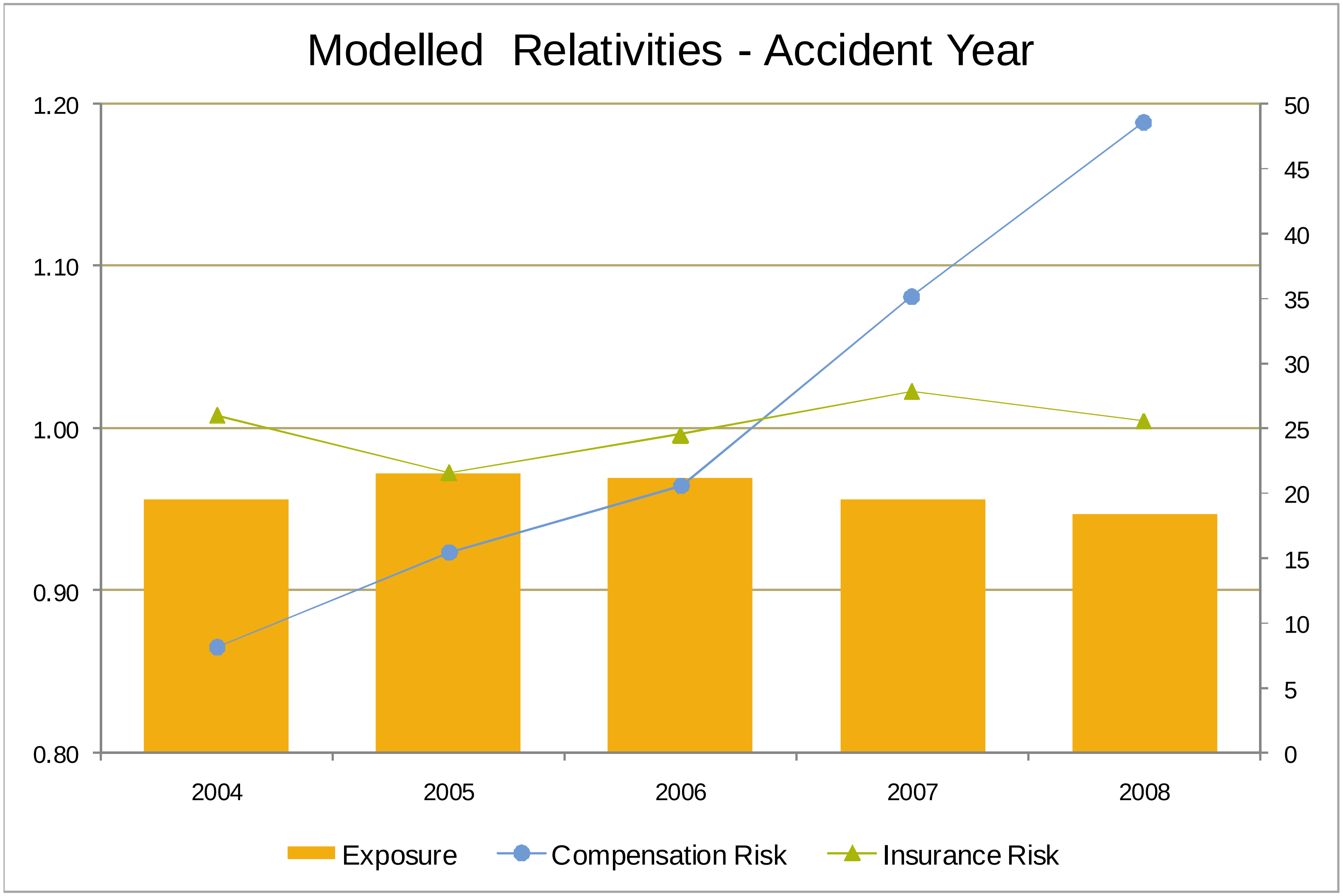
Machine vs man

I reckon that lots of recent bodily injury changes are down to new types of claims

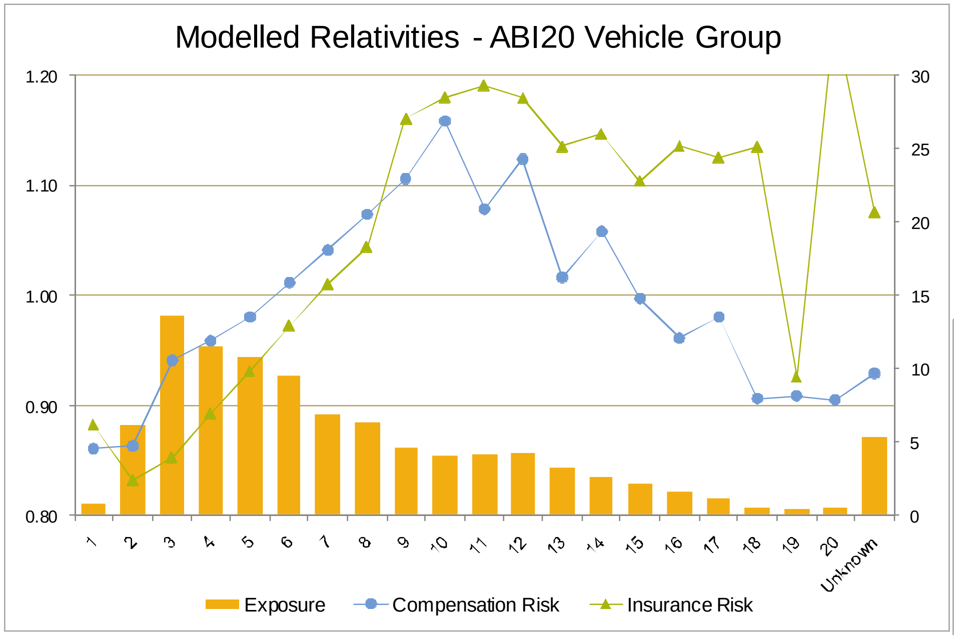
Claims management companies

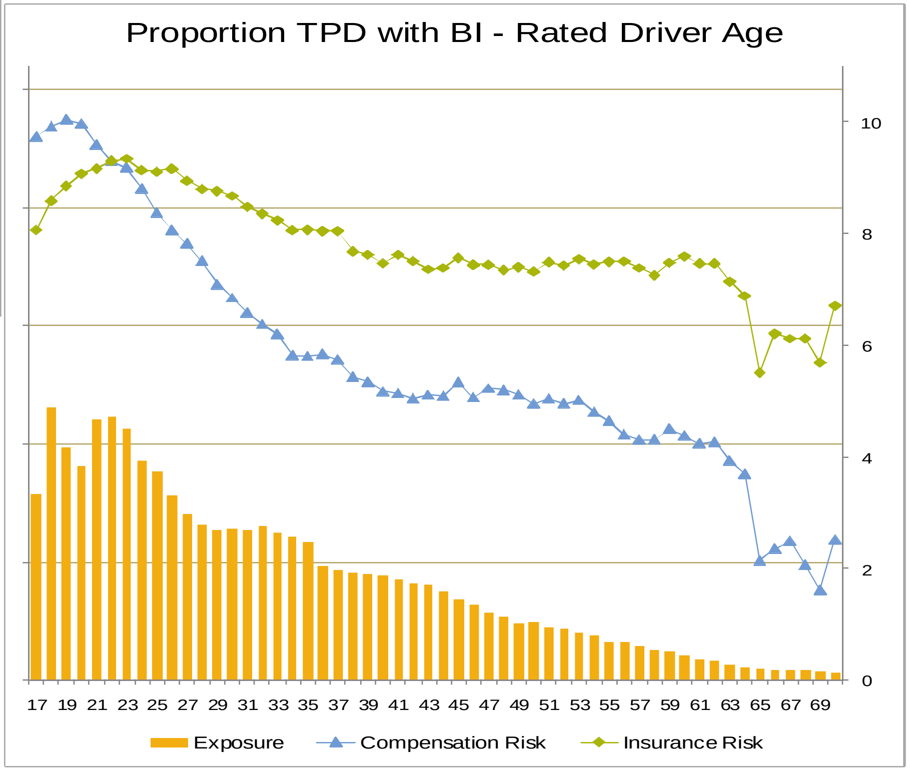


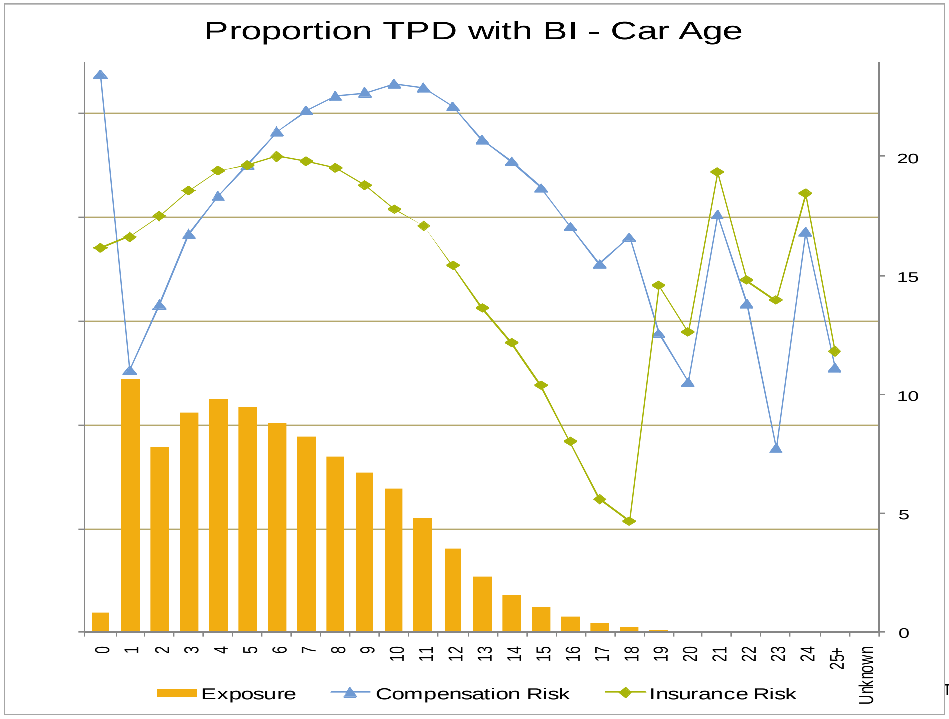
BI models - "insurance" and "compensation" risk



towerswatson.com


Modelled Relativities - ABI20 Vehicle Group
1.20
1.10
1.00
0.90
0.80
30
25
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 Unknown
Exposure
Compensation Risk
Insurance Risk


Proportion TPD with BI - Rated Driver Age
10
8
6
4
2
0
17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69
Exposure
Compensation Risk
Insurance Risk


Proportion TPD with BI - Car Age
20
15
10
5
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25+ Unknown
Exposure
Compensation Risk
Insurance Risk

GLM III - The Matrix Reloaded

Brent Petzoldt

TOWERS WATSON