EMB

GLM III

Duncan Anderson MA FIA
Partner, EMB Consultancy LLP

Agenda

- Testing the link function
- The Tweedie distribution
- Regression splines
- Reference models
- Aliasing/near-aliasing
- Combining models across claim types
- Restricted models
- Model validation
- Modeling elasticity / GNMs

Agenda

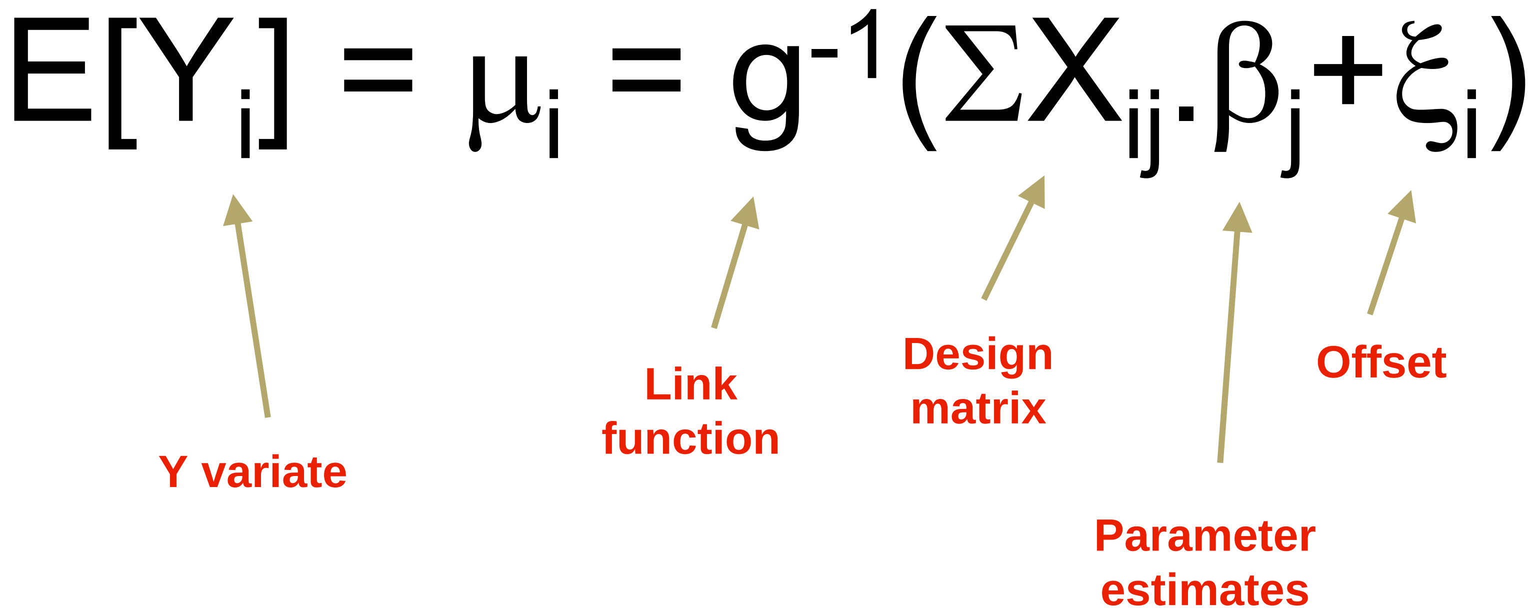
- Testing the link function
- The Tweedie distribution
- Regression splines
- Reference models
- Aliasing/near-aliasing
- Combining models across claim types
- Restricted models
- Model validation
- Modeling elasticity / GNMs

Formularization of GLMs

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

$$Var[Y_i] = \phi.V(\mu_i)/\omega_i$$

Formularization of GLMs



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

Y variate

Link function

Design matrix

Parameter estimates

Offset

Formularization of GLMs

Scale parameter

Variance function

Prior weights

$$\mathrm{Var}[Y_i] = \overset{\substack{\text{Scale parameter}\\ \searrow}}{\phi}.\overset{\substack{\text{Variance function}\\ \downarrow}}{V}(\mu_i)/\overset{\substack{\text{Prior weights}\\ \swarrow}}{\omega_i}$$

Formularization of GLMs

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

$$Var[Y_i] = \phi.V(\mu_i)/\omega_i$$

Link function

Eg if $\Sigma X_{ij}.\beta_j =$

$\alpha + \beta$ if male $+ \gamma$ if small car $+ \delta$ if big car

$g(x) = x \Rightarrow E[Y_i] = \alpha + \beta + \gamma + \delta$

$$g(x) = \ln(x) \Rightarrow E[Y_i] = e^{\alpha + \beta + \gamma + \delta}$$

$$= e^{\alpha}.e^{\beta}.e^{\gamma}.e^{\delta}$$

$$= A . B . C . D$$

Box-Cox link function test

$E[Y_i] = \mu_i = g^{-1}(\Sigma X_{ij}.\beta_j+\xi_i)$ $\quad Var[Y_i] = \phi.V(\mu_i)/\omega_i$

Box-Cox link function defined as:

$g(x) = (x^\lambda - 1) / \lambda$ for $\lambda \neq 0$; $\ln(x)$ for $\lambda = 0$

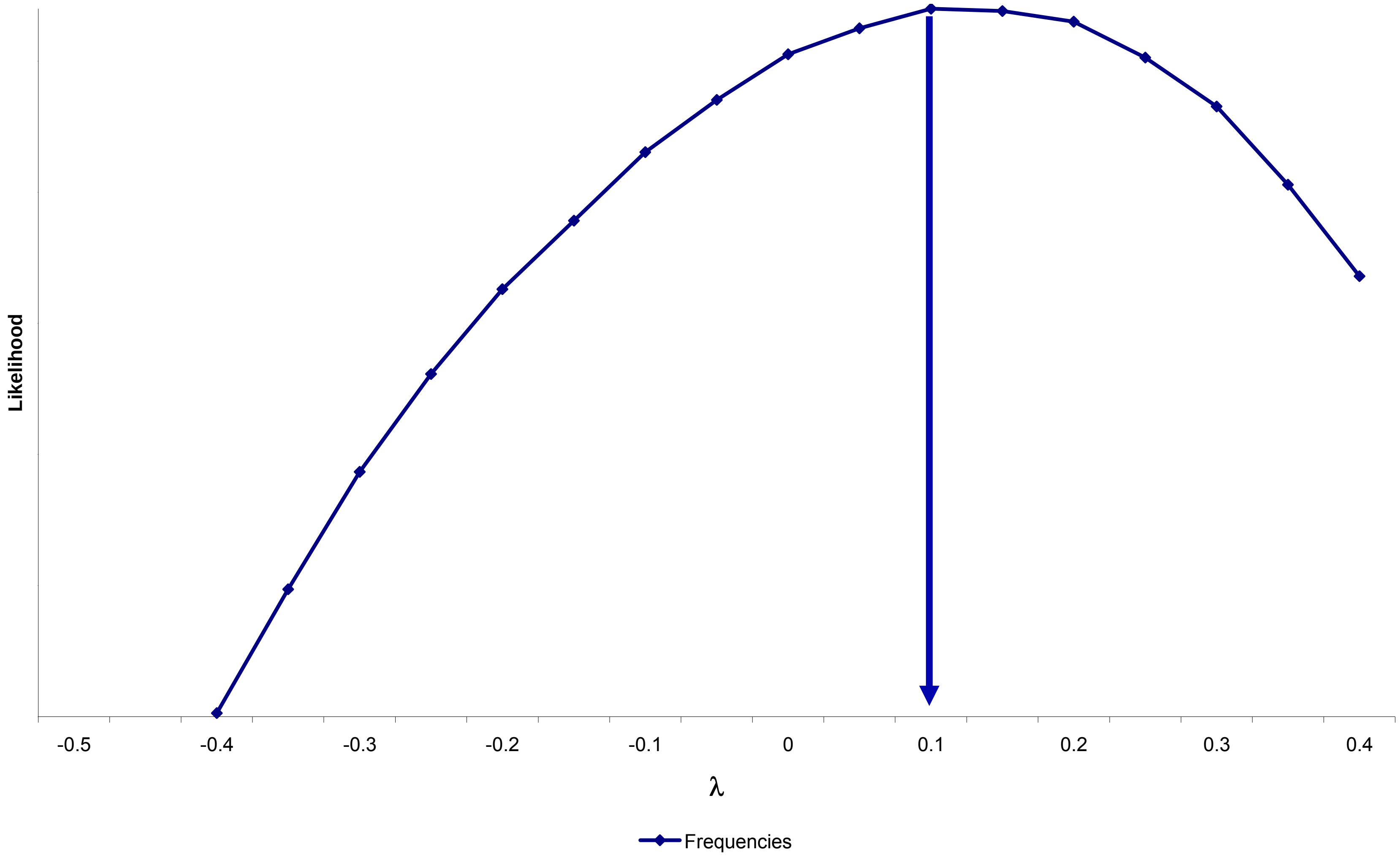
$\lambda = 1 \quad \Rightarrow g(x) = (x - 1) \quad \Rightarrow$ additive (with a base level shift)

$\lambda \to 0 \quad \Rightarrow g(x) \to \ln(x) \quad \Rightarrow$ multiplicative (via l'Hôpital)

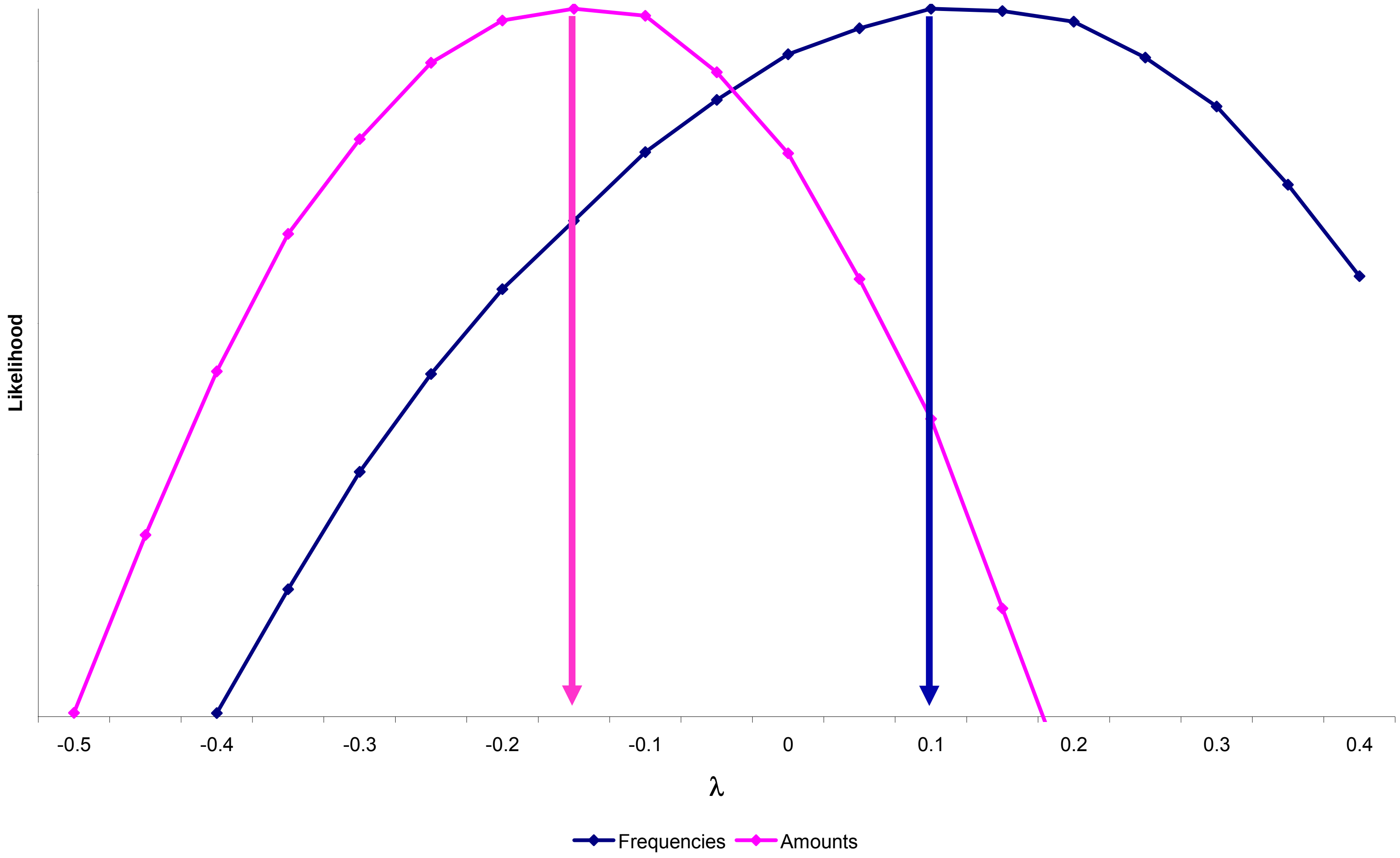
$\lambda = -1 \quad \Rightarrow g(x) = 1-1/x \quad \Rightarrow$ inverse (with a base level shift)

Test a range of values of λ and see which maximizes likelihood

Box-Cox link function test



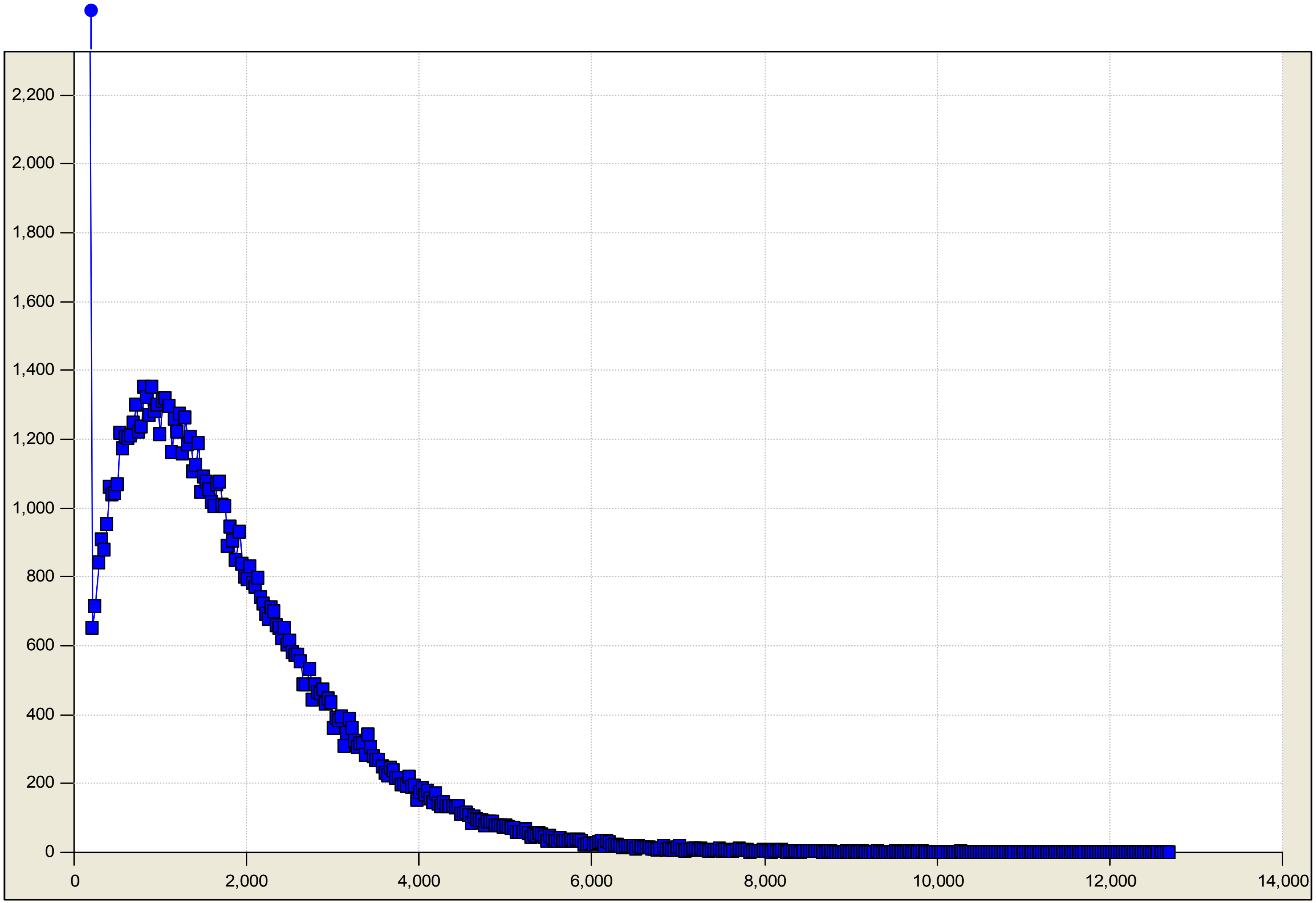
Box-Cox link function test



Agenda

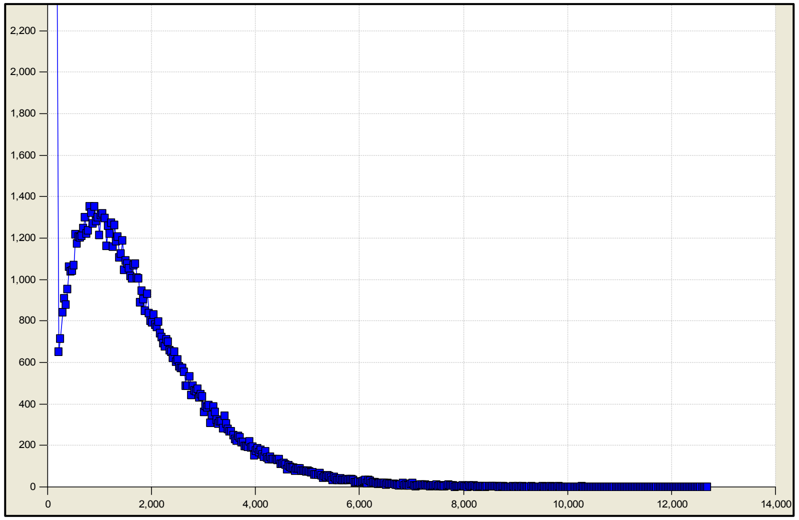
- Testing the link function
- The Tweedie distribution
- Regression splines
- Reference models
- Aliasing/near-aliasing
- Combining models across claim types
- Restricted models
- Model validation
- Modeling elasticity / GNMs

Tweedie GLMs



Tweedie GLMs

- 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



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

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

Formularization of GLMs

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

$$Var[Y_i] = \phi.V(\mu_i)/\omega_i$$

$$\begin{array}{llll}
\text{Normal:} & \phi=\sigma^2, & V[x]=1 & \Rightarrow Var[Y_i] = \sigma^2 \\
\text{Poisson:} & \phi=1, & V[x]=x & \Rightarrow Var[Y_i] = \mu_i \\
\text{Gamma:} & \phi=k, & V[x]=x^2 & \Rightarrow Var[Y_i] = k\mu_i^2 \\
 & & & \\
\text{Tweedie:} & \phi=k, & V[x]=x^p & \Rightarrow Var[Y_i] = k\mu_i^p
\end{array}$$

Tweedie GLMs

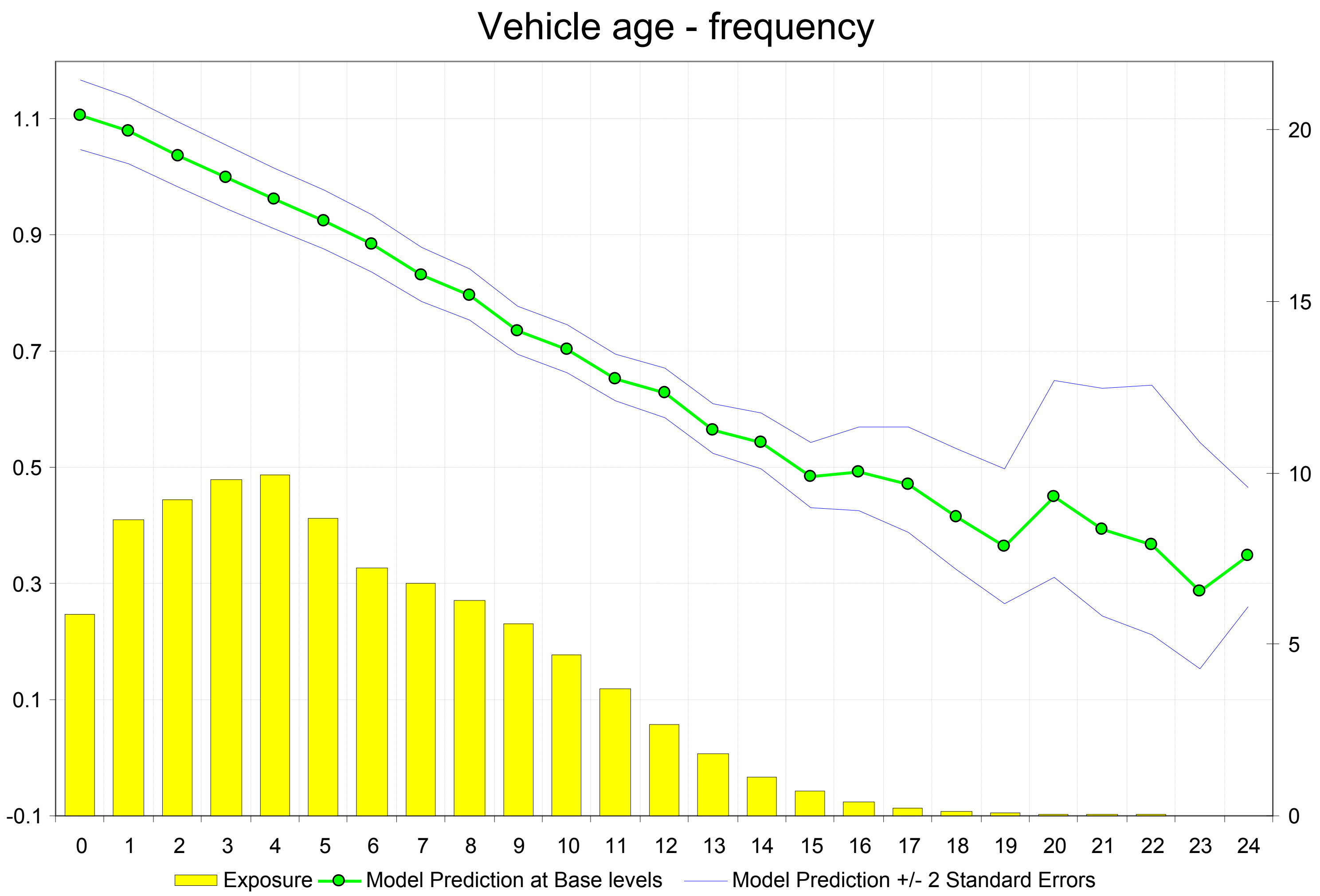
Tweedie: $\phi=k$, $V[x]=x^p \Rightarrow \text{Var}[Y_i] = k\mu_i^p$

p=1	Poisson
p=2	gamma
1<p<2	Poisson/gamma process

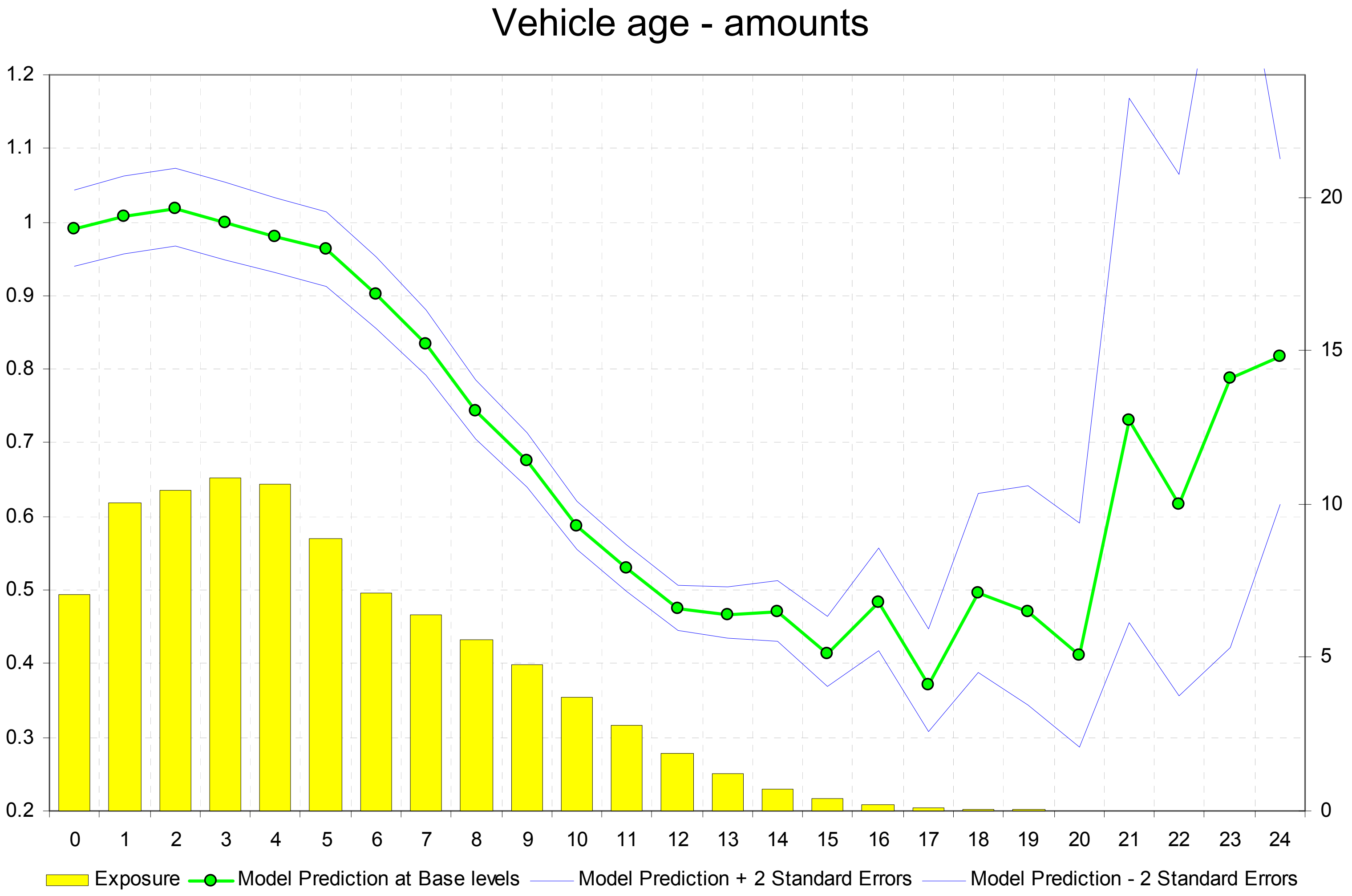
(can also be <0 or >2)

- Need to estimate both k and p when fitting models
- Typically p≈1.5 for incurred claims

Example 1

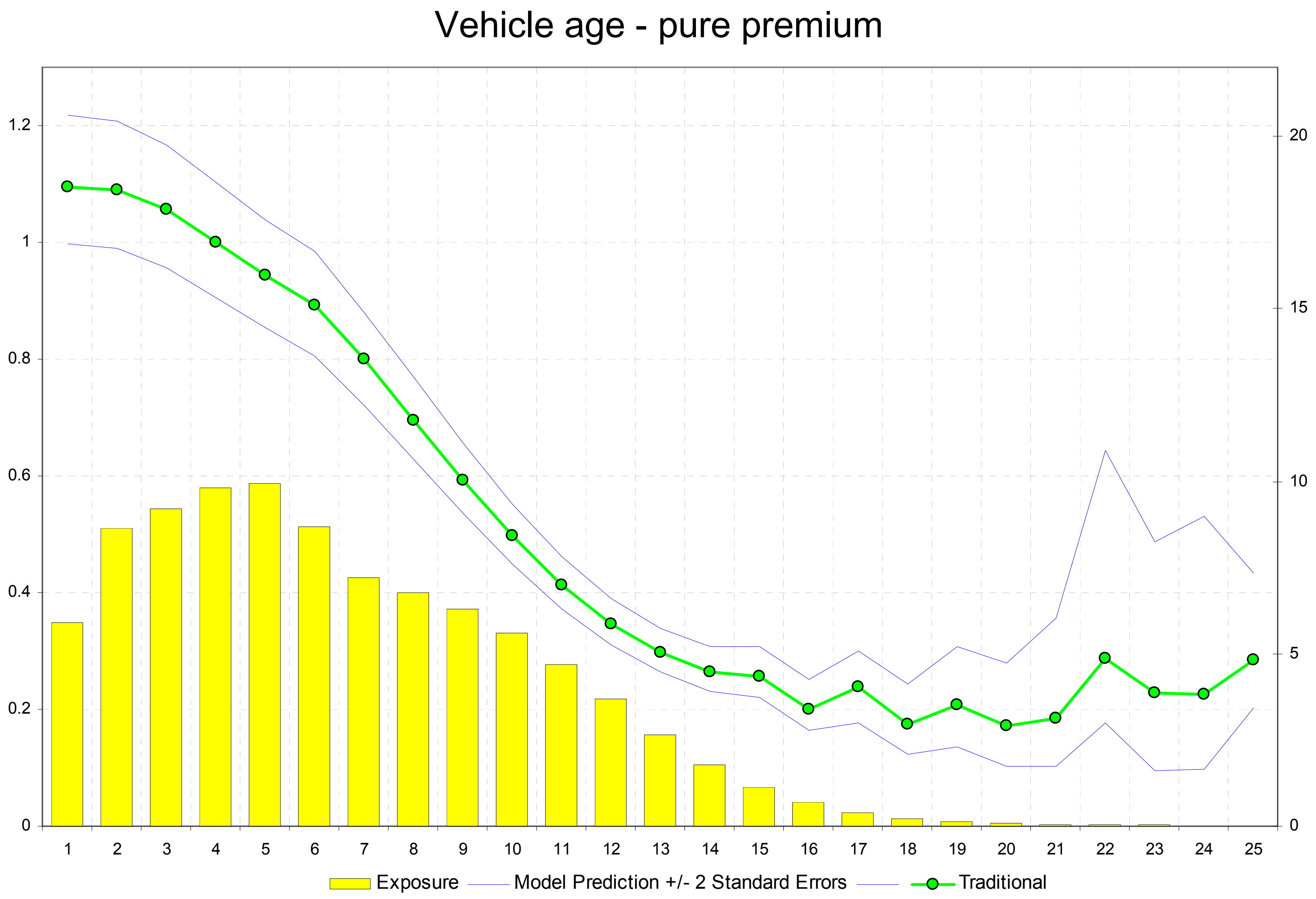

Vehicle age - frequency
1.1
0.9
0.7
0.5
0.3
0.1
-0.1
20
15
10
5
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
Exposure
Model Prediction at Base levels
Model Prediction +/- 2 Standard Errors

Example 1

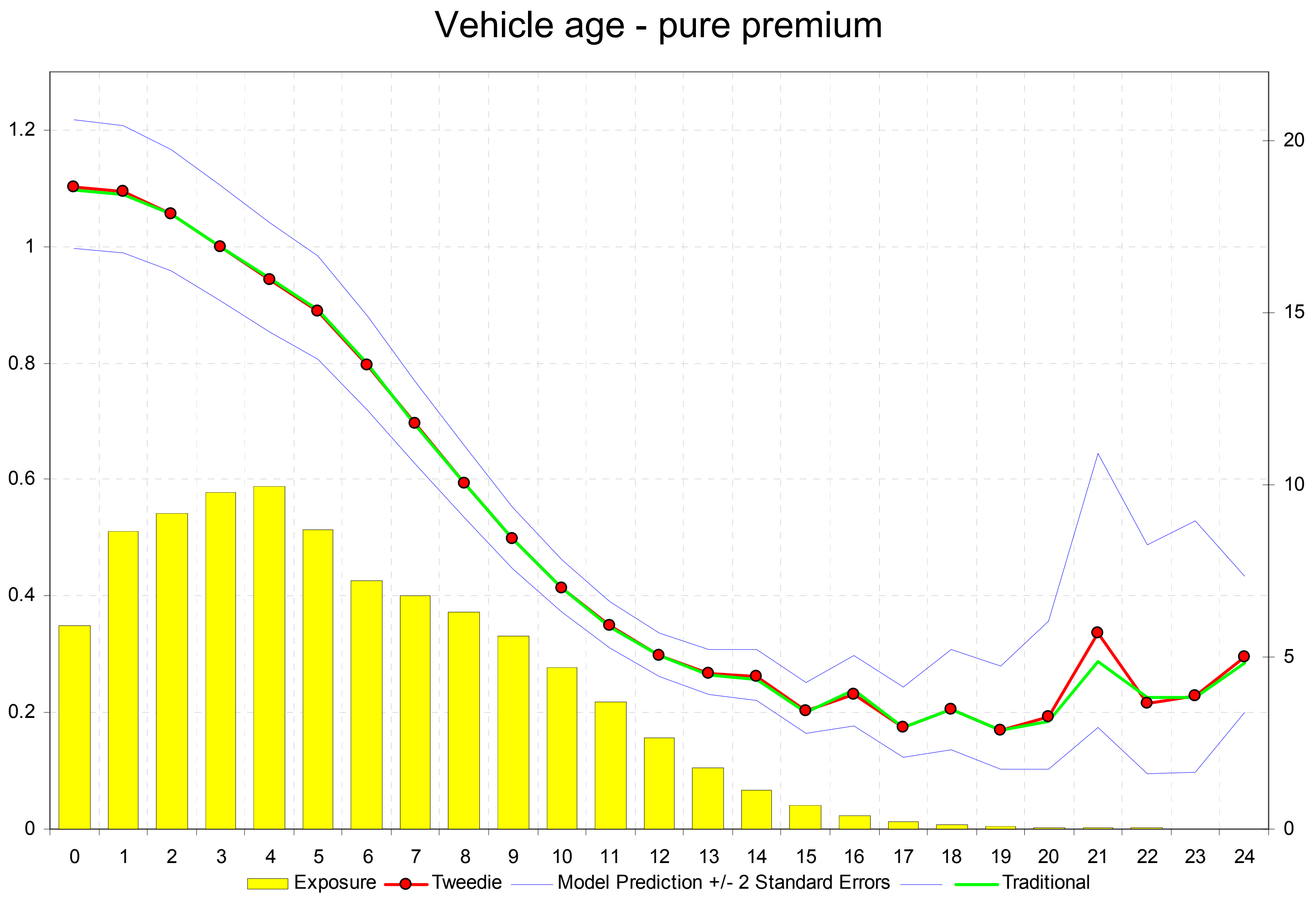

Vehicle age - amounts
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
20
15
10
5
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
Exposure
Model Prediction at Base levels
Model Prediction + 2 Standard Errors
Model Prediction - 2 Standard Errors

Example 1

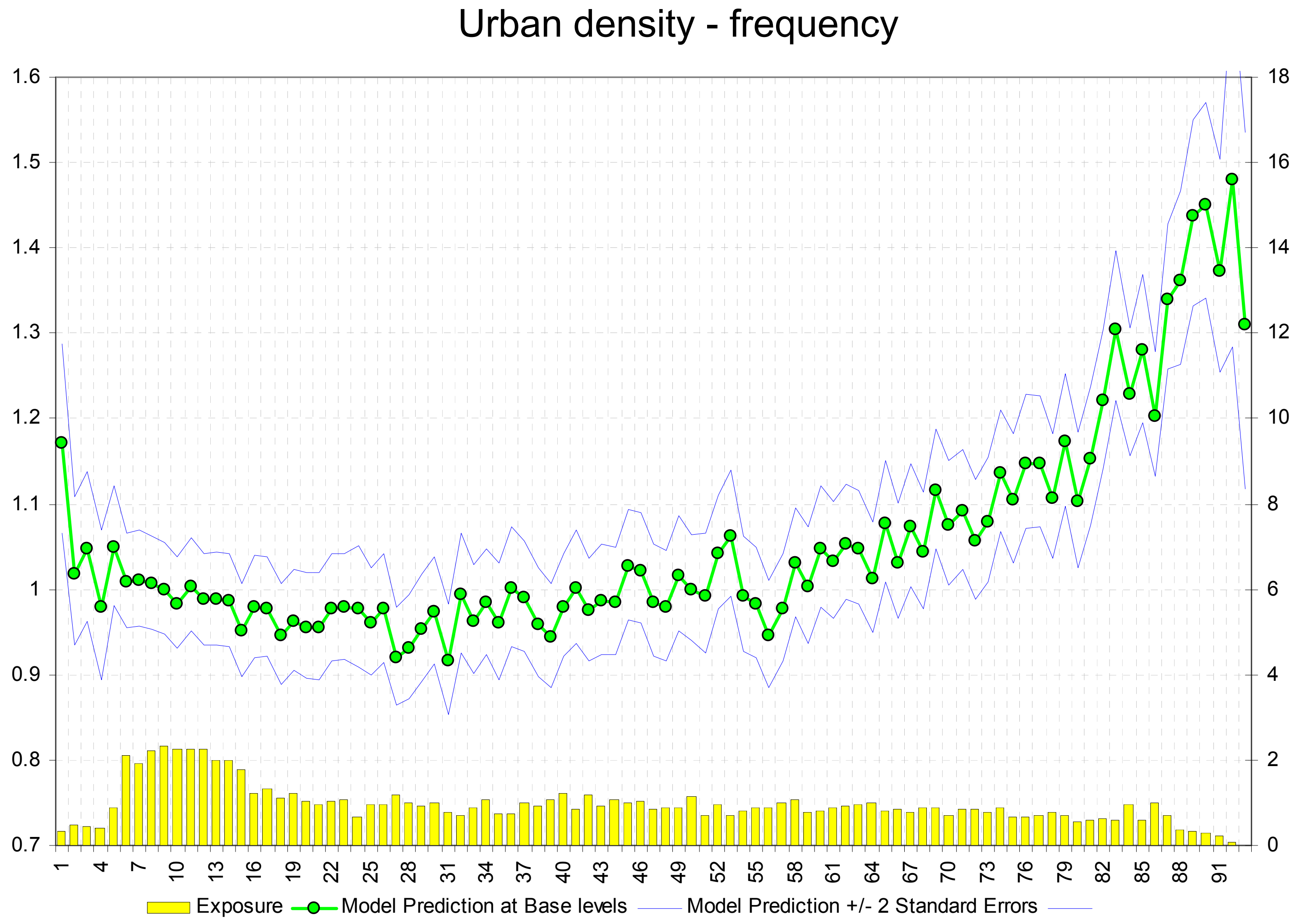
EMB



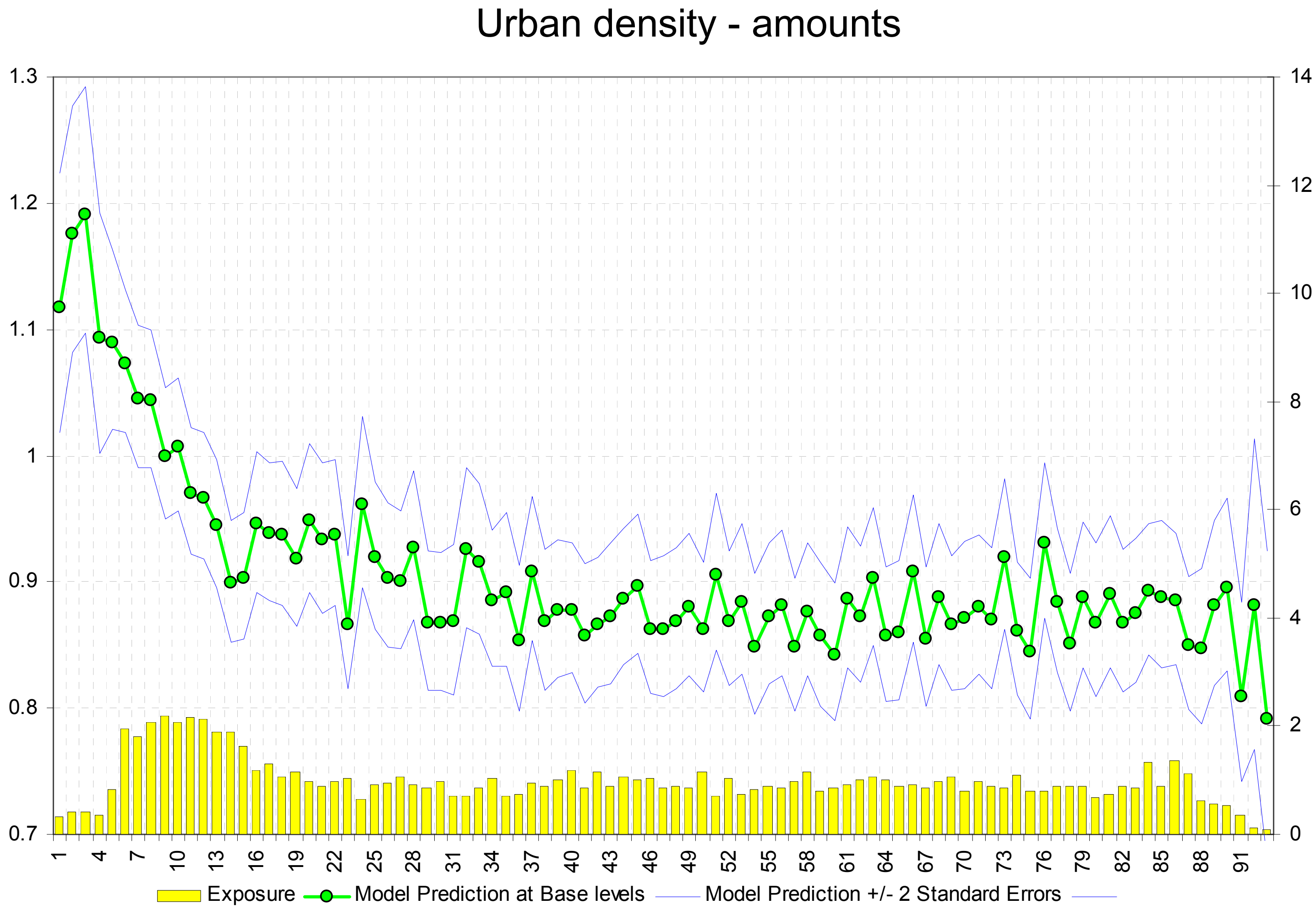
Example 1


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

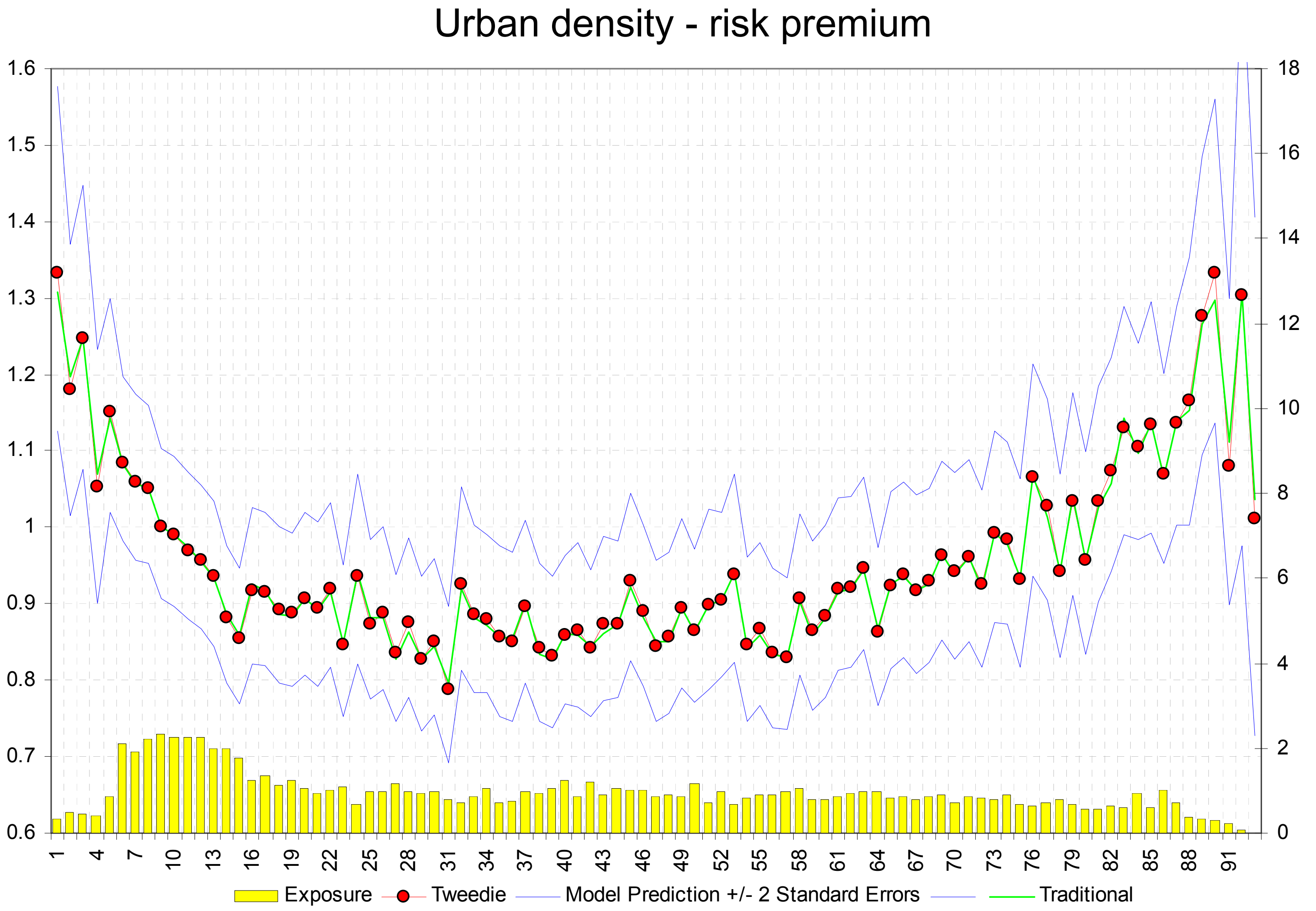
Example 2


Urban density - frequency
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
18
16
14
12
10
8
6
4
2
0
1
4
7
10
13
16
19
22
25
28
31
34
37
40
43
46
49
52
55
58
61
64
67
70
73
76
79
82
85
88
91
Exposure
Model Prediction at Base levels
Model Prediction +/- 2 Standard Errors

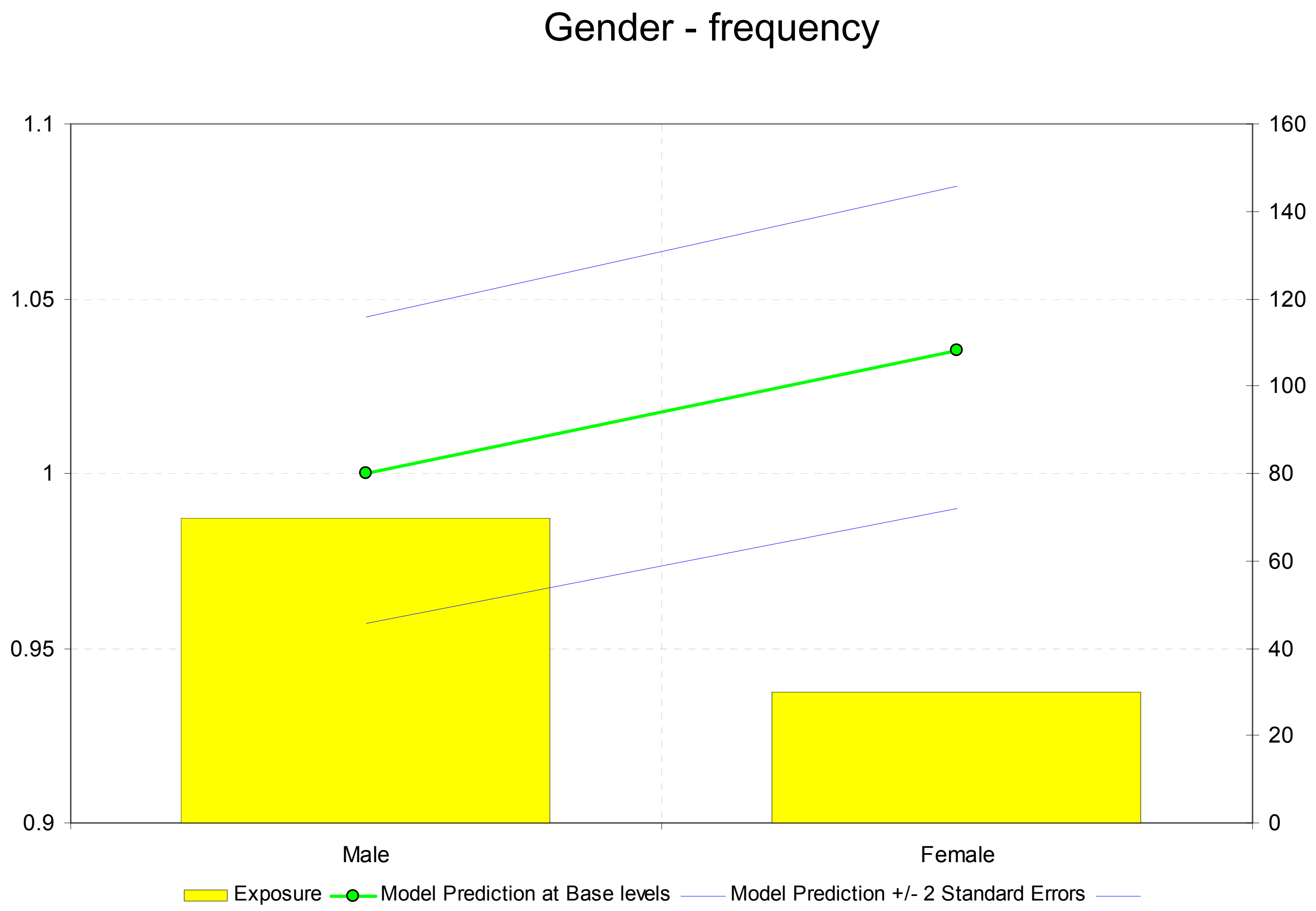
Example 2


Urban density - amounts
1.3
1.2
1.1
1
0.9
0.8
0.7
14
12
10
8
6
4
2
0
1
4
7
10
13
16
19
22
25
28
31
34
37
40
43
46
49
52
55
58
61
64
67
70
73
76
79
82
85
88
91
Exposure
Model Prediction at Base levels
Model Prediction +/- 2 Standard Errors

Example 2

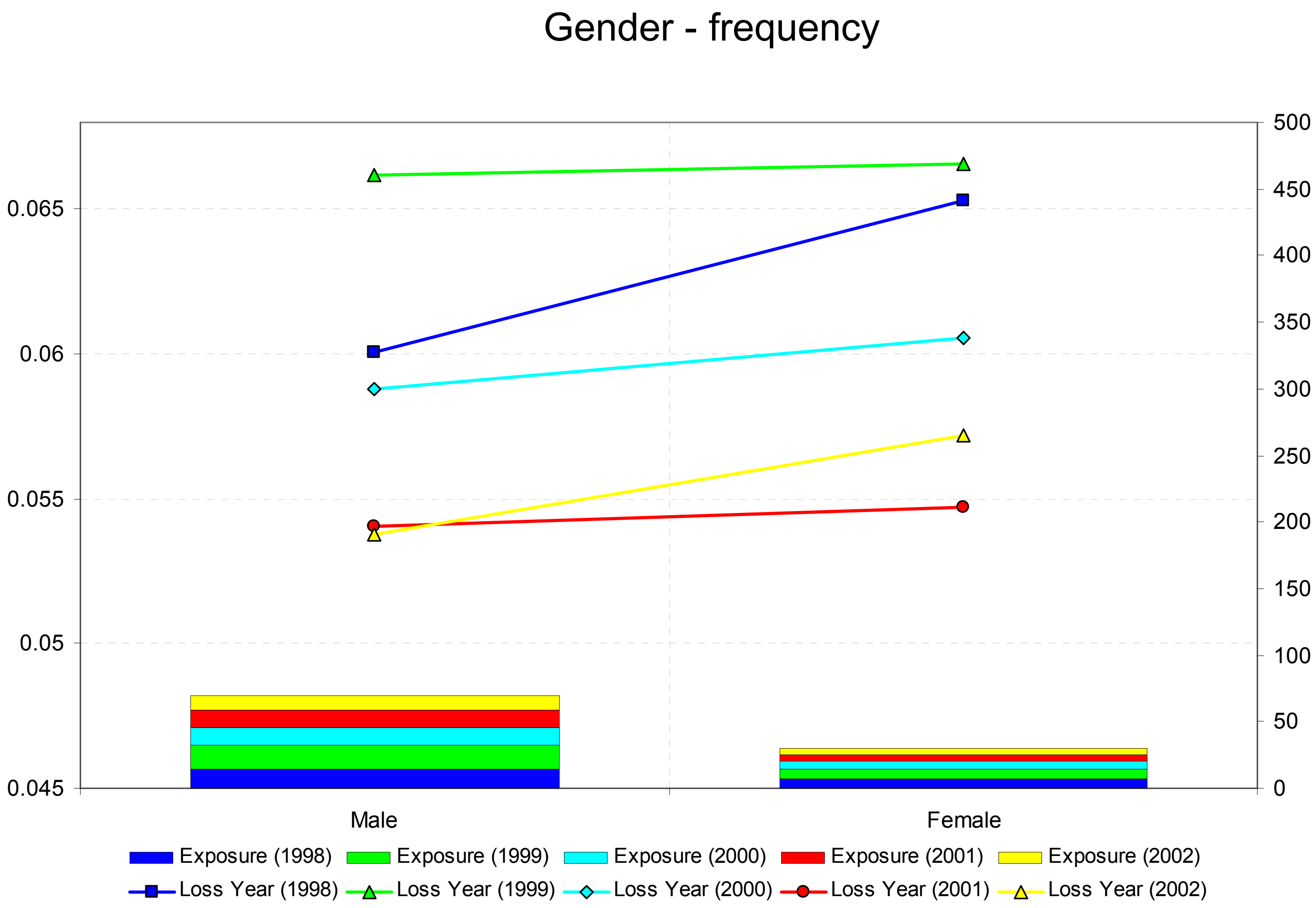

Urban density - risk premium
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
18
16
14
12
10
8
6
4
2
0
1
4
7
10
13
16
19
22
25
28
31
34
37
40
43
46
49
52
55
58
61
64
67
70
73
76
79
82
85
88
91
Exposure
Tweedie
Model Prediction +/- 2 Standard Errors
Traditional

Example 3

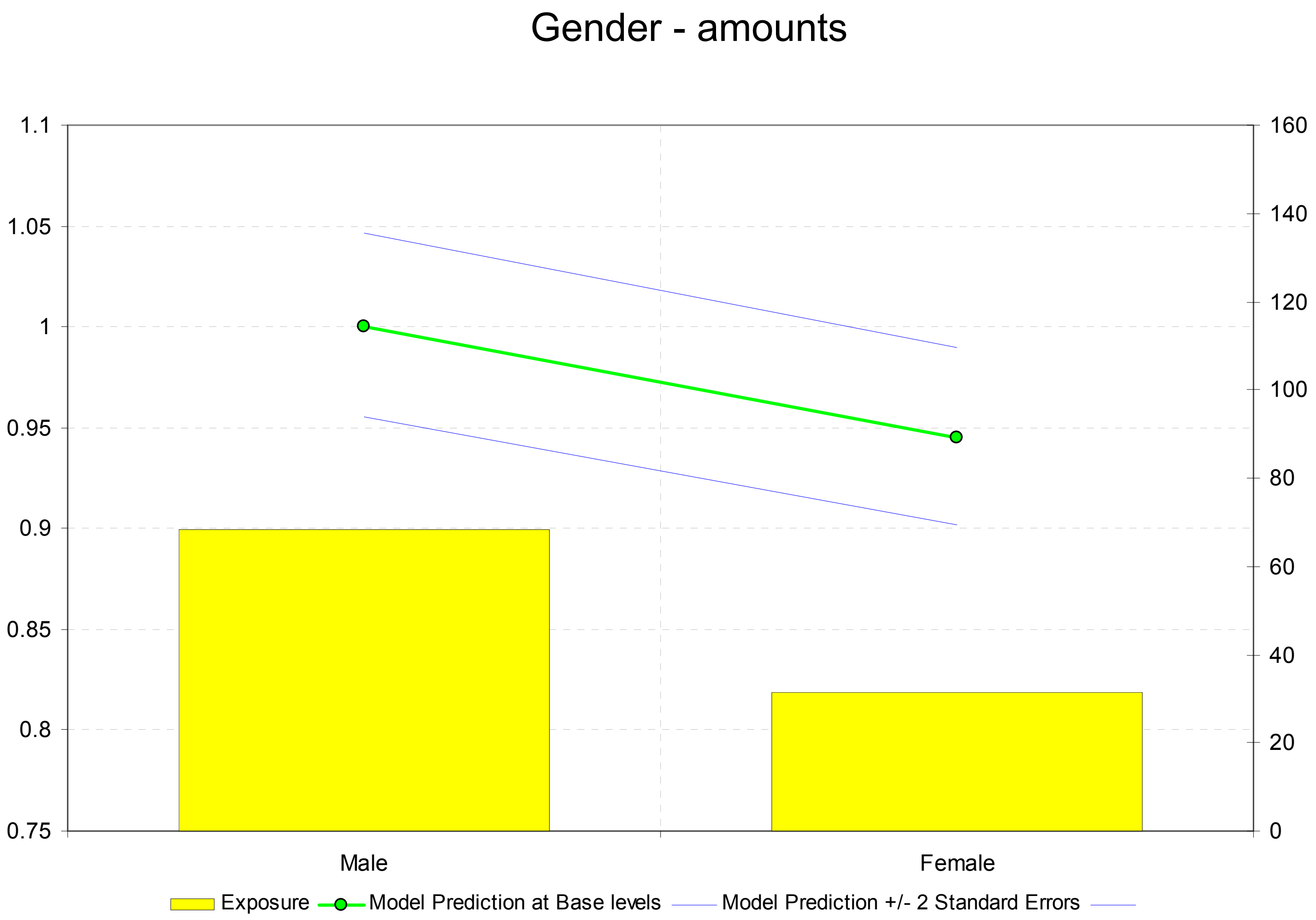


Gender - frequency
1.1
1.05
1
0.95
0.9
160
140
120
100
80
60
40
20
0
Male
Female
Exposure
Model Prediction at Base levels
Model Prediction +/- 2 Standard Errors

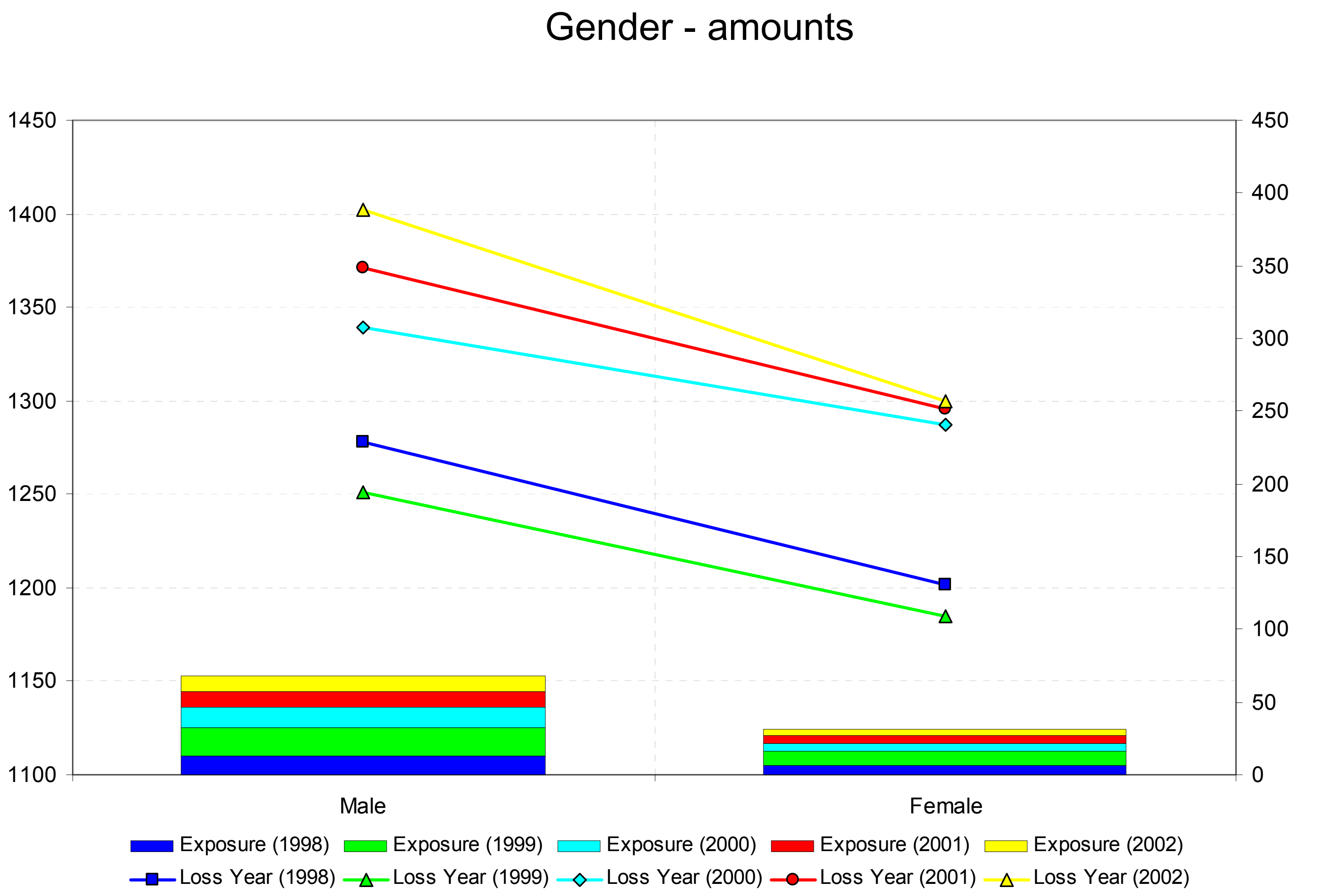
Example 3


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)

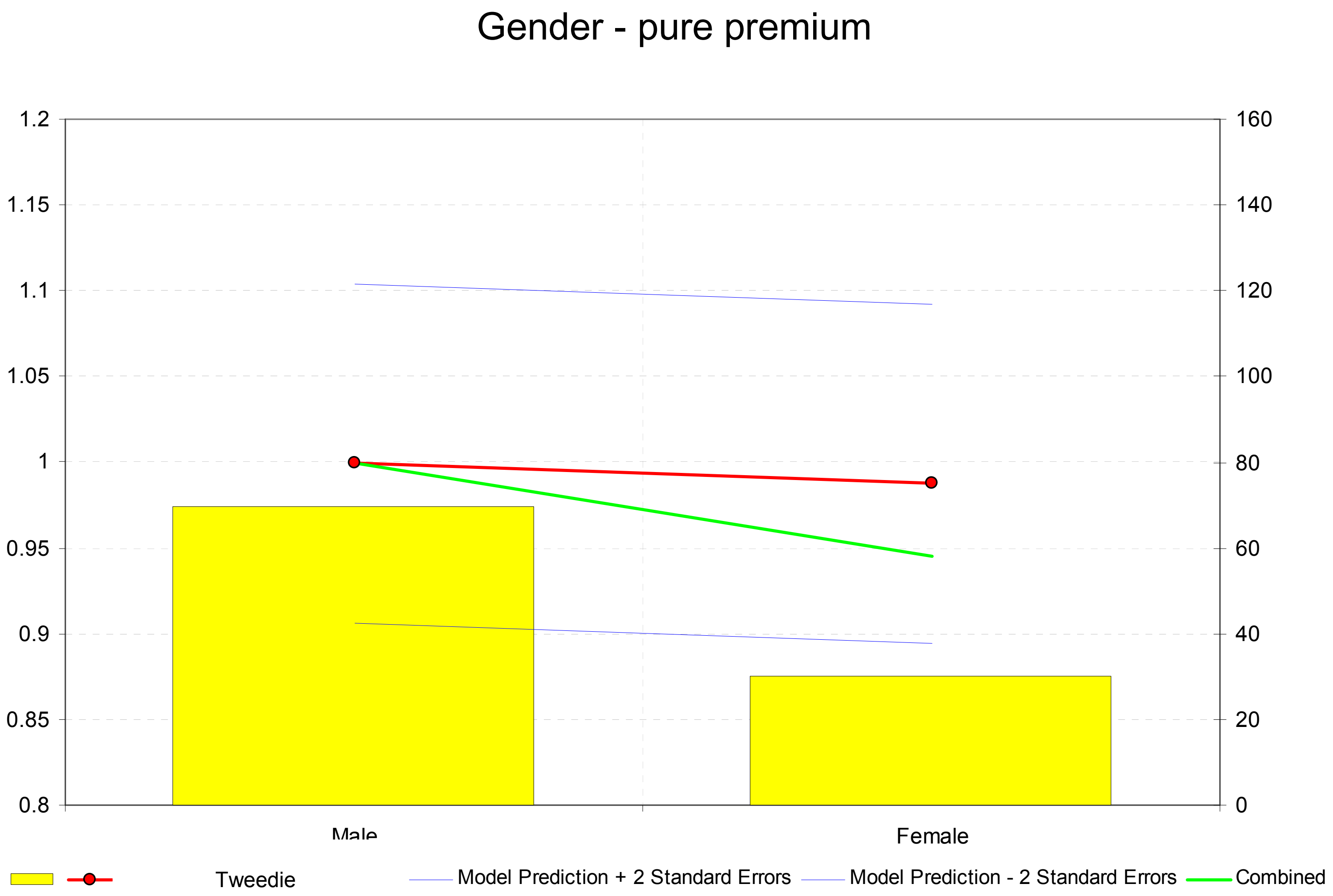
Example 3


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

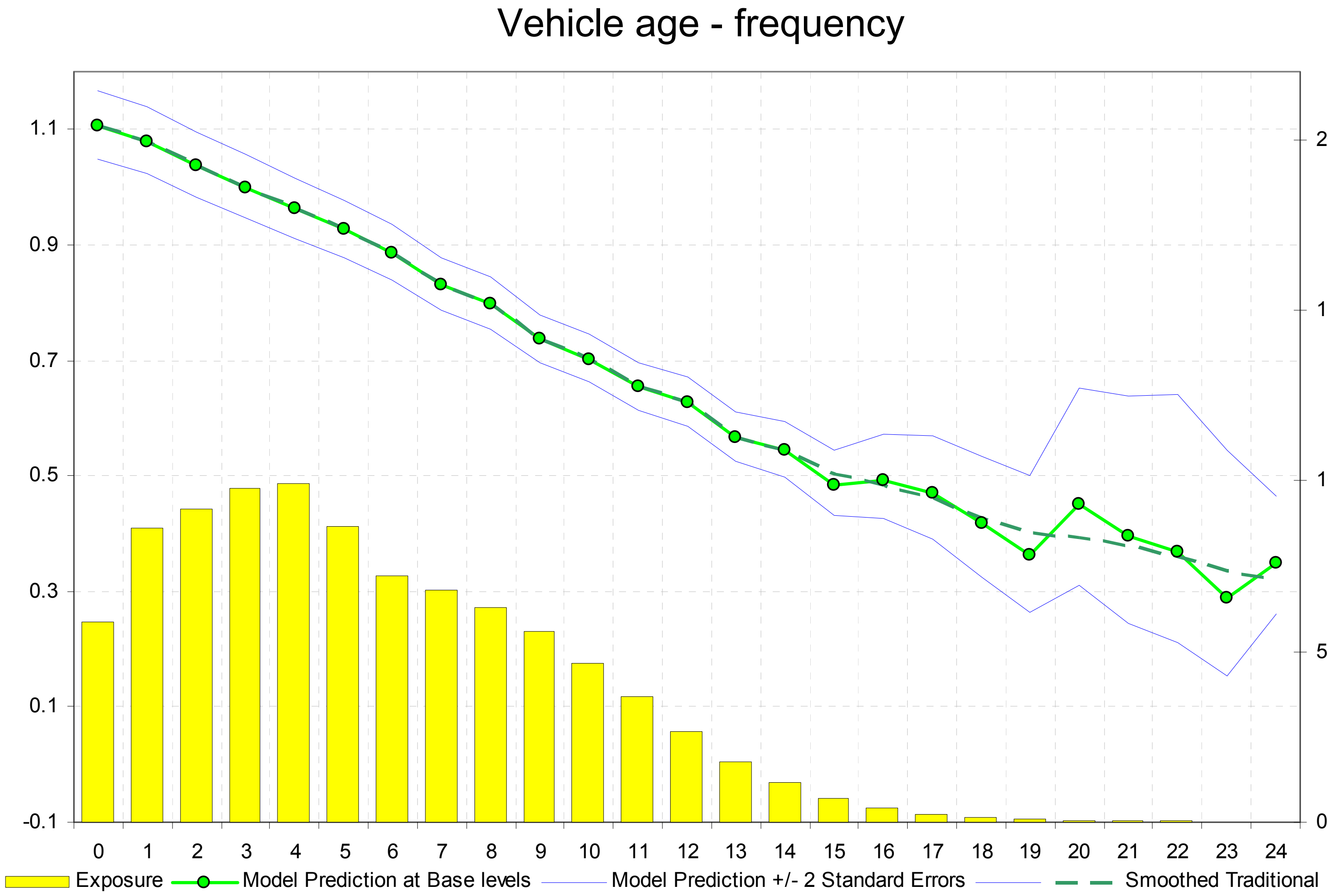
Example 3


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)

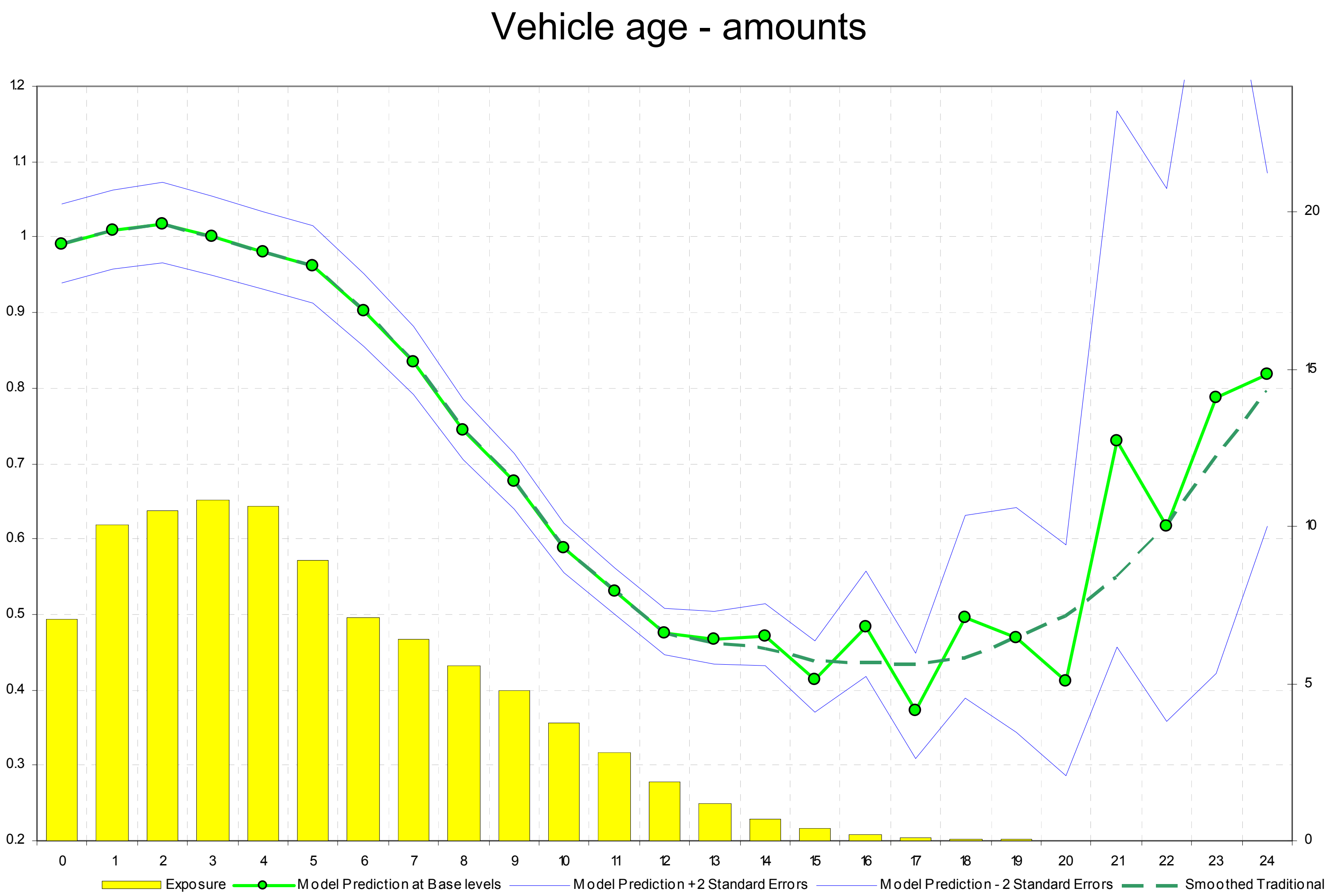
Example 3


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

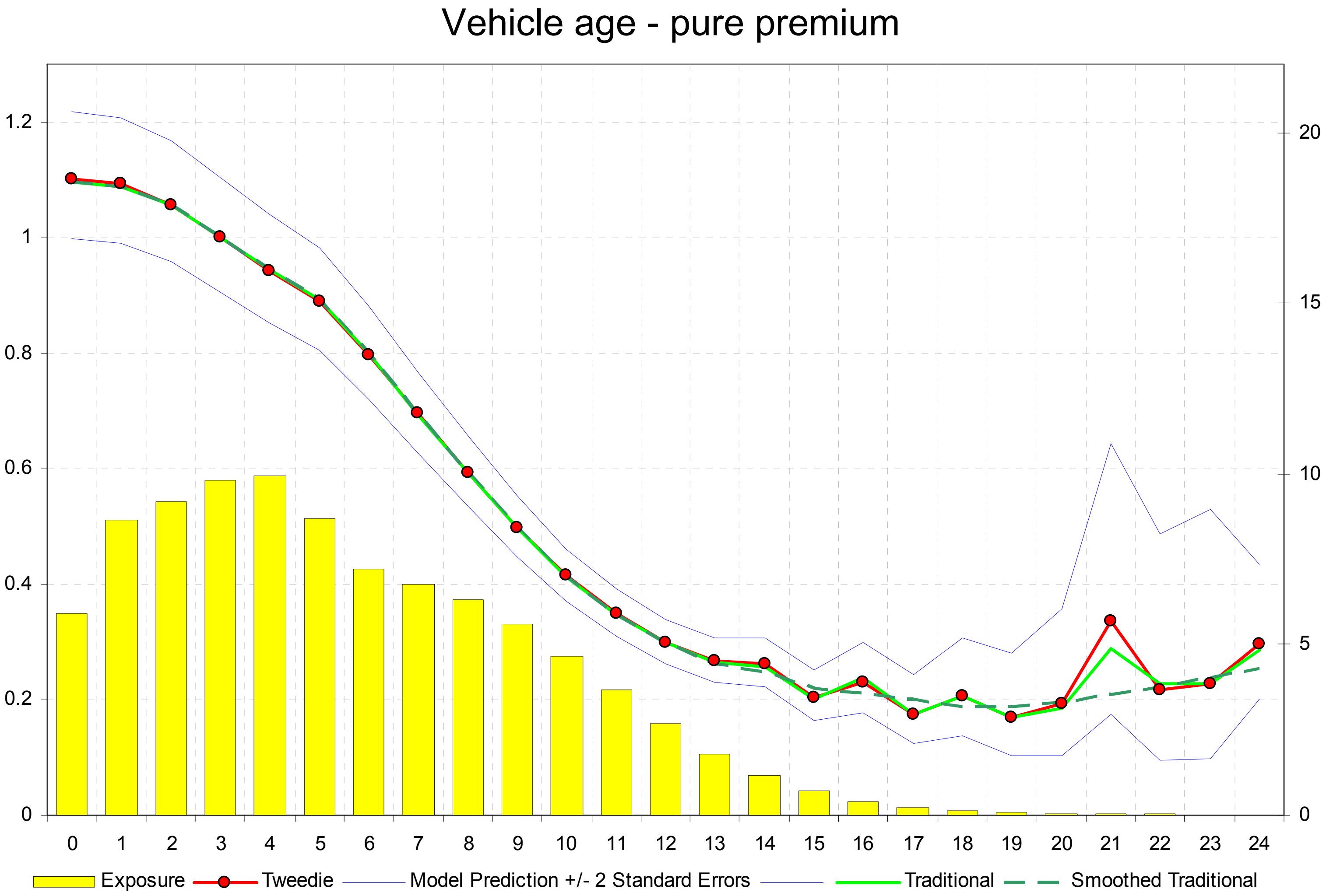
Example 4


Vehicle age - frequency
1.1
0.9
0.7
0.5
0.3
0.1
-0.1
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
Smoothed Traditional

Example 4


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
Smoothed Traditional

Example 4


Vehicle age - pure premium
0
0.2
0.4
0.6
0.8
1
1.2
0
5
10
15
20
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
Smoothed Traditional

Tweedie GLMs

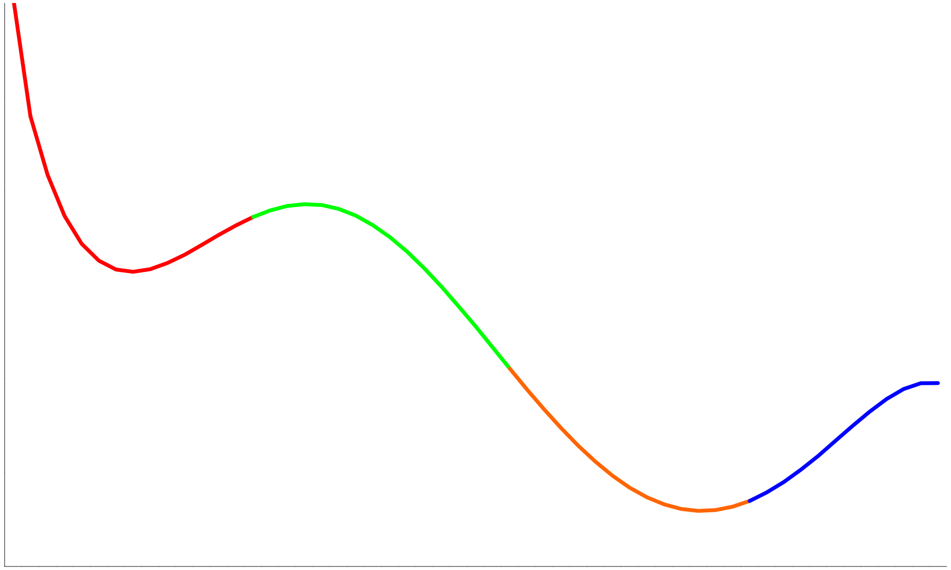
- Helpful when it's important to fit to incurred costs directly
- Similar results to frequency/severity traditional approach if frequency and amounts effects are clearly weak or clearly strong
- Distorted by large insignificant effects
- Removes understanding of what is driving results
- Smoothing harder

Agenda

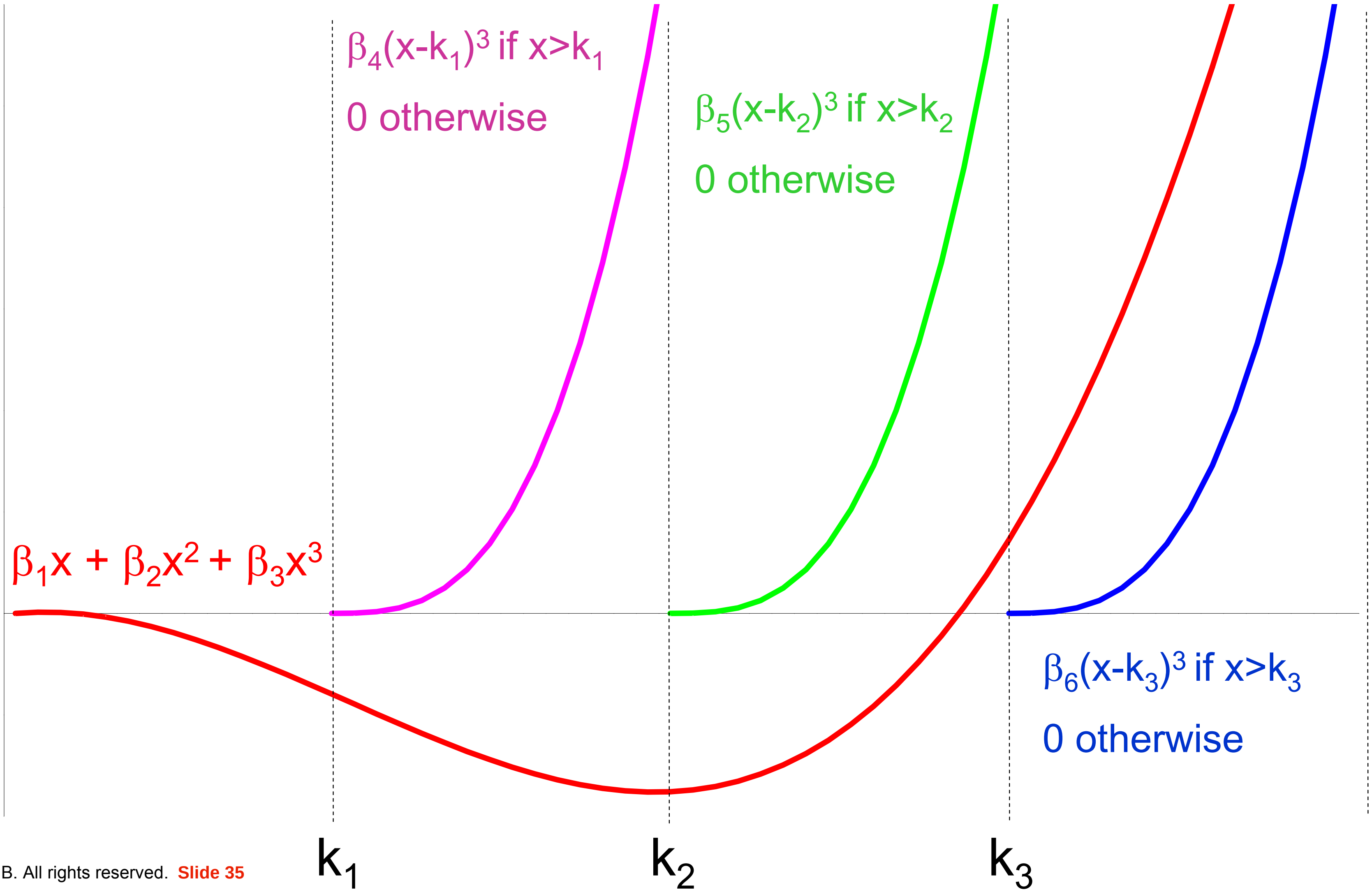
- Testing the link function
- The Tweedie distribution
- Regression splines
- Reference models
- Aliasing/near-aliasing
- Combining models across claim types
- Restricted models
- Model validation
- Modeling elasticity / GNMs

Splines

- A spline is
  - a series of polynomials…
  - …joining at "knots"…
  - …"smoothly"
  - (k "internal" knots and 2 extra knots at end of data range)
- A cubic spline is
  - a spline made up of cubic polymonials
  - continuous at each knot
  - first derivative continuous at each knot
  - second derivative continuous at each knot
- A regression spline is
  - a formularization which allows splines to be fitted within a GLM framework
  - requires manual selection of knots



Regression splines



$\beta_4(x-k_1)^3$ if $x>k_1$

0 otherwise

$\beta_5(x-k_2)^3$ if $x>k_2$

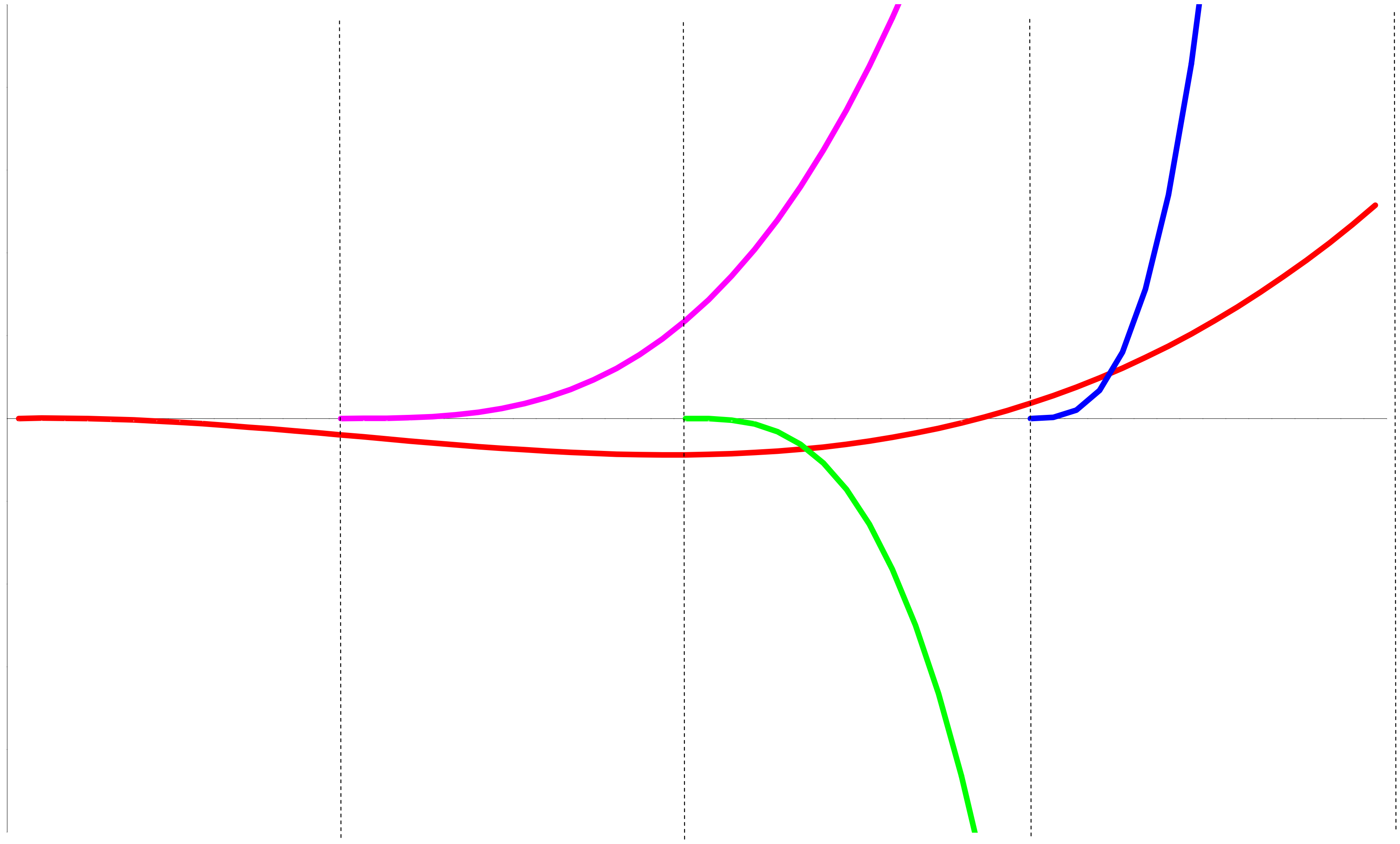
0 otherwise

$\beta_1 x + \beta_2 x^2 + \beta_3 x^3$

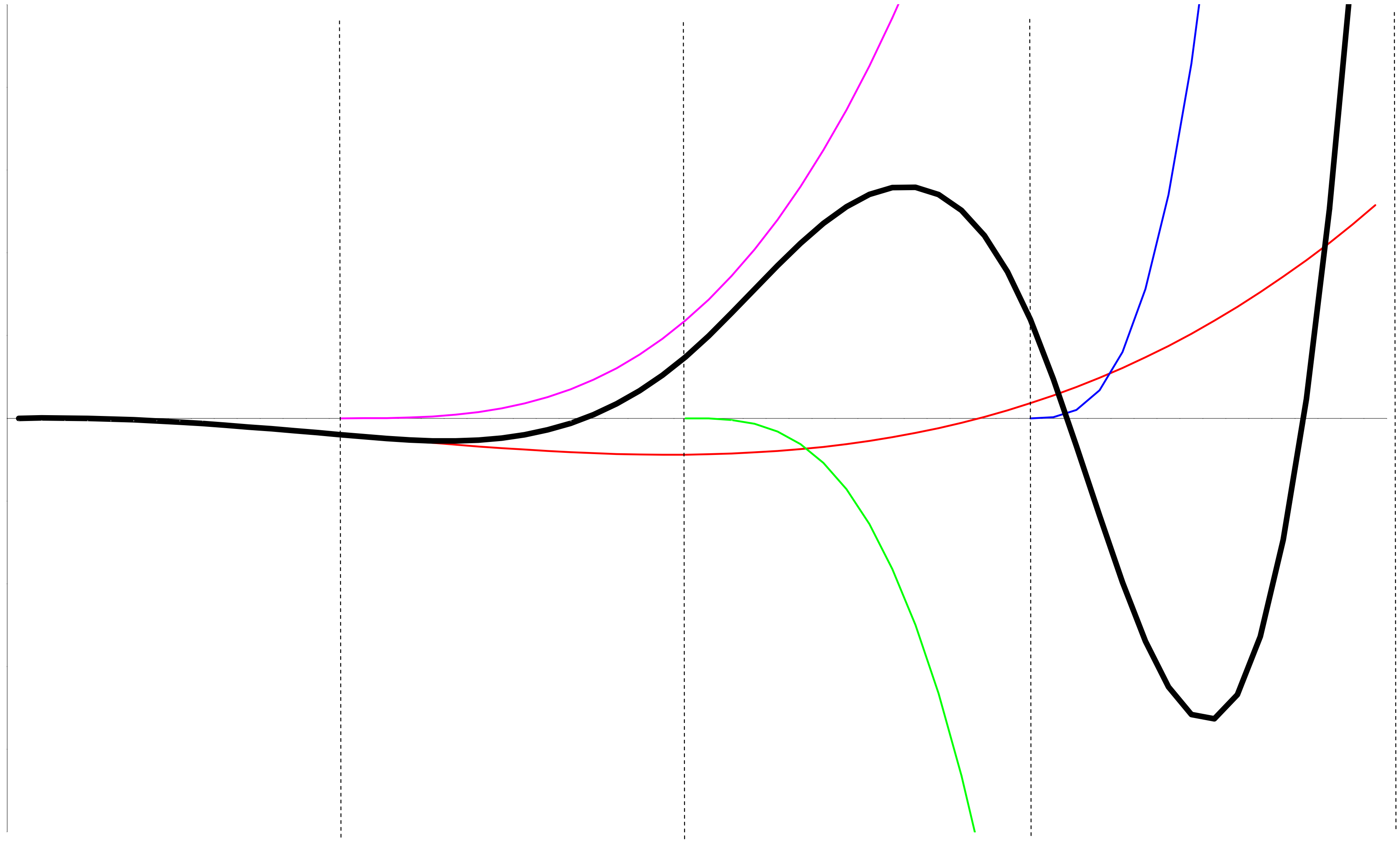
$\beta_6(x-k_3)^3$ if $x>k_3$

0 otherwise

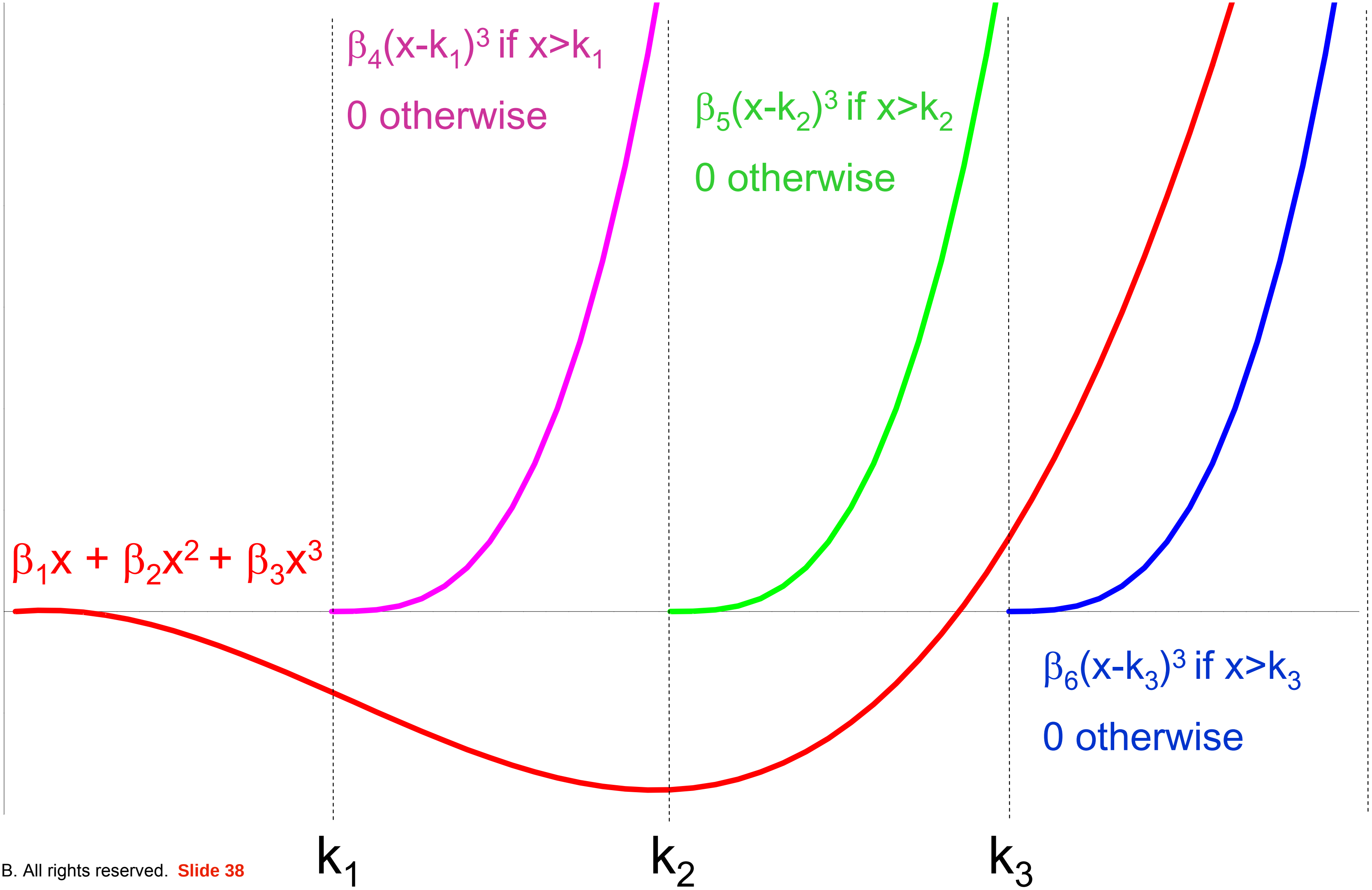
Regression splines



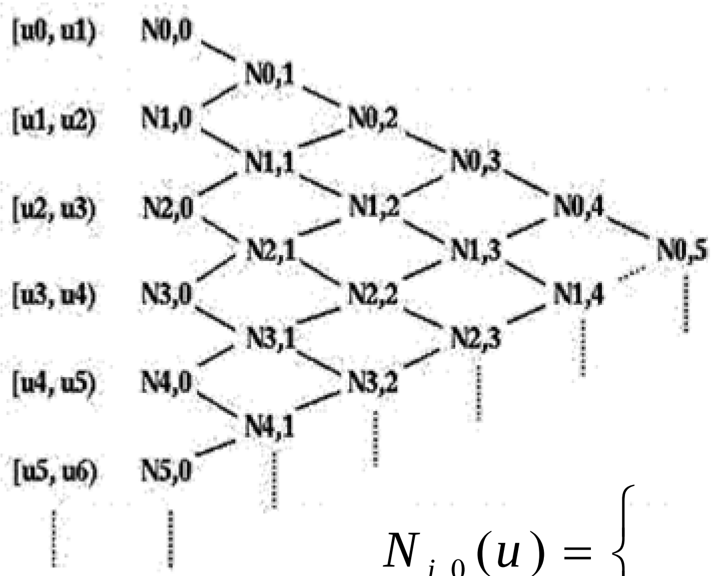
Regression slines



Regression splines


$\beta_4(x-k_1)^3$ if $x>k_1$
0 otherwise
$\beta_5(x-k_2)^3$ if $x>k_2$
0 otherwise
$\beta_1 x + \beta_2 x^2 + \beta_3 x^3$
$\beta_6(x-k_3)^3$ if $x>k_3$
0 otherwise
k_1
k_2
k_3

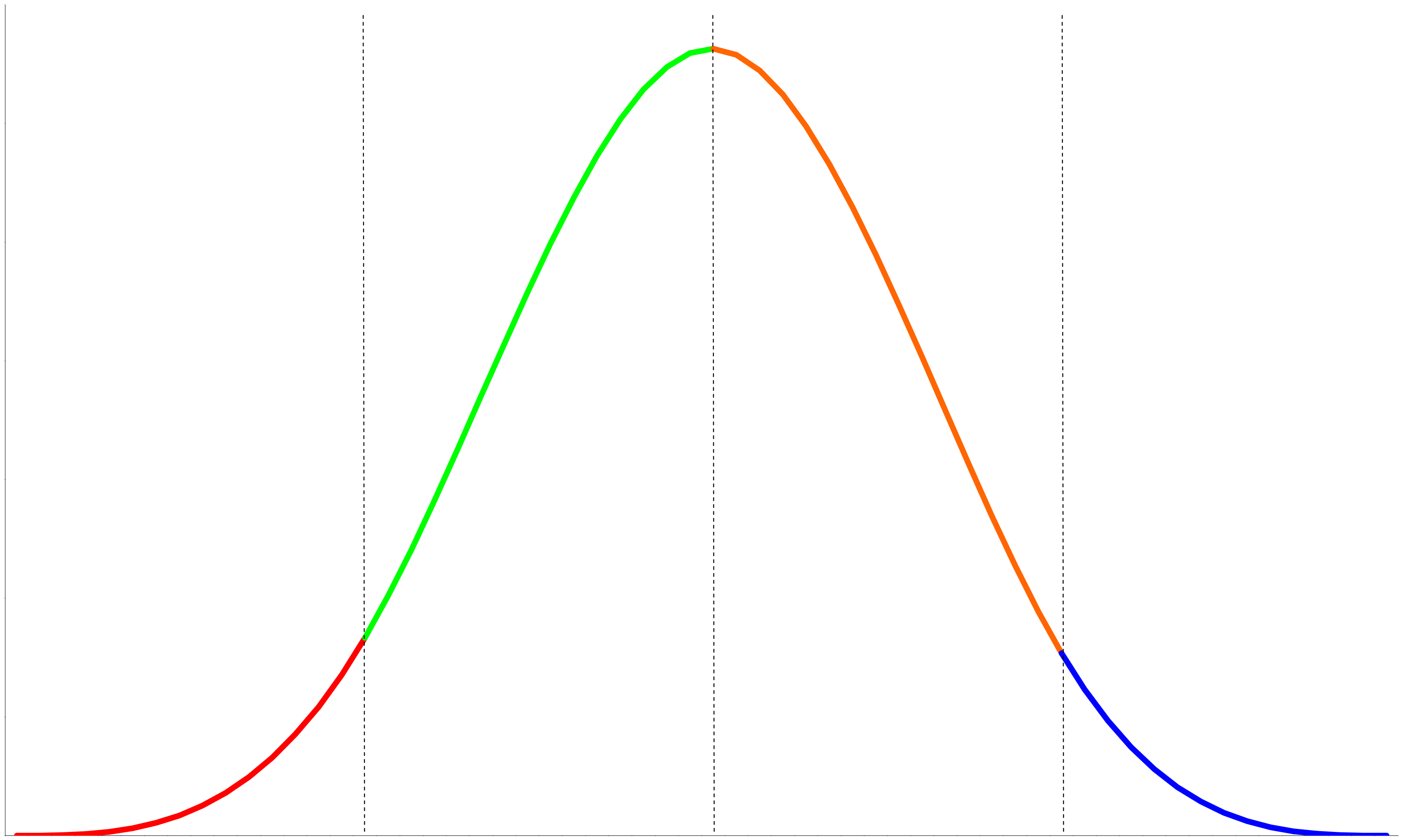
B-splines



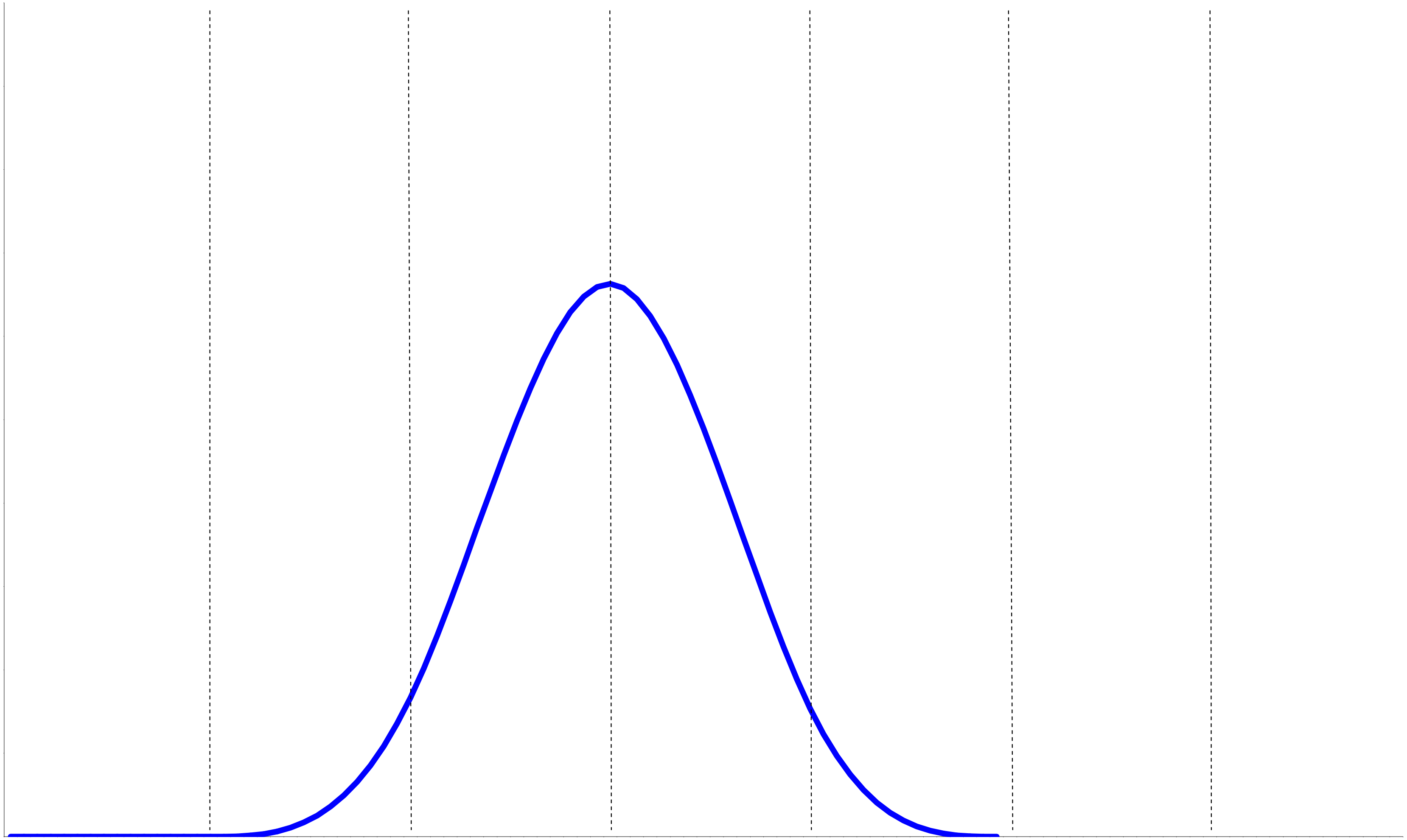
$$N_{i,0}(u) = \begin{cases} 1 & \text{if } u_i \le u < u_{i+1} \\ 0 & \text{otherwise} \end{cases}$$

$$N_{i,p}(u) = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u)$$

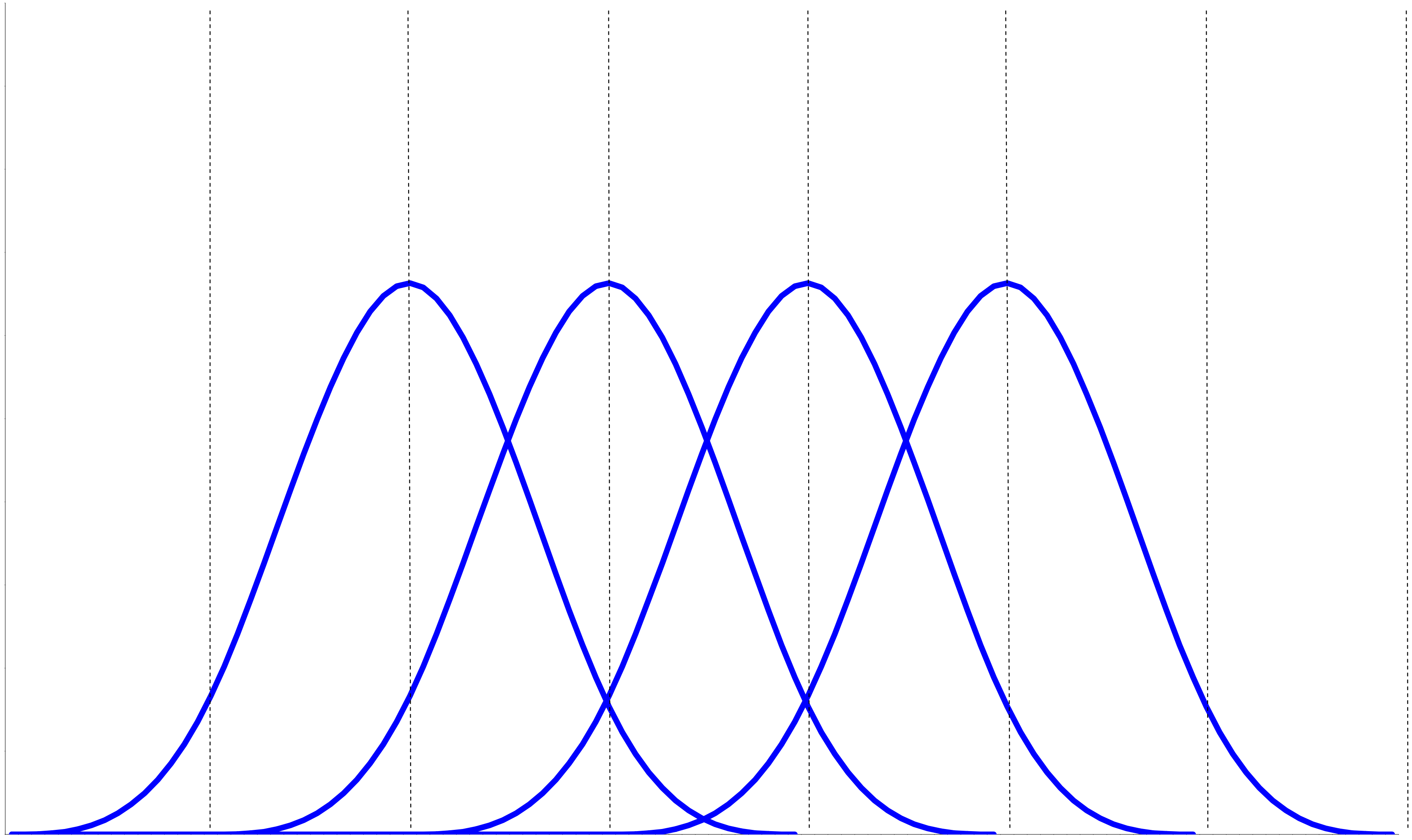
B-splines



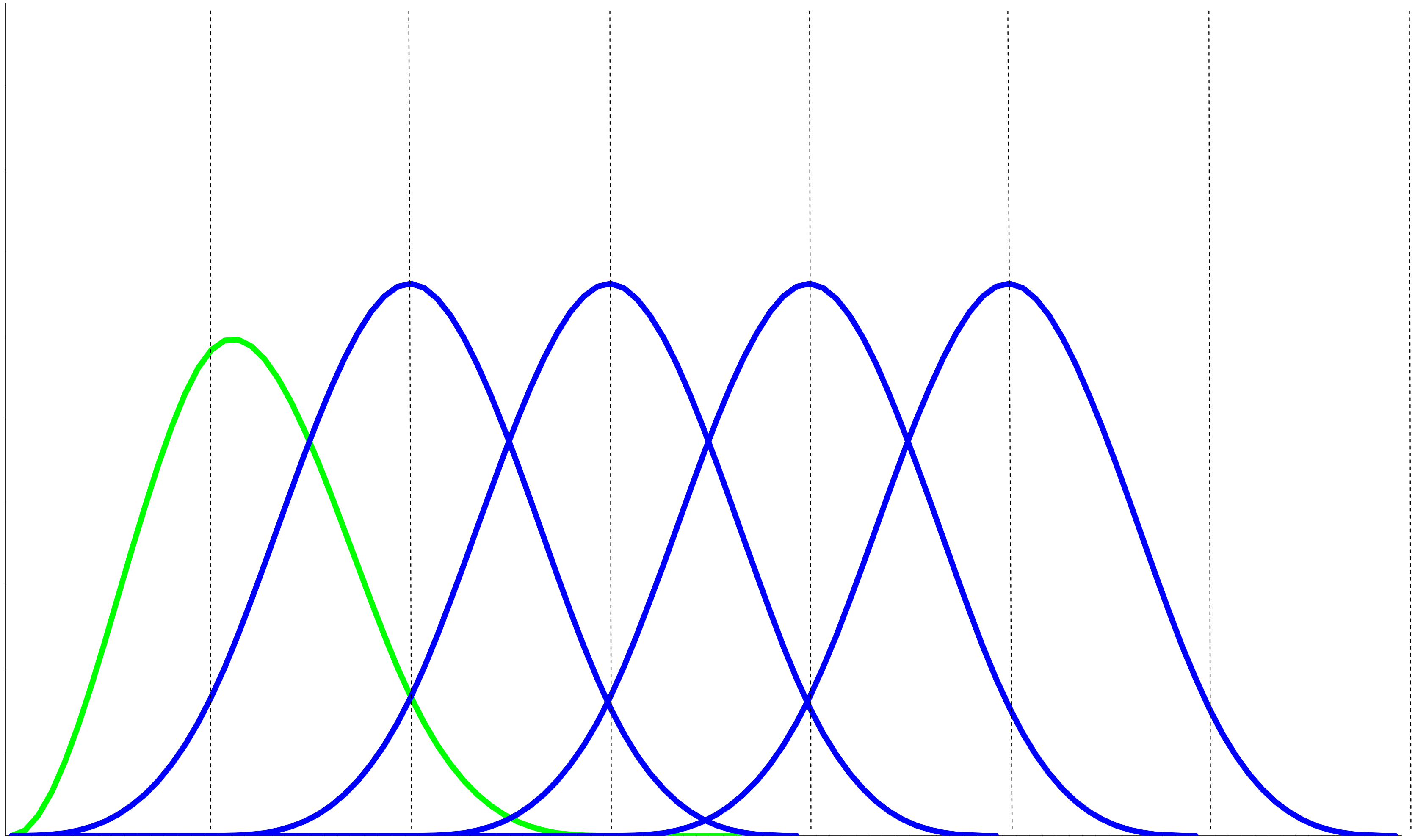
B-splines



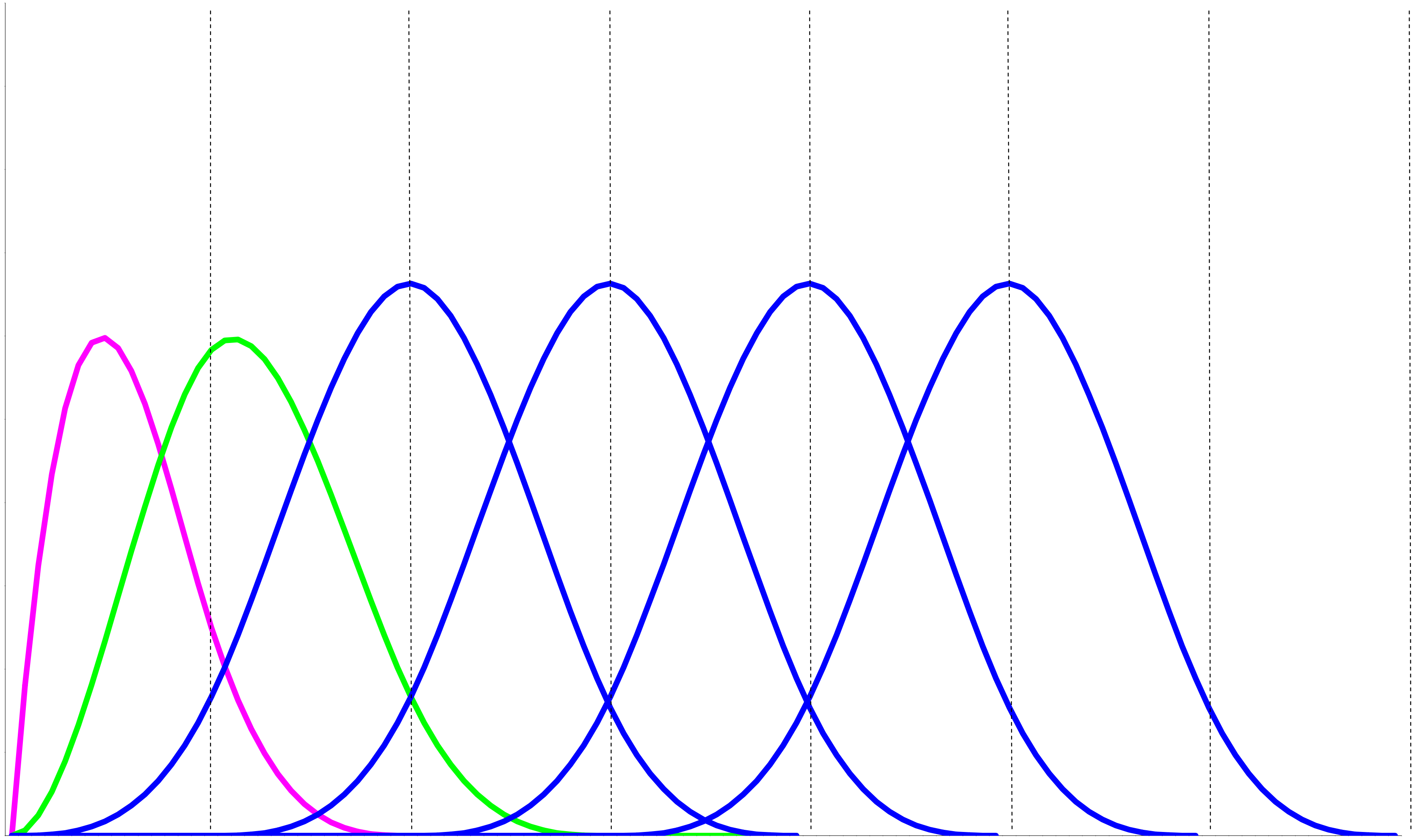
B-splines



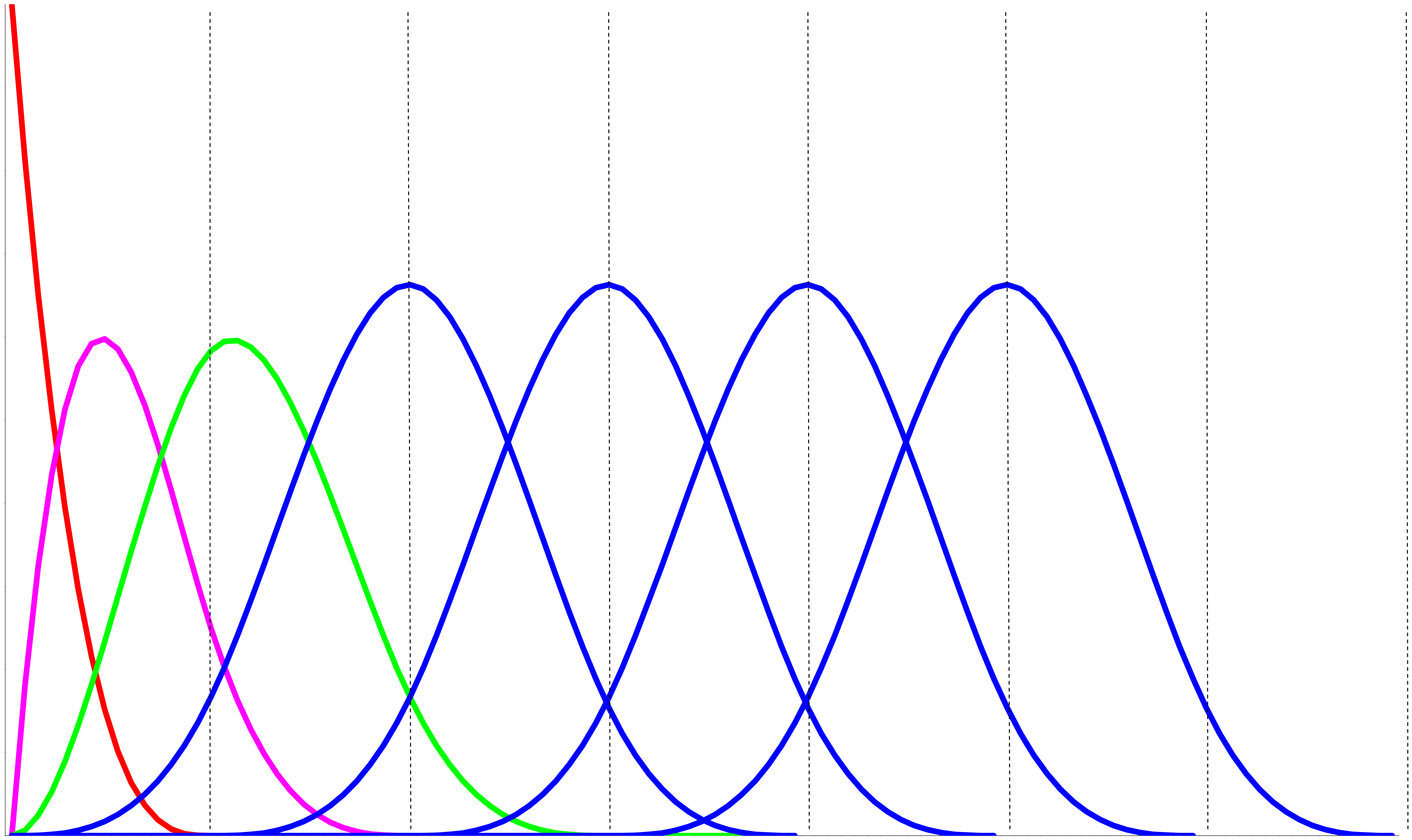
B-splines



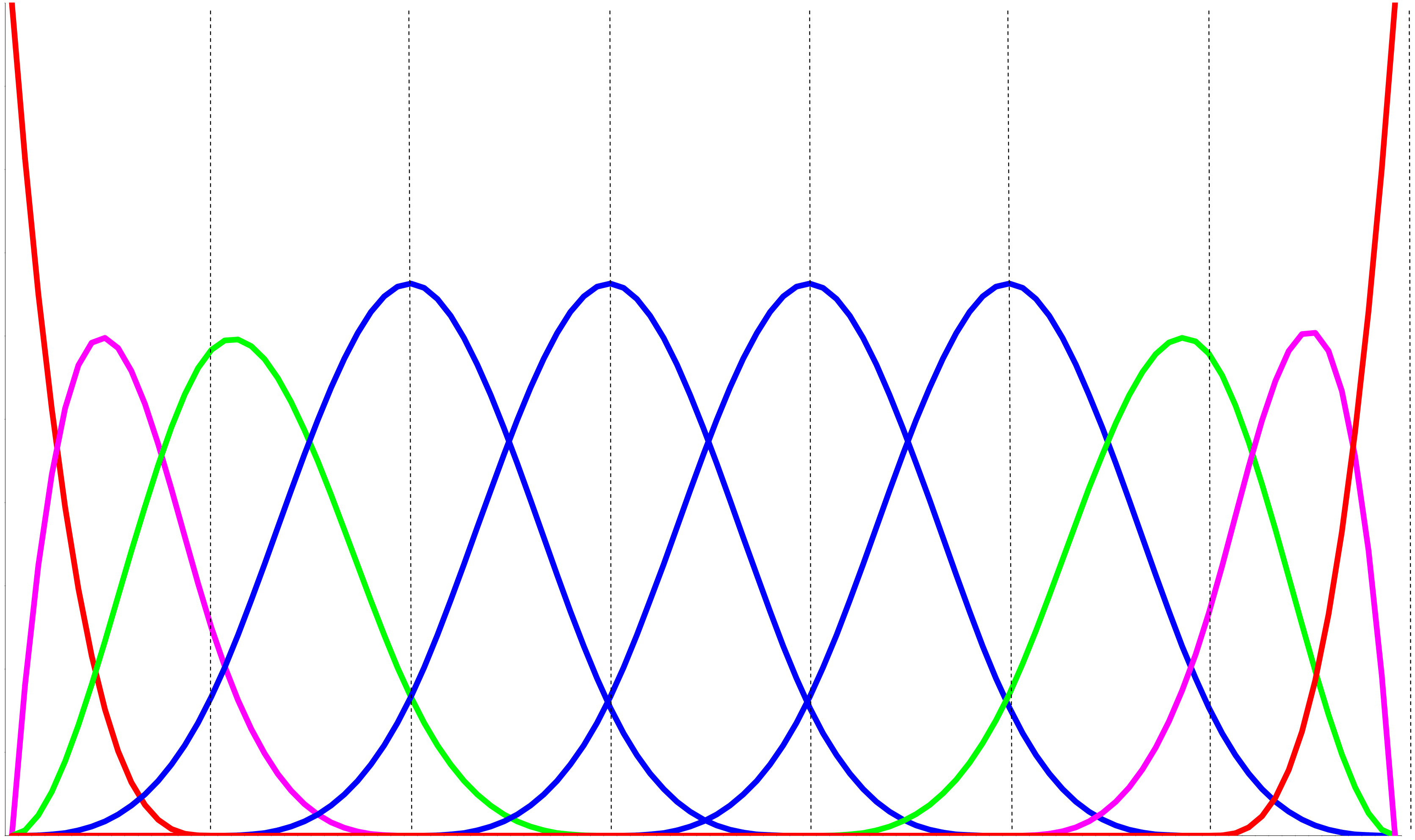
B-splines



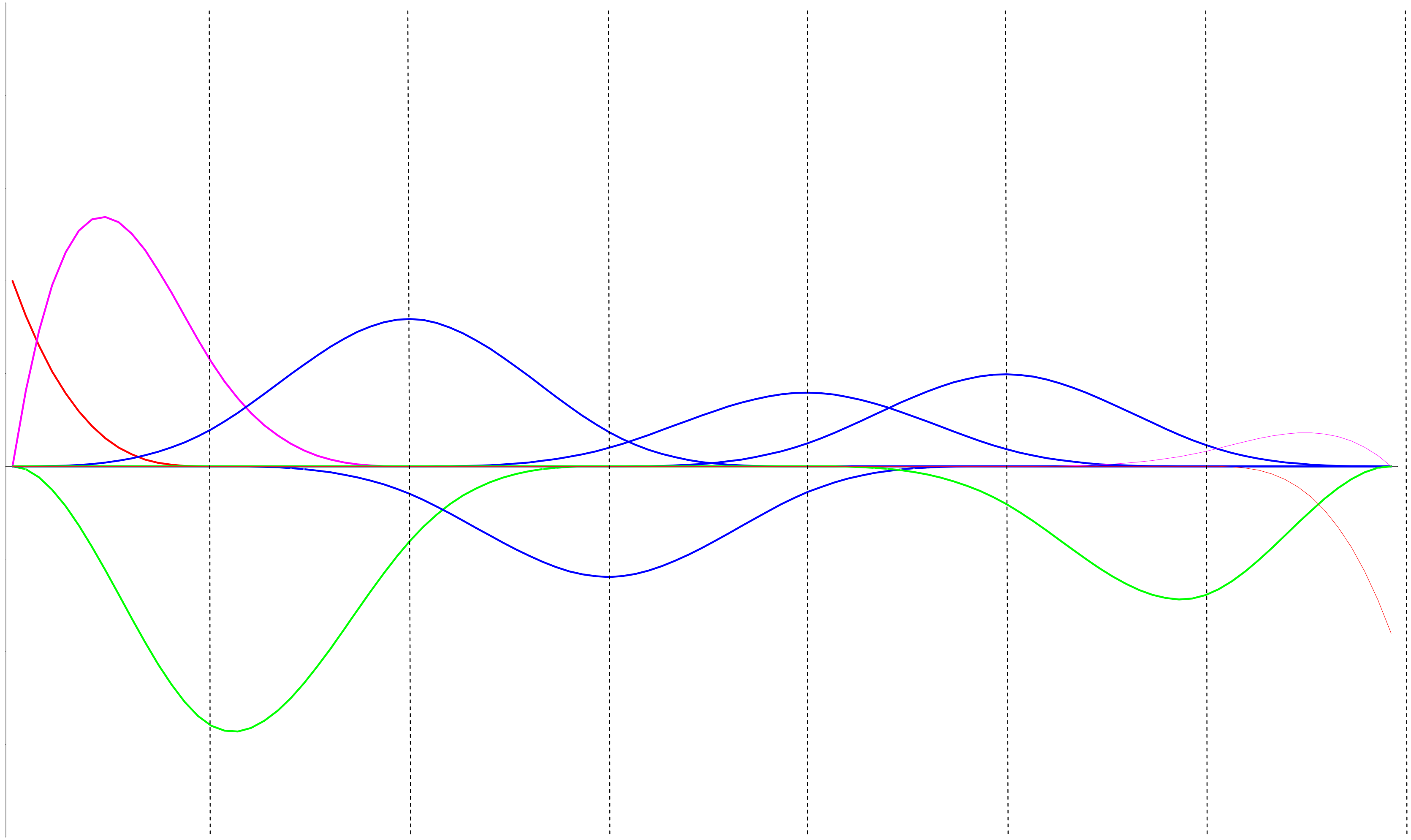
B-splines



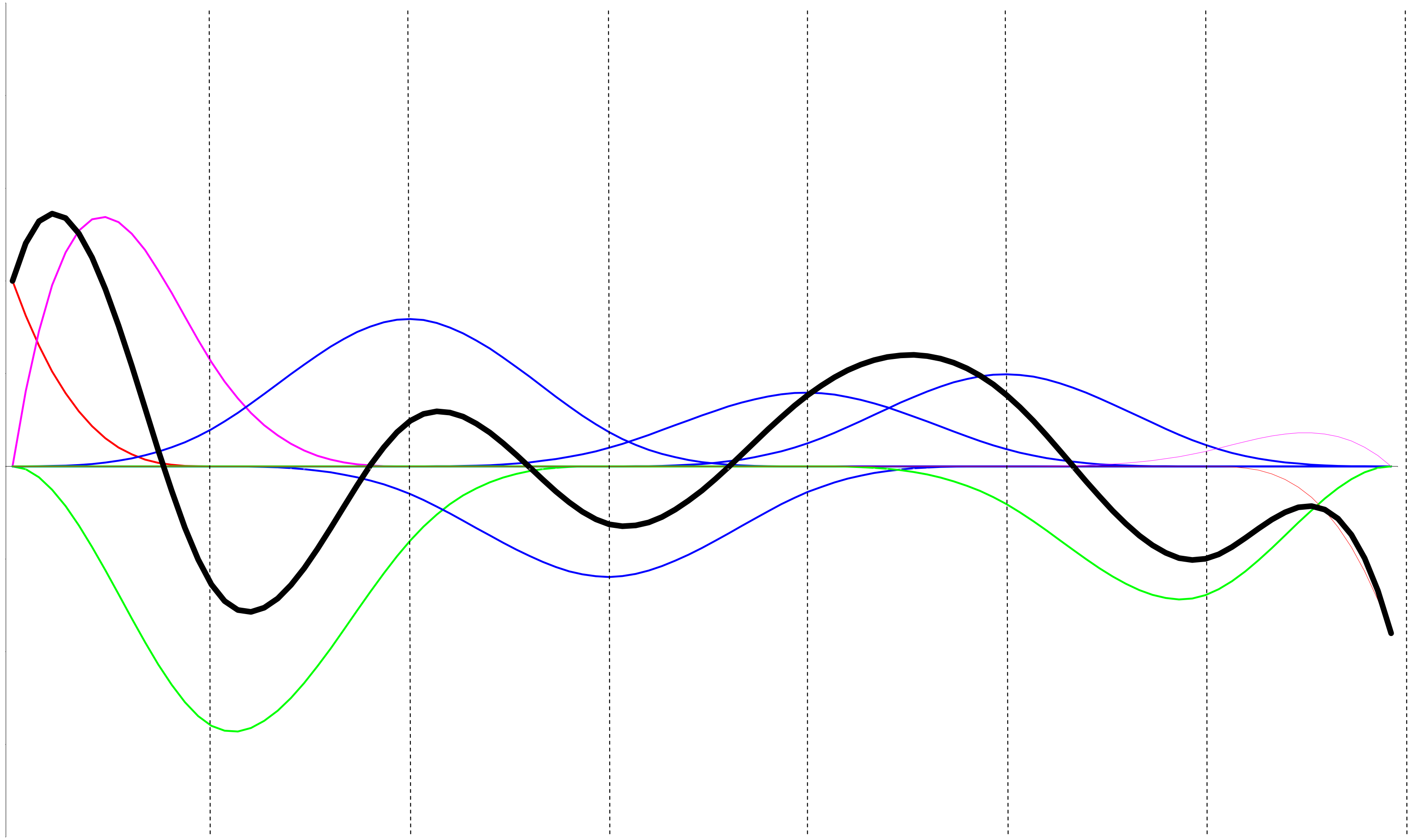
B-splines



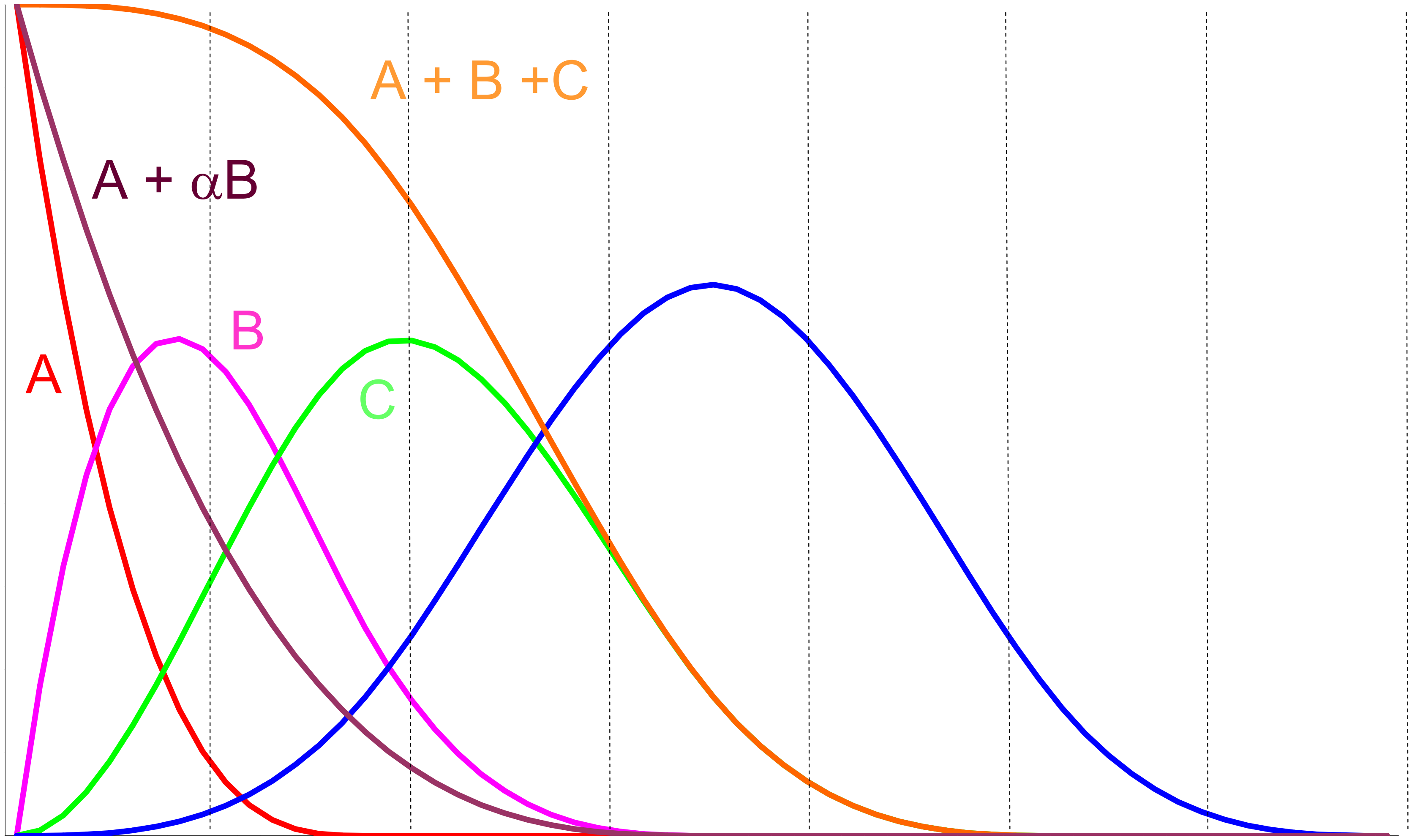
B-splines



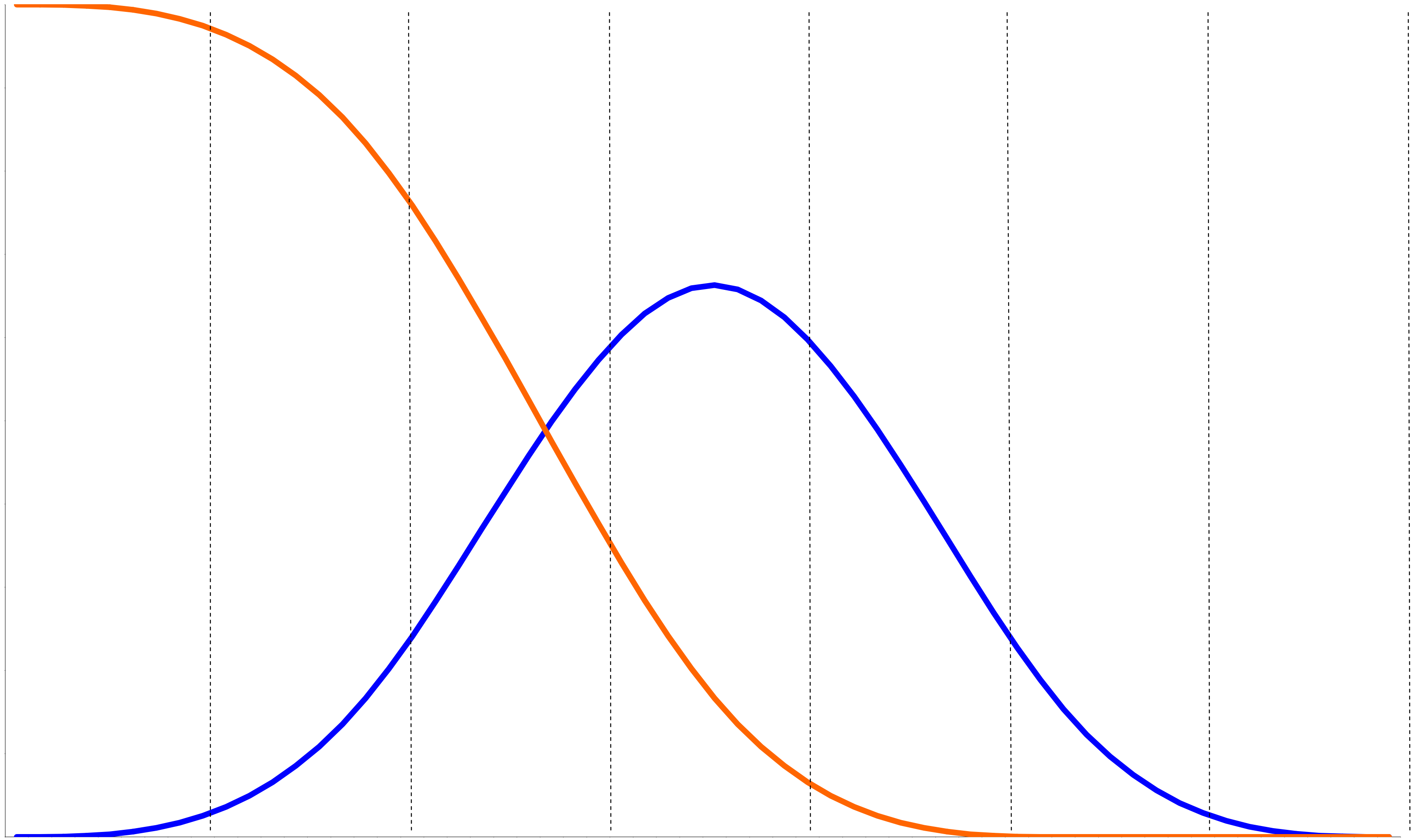
B-splines



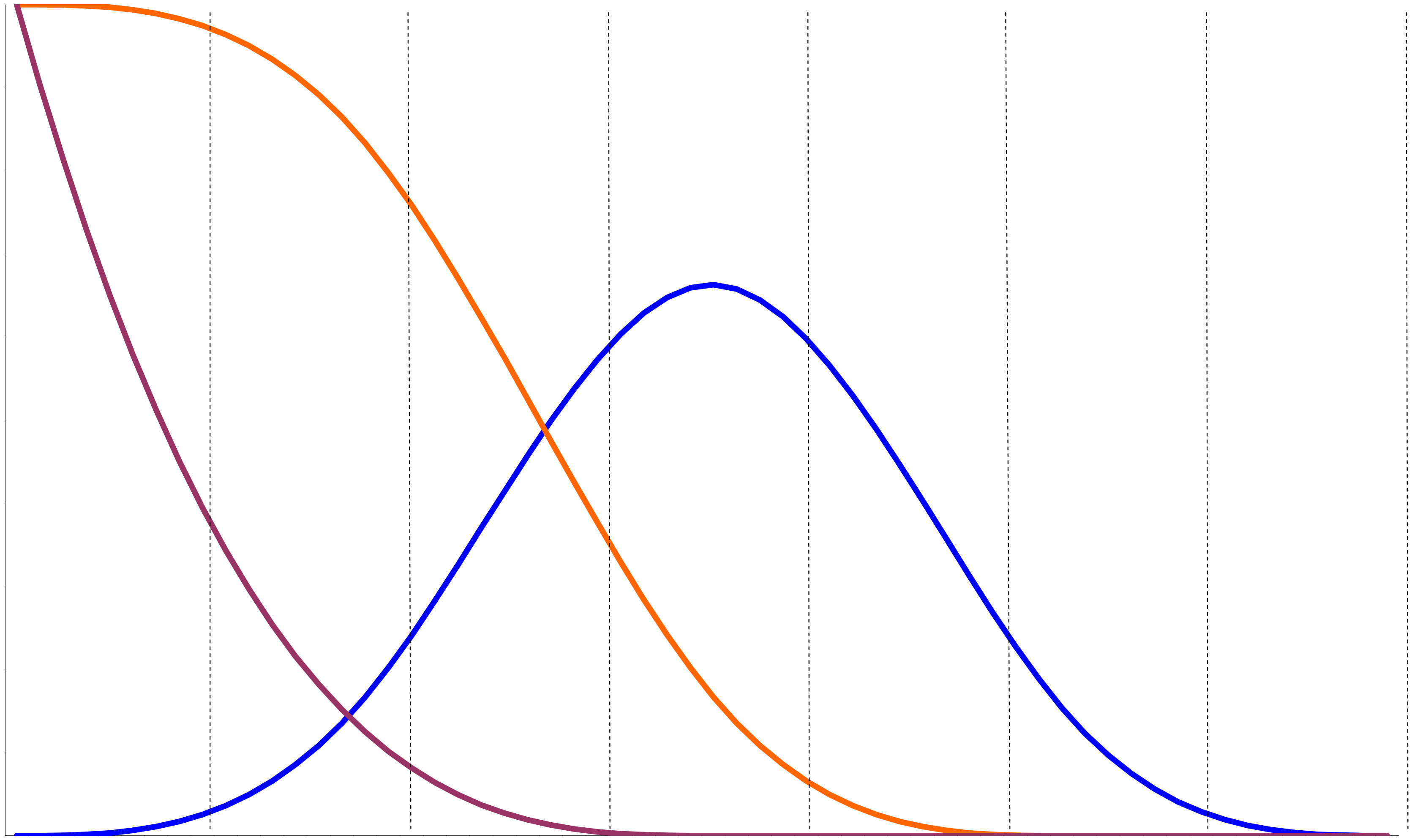
B-splines - extrapolation



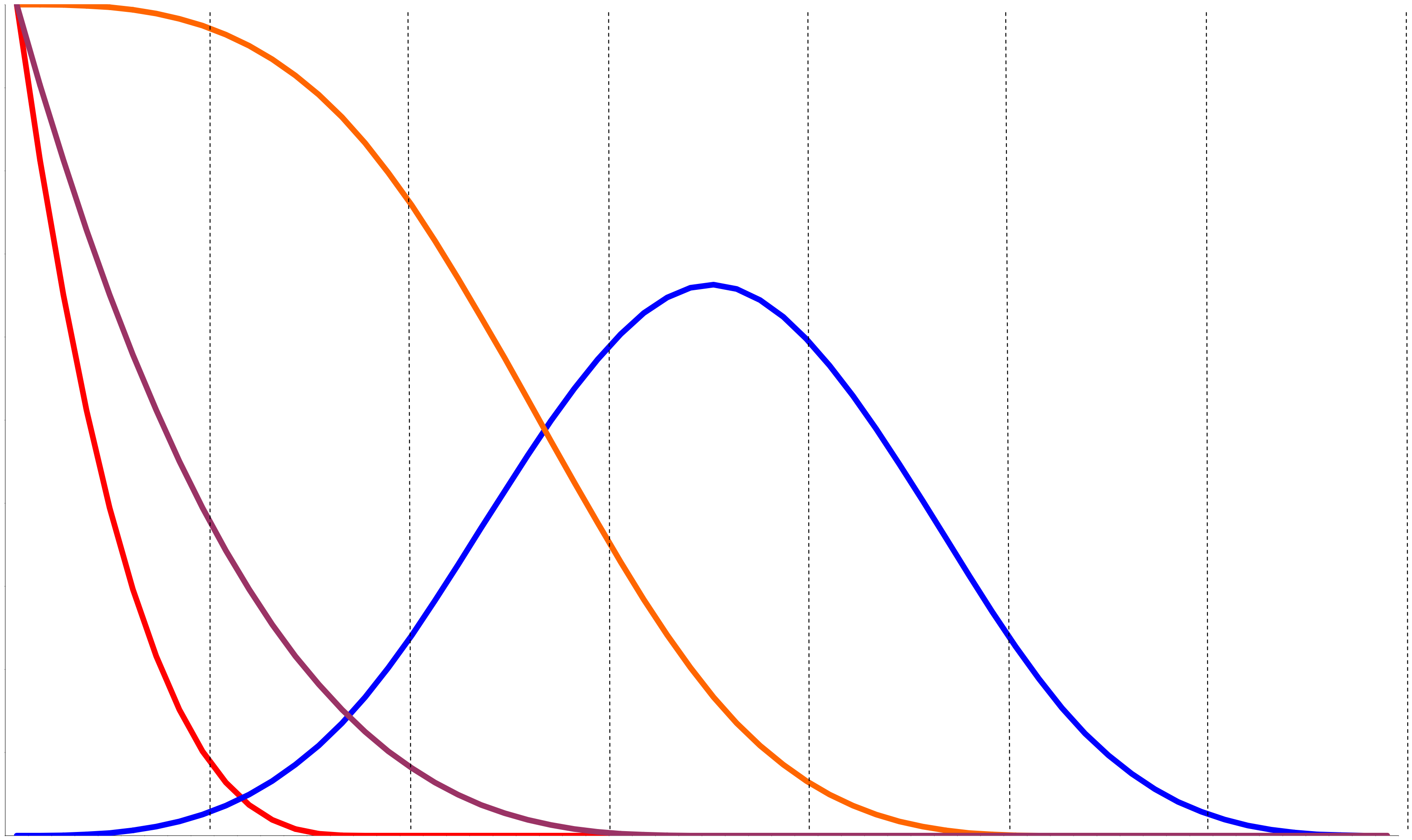
B-splines - constant extrapolation



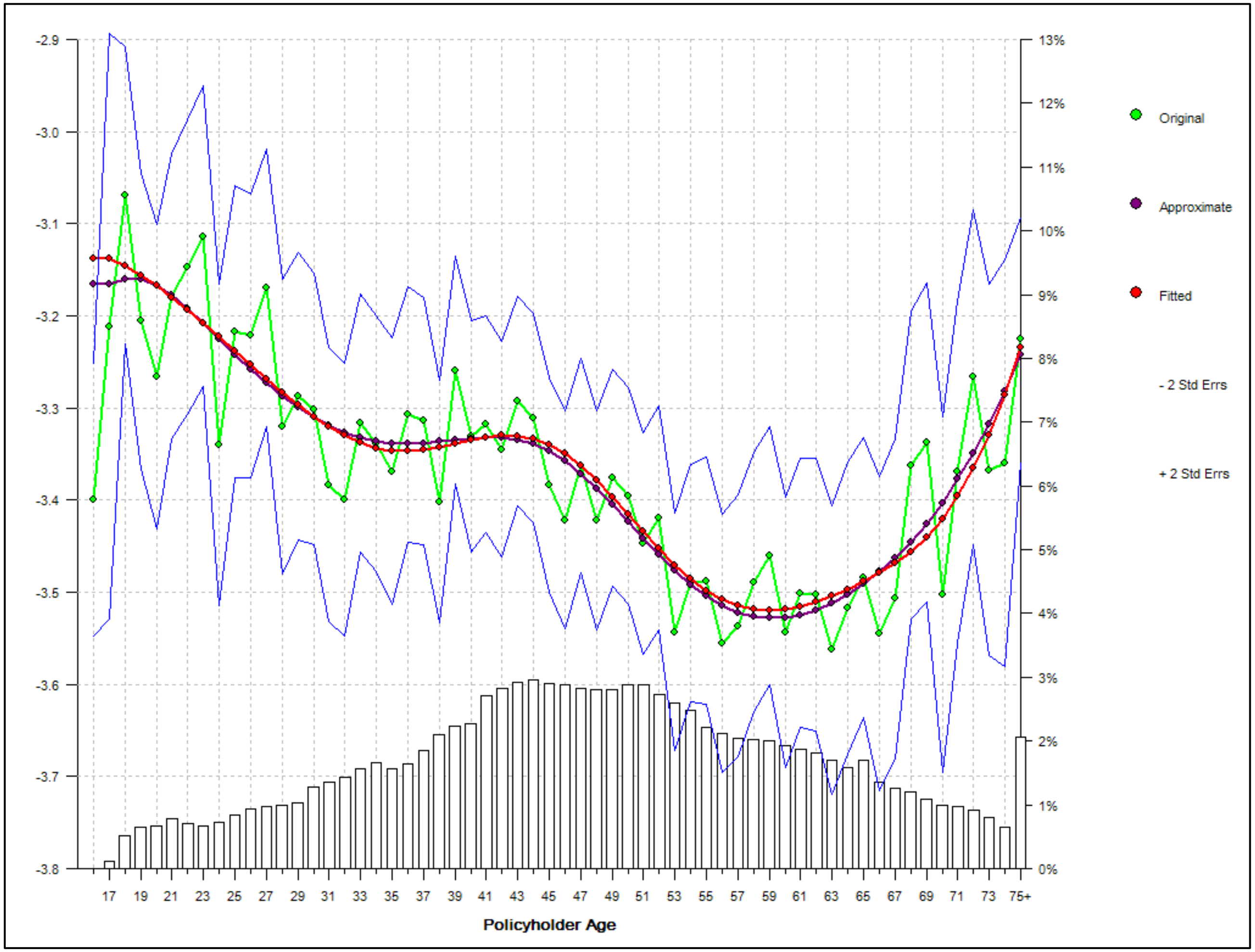
B-splines - linear extrapolation



B-splines - quadratic extrapolation



Example



-2.9
-3.0
-3.1
-3.2
-3.3
-3.4
-3.5
-3.6
-3.7
-3.8
13%
12%
11%
10%
9%
8%
7%
6%
5%
4%
3%
2%
1%
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 71 73 75+
Policyholder Age
Original
Approximate
Fitted
- 2 Std Errs
+ 2 Std Errs

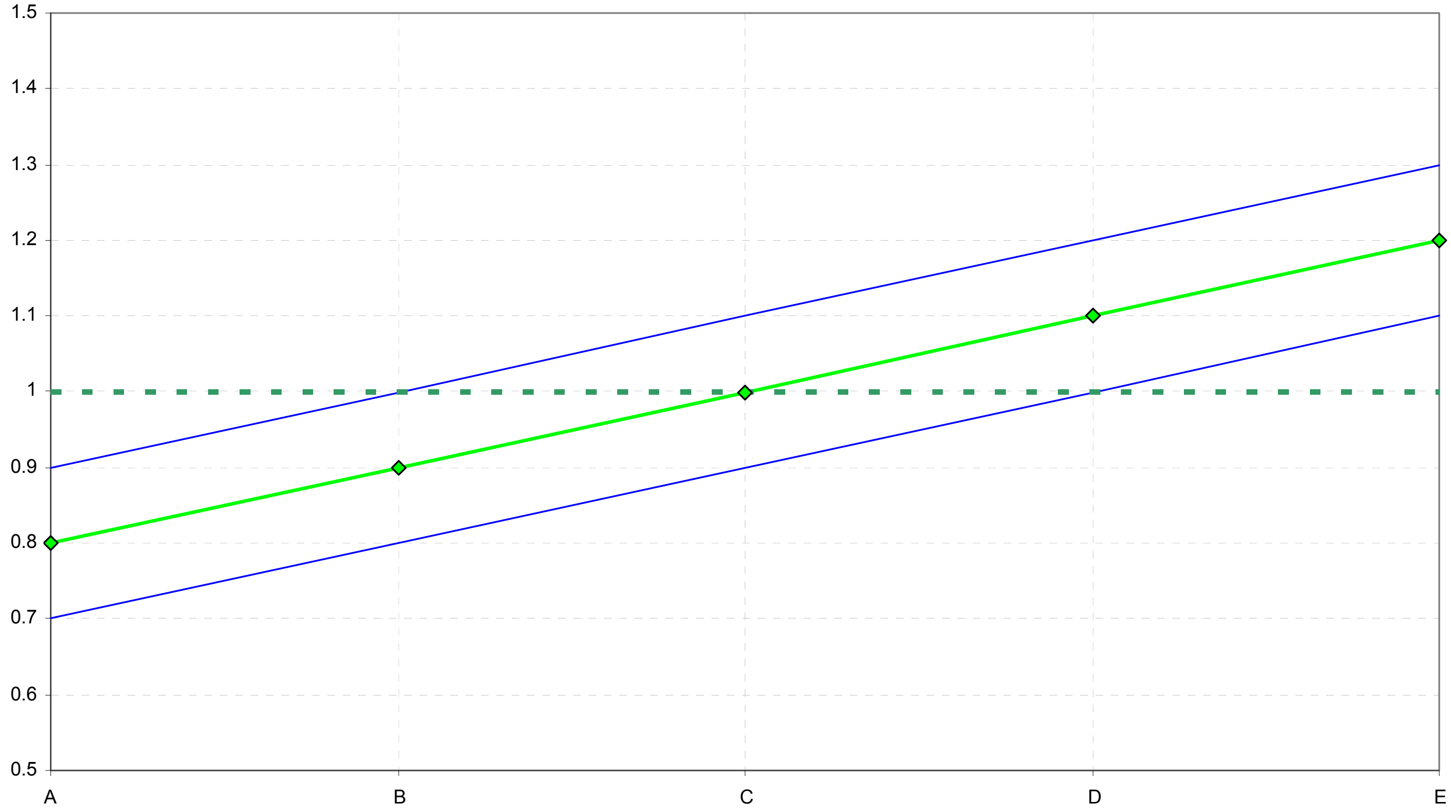
Splines

- Can be useful when continuous effect required
- Increases complexity
- Knot selection important and iterative
  - interactively select design of knot placement on curve fitted to parameter estimates and then incorporate within model
- Can be helpful when simplifying interactions

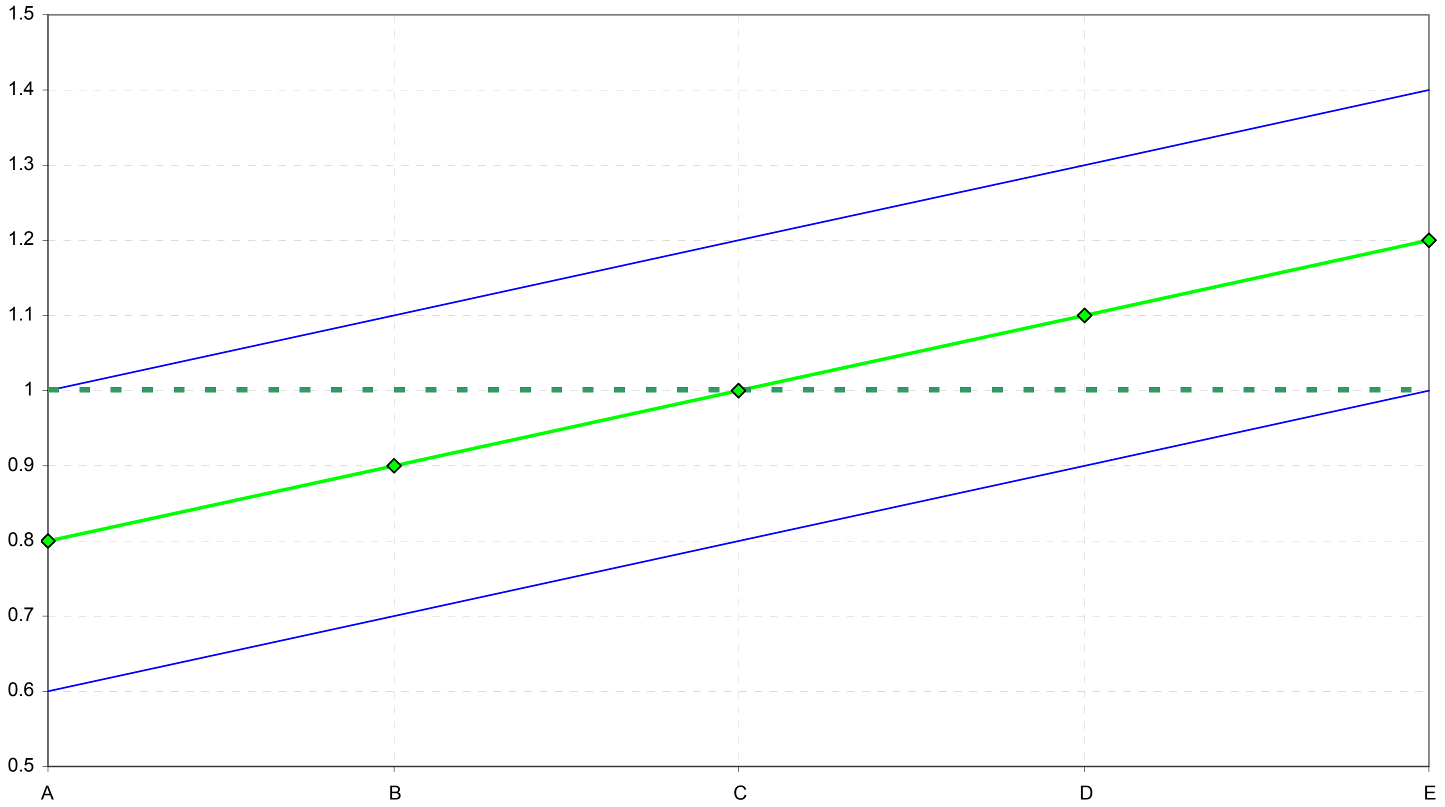
Agenda

- Testing the link function
- The Tweedie distribution
- Regression splines
- Reference models
- Aliasing/near-aliasing
- Combining models across claim types
- Restricted models
- Model validation
- Modeling elasticity / GNMs

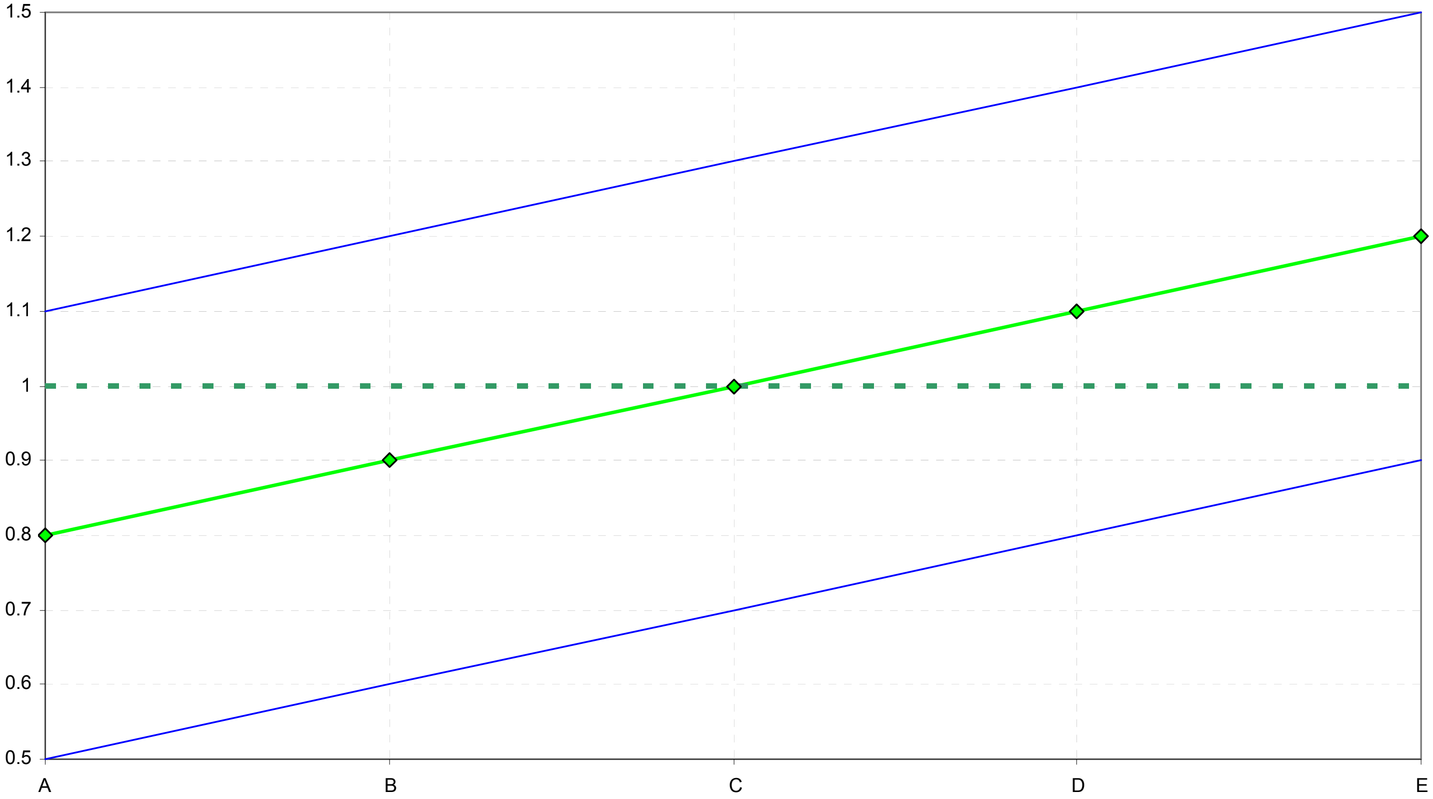
Reference models



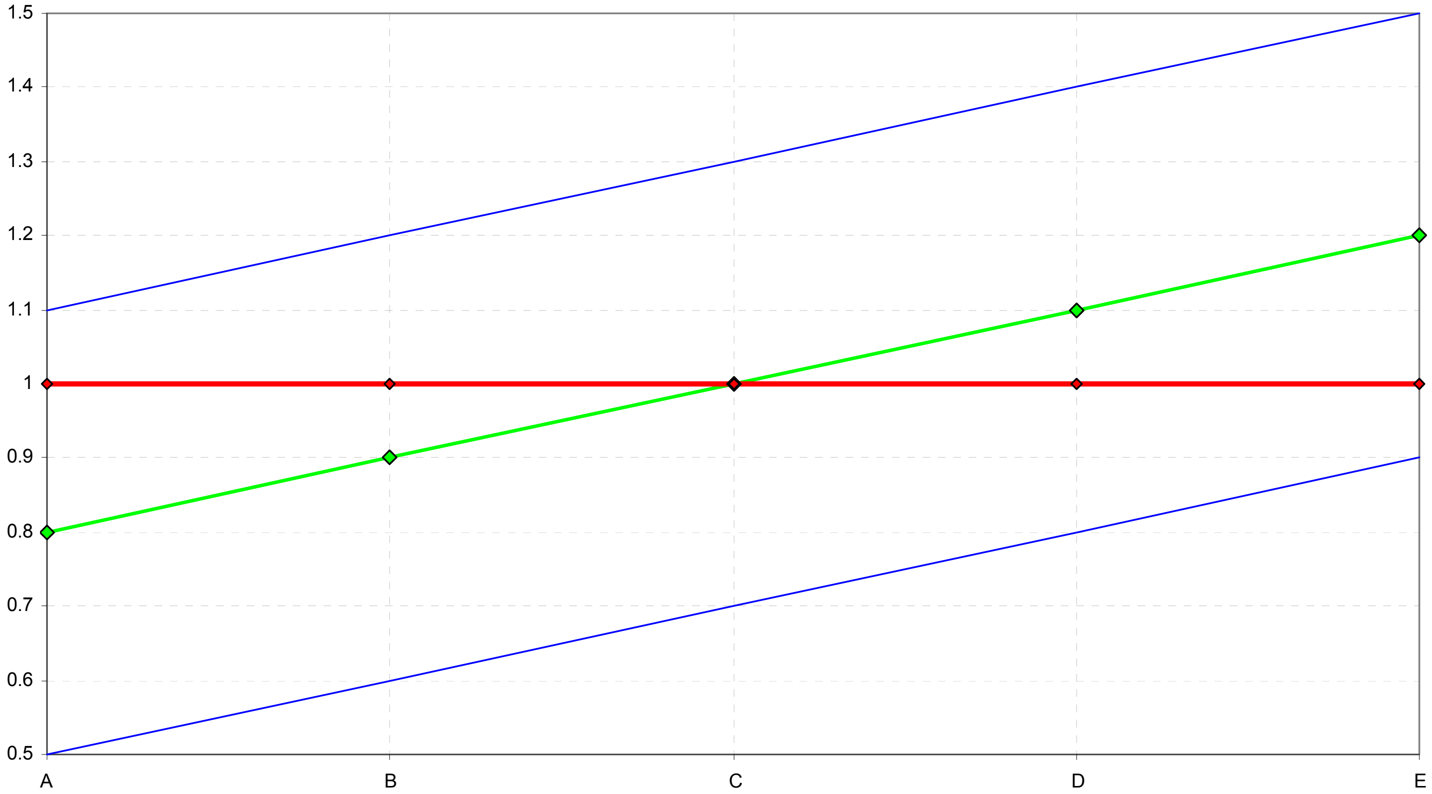
Reference models



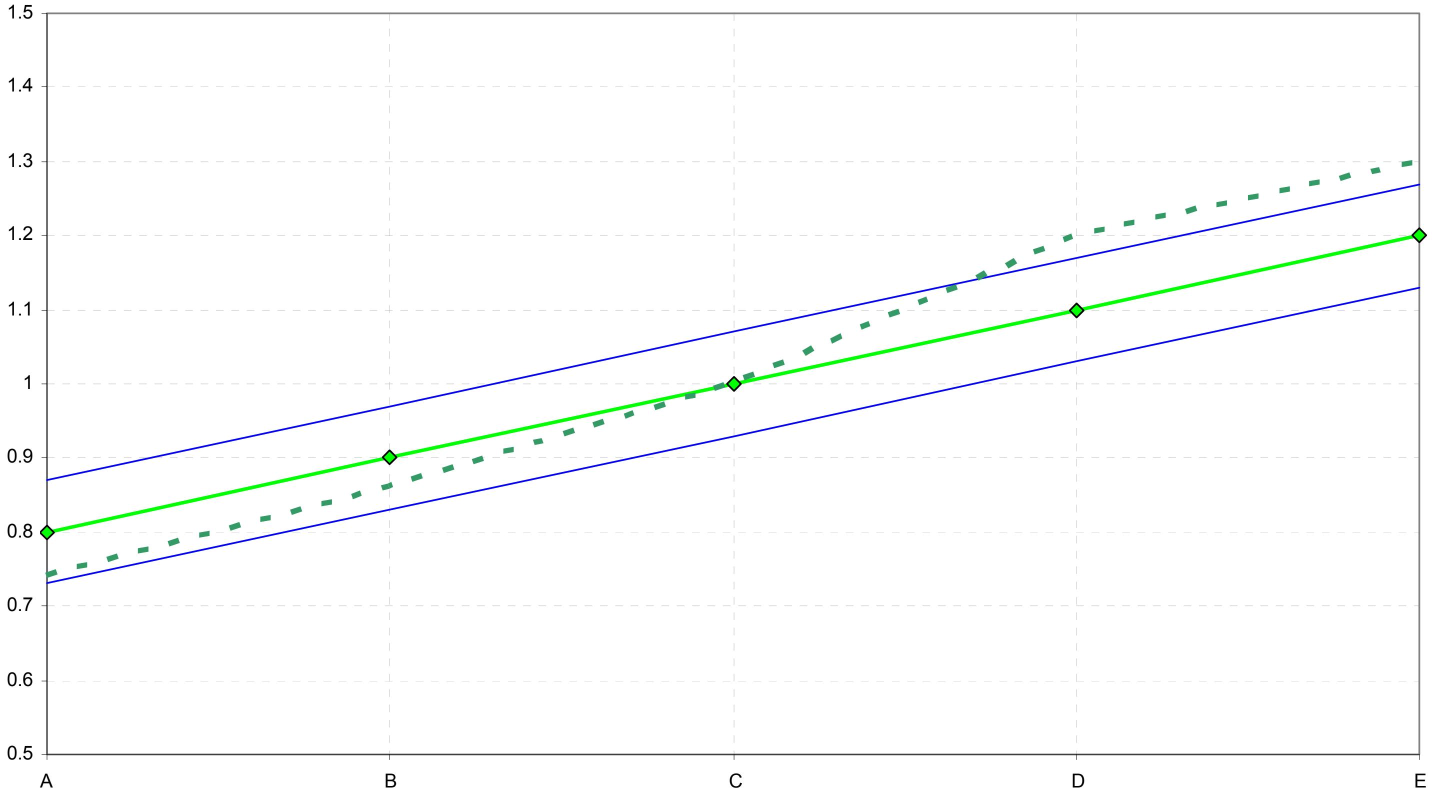
Reference models



Reference models

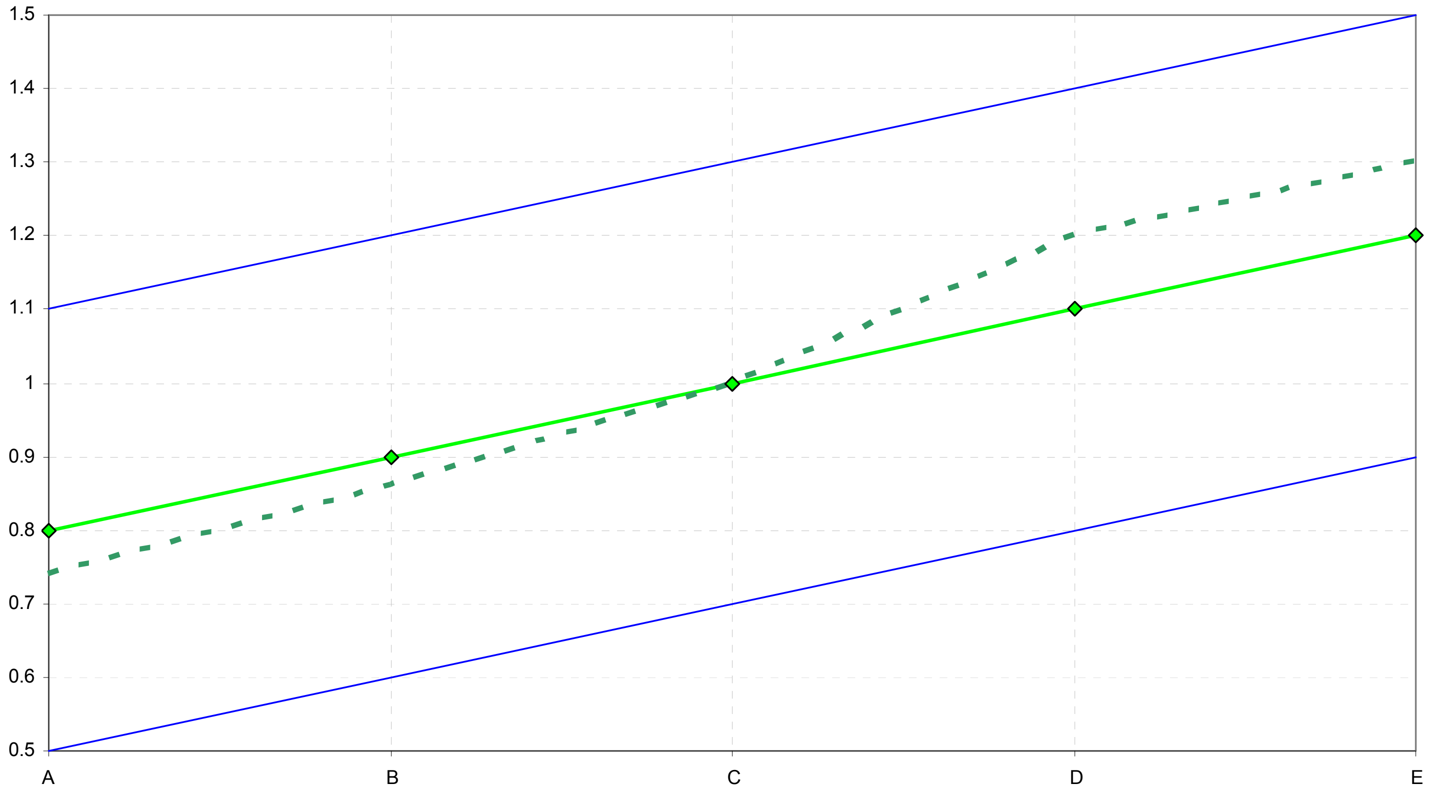


Reference models

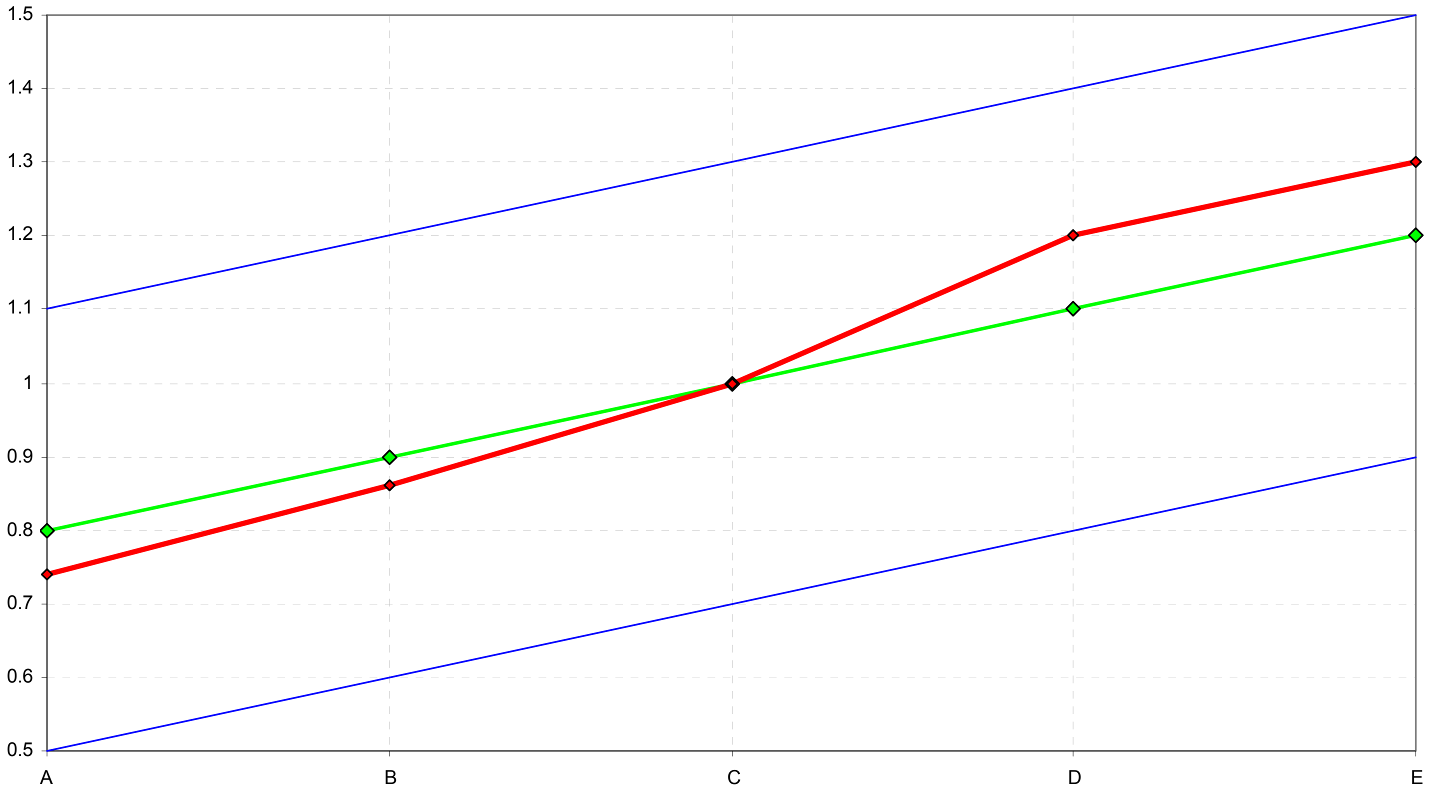


© 2008 EMB. All rights reserved.

Reference models



Reference models



Reference models - approach 1

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

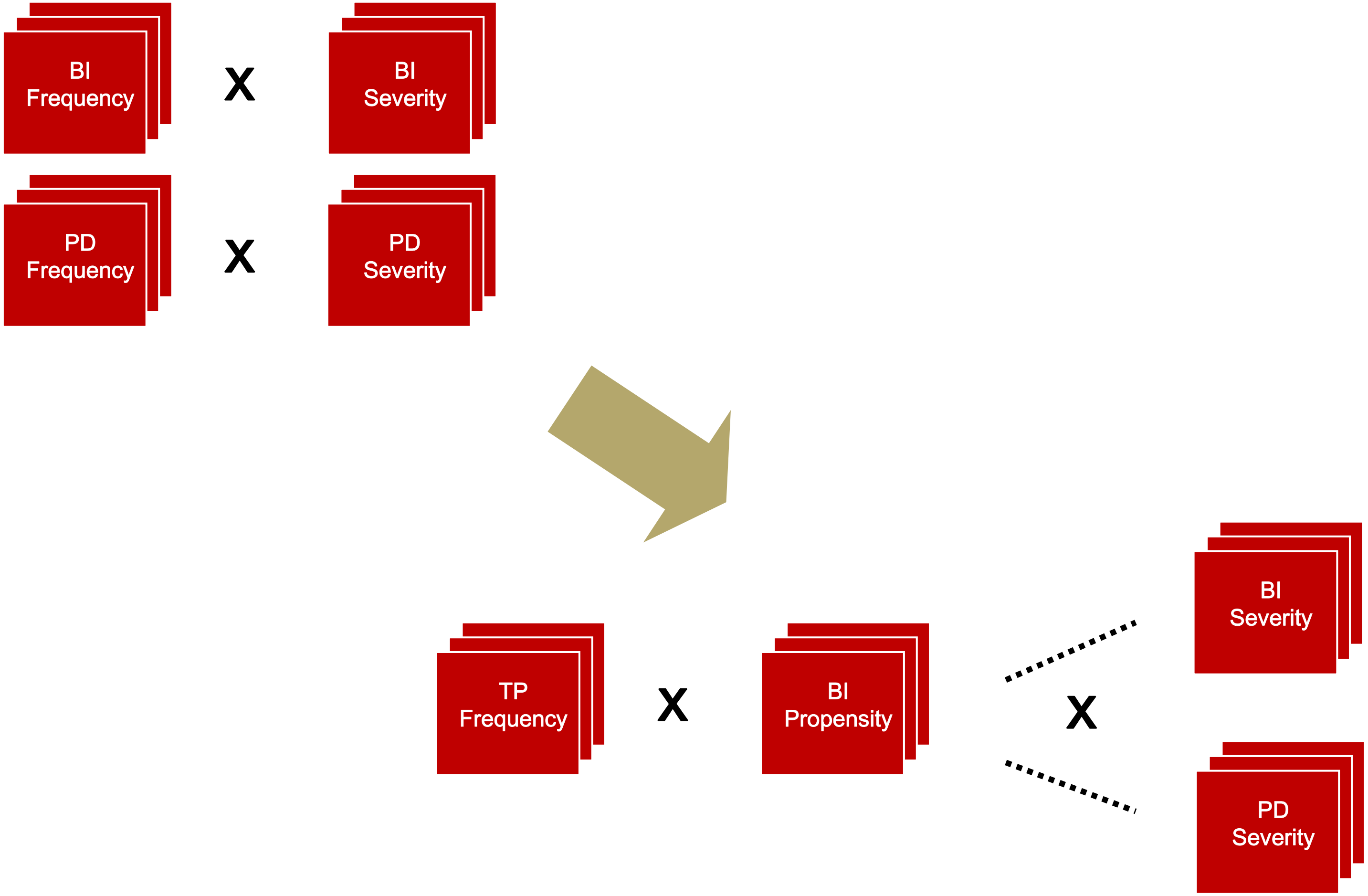
Offset term

When modeling BI

Set PD fitted values to be offset term

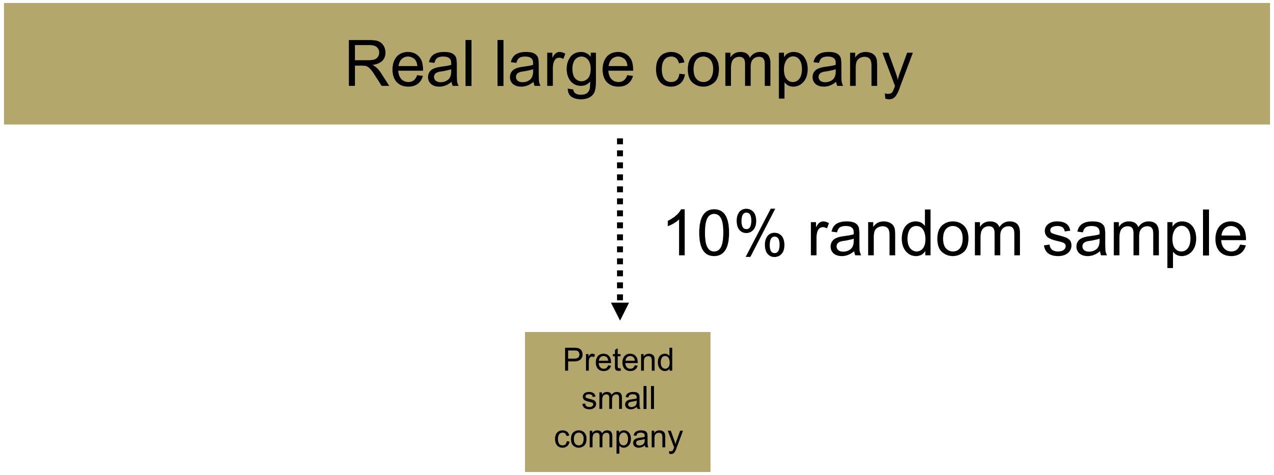
GLM will seek effects over and above assumed PD effect

Reference models - approach 2



Experiment

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



Real large company

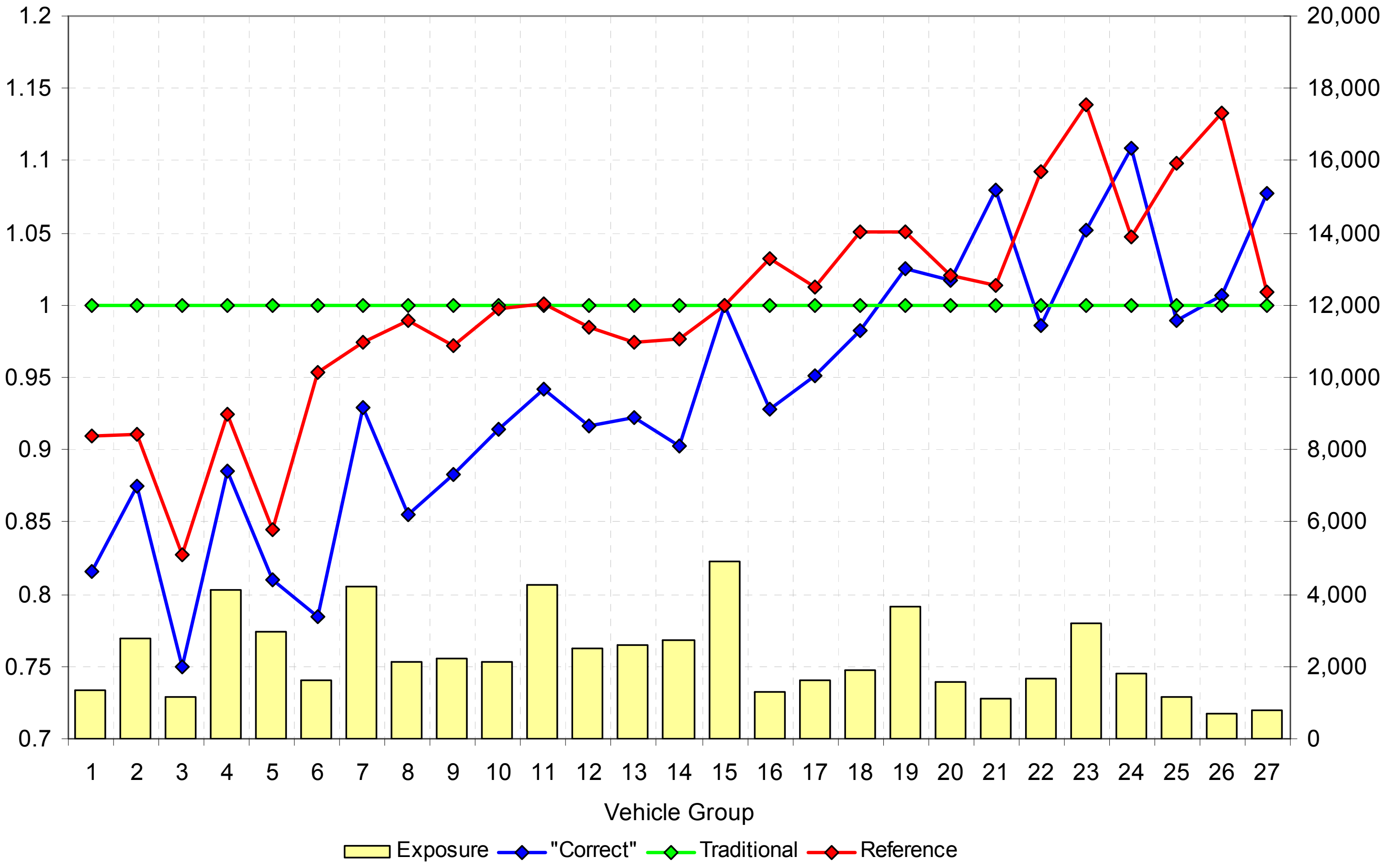
10% random sample

Pretend small company

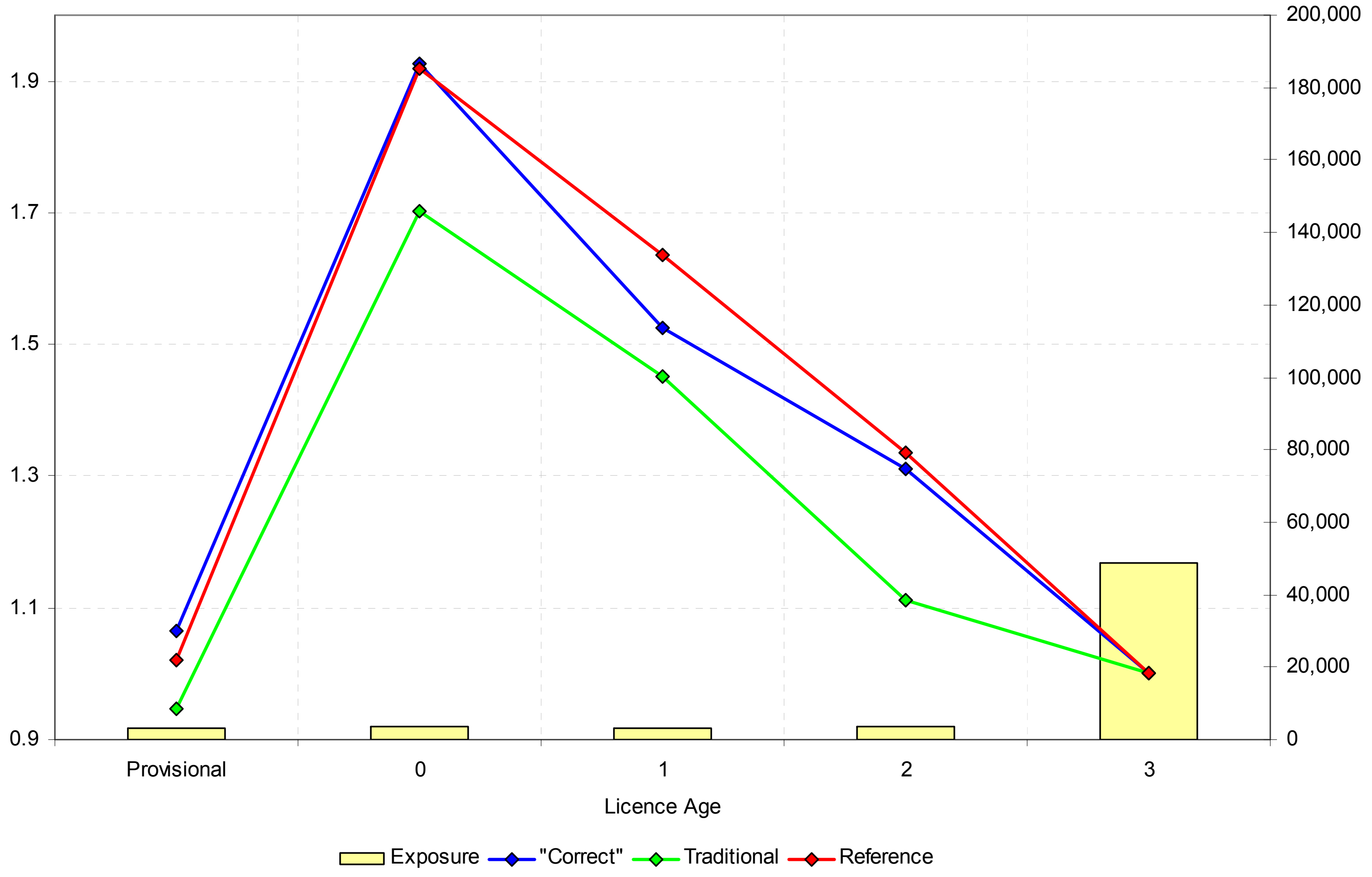
(2) Traditional GLM on BI claims on the "small company"

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

Example result



Example result



Agenda

- Testing the link function
- The Tweedie distribution
- Regression splines
- Reference models
- Aliasing/near-aliasing
- Combining models across claim types
- Restricted models
- Model validation
- Modeling elasticity / GNMs

Aliasing

- The removal of unwanted and unnecessary parameters
  - formally, linear dependencies in the design matrix
- **Intrinsic** aliasing
  - happens naturally because of the way the model is designed
- **Extrinsic** aliasing
  - happens "accidentally" because of some quirk in the data

Intrinsic aliasing

Consider model of form

$\mu_i =$ β_1 (base level)

$+ \beta_2$ if observation i is male

$+ \beta_3$ if observation i is female

$+ \beta_4$ if observation i is a small car

$+ \beta_5$ if observation i is a medium car

$+ \beta_6$ if observation i is a big car

Intrinsic aliasing - X.β

$$\left(\begin{array}{cccccc} \text{Base} & \text{Male} & \text{Female} & \text{Small} & \text{Med} & \text{Large} \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ & \multicolumn{5}{c}{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots} \\ & \multicolumn{5}{c}{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots} \end{array}\right) . \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{pmatrix}$$

Intrinsic aliasing

Consider model of form

$\mu_i =$ β_1 (base level)

~~$+ \beta_2$ if observation i is male~~

$+ \beta_3$ if observation i is female

$+ \beta_4$ if observation i is a small car

~~$+ \beta_5$ if observation i is a medium car~~

$+ \beta_6$ if observation i is a big car

"Base levels"

Intrinsic aliasing - X.β

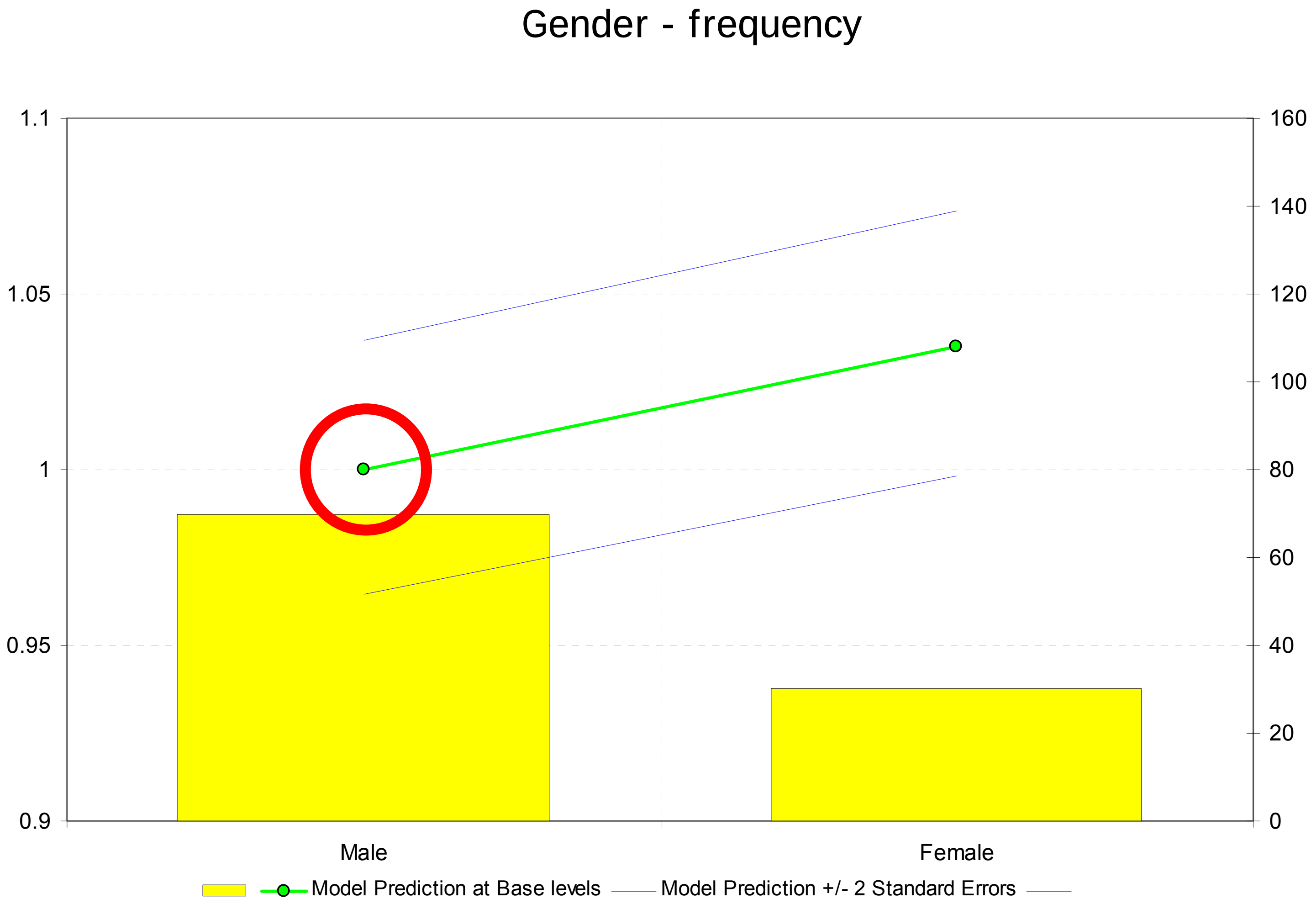
$$\begin{array}{cccccc} \text{Base} & \text{Male} & \text{Female} & \text{Small} & \text{Med} & \text{Large} \\ \end{array}$$

$$\left(\begin{array}{cccccc} 1 & \cancel{1} & 0 & 0 & \cancel{1} & 0 \\ 1 & \cancel{1} & 0 & 1 & \cancel{0} & 0 \\ 1 & \cancel{0} & 1 & 0 & \cancel{1} & 0 \\ 1 & \cancel{1} & 0 & 0 & \cancel{0} & 1 \\ 1 & \cancel{0} & 1 & 0 & \cancel{0} & 1 \\ 1 & \cancel{1} & 0 & 0 & \cancel{1} & 0 \\ & \multicolumn{5}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\ & \multicolumn{5}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \end{array}\right) . \begin{pmatrix} \beta_1 \\ \cancel{\beta_2} \\ \beta_3 \\ \beta_4 \\ \cancel{\beta_5} \\ \beta_6 \end{pmatrix}$$

Intrinsic aliasing - X.β

$$
\begin{array}{cccc}
\text{Base} & \text{Female} & \text{Small} & \text{Large} \\
\left(\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \\ \\ \end{array}\right. &
\begin{array}{c} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ \cdots\cdots \\ \cdots\cdots \end{array} &
\begin{array}{c} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ \cdots\cdots \\ \cdots\cdots \end{array} &
\left.\begin{array}{c} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ \\ \\ \end{array}\right)
\end{array}
.
\begin{pmatrix} \beta_1 \\ \\ \beta_3 \\ \beta_4 \\ \\ \beta_6 \end{pmatrix}
$$

Example intrinsic aliasing

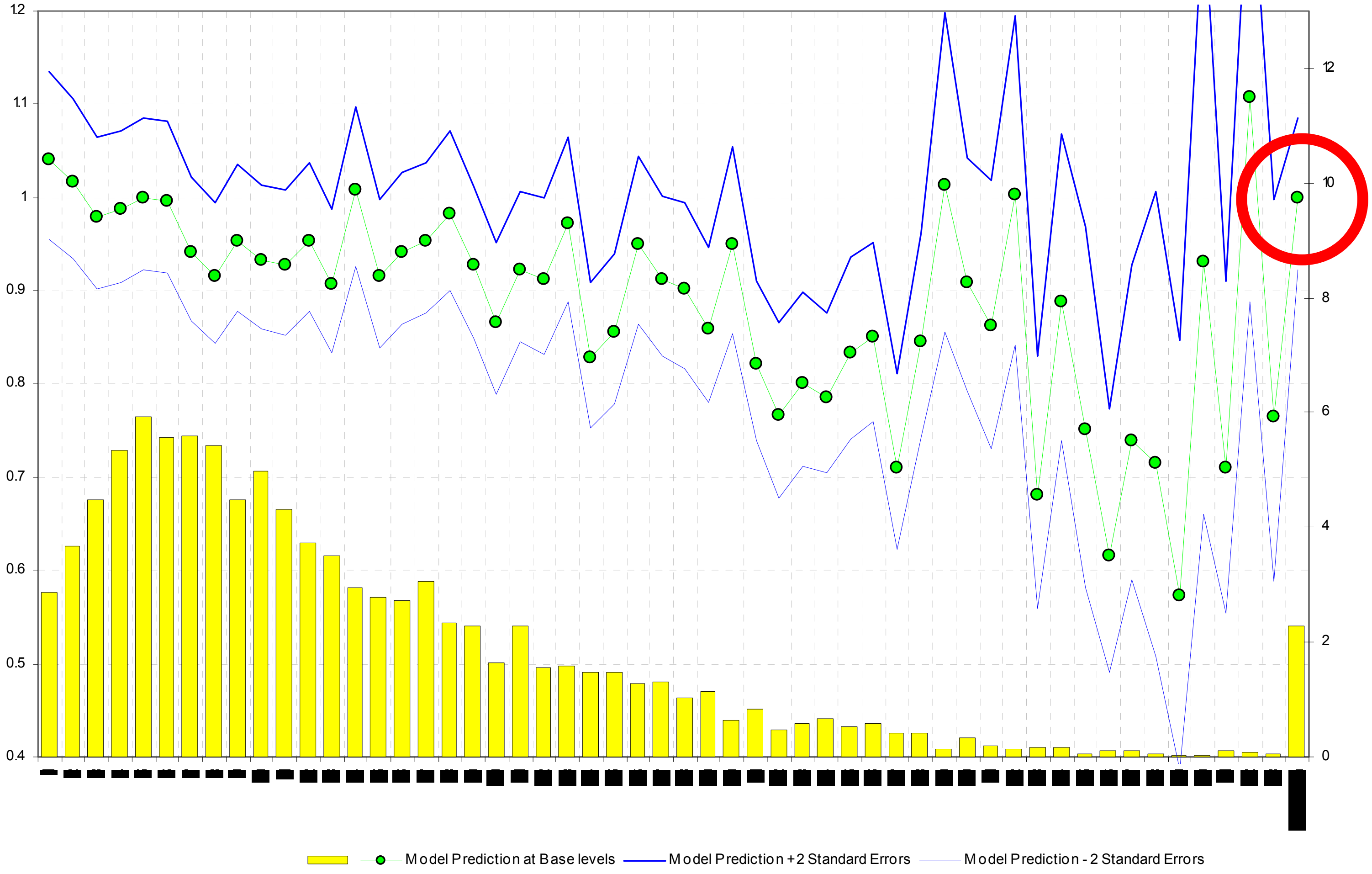


Extrinsic aliasing

Exposure

Density→ ↓ Wealth	Very urban	Urban	Rural Intrinsic aliasing	Very rural	Unknown Extrinsic aliasing
Very rich	12,123	14,673	25,353	22,342	0
Rich Intrinsic aliasing	32,343	36,945	40,236	32,234	0
Poor	29,454	28,343	33,324	26,954	0
Very poor	14,343	12,456	18,343	9,934	0
Unknown	0	0	0	0	1,235

Example extrinsic aliasing



"Near" aliasing

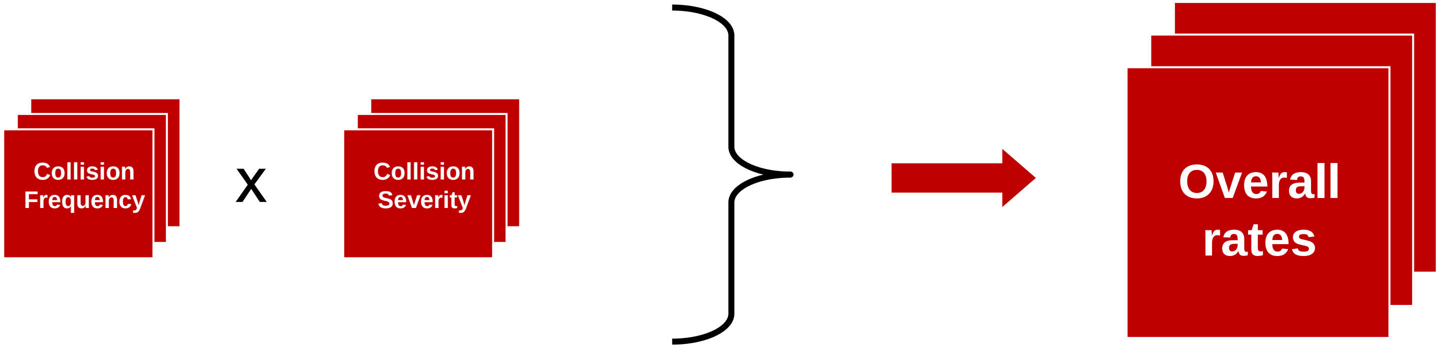
Exposure

Density→ ↓ Wealth	Very urban	Urban	~~Rural~~ Intrinsic aliasing	Very rural	Unknown
Very rich	12,123	14,673	25,353	22,342	0
~~Rich~~ Intrinsic aliasing	32,343	36,945	40,236	32,234	0
Poor	29,454	28,343	33,324	26,954	0
Very poor	14,343	12,456	18,343	9,934	0
Unknown	0	0	0	22	1,235

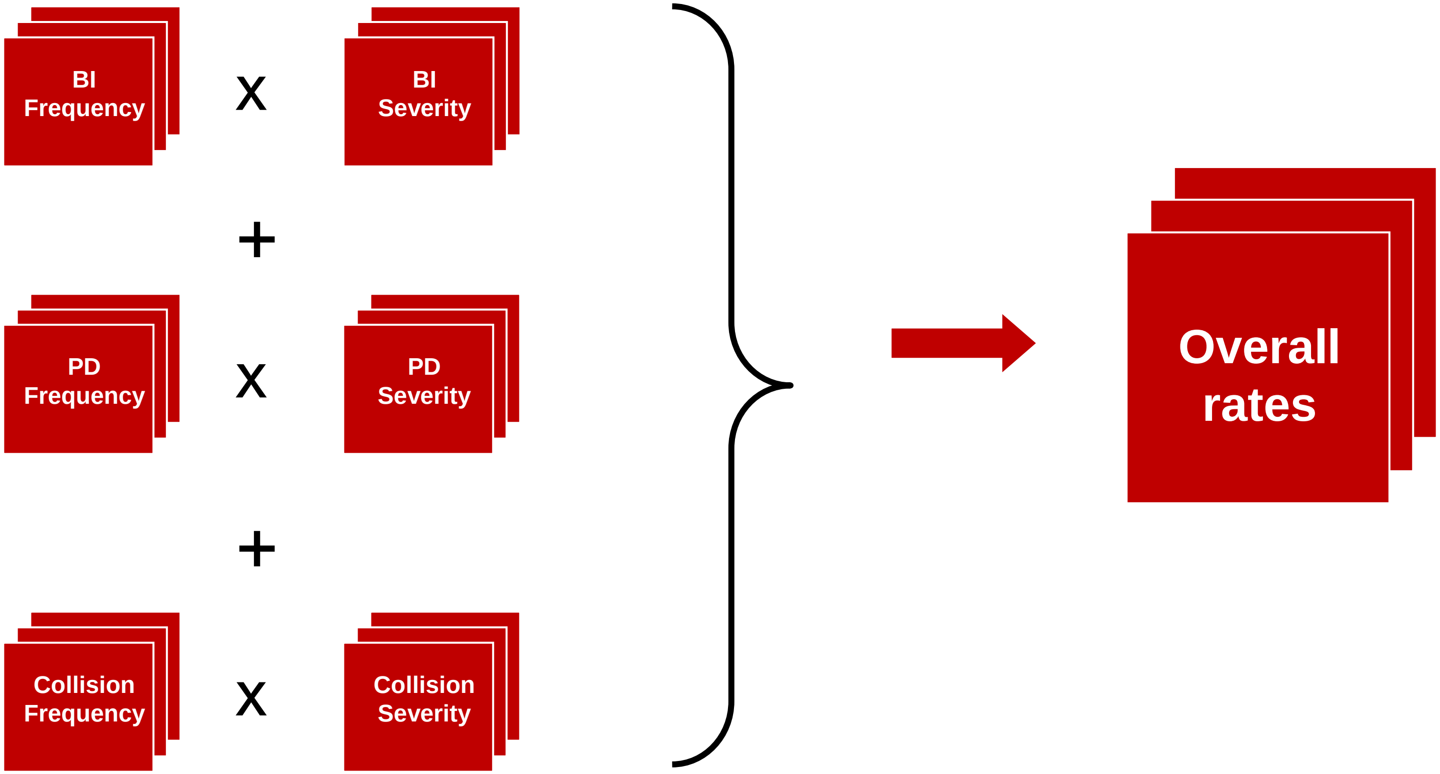
Agenda

- Testing the link function
- The Tweedie distribution
- Regression splines
- Reference models
- Aliasing/near-aliasing
- Combining models across claim types
- Restricted models
- Model validation
- Modeling elasticity / GNMs

Combining models



Combining models



Combining models

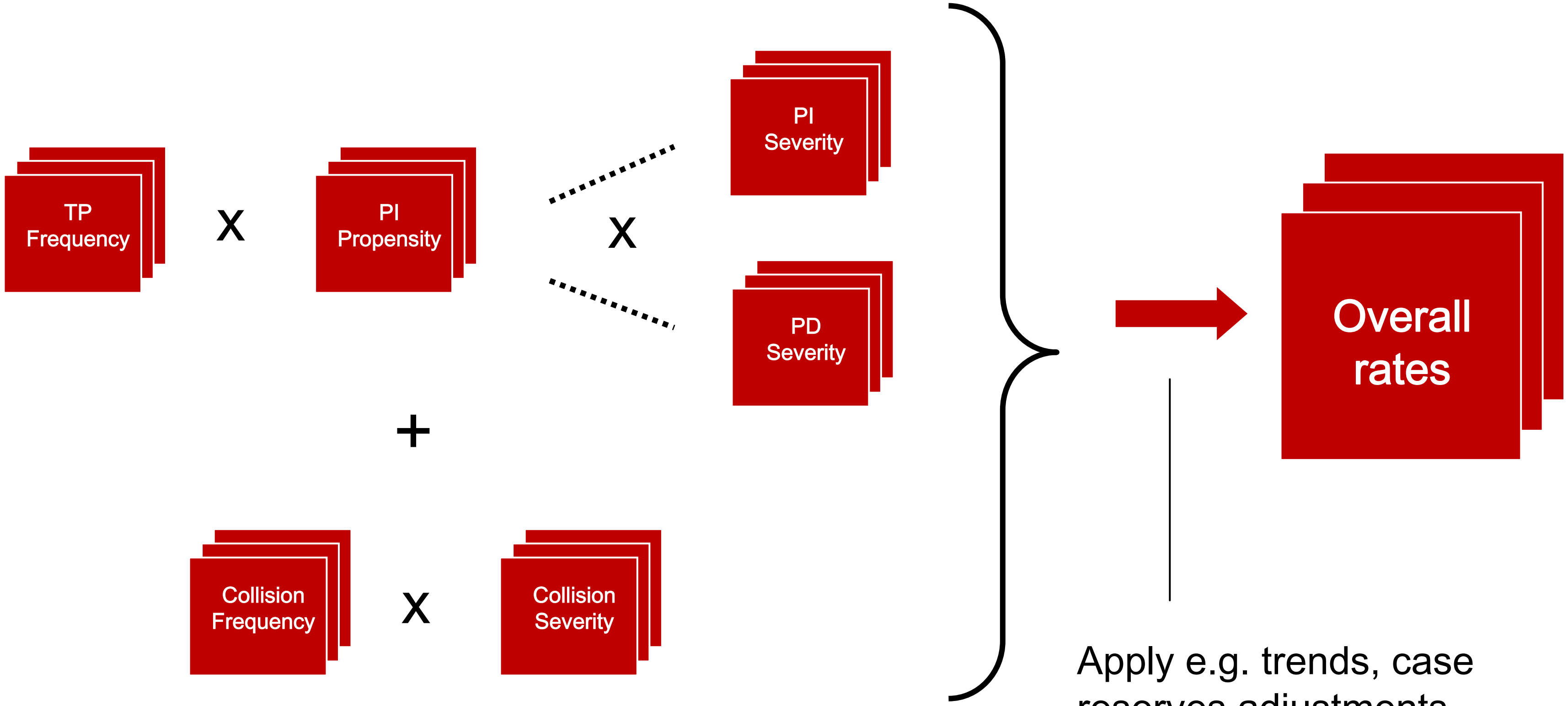
- Take models
- Take relevant mix of business
  - eg current in force policies
- For each record calculate expected frequencies and severities according to the models
- For each record, calculate expected total cost of claims "C"
- Fit a GLM to "C" using all available factors

Combining models

	PD Freq	PD Sev	PI Freq	PI Sev
Base	10%	$1500	2%	$5000
Male	1	1	1	1
Female	0.9	0.85	0.95	0.88
Small	1.1	0.8	1.15	0.7
Medium	1	1	1	1
Large	0.9	1.3	0.95	1.25

Policy	Gender	Car	PD F	PD S	PI F	PI S	Cost
…	…	…	…	…	…	…	…
762374	Male	Large	9%	$1,950	1.9%	$6,250	294.25
762375	Male	Small	11%	$1,200	2.3%	$3,500	212.50
762376	Female	Medium	9%	$1,275	1.9%	$4,400	198.35
762377	Male	Medium	10%	$1,500	2.0%	$5,000	250.00
…	…	…	…	…	…	…	…

Combining models



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

Agenda

- Testing the link function
- The Tweedie distribution
- Regression splines
- Reference models
- Aliasing/near-aliasing
- Combining models across claim types
- Restricted models
- Model validation
- Modeling elasticity / GNMs

Formularization of GLMs

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

Offset

Formularization of GLMs

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

$$\begin{array}{cccc} \text{Base} & \text{Female} & \text{Small} & \text{Large} \\ \left(\begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \cdots \\ \cdots \end{array}\right. & \begin{array}{c} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ \cdots \\ \cdots \end{array} & \begin{array}{c} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ \cdots \\ \cdots \end{array} & \left.\begin{array}{c} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ \cdots \\ \cdots \end{array}\right) \end{array} . \begin{pmatrix} \beta_1 \\ \\ \beta_3 \\ \beta_4 \\ \\ \beta_6 \end{pmatrix}$$

Formularization of GLMs

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

$$\begin{array}{cccc} \text{Base} & \text{Female} & \text{Small} & \text{Large} \\ \end{array}$$

$$\left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{array}\right) . \left(\begin{array}{c} \beta_1 \\ \\ \beta_3 \\ \beta_4 \\ \\ \beta_6 \end{array}\right)$$

Formularization of GLMs

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

$$\left(\begin{array}{cccc} \text{Base} & \text{Female} & \text{Small} & \text{Large} \\ 1 & & 0 & 0 \\ 1 & & 1 & 0 \\ 1 & & 0 & 0 \\ 1 & & 0 & 1 \\ 1 & & 0 & 1 \\ 1 & & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{array}\right) . \begin{pmatrix} \beta_1 \\ \\ \\ \beta_4 \\ \beta_6 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0.1 \\ 0 \\ 0.1 \\ 0 \\ \cdots \\ \cdots \end{pmatrix}$$

Formularization of GLMs

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

$$\begin{array}{c} \begin{matrix} \text{Base} & \text{Female} & \text{Small} & \text{Large} \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} \end{array} . \begin{pmatrix} \beta_1 \\ \\ 0.1 \\ \beta_4 \\ \\ \beta_6 \end{pmatrix}$$

Offset example

No Claims Discount

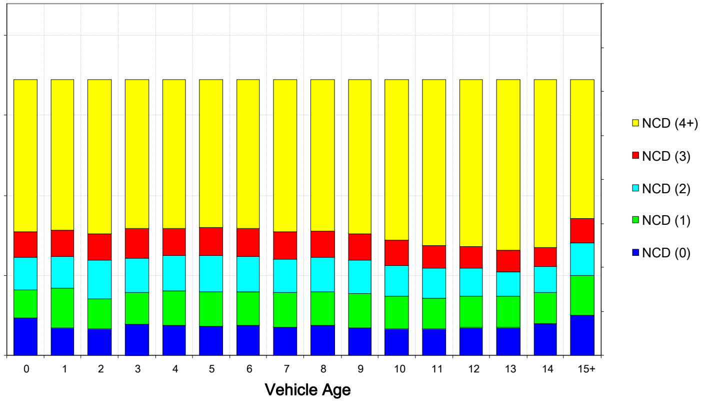
Cramer’s V measures exposure correlation

Factor (#Levels)	Gender	Rating Area	Vehicle Category	Age	No Claims Discount	Driving Restriction	Vehicle Age	LossYear
Gender	-	-	-	-	-	-	-	-
Rating Area	0.017	-	-	-	-	-	-	-
Vehicle Category	0.297	0.017	-	-	-	-	-	-
Age	0.182	0.035	0.087	-	-	-	-	-
No Claims Discount	0.126	0.021	0.139	0.253	-	-	-	-
Driving Restriction	0.076	0.034	0.088	0.224	0.112	-	-	-
Vehicle Age	0.044	0.016	0.068	0.025	0.025	0.041	-	-
LossYear	0.006	0.014	0.064	0.126	0.124	0.055	0.049	-

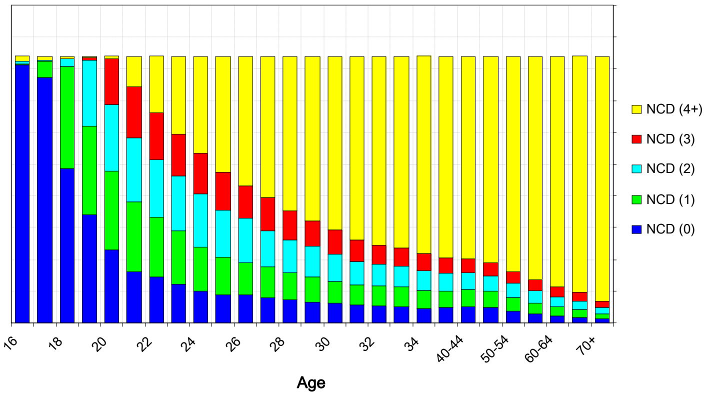
High

Low

Vehicle Age x NCD



Age x NCD



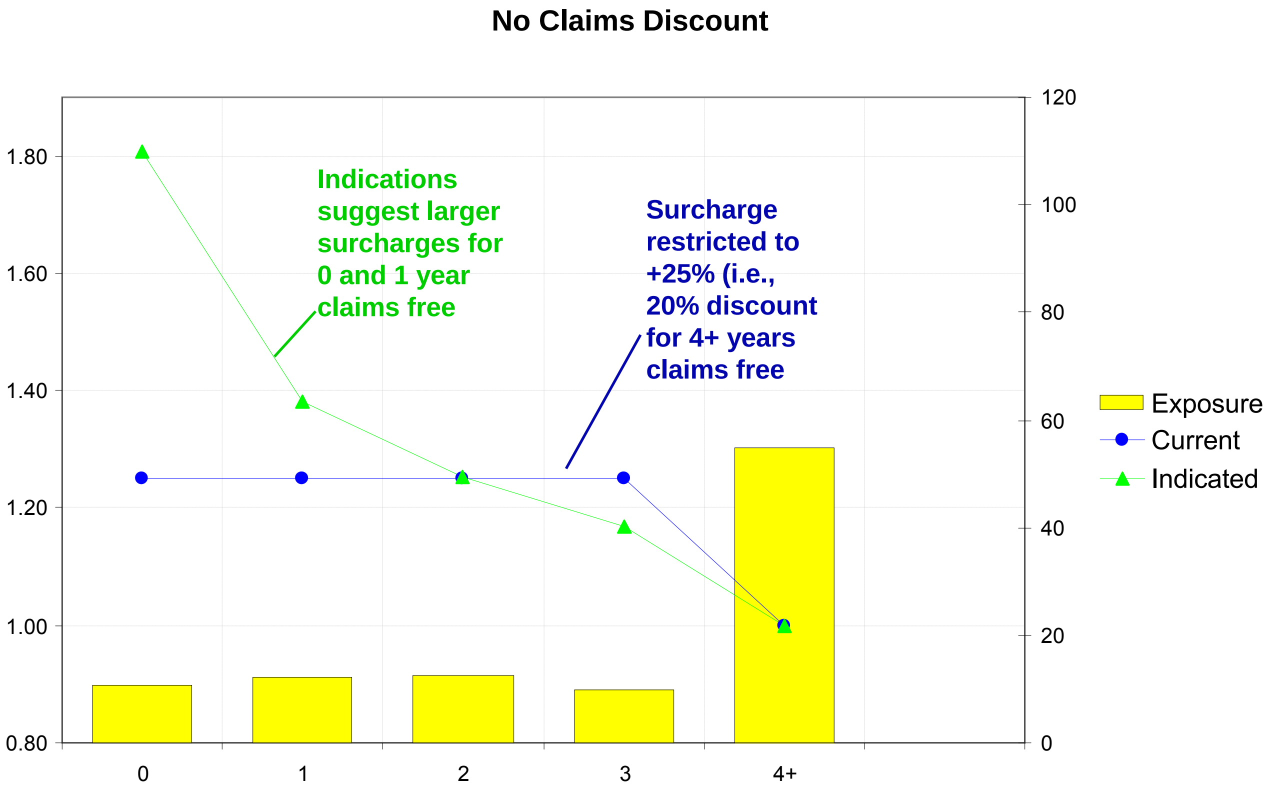
0.025 implies low correlation

0.253 implies high correlation

Offset Example

No Claims Discount

Company decides to maintain current NCD relativities



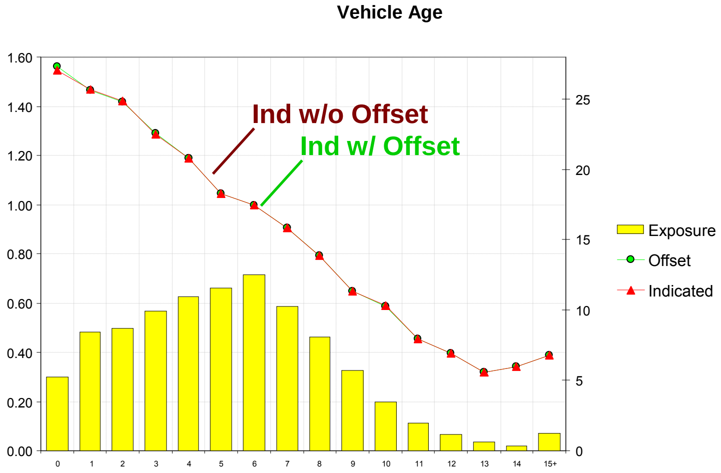
Indications suggest larger surcharges for 0 and 1 year claims free

Surcharge restricted to +25% (i.e., 20% discount for 4+ years claims free

Impact of offsetting on indications of other variables depends on exposure correlation with NCD

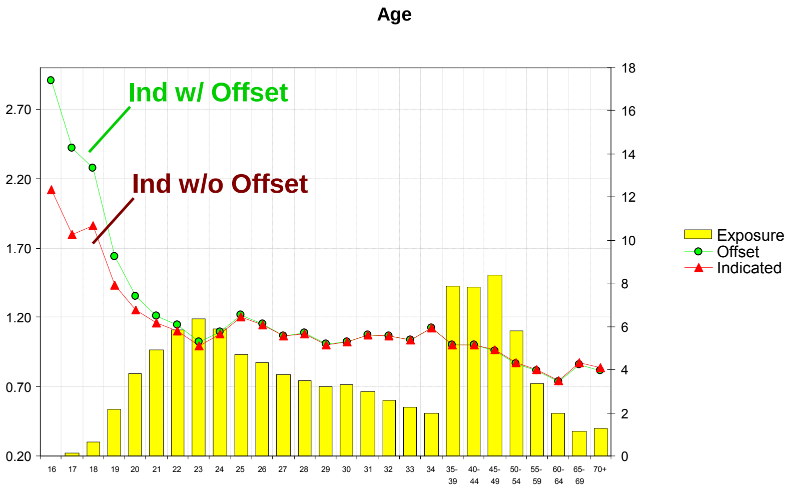
Offset Example

No Claims Discount



Cramers V=.025 (Low)

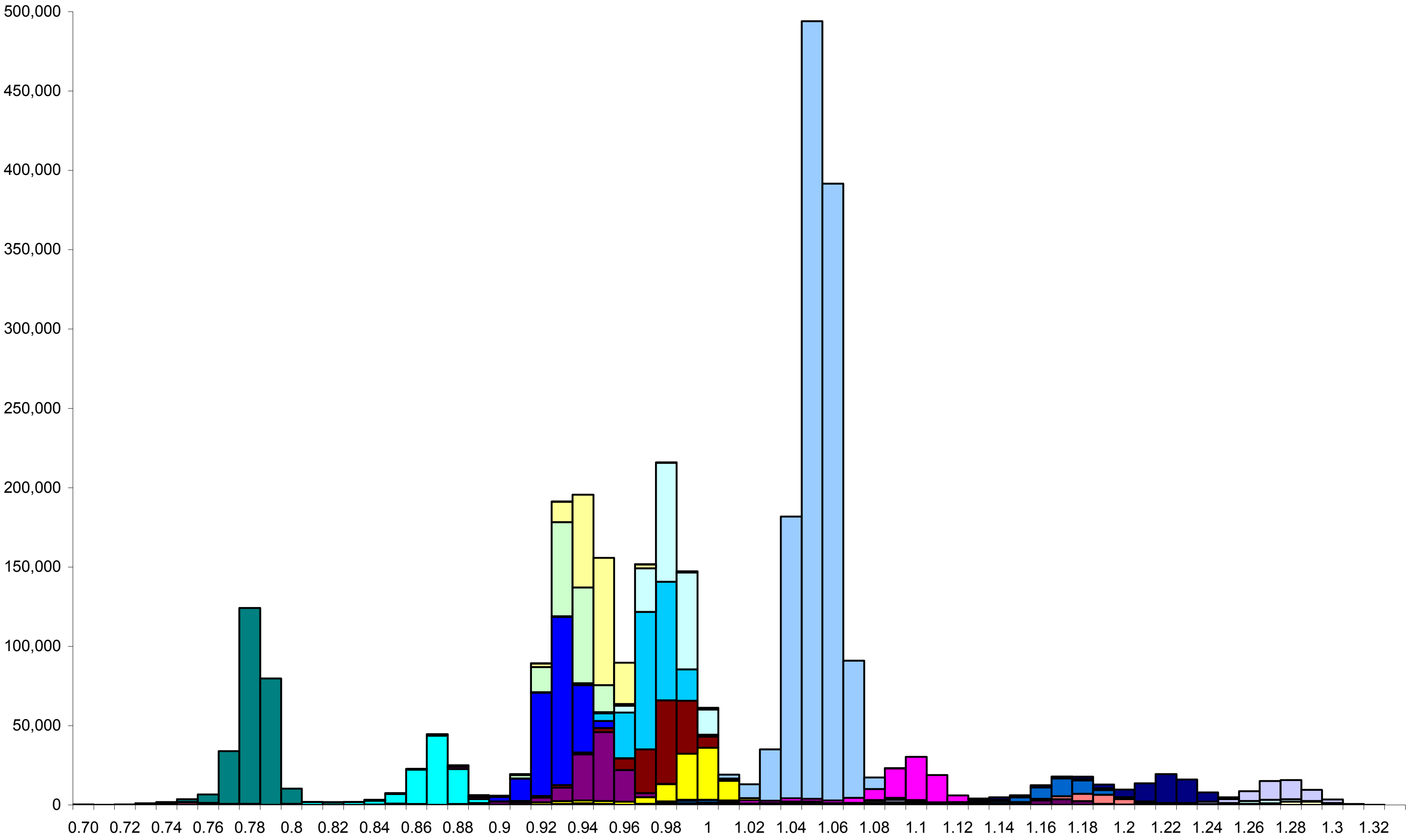
No material difference between model with and without the offset for “NCD”



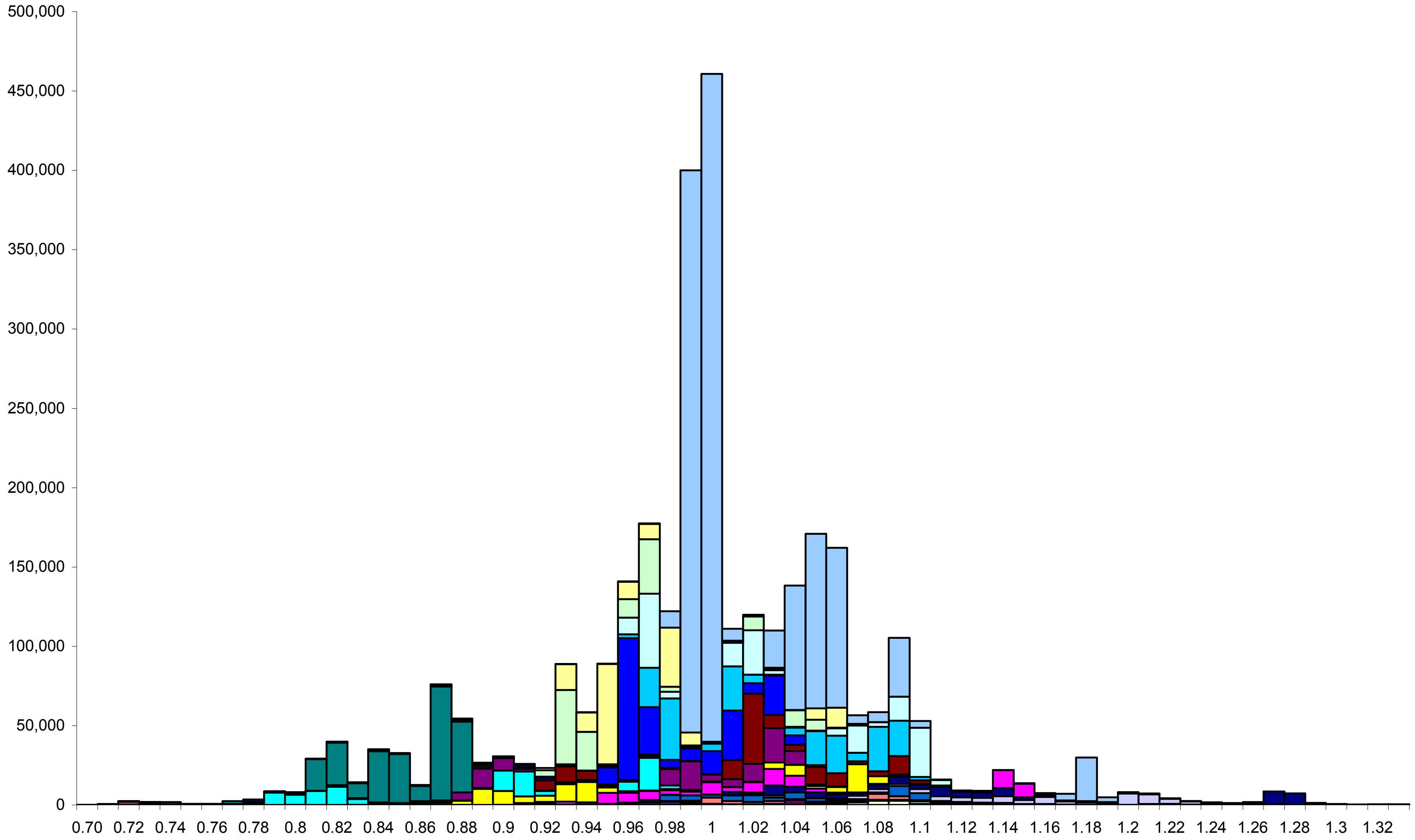
Cramers V=.253 (High)

Youthful relativities increased to account for premiums lost by dampening surcharges for policies with less than 4 years clean

Checking the effectiveness of compensating factors



Checking the effectiveness of compensating factors



Using restrictions

- Apply at risk premium (model combining) stage
- Other factors will compensate - use to restrict the multivariate effect, not the overall effect

	Desirable Subsidy	**Undesirable Subsidy**
Example	Sr. Mgmt wants subsidy to attract drivers 65+	Regulators force subsidy of drivers 65+
Result of Offset	Correlated factors will adjust to make up for the difference. For example, territories with retirement communities will increase	
Recommendation	Do Not Offset	Offset

Agenda

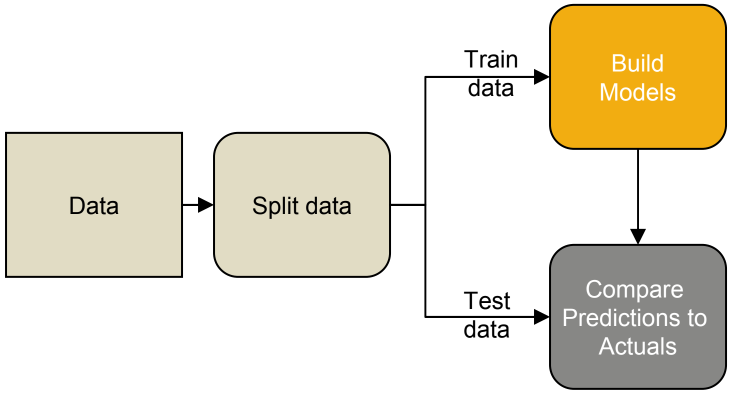
- Testing the link function
- The Tweedie distribution
- Regression splines
- Reference models
- Aliasing/near-aliasing
- Combining models across claim types
- Restricted models
- Model validation
- Modeling elasticity / GNMs

Model validation: holdout samples

Hold-out samples are effective at validating model

- Determine estimates based on part of dataset
- Uses estimates to predict other part of dataset

Test/Training

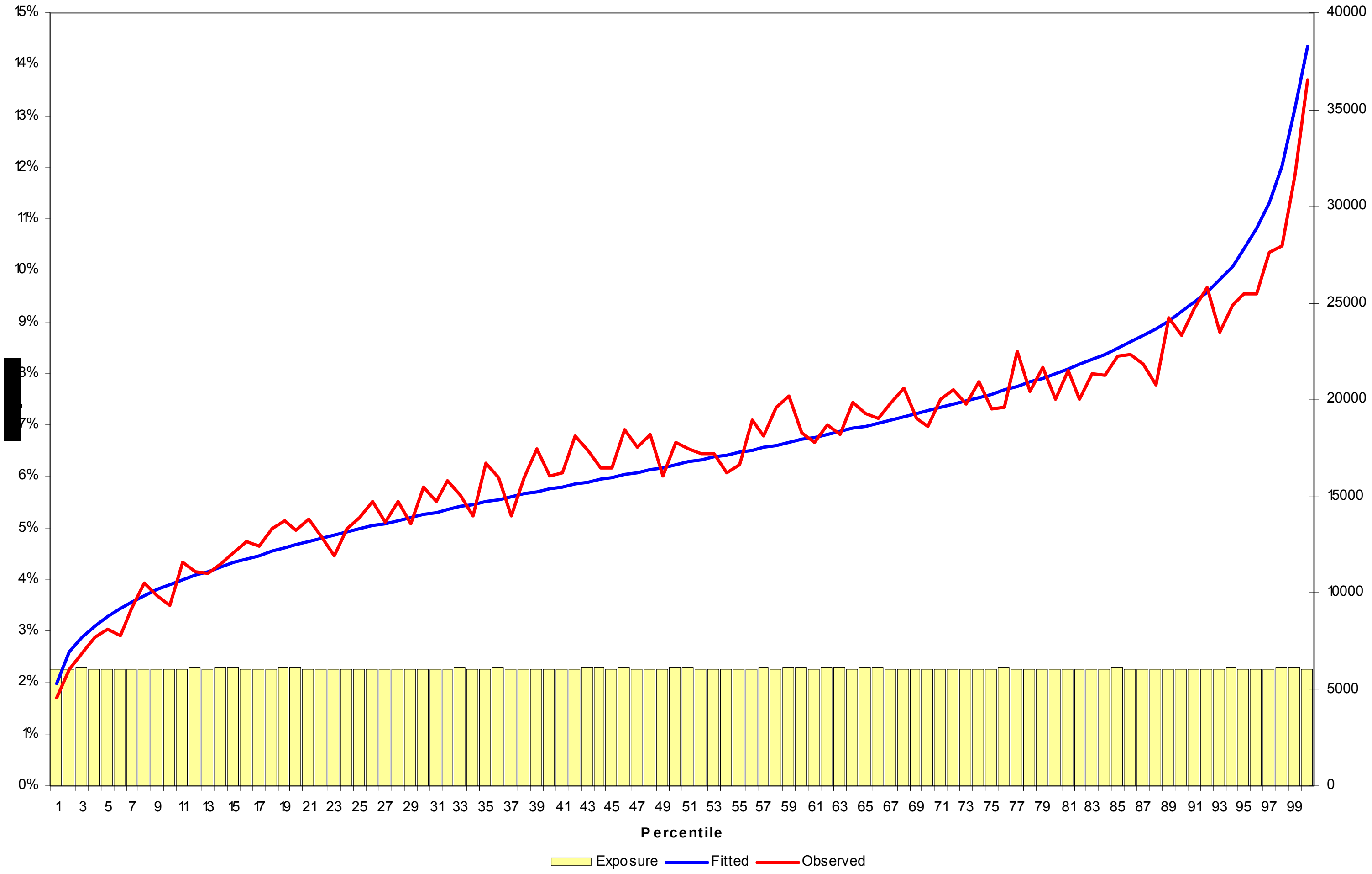


Larger companies may consider 3 splits

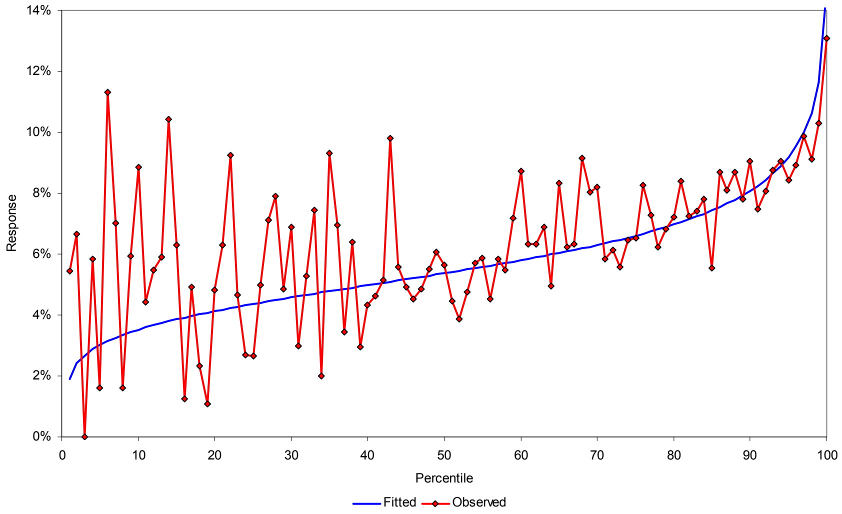
1. Build models
2. Fit parameters
3. Validate models/parameters

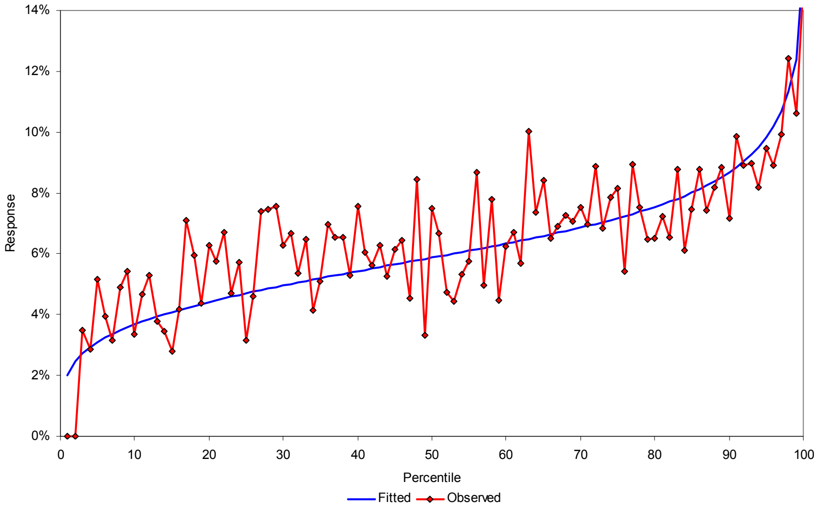
Predictions should be close to actuals for populated cells

Model validation



Model validation

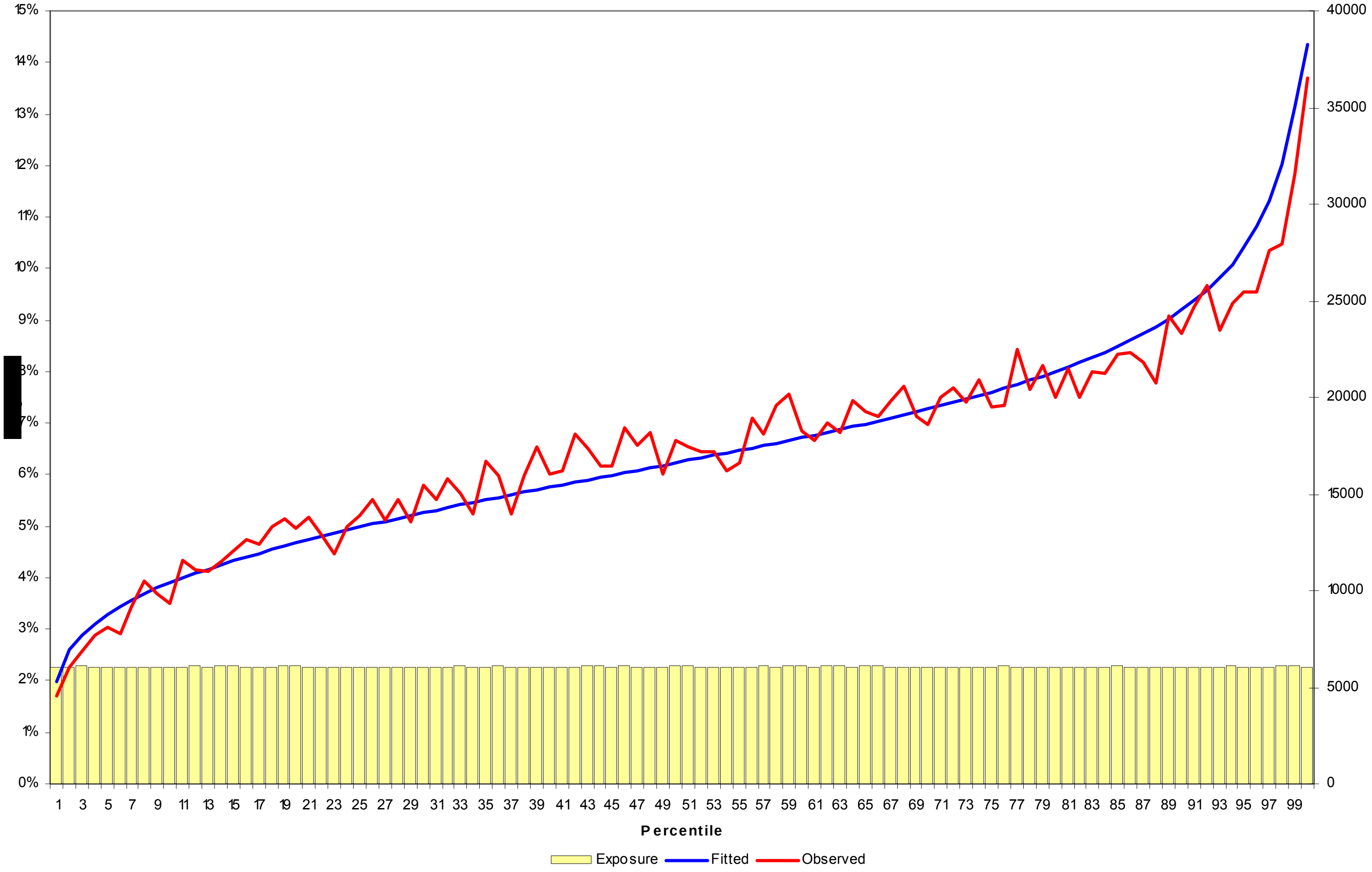




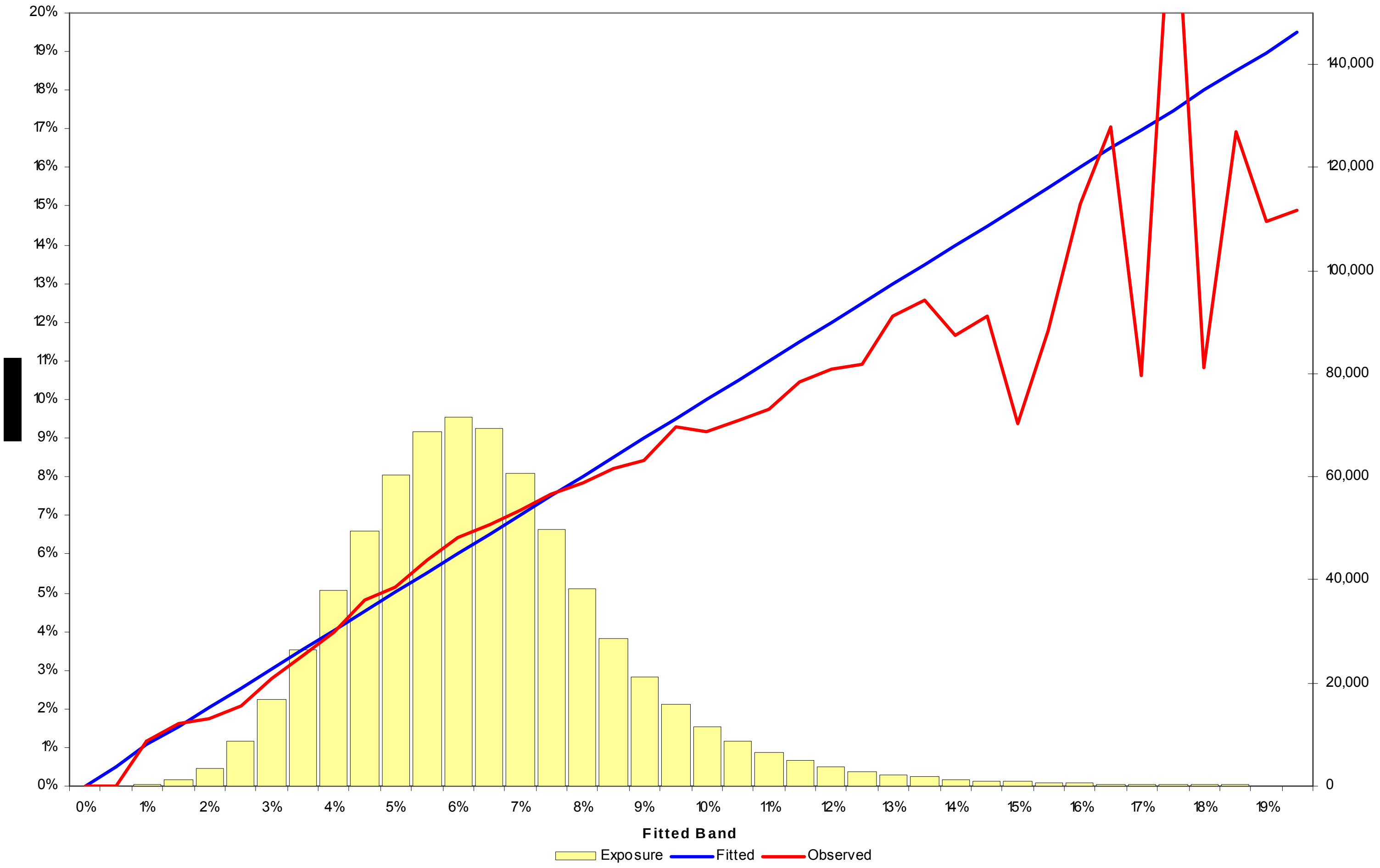
- Auto own damage frequency
- Many rating factors
- Just a few interactions
- For under 30s segment, model is not predictive in the future

- Auto own damage frequency
- Many rating factors
- Many interactions
- Model can predict well in the future, even for small segments

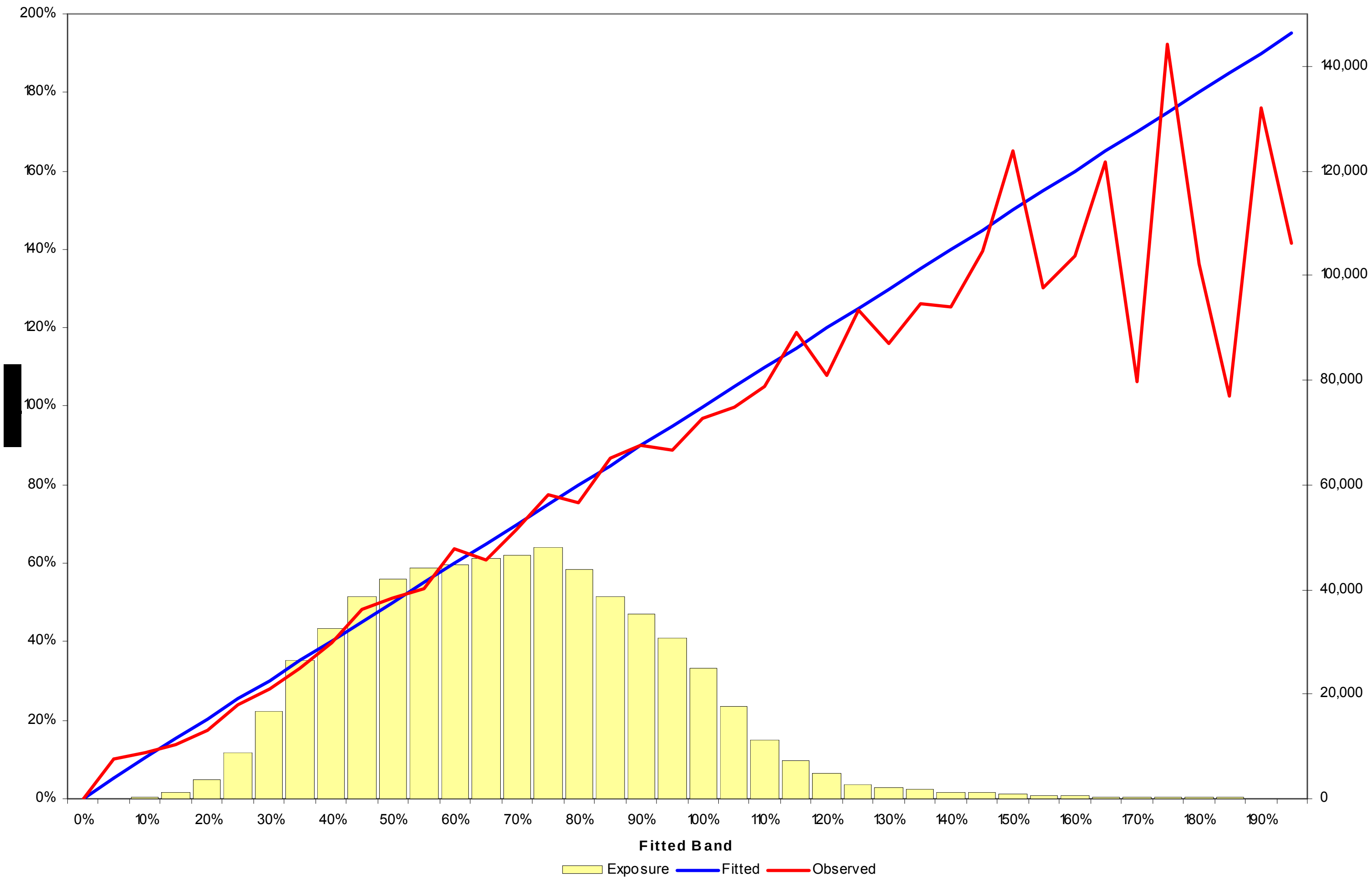
Model validation



Model validation
Test of statistical validity



Model validation
Demonstration of financial materiality

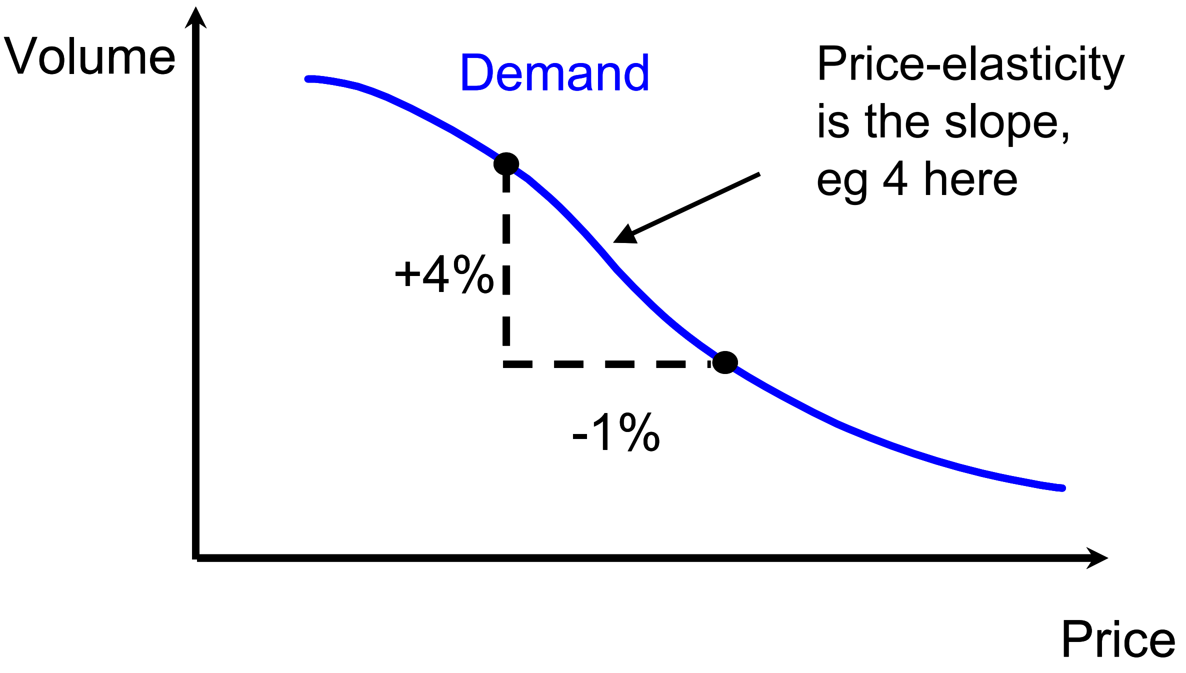


Agenda

- Testing the link function
- The Tweedie distribution
- Regression splines
- Reference models
- Aliasing/near-aliasing
- Combining models across claim types
- Restricted models
- Model validation
- Modeling elasticity / GNMs

Demand models

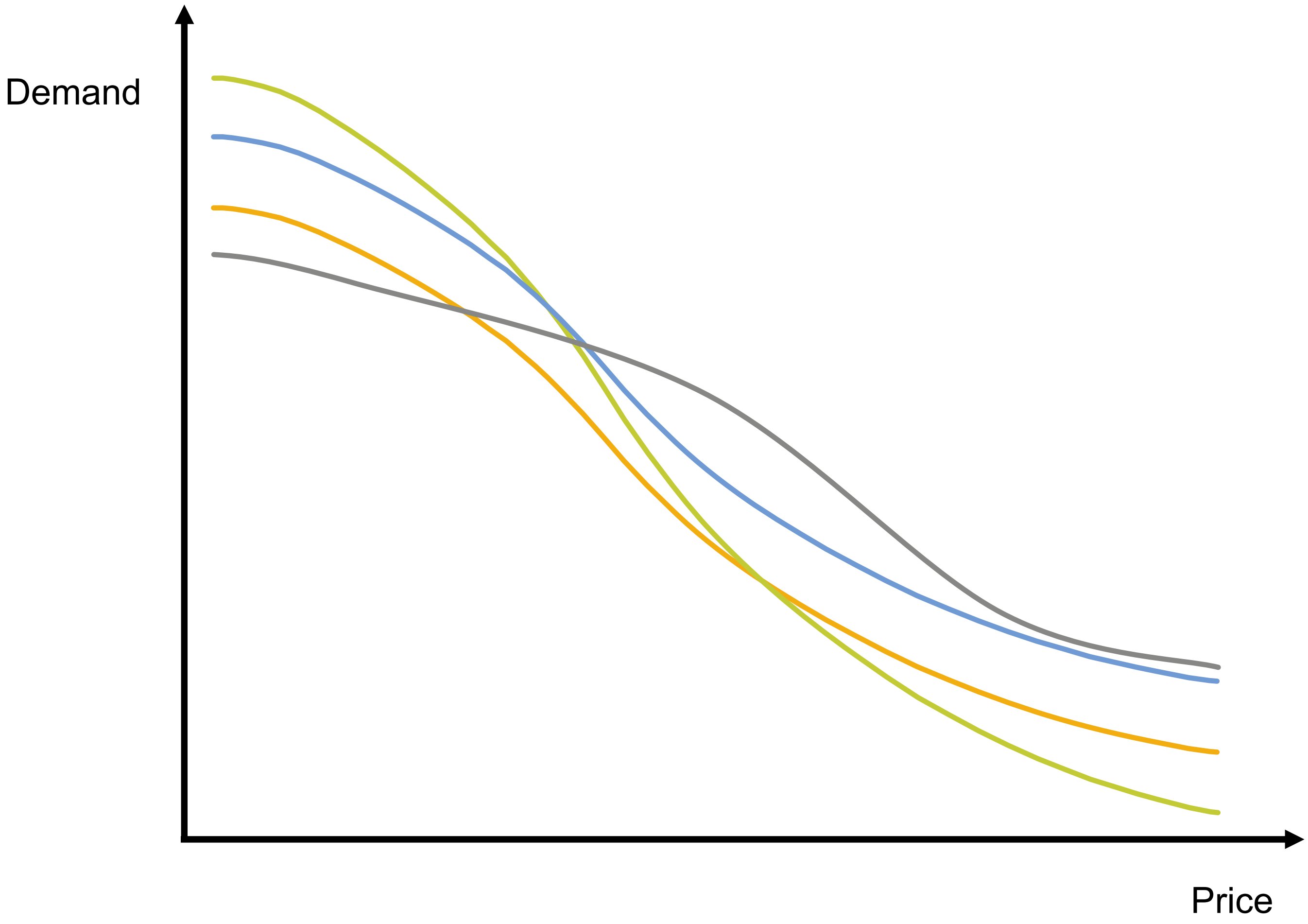
- Demand models are a key ingredient to price optimization
- Elasticity is (minus) the slope of the curve



Price-elasticity is the slope, eg 4 here

© 2008 EMB. All rights reserved.

Price demand elasticity



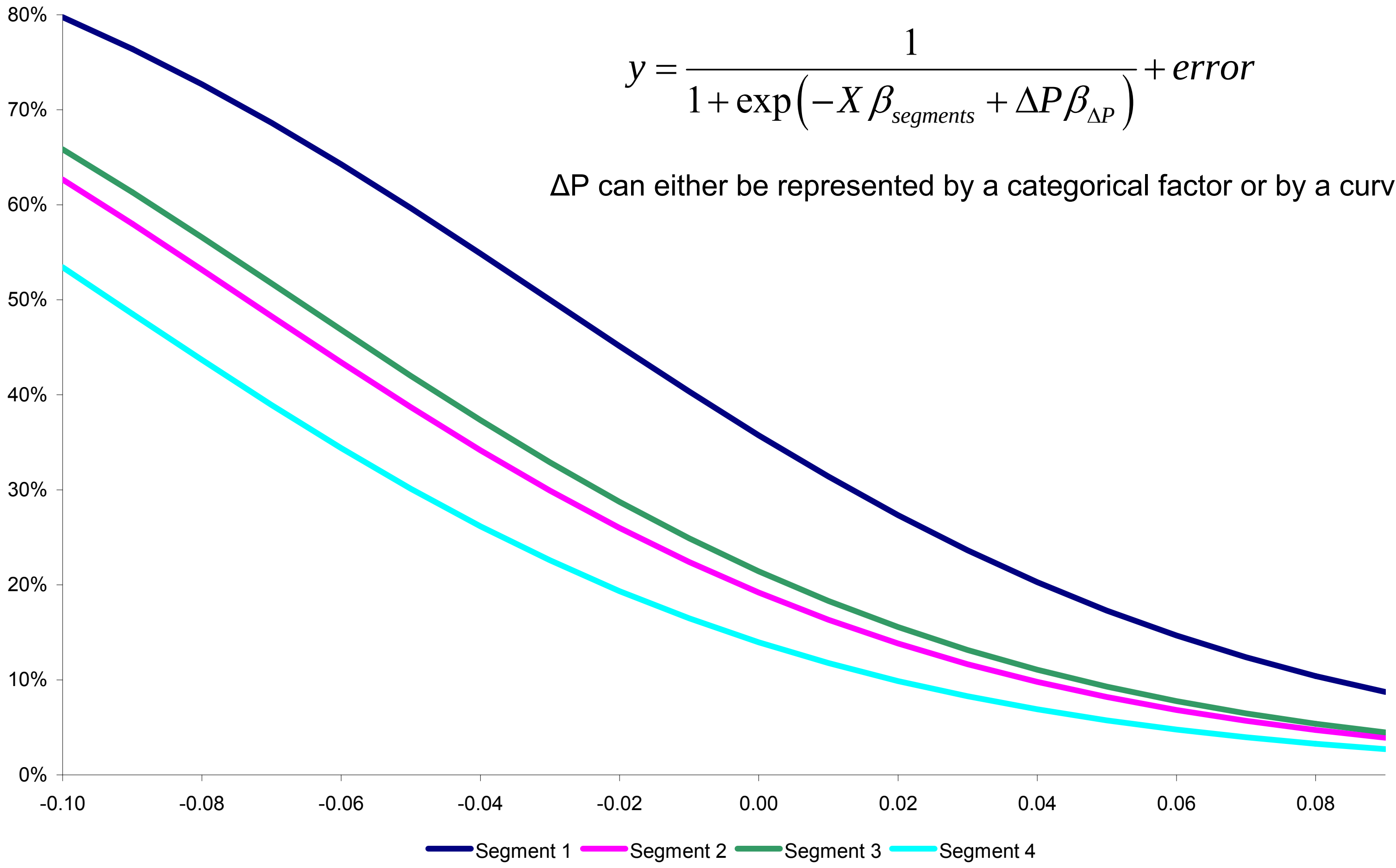
Model Forms - Simple GLMs

0/1 response

Logit link function with binomial response

$$y = \frac{1}{1+\exp\left(-X\beta_{segments} + \Delta P \beta_{\Delta P}\right)} + error$$

ΔP can either be represented by a categorical factor or by a curve

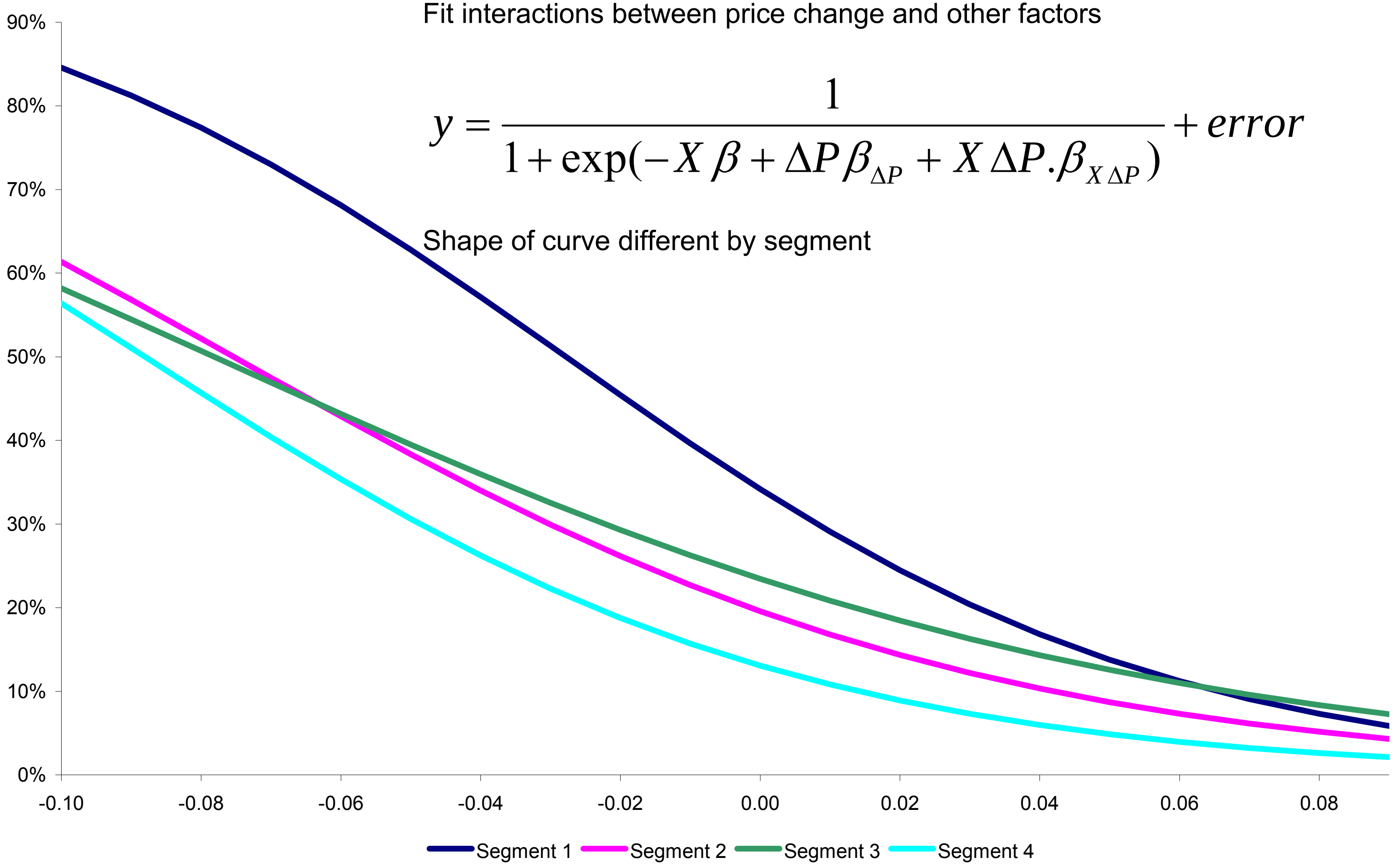


GLMs with Interactions

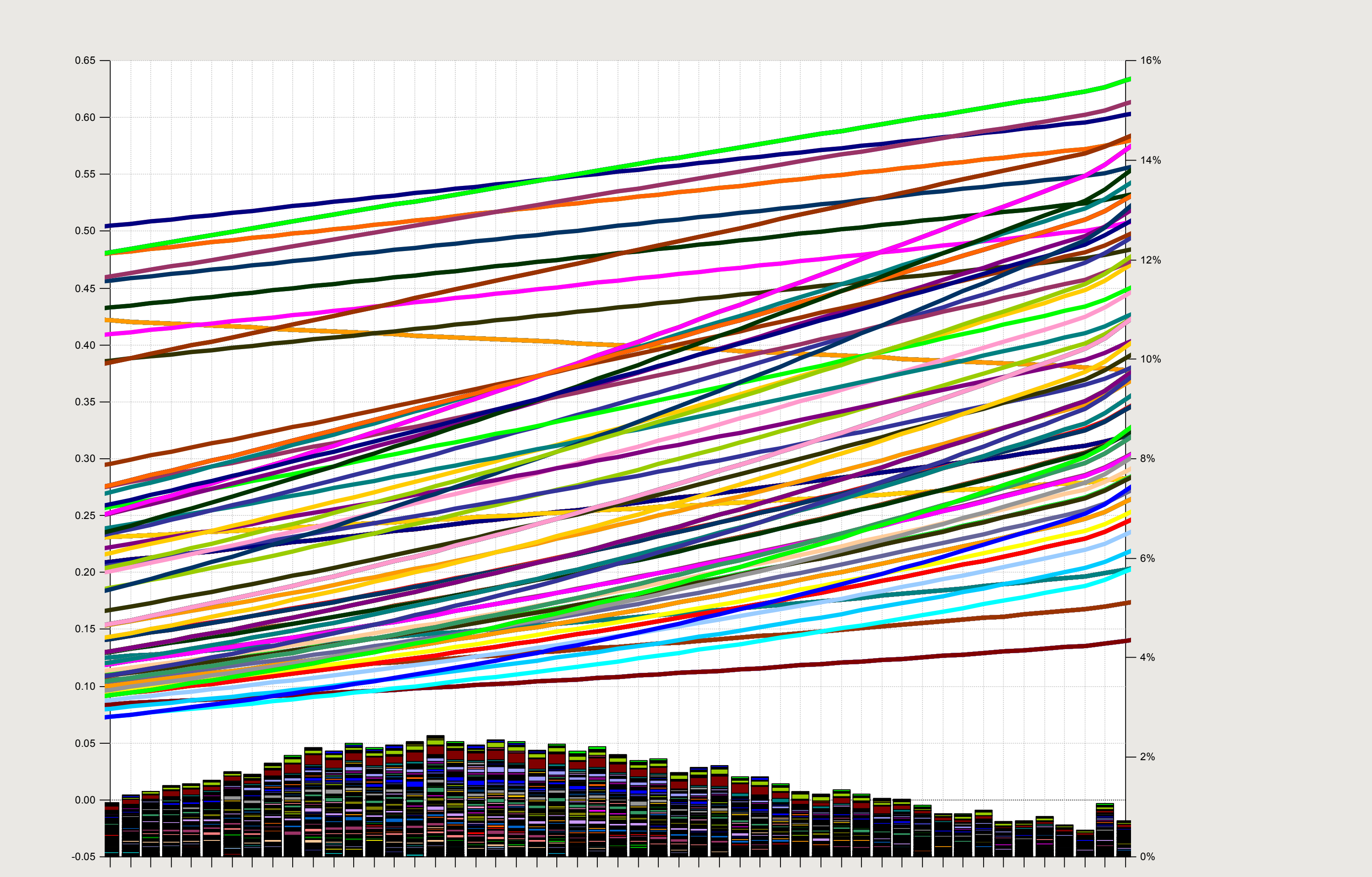
Fit interactions between price change and other factors

$$y = \frac{1}{1 + \exp(-X\beta + \Delta P \beta_{\Delta P} + X\,\Delta P.\beta_{X\,\Delta P})} + error$$

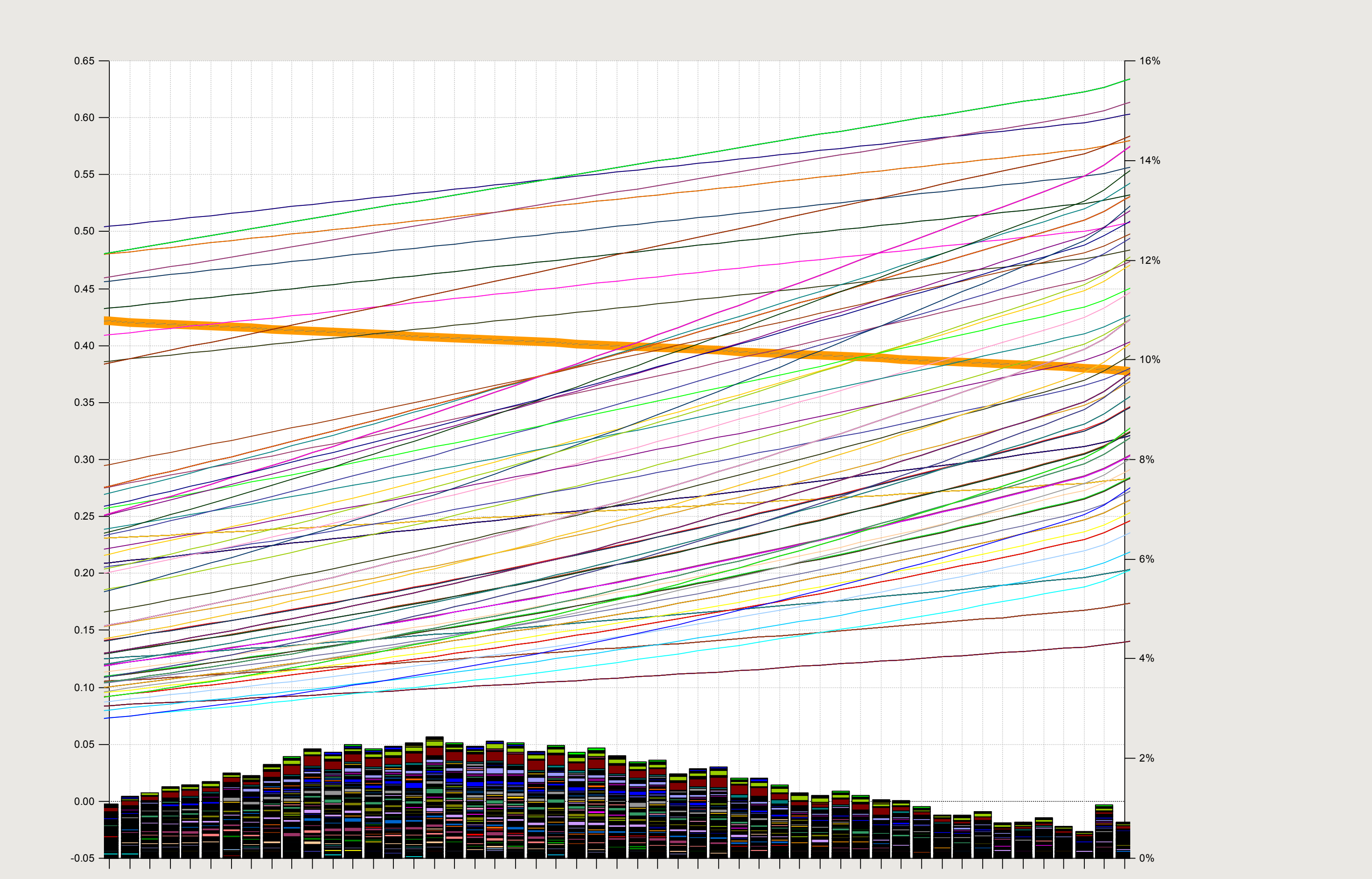
Shape of curve different by segment



GLM with interactions



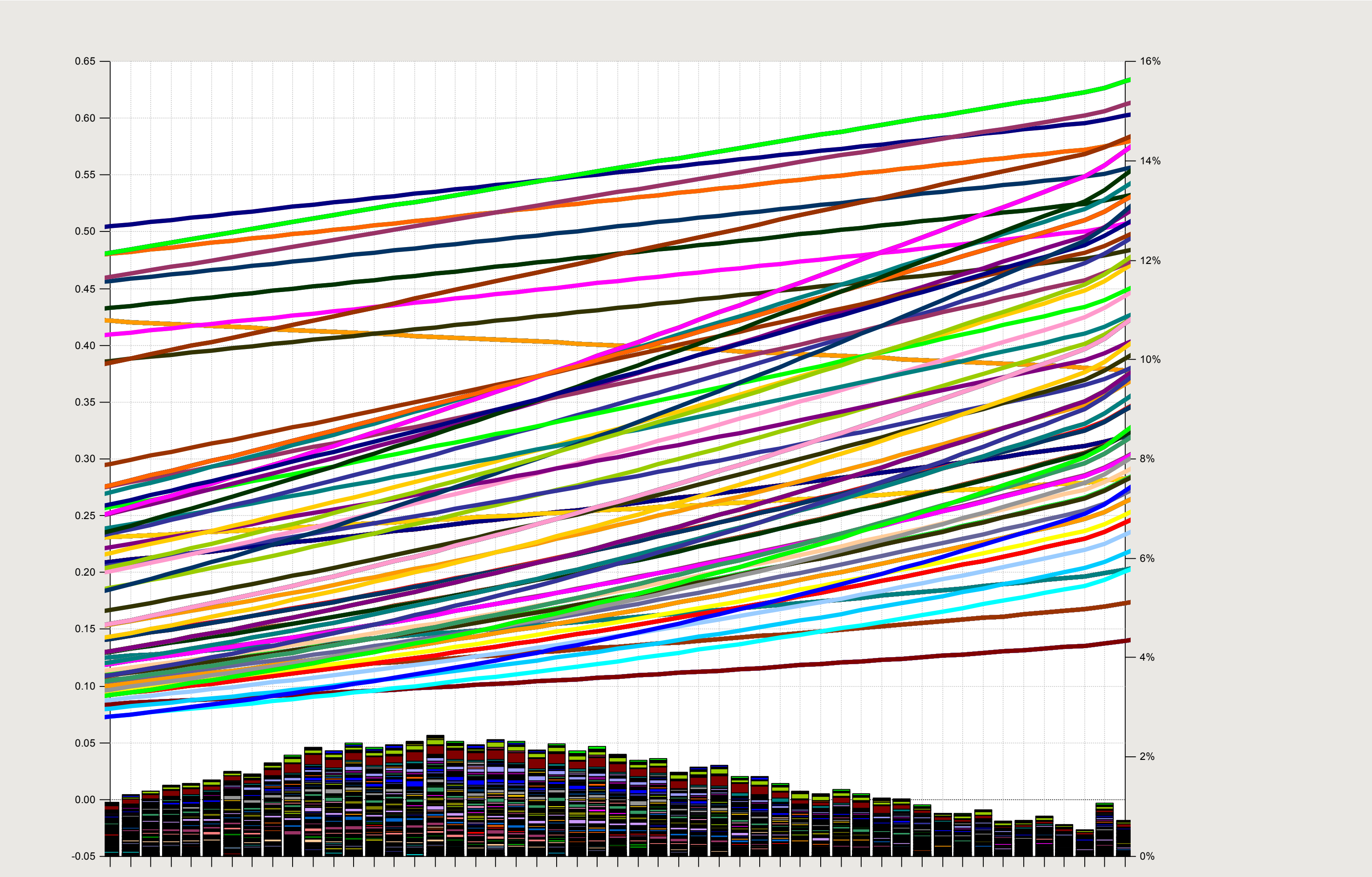
GLM with interactions



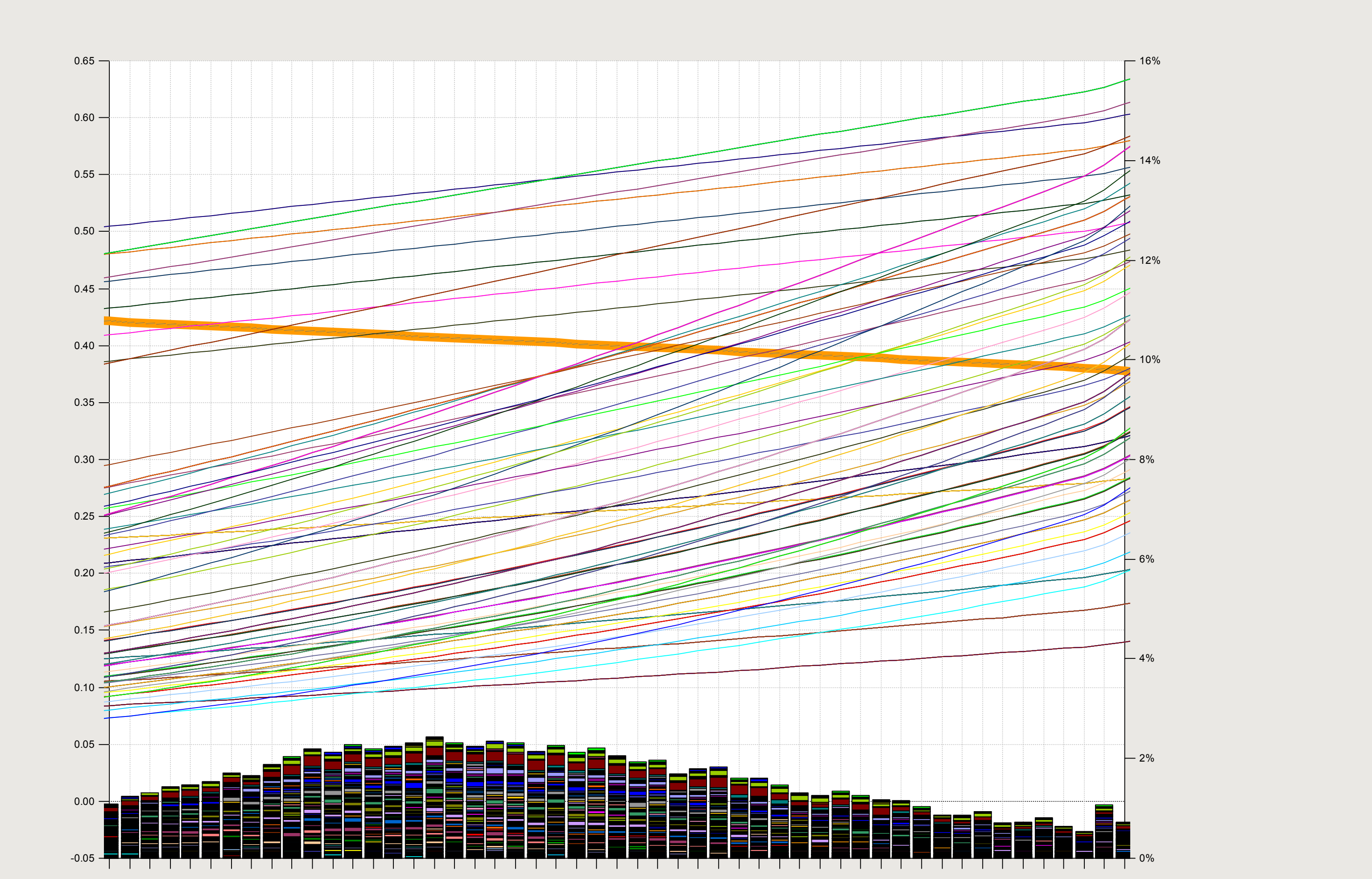
Generalized Non-Linear Models

- GLM
  - $E[\underline{Y}] = \underline{\mu} = g^{-1}(\mathbf{X}.\underline{\beta} + \underline{\xi})$
- GNM
  - many forms, eg
  - $E[\underline{Y}] = \underline{\mu} = g^{-1}(\mathbf{X}.\underline{\beta} + e^{\mathbf{Z}.\underline{\gamma}})$
  - $E[\underline{Y}] = \underline{\mu} = g^{-1}(\mathbf{X}.\underline{\beta} + Y.\underline{\zeta}.e^{\mathbf{Z}.\underline{\gamma}})$
- A potentially useful form for demand modeling:
  - $E[\underline{Y}] = \underline{\mu} = 1 \,/\, (\, 1 + \exp(\, \mathbf{X}.\underline{\beta} + \underbrace{\Delta P.e^{\mathbf{Z}.\underline{\gamma}}}_{\text{Forces elasticity to be positive}}) \,)$

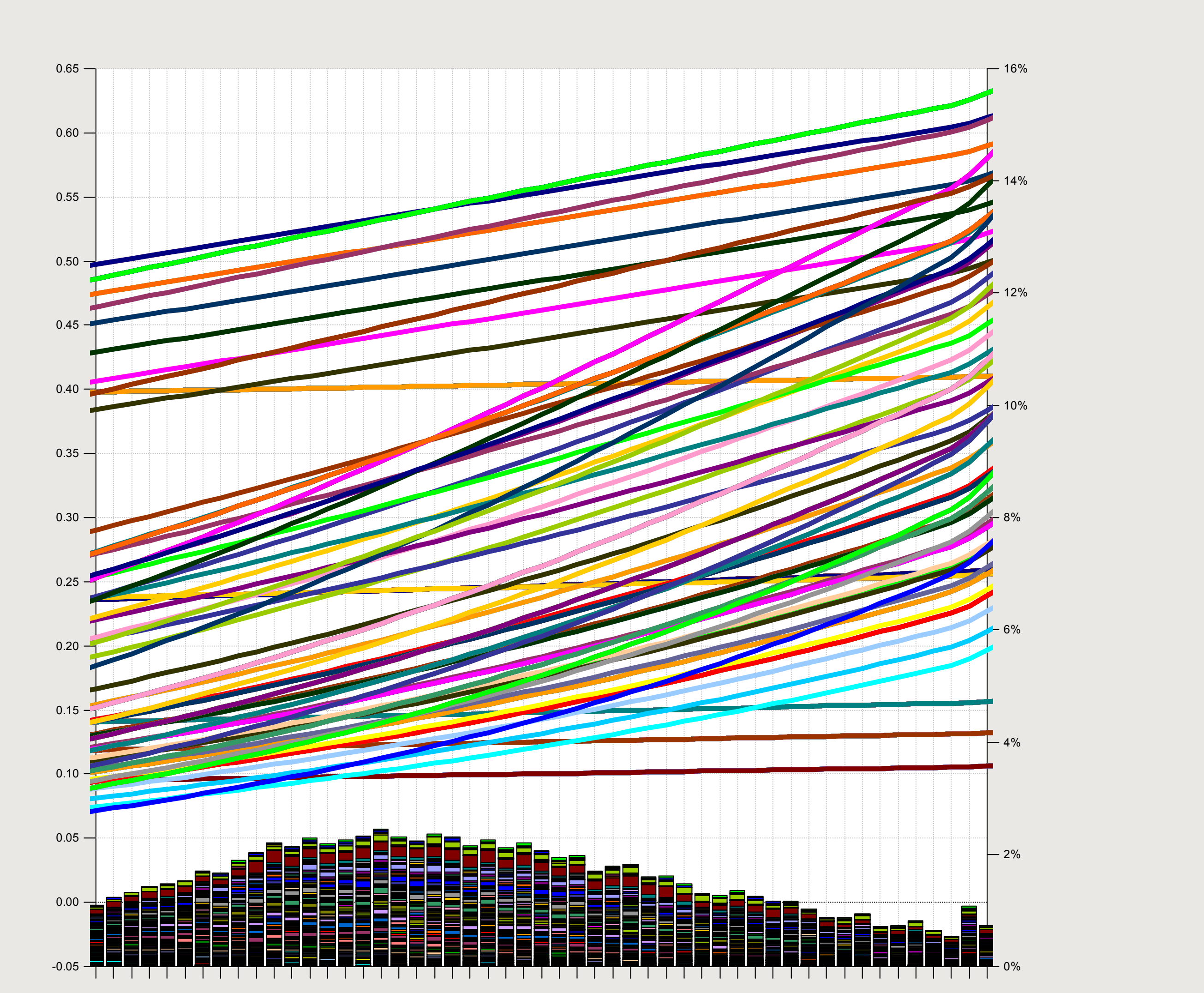
GLM with interactions



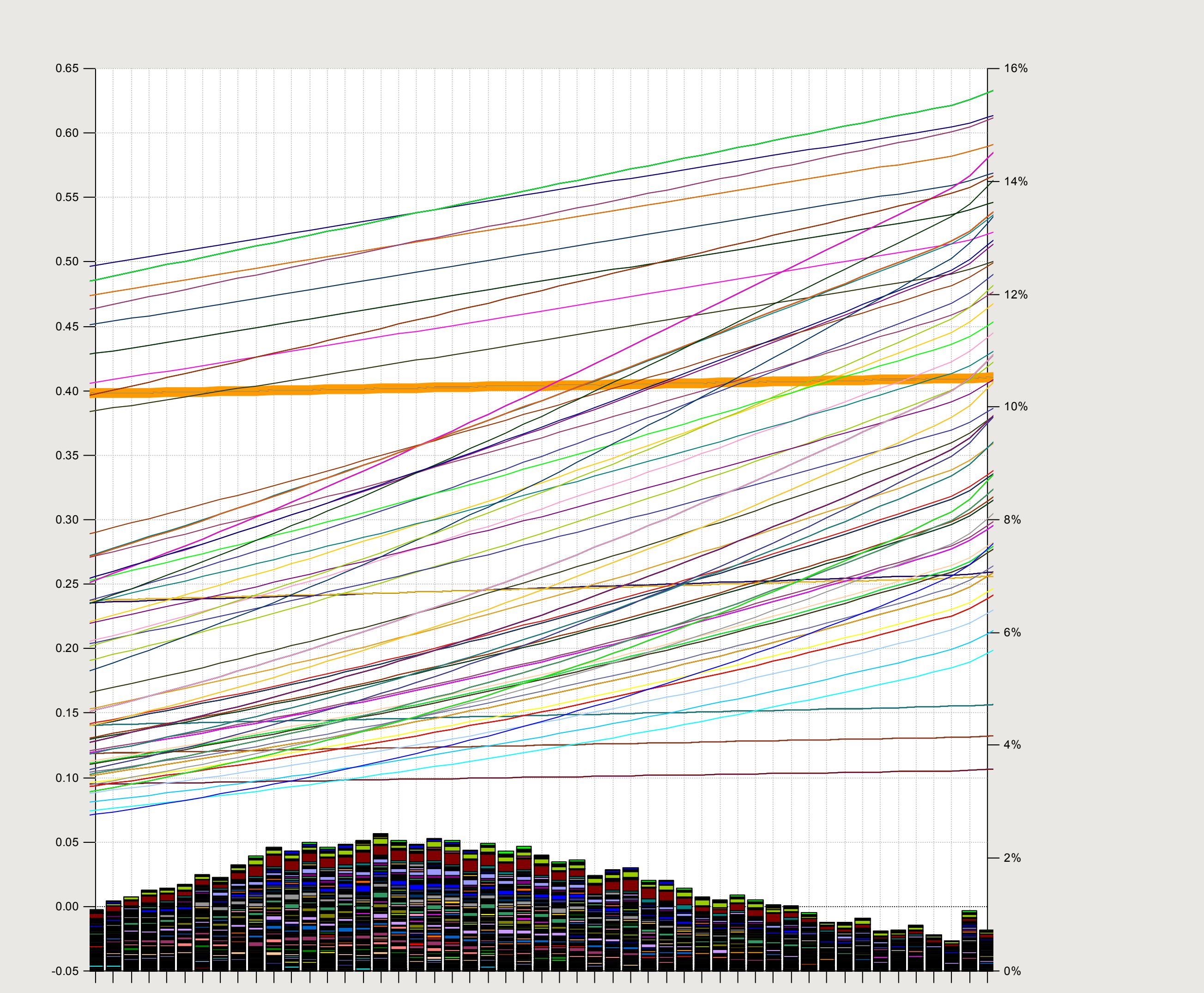
GLM with interactions



Generalized Non-Linear Model



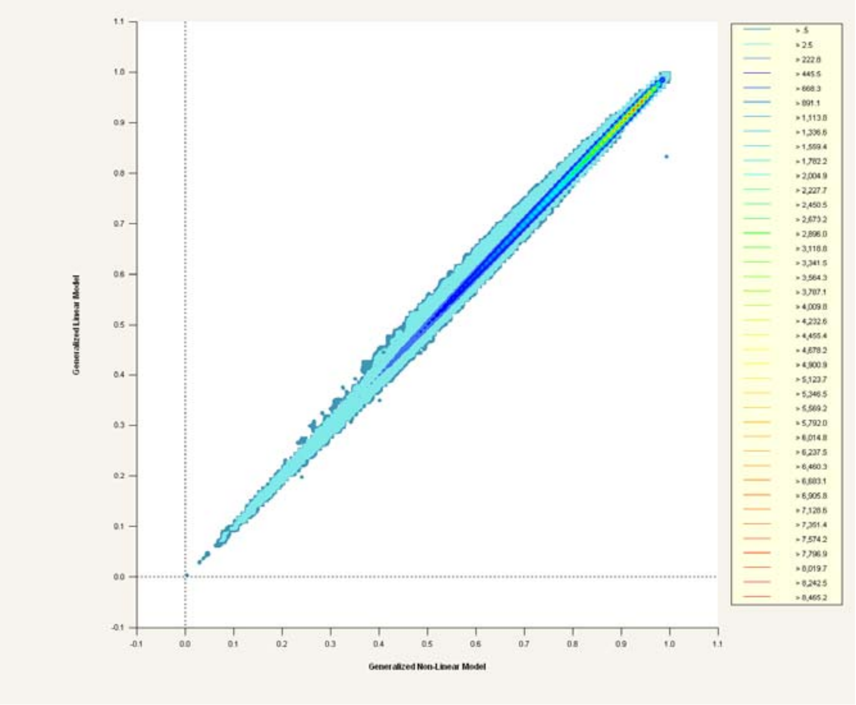
Generalized Non-Linear Model



Generalized Non-Linear Models

Often only relevant if models are complex

Number of interactions	% records with GLM negative elasticity
0	0%
1	0.04%
2	0.3%
3	0.8%
4	1.5%



Agenda

- Testing the link function
- The Tweedie distribution
- Regression splines
- Reference models
- Aliasing/near-aliasing
- Combining models across claim types
- Restricted models
- Model validation
- Modeling elasticity / GNMs

GLM III

Duncan Anderson MA FIA
Partner, EMB Consultancy LLP