watsonwyatt.com

PM-2
An Introduction to GLM Theory

CAS Seminar on Ratemaking
Boston, March 17, 2008

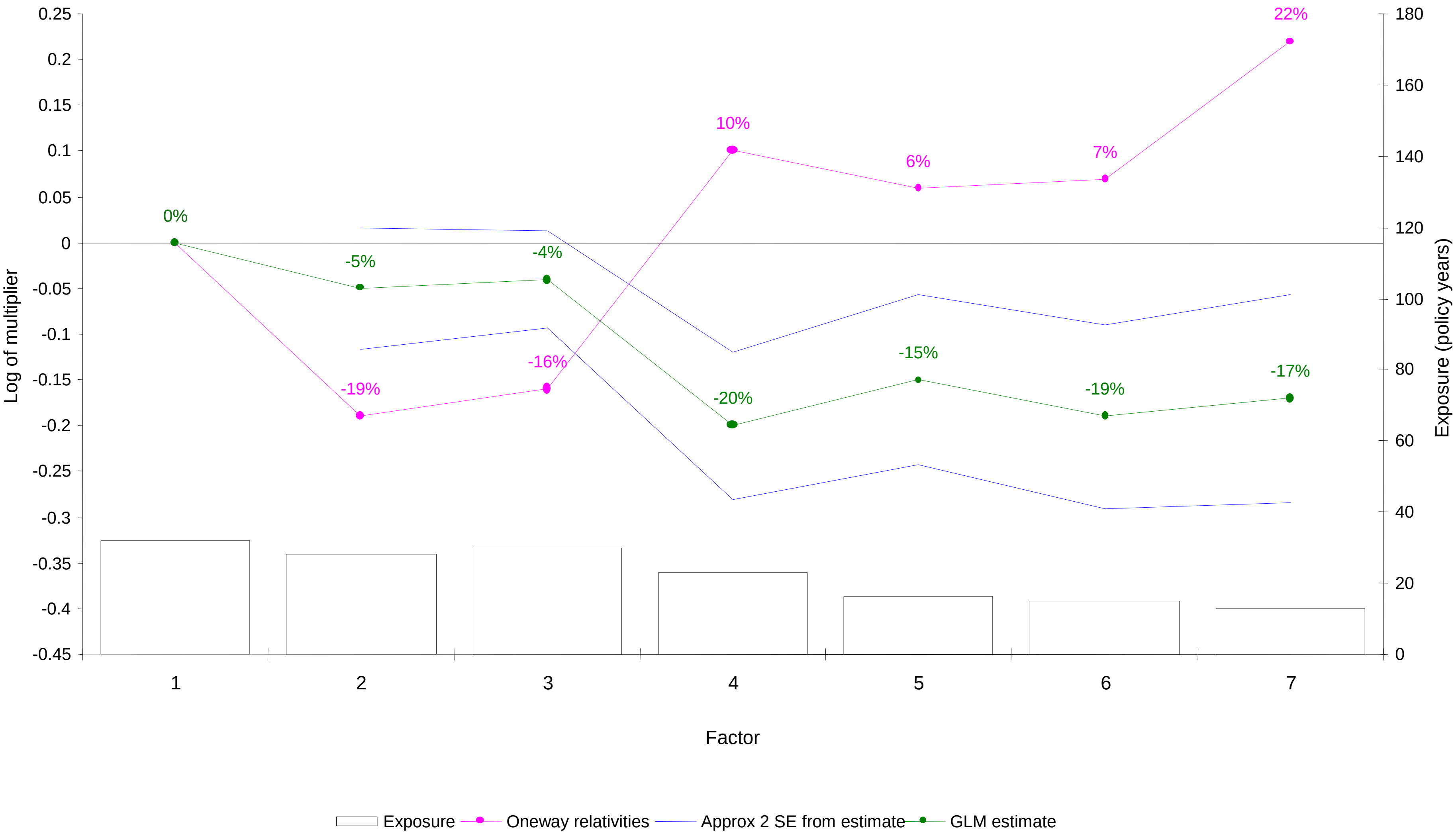
Claudine Modlin, FCAS, MAAA

Watson Wyatt
Worldwide

Generalized linear model benefits

- Consider all factors simultaneously
- Allow for nature of random process
- Provide diagnostics
- Robust and transparent

Example of GLM output



Agenda

- GLM formulae
- Model testing
  - use only variables that are predictive
  - make sure model is reasonable
- Aliasing
- Refinements

Linear models

- Linear model $Y_i = \mu_i$ + error
- μ_i based on linear combination of measured factors
- Which factors, and how they are best combined is to be derived

$$\mu_i = \alpha + \beta.\text{age}_i + \gamma.\text{age}_i^2 + \delta.\text{height}_i.\text{age}_i$$ ✓

$$\mu_i = \alpha + \beta.\text{age}_i + \gamma.(\text{sex}_i\text{=female})$$ ✓

$$\mu_i = (\alpha + \beta.\text{age}_i) * \exp(\delta.\text{height}_i.\text{age}_i)$$ ✗

Linear models - formularization

$$E[Y_i] = \mu_i = \Sigma X_{ij}\beta_j$$

$$Var[Y_i] = \sigma^2$$

$$Y_i \sim N(\mu_i, \sigma^2)$$

What is $\Sigma X_{ij}\beta_j$?

- **X** defines the explanatory variables to be included in the model
  - could be continuous variables - "variates"
  - could be categorical variables - "factors"
- $\underline{\beta}$ contains the parameter estimates which relate to the factors / variates defined by the structure of **X**
  - "the answer"

What is $\mathbf{X}.\underline{\beta}$?

- Write $\Sigma X_{ij}\beta_j$ as $\mathbf{X}.\underline{\beta}$
- Consider 3 rating factors

  - age of driver ("age")

  - sex of driver ("sex")

  - age of vehicle ("car")
- Represent $\underline{\beta}$ by α, β, γ, δ, ...

What is **X**.$\underline{\beta}$?

- Suppose we wanted a model of the form:

  $\underline{\mu} = \alpha + \beta.\underline{age} + \gamma.\underline{age}^2 + \delta.\underline{car}^{27}.\underline{age}^{52\frac{1}{2}}$

- **X**.$\underline{\beta}$ would need to be defined as:

$$\begin{pmatrix} 1 & age_1 & age_1^{2} & car_1^{27}.age_1^{52\frac{1}{2}} \\ 1 & age_2 & age_2^{2} & car_2^{27}.age_2^{52\frac{1}{2}} \\ 1 & age_3 & age_3^{2} & car_3^{27}.age_3^{52\frac{1}{2}} \\ 1 & age_4 & age_4^{2} & car_4^{27}.age_4^{52\frac{1}{2}} \\ 1 & age_5 & age_5^{2} & car_5^{27}.age_5^{52\frac{1}{2}} \\ & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ & \cdots\cdots & \cdots\cdots & \cdots\cdots \end{pmatrix} . \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix}$$

What is $\mathbf{X}.\underline{\beta}$?

- Suppose we wanted a model of the form:

$$\begin{aligned}\underline{\mu} = \alpha &+ \beta_1 \text{ if } \underline{\text{age}} < 30 \\ &+ \beta_2 \text{ if } \underline{\text{age}} \text{ 30 - 40} \\ &+ \beta_3 \text{ if } \underline{\text{age}} > 40 \\ &+ \gamma_1 \text{ if } \underline{\text{sex}} \text{ male} \\ &+ \gamma_2 \text{ if } \underline{\text{sex}} \text{ female}\end{aligned}$$

What is X.$\underline{\beta}$?

$$\begin{array}{c|ccccccc|}
 & & & \text{Age} & & & \text{Sex} & \\
 & & <30 & 30\text{-}40 & >40 & M & F \\
1 & 1 & 0 & 1 & 0 & 1 & 0 \\
2 & 1 & 1 & 0 & 0 & 1 & 0 \\
3 & 1 & 1 & 0 & 0 & 0 & 1 \\
4 & 1 & 0 & 0 & 1 & 1 & 0 \\
5 & 1 & 0 & 1 & 0 & 0 & 1 \\
 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{array} \cdot \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \gamma_1 \\ \gamma_2 \end{pmatrix}$$

What is $\mathbf{X}.\underline{\beta}$?

- Suppose we wanted a model of the form:

$$
\begin{aligned}
\underline{\mu} = \alpha &+ \beta_1 \text{ if } \underline{\text{age}} < 30 \\
&+ \beta_2 \text{ if } \underline{\text{age}}\ 30 - 40 \\
&+ \beta_3 \text{ if } \underline{\text{age}} > 40 \\
&+ \gamma_1 \text{ if } \underline{\text{sex}} \text{ male} \\
&+ \gamma_2 \text{ if } \underline{\text{sex}} \text{ female}
\end{aligned}
$$

What is **X**.$\underline{\beta}$?

- Suppose we wanted a model of the form:

$$\underline{\mu} = \alpha + \beta_1 \text{ if } \underline{\text{age}} < 30$$

$$\cancel{+ \beta_2 \text{ if } \underline{\text{age}}\ 30 - 40}$$

"Base levels"

$$+ \beta_3 \text{ if } \underline{\text{age}} > 40$$

$$\cancel{+ \gamma_1 \text{ if } \underline{\text{sex}} \text{ male}}$$

$$+ \gamma_2 \text{ if } \underline{\text{sex}} \text{ female}$$

X.$\underline{\beta}$ having adjusted for base levels

$$
\begin{array}{c|c|ccc|cc|}
 & & & \text{Age} & & \text{Sex} & \\
 & & <30 & 30\text{-}40 & >40 & M & F \\
1 & 1 & 0 & \cancel{1} & 0 & \cancel{1} & 0 \\
2 & 1 & 1 & \cancel{0} & 0 & \cancel{1} & 0 \\
3 & 1 & 1 & \cancel{0} & 0 & \cancel{0} & 1 \\
4 & 1 & 0 & \cancel{0} & 1 & \cancel{1} & 0 \\
5 & 1 & 0 & \cancel{1} & 0 & \cancel{0} & 1 \\
 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{array}
\;.\;
\begin{pmatrix}
\alpha \\ \beta_1 \\ \cancel{\beta_2} \\ \beta_3 \\ \cancel{\gamma_1} \\ \gamma_2
\end{pmatrix}
$$

Linear models - formularization

$$E[Y_i] = \mu_i = \Sigma X_{ij}\beta_j$$

$$Var[Y_i] = \sigma^2$$

$$Y_i \sim N(\mu_i, \sigma^2)$$

Generalized linear models

$$\mu_i = f(\ \alpha + \beta.\text{age}_i + \gamma.\text{age}_i^2 + \delta.\text{height}_i.\text{age}_i\ )$$

$$\mu_i = f(\ \alpha + \beta.\text{age}_i + \gamma.(\text{sex}_i\text{=female})\ )$$

Generalized linear models

$$\mu_i = g^{-1}(\alpha + \beta.\mathrm{age}_i + \gamma.\mathrm{age}_i^2 + \delta.\mathrm{height}_i.\mathrm{age}_i)$$

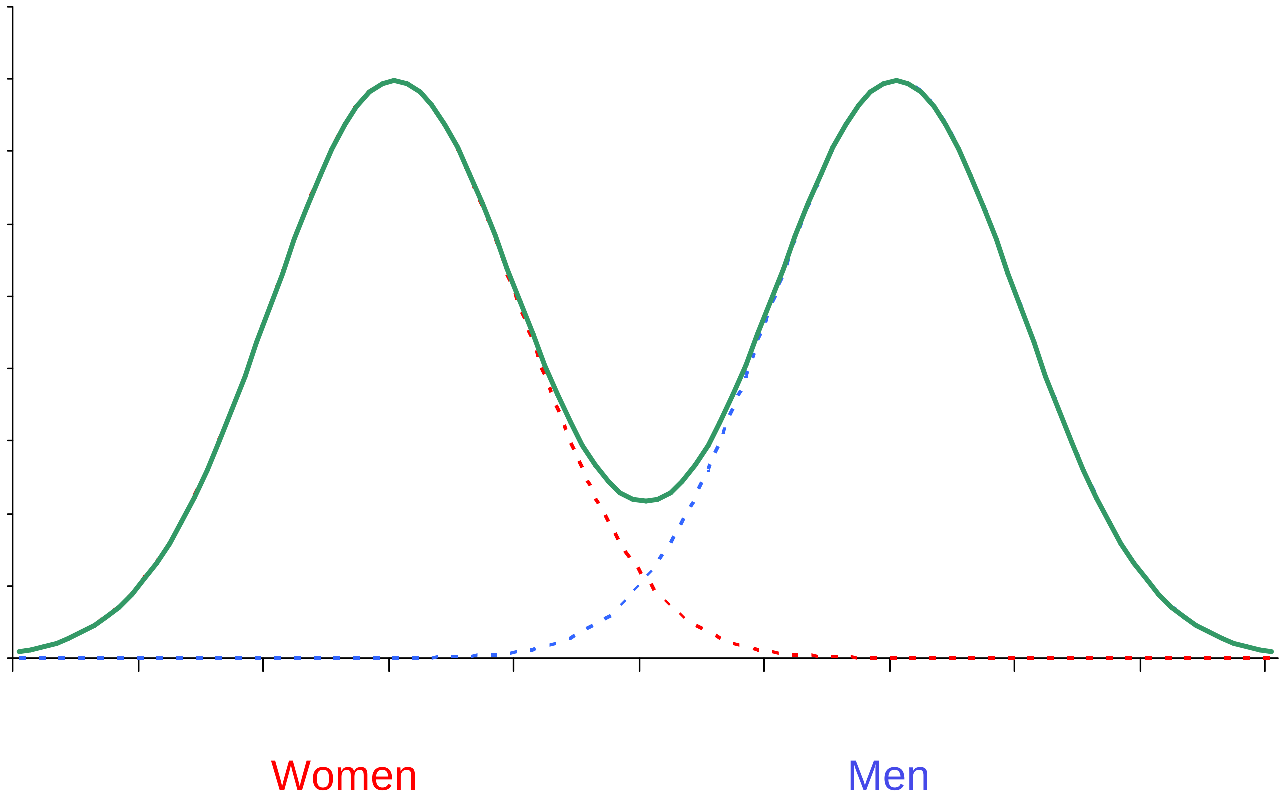
$$\mu_i = g^{-1}(\alpha + \beta.\mathrm{age}_i + \gamma.(\mathrm{sex}_i\mathrm{=female}))$$

Generalized linear models

Linear Models	Generalized Linear Models
$E[Y_i] = \mu_i = \Sigma X_{ij}\beta_j$	$E[Y_i] = \mu_i = g^{-1}(\Sigma X_{ij}\beta_j + \xi_i)$
$Var[Y_i] = \sigma^2$	$Var[Y_i] = \phi V(\mu_i)/\omega_i$
Y from Normal distribution	Y from a distribution from the exponential family

Generalized linear models

- Each observation i from distribution with mean μ_i

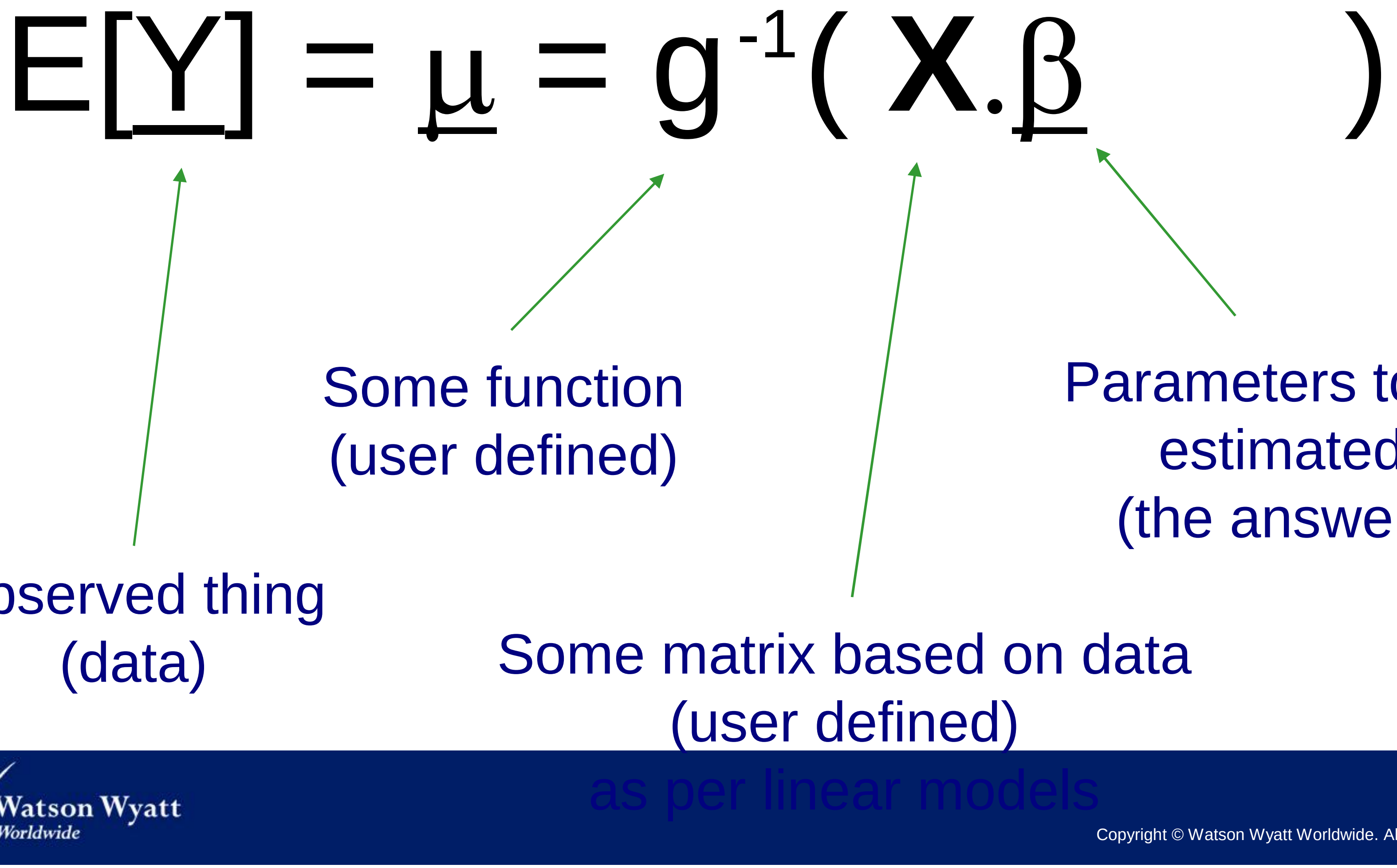


Generalized linear models

$$E[\underline{Y}] = \underline{\mu} = g^{-1}(\, \mathbf{X}.\underline{\beta} + \underline{\xi}\,)$$

$$Var[\underline{Y}] = \phi.V(\underline{\mu}) \,/\, \underline{\omega}$$

Generalized linear models



$$E[\underline{Y}] = \underline{\mu} = g^{-1}(\,\mathbf{X}.\underline{\beta}\qquad\qquad)$$

Observed thing (data)

Some function (user defined)

Some matrix based on data (user defined) as per linear models

Parameters to be estimated (the answer!)

What is $g^{-1}(\mathbf{X}.\underline{\beta})$?

$$\underline{Y} = g^{-1}(\mathbf{X}.\underline{\beta}) + \text{error}$$

Assuming a model with three categorical factors, each observation can be expressed as:

$$Y_{ijk} = g^{-1}(\alpha + \beta_i + \gamma_j + \delta_k) + \text{error}$$

$$\beta_2 = \gamma_1 = \delta_3 = 0$$

age is in group i

sex is in group j

car is in group k

What is $g^{-1}(\mathbf{X}.\underline{\beta})$?

- $g(x) = x \quad \Rightarrow \quad Y_{ijk} = \alpha + \beta_i + \gamma_j + \delta_k + \text{error}$

- $g(x) = \ln(x) \quad \Rightarrow \quad Y_{ijk} = e^{(\alpha + \beta_i + \gamma_j + \delta_k)} + \text{error}$

  $= A.B_i.C_j.D_k \quad + \text{error}$

  where $B_i = e^{\beta_i}$ etc

- Multiplicative form common for frequency and amounts

Multiplicative model

$ 207.10 x

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

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

Sex	Factor
Male	1.00
Female	1.25

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

E(losses) = $ 207.10 x 1.42 x 0.92 x 1.00 x 1.15 = $ 311.14

Generalized linear models

$$E[\underline{Y}] = \underline{\mu} = g^{-1}(\,\mathbf{X}.\underline{\beta} + \underline{\xi}\,)$$

"Offset"

Eg $\underline{Y}$ = claim *numbers*

Smith: Male, 30, Ford, 1 years, 2 claims

Jones: Female, 40, VW, ½ year, 1 claim

What is ξ ?

- $g(x) = \ln(x)$
- $\xi_{ijk} = \ln(\text{exposure}_{ijk})$
- $E[Y_{ijk}] = e^{(\alpha + \beta_i + \gamma_j + \delta_k + \xi_{ijk})}$

$$= A.B_i.C_j.D_k.\ e^{(\ln(\text{exposure}_{ijk}))}$$

$$= A.B_i.C_j.D_k.\ \text{exposure}_{ijk}$$

Restricted models

$$E[\underline{Y}] = \underline{\mu} = g^{-1}(\, \mathbf{X}.\underline{\beta} + \color{red}{\underline{\xi}} \,)$$

Offset

- Constrain model (eg increased limits, territory, amount of insurance, discounts)
- Other factors adjusted to compensate

Generalized linear models

$$E[\underline{Y}] = \underline{\mu} = g^{-1}(\mathbf{X}.\underline{\beta} + \underline{\xi})$$

$$\mathrm{Var}[\underline{Y}] = \phi.V(\underline{\mu}) / \underline{\omega}$$

Generalized linear models

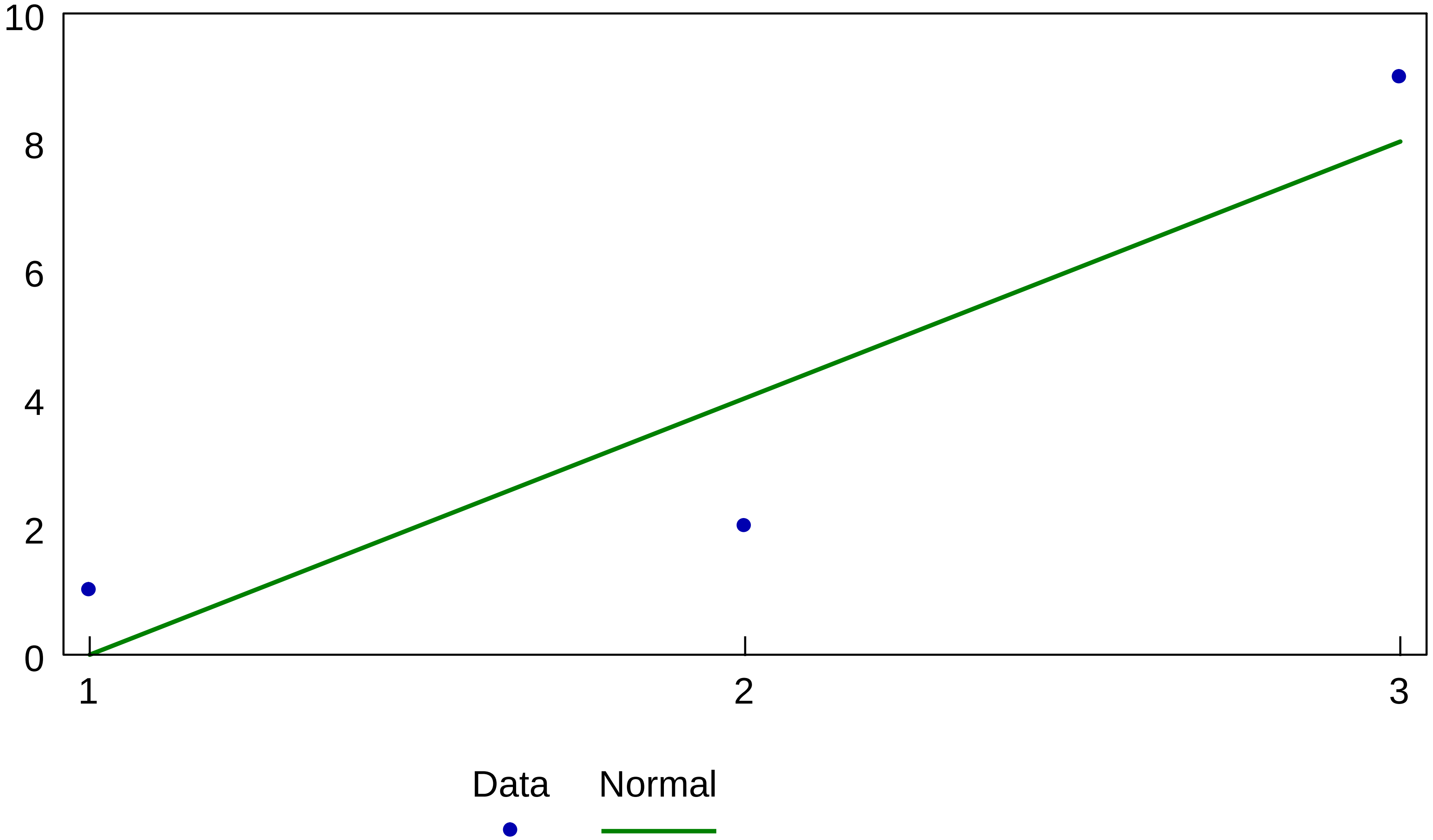
$$\mathrm{Var}[\underline{Y}] = \phi.V(\underline{\mu}) \;/\; \underline{\omega}$$

Normal: $\phi = \sigma^2$, $V(x) = 1 \Rightarrow \mathrm{Var}[\underline{Y}] = \sigma^2.\underline{1}$

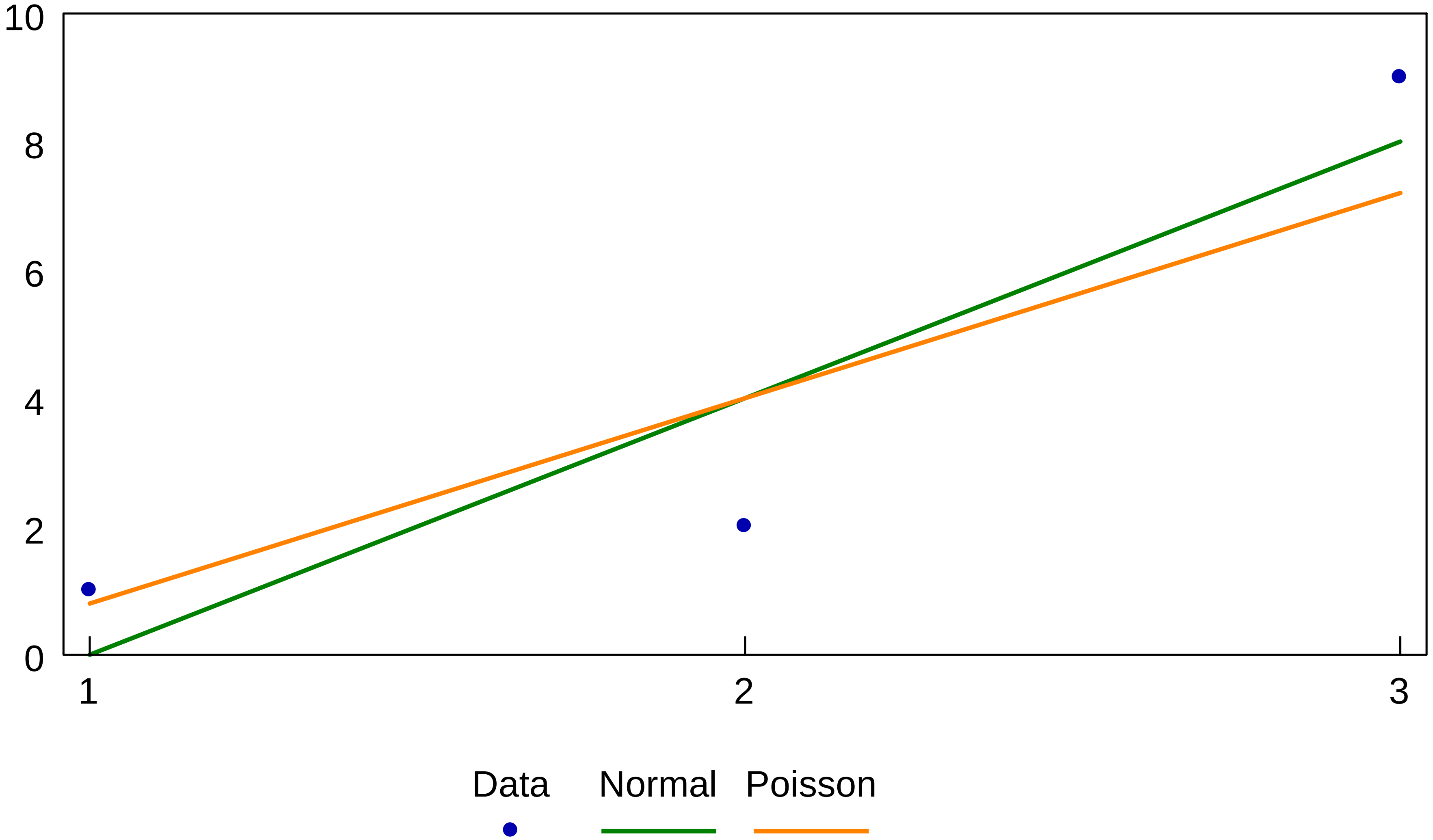
Poisson: $\phi = 1$, $V(x) = x \Rightarrow \mathrm{Var}[\underline{Y}] = \underline{\mu}$

Gamma: $\phi = k$, $V(x) = x^2 \Rightarrow \mathrm{Var}[\underline{Y}] = k\underline{\mu}^2$

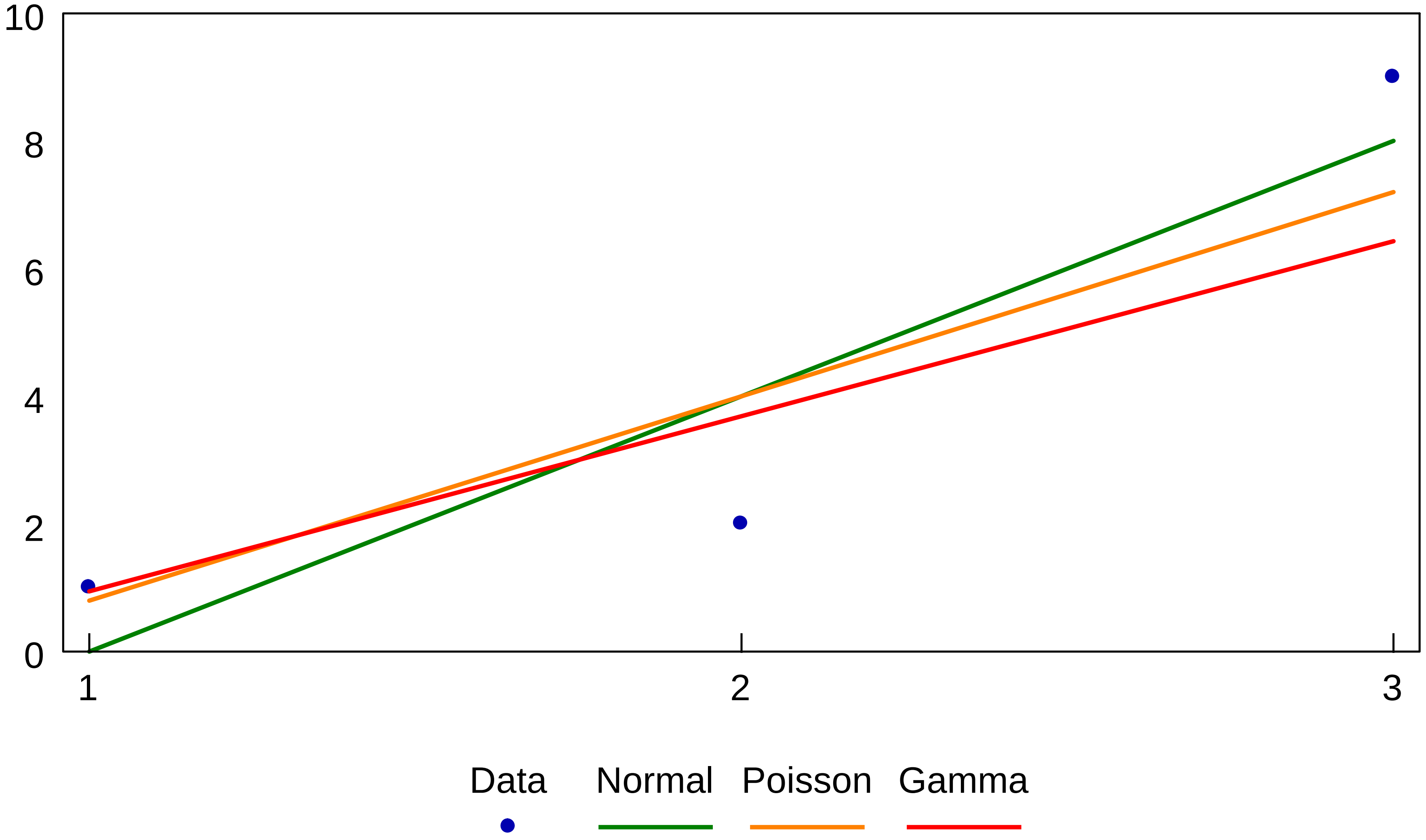
Example of effect of changing assumed error - 1



Example of effect of changing assumed error - 1



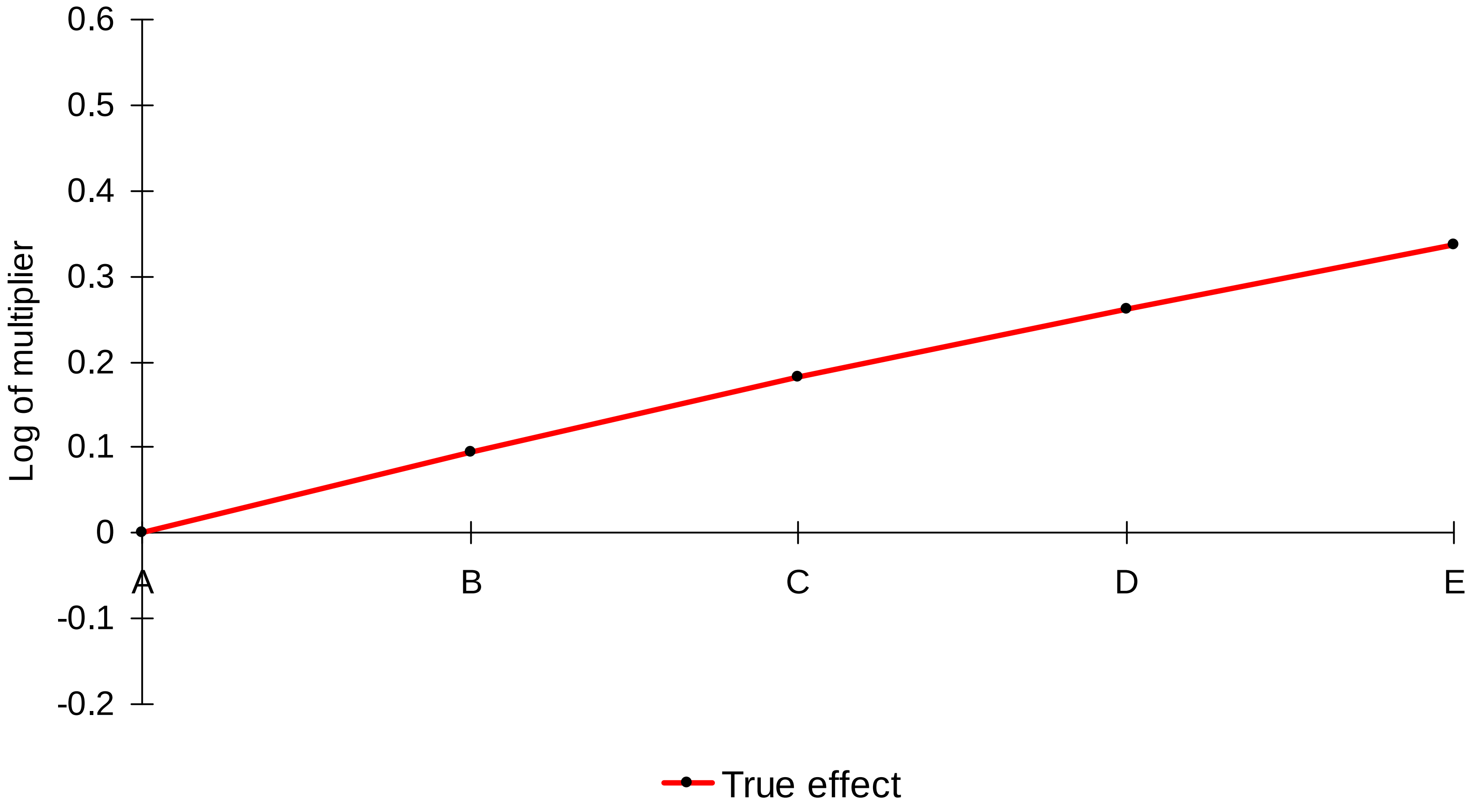
Example of effect of changing assumed error - 1



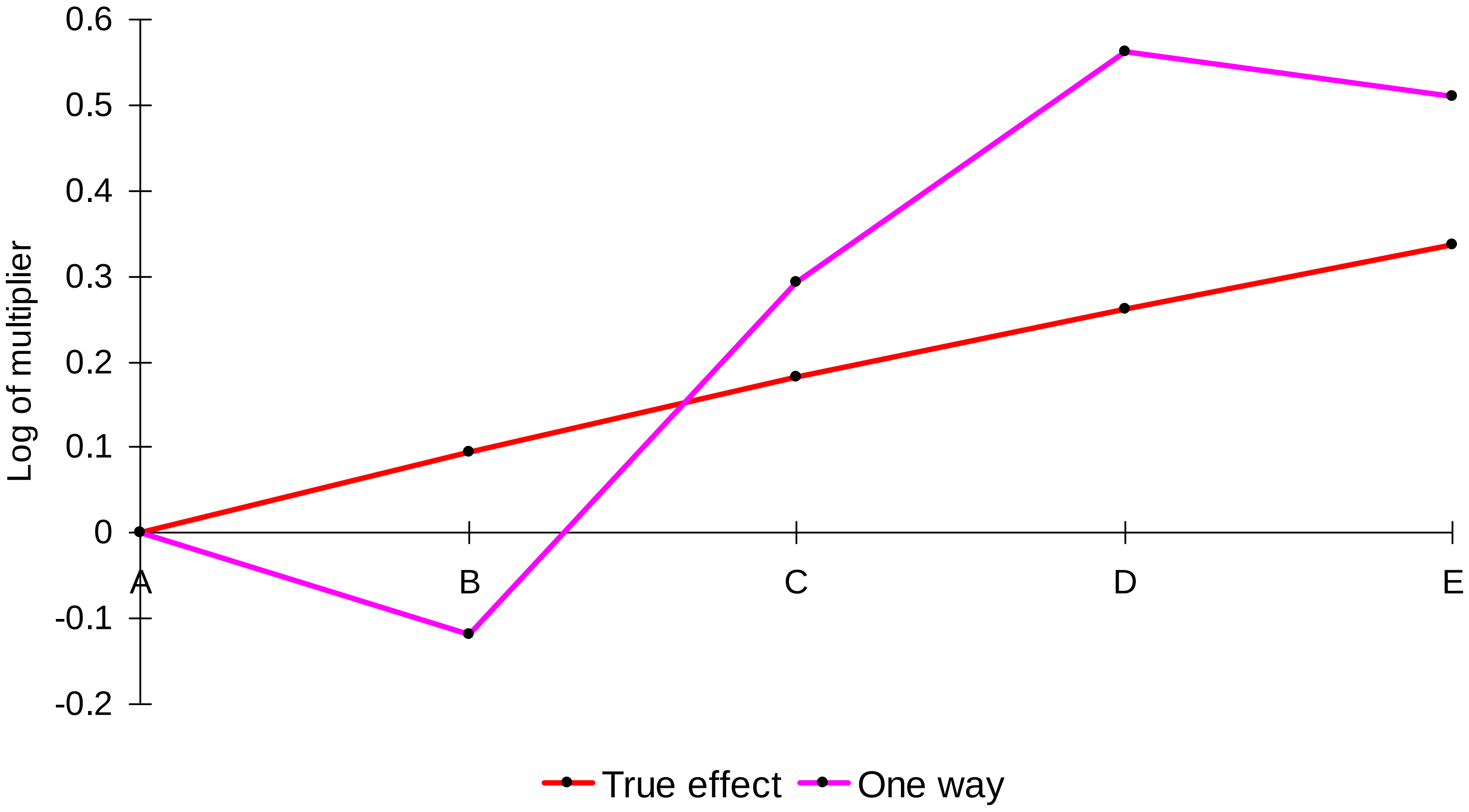
Example of effect of changing assumed error - 2

- Example portfolio with five rating factors, each with five levels A, B, C, D, E
- Typical correlations between those rating factors
- Assumed true effect of factors
- Claims randomly generated (with Gamma)
- Random experience analyzed by three models

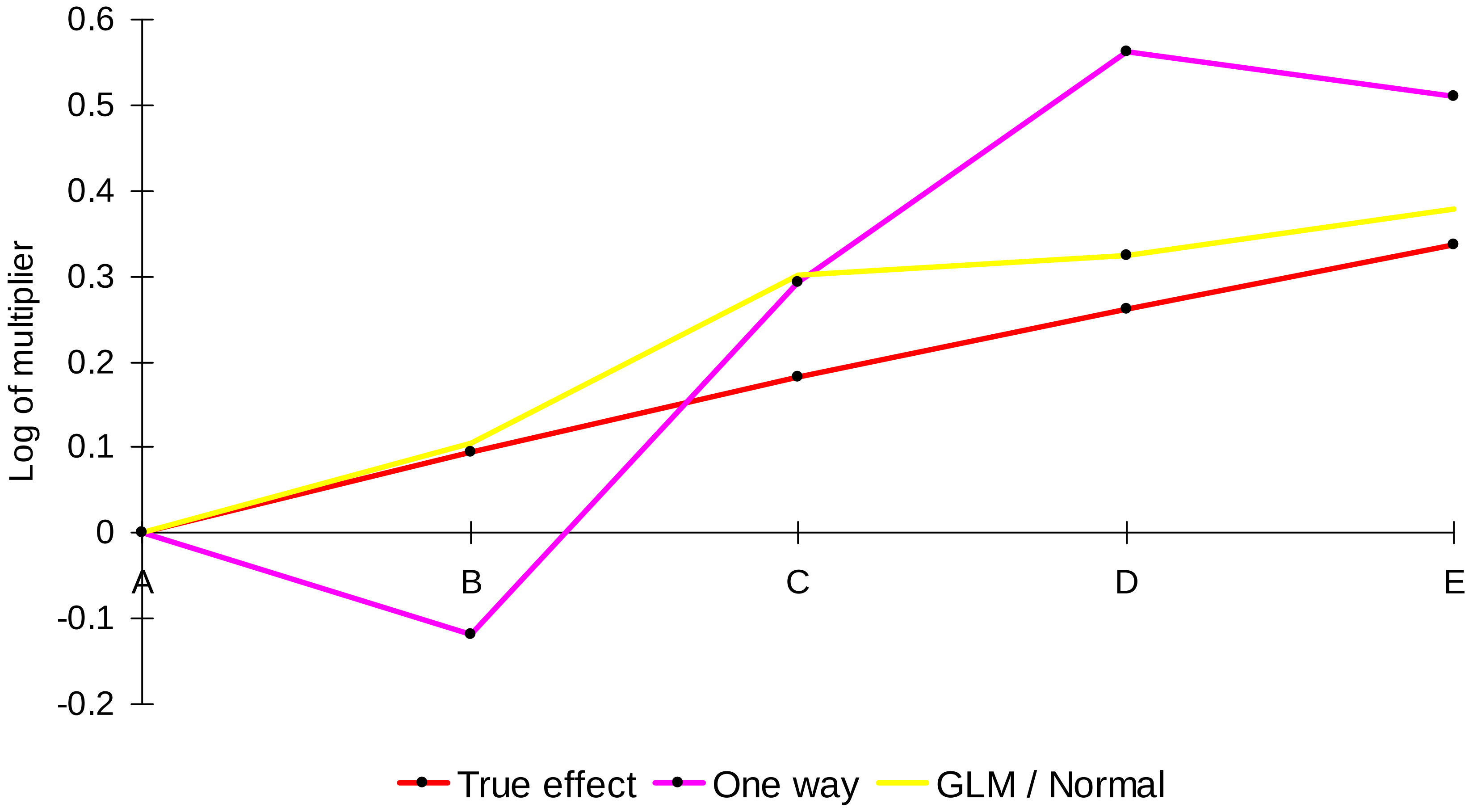
Example of effect of changing assumed error - 2



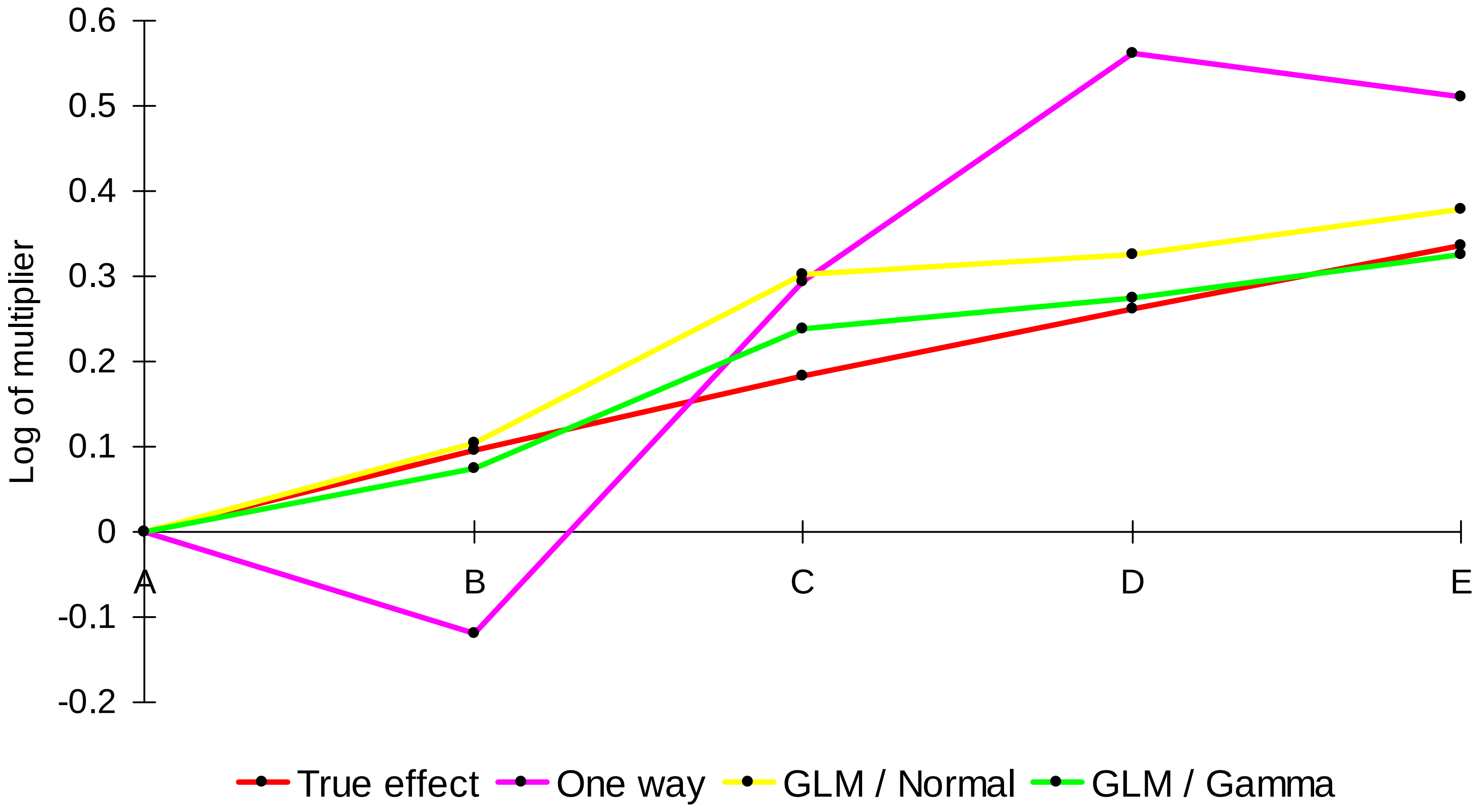
Example of effect of changing assumed error - 2



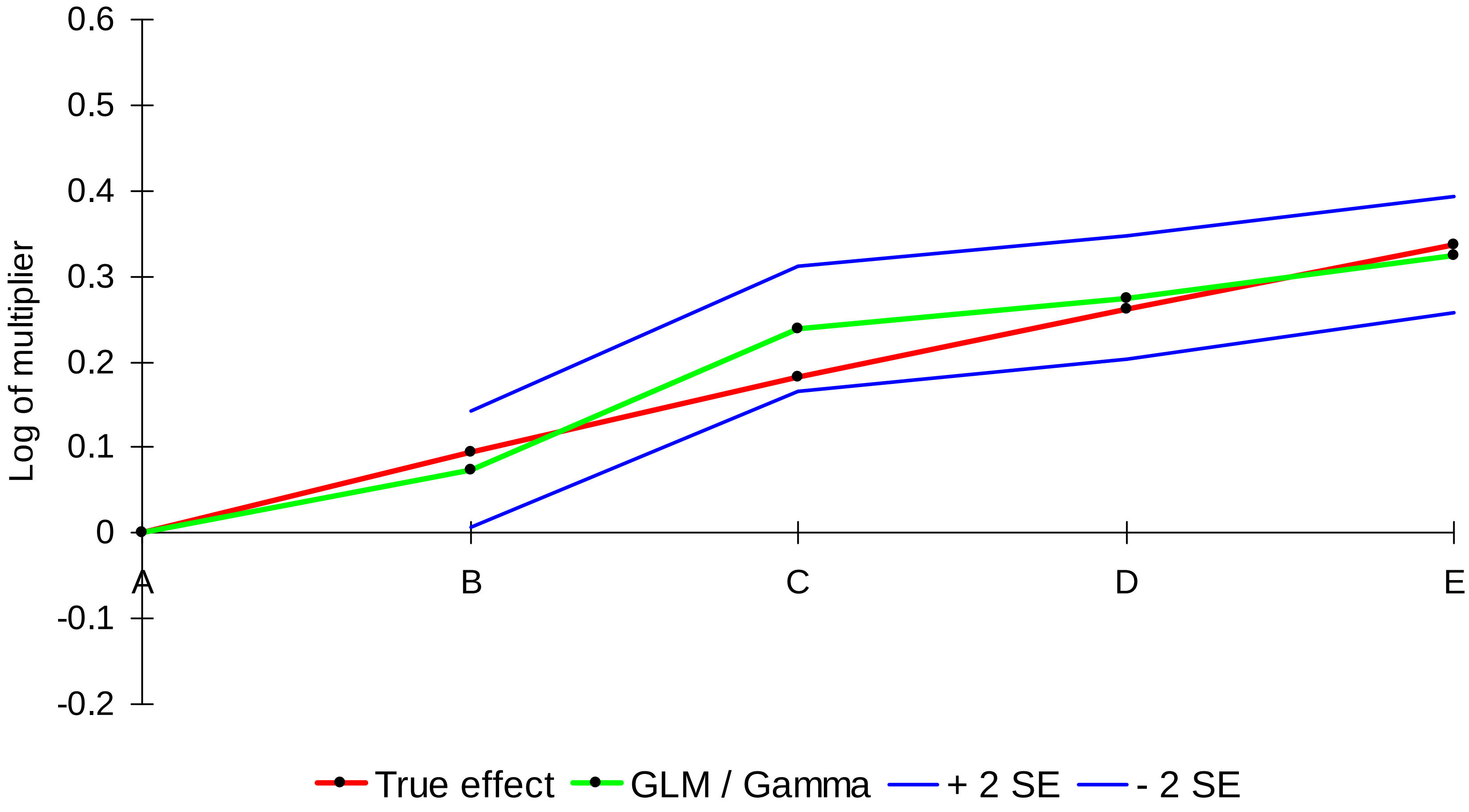
Example of effect of changing assumed error - 2



Example of effect of changing assumed error - 2



Example of effect of changing assumed error - 2



Prior weights

$$\mathrm{Var}[\underline{Y}] = \phi . V(\underline{\mu}) \;/\; \underline{\omega}$$

- Exposure
- Other credibility

Eg $\underline{Y}$ = claim *frequency*

Smith: Male, 30, Ford, 1 years, 2 claims, 100%

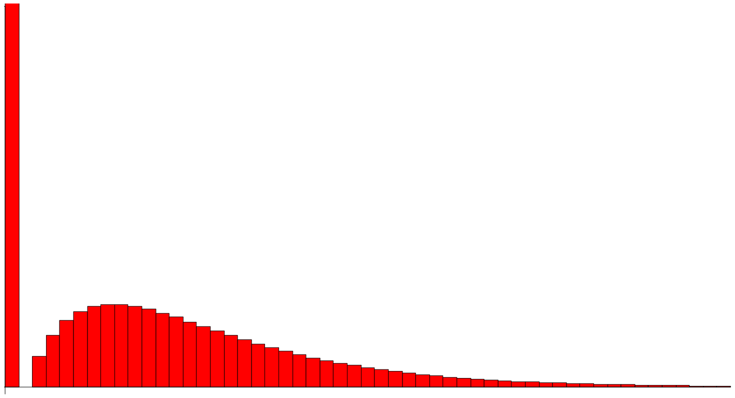
Jones: Female, 40, VW, ½ year, 1 claim, 100%

Typical model forms

Y	Claim frequency	Claim number	Average claim amount	Probability (eg lapses)
g(x)	ln(x)	ln(x)	ln(x)	ln(x/(1-x))
Error	Poisson	Poisson	Gamma	Binomial
ϕ V(x)	1 x	1 x	estimate x^2	1 x(1-x)
ω	exposure	1	# claims	1
ξ	0	ln(exposure)	0	0

Tweedie distributions

- 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 and parameters which can alter the shape to be like Poisson and gamma above zero

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

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

Generalized linear models

$$\mathrm{Var}[\underline{Y}] = \phi.V(\underline{\mu}) \,/\, \underline{\omega}$$

Normal: $\phi = \sigma^2$, $V(x) = 1 \Rightarrow \mathrm{Var}[\underline{Y}] = \sigma^2.\underline{1}$

Poisson: $\phi = 1$, $V(x) = x \Rightarrow \mathrm{Var}[\underline{Y}] = \underline{\mu}$

Gamma: $\phi = k$, $V(x) = x^2 \Rightarrow \mathrm{Var}[\underline{Y}] = k\underline{\mu}^2$

Tweedie: $\phi = k$, $V(x) = x^p \Rightarrow \mathrm{Var}[\underline{Y}] = k\underline{\mu}^p$

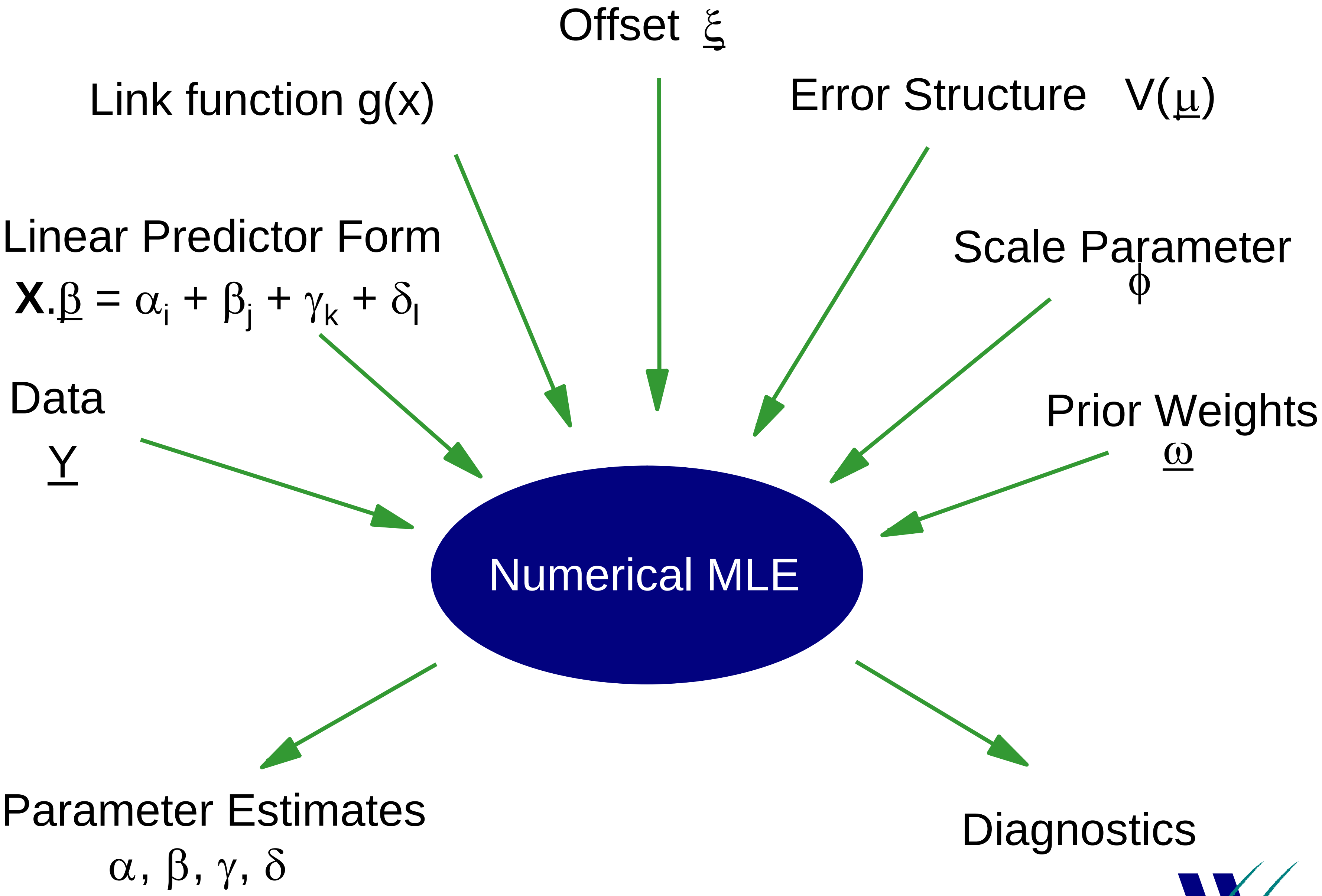
Tweedie distributions

Tweedie: $\phi = k,\ V(x) = x^p \Rightarrow \mathrm{Var}[\underline{Y}] = k\underline{\mu}^p$

- Defines a valid distribution for $p<0$, $1<p<2$, $p>2$
- Can be considered as Poisson/gamma process for $1<p<2$
- Typical values of p for insurance incurred claims around, or just under, 1.5

Tweedie distributions

- Helpful when important to fit to pure premium
- Often similar results to traditional approach but differences may occur if numbers and amounts models have effects which are both large and insignificant
- No information about whether frequencies or amounts are driving result



Offset $\underline{\xi}$

Link function g(x)

Error Structure V($\underline{\mu}$)

Linear Predictor Form

$\mathbf{X}.\underline{\beta} = \alpha_i + \beta_j + \gamma_k + \delta_l$

Scale Parameter

ϕ

Data

$\underline{Y}$

Prior Weights

$\underline{\omega}$

Numerical MLE

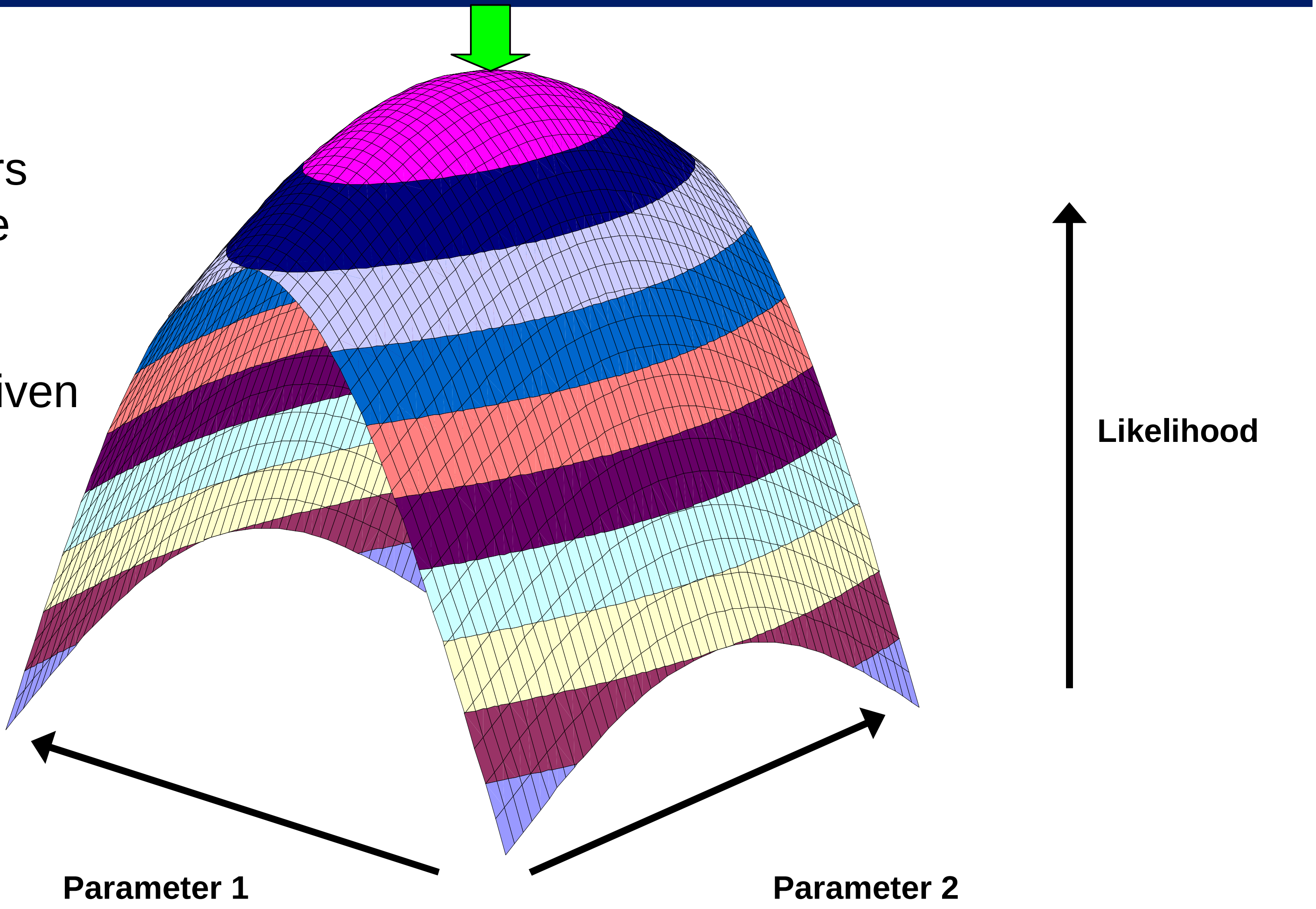
Parameter Estimates

$\alpha, \beta, \gamma, \delta$

Diagnostics

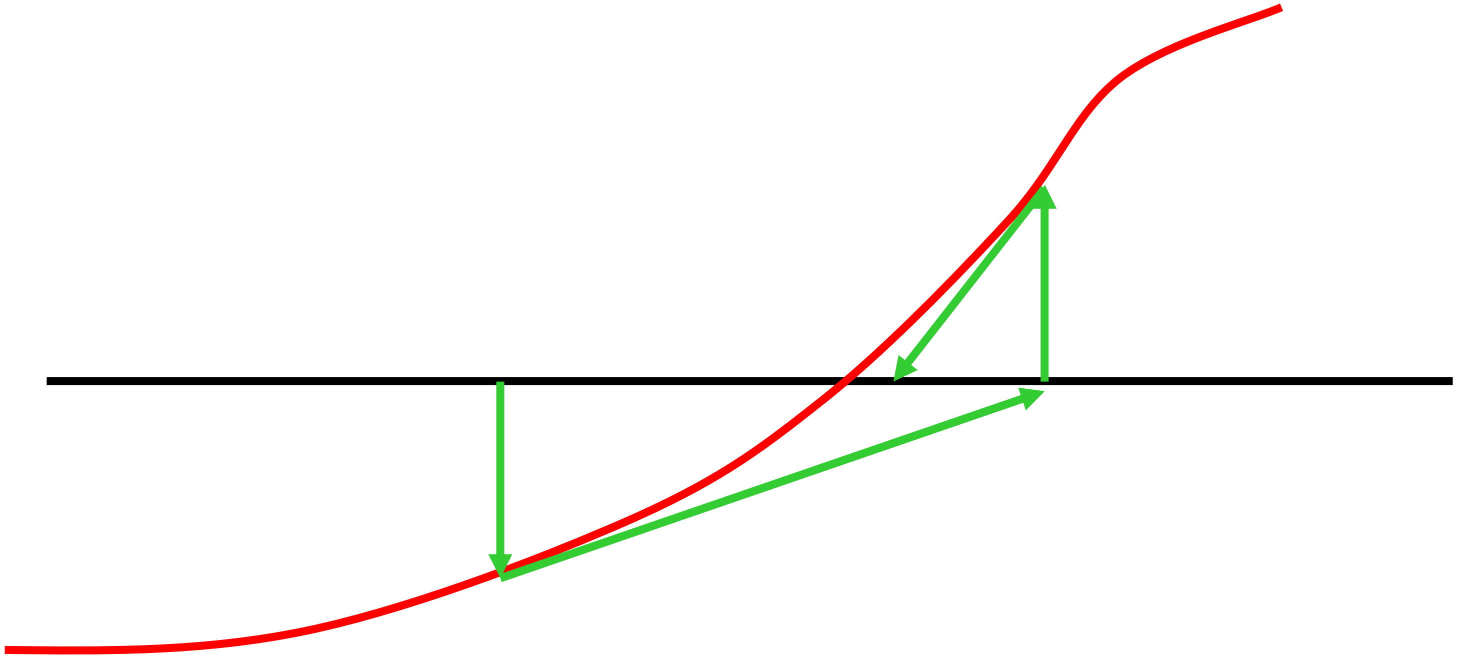
Maximum likelihood estimation

- Seek parameters which give highest likelihood function given data



Newton-Raphson

- In one dimension: $x_{n+1} = x_n - f'(x_n) / f''(x_n)$



- In n dimensions: $\underline{\beta}_{n+1} = \underline{\beta}_n - \mathbf{H}^{-1}.\underline{s}$

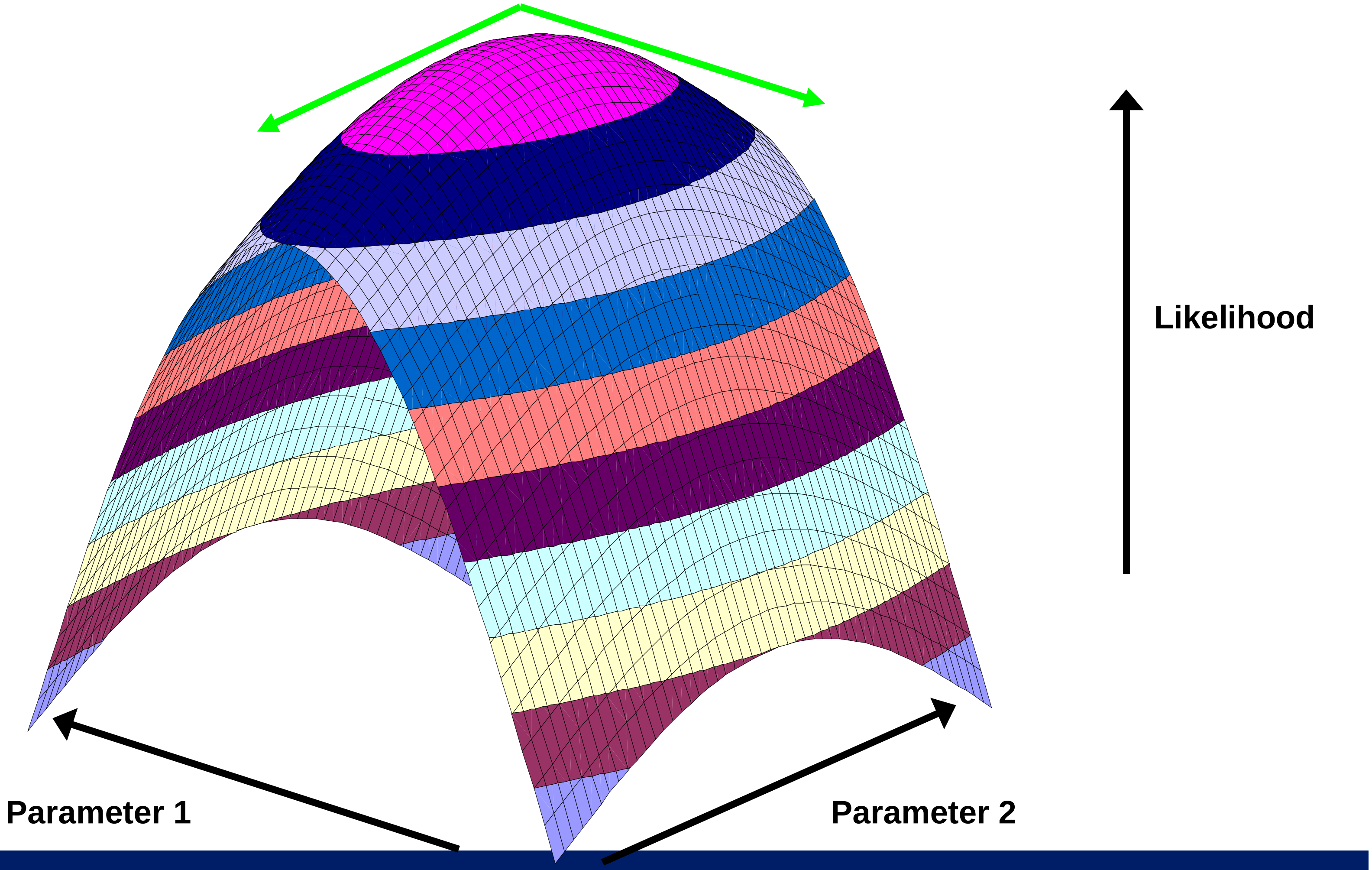
where $\underline{\beta}$ is the vector of the parameter estimates (with *p* elements), $\underline{s}$ is the vector of the first derivatives of the log-likelihood and **H** is the (*p*p*) matrix containing the second derivatives of the log-likelihood

Agenda

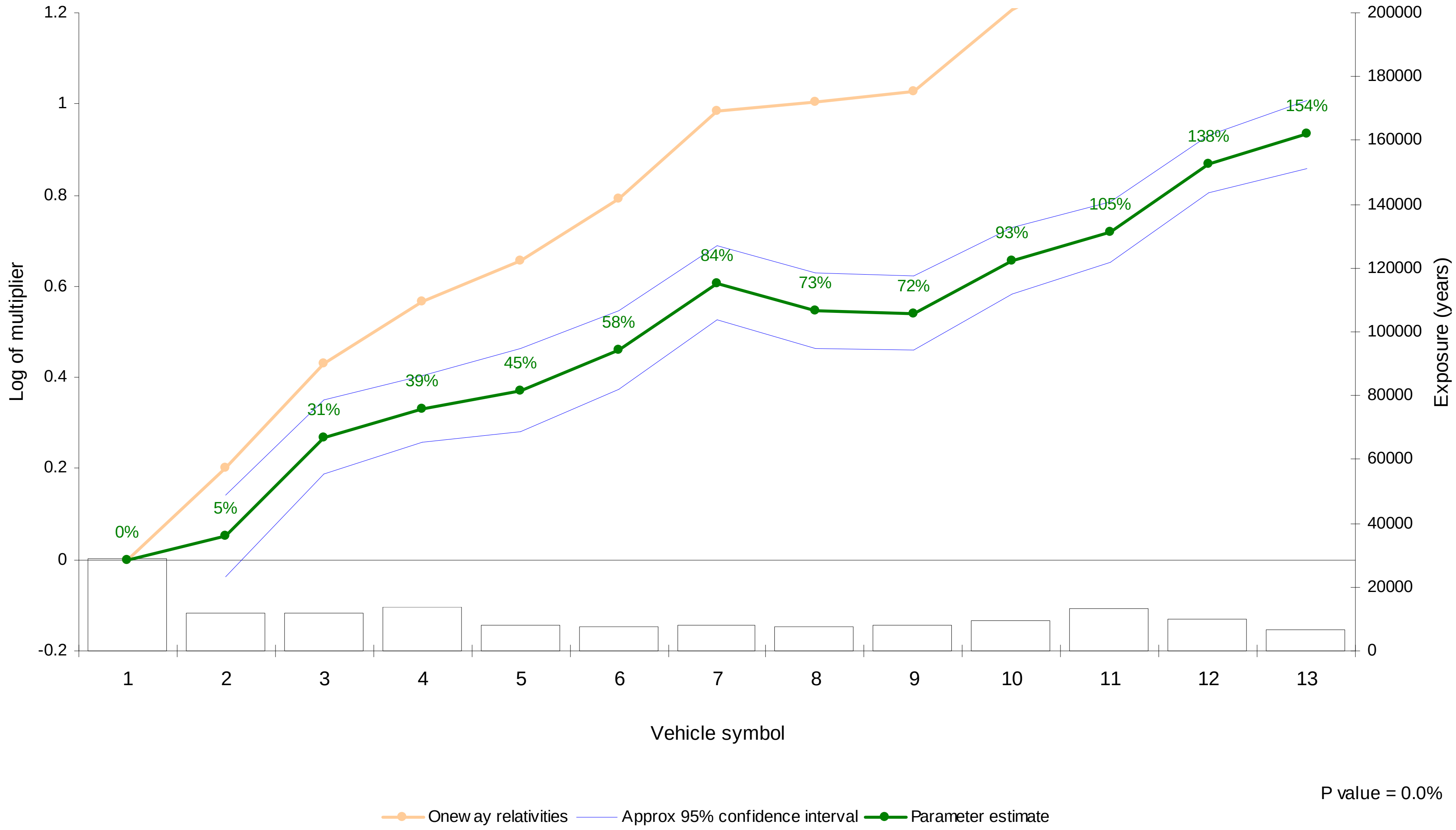
- GLM formulae
- Model testing
  - – use only variables that are predictive
  - – make sure model is reasonable
- Aliasing
- Model refinements

Standard errors

- Roughly speaking, for a parameter p: $SE = -1 / (\partial^2 / \partial p^2 \text{ Likelihood})$

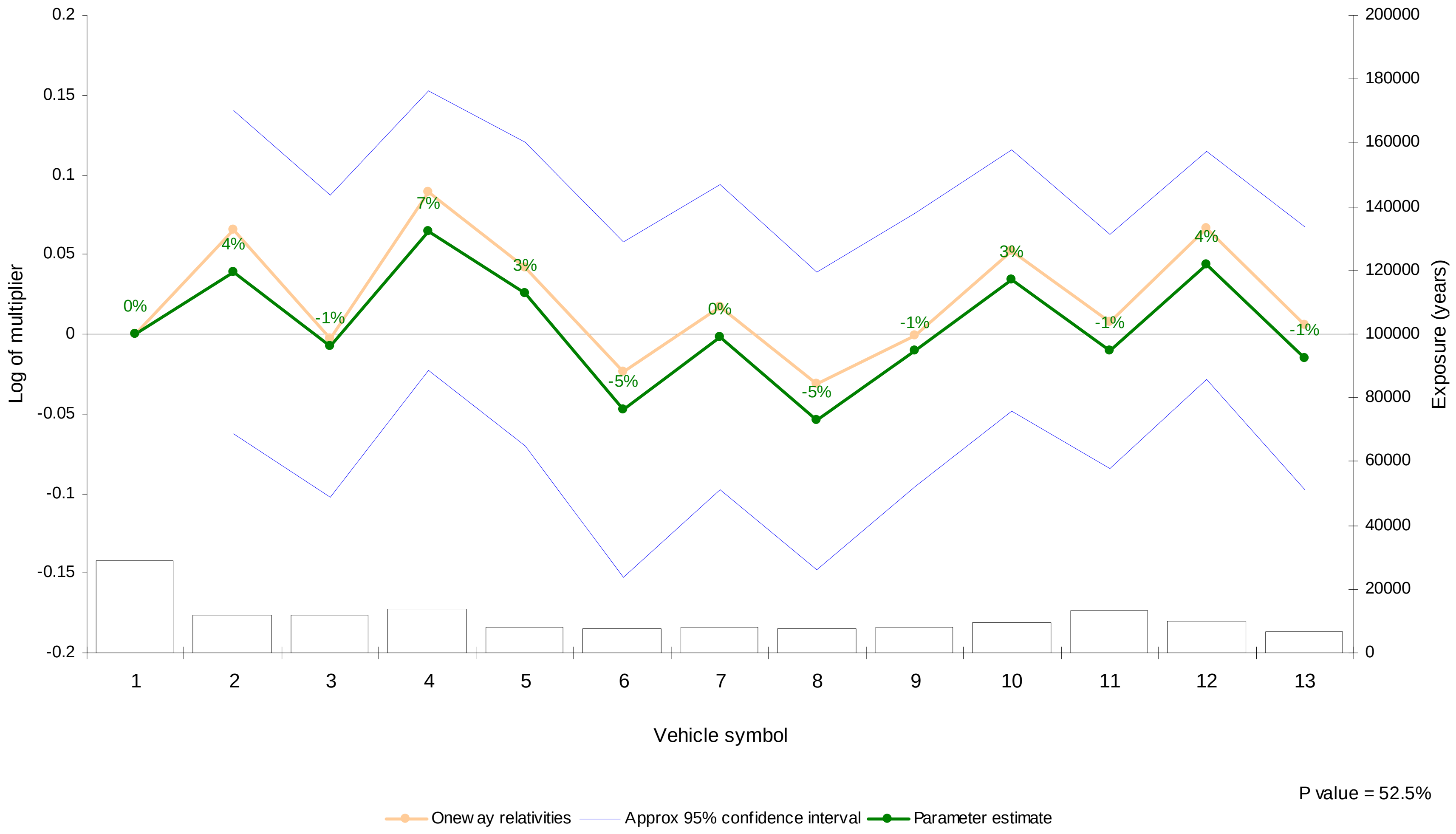


GLM output (significant factor)



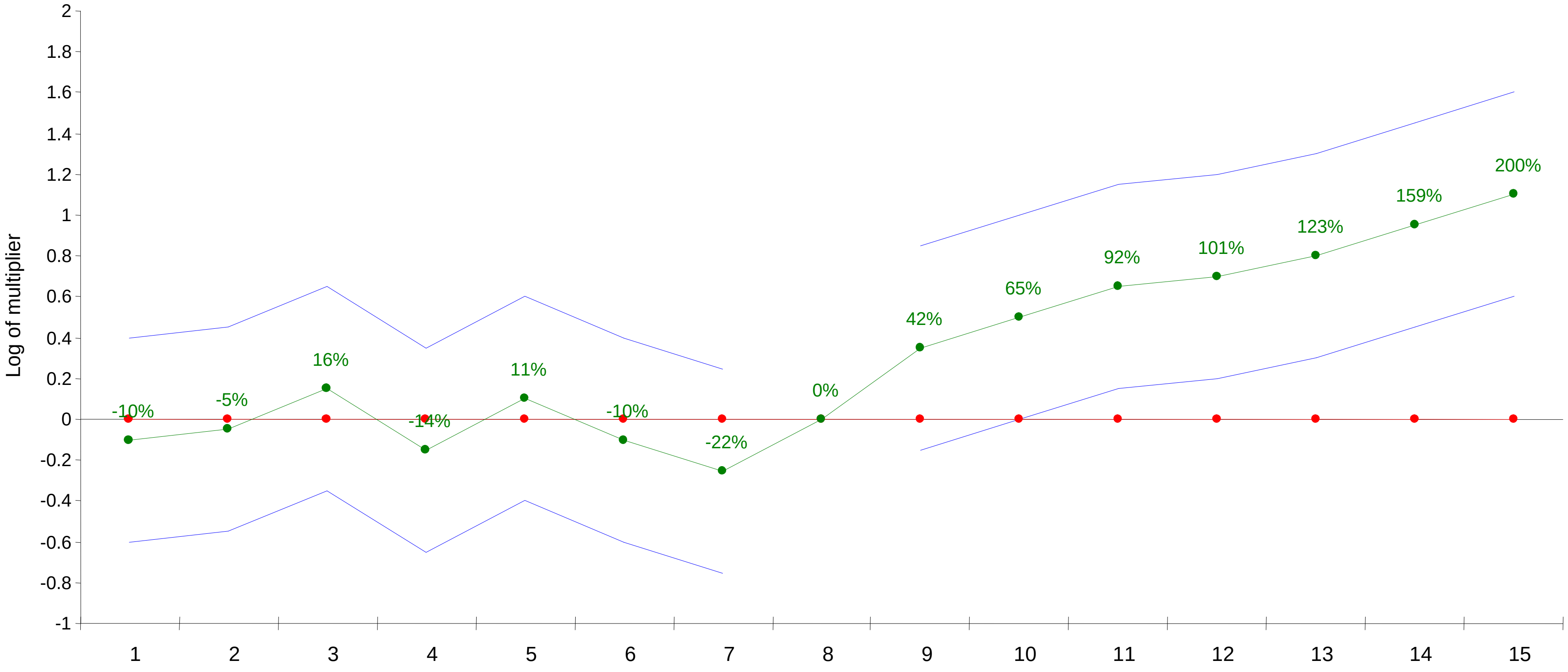
P value = 0.0%

GLM output (insignificant factor)



P value = 52.5%

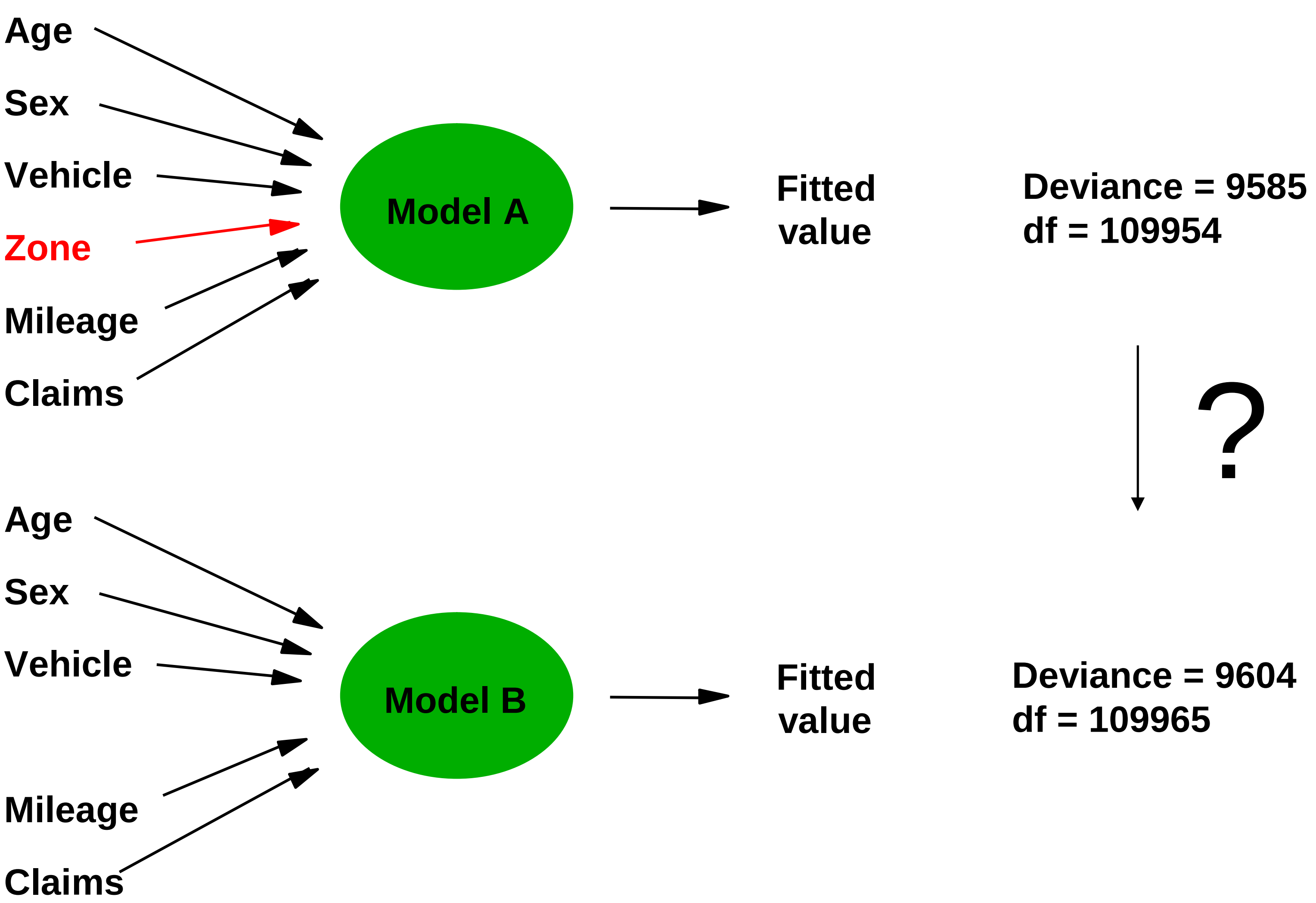
Awkward cases



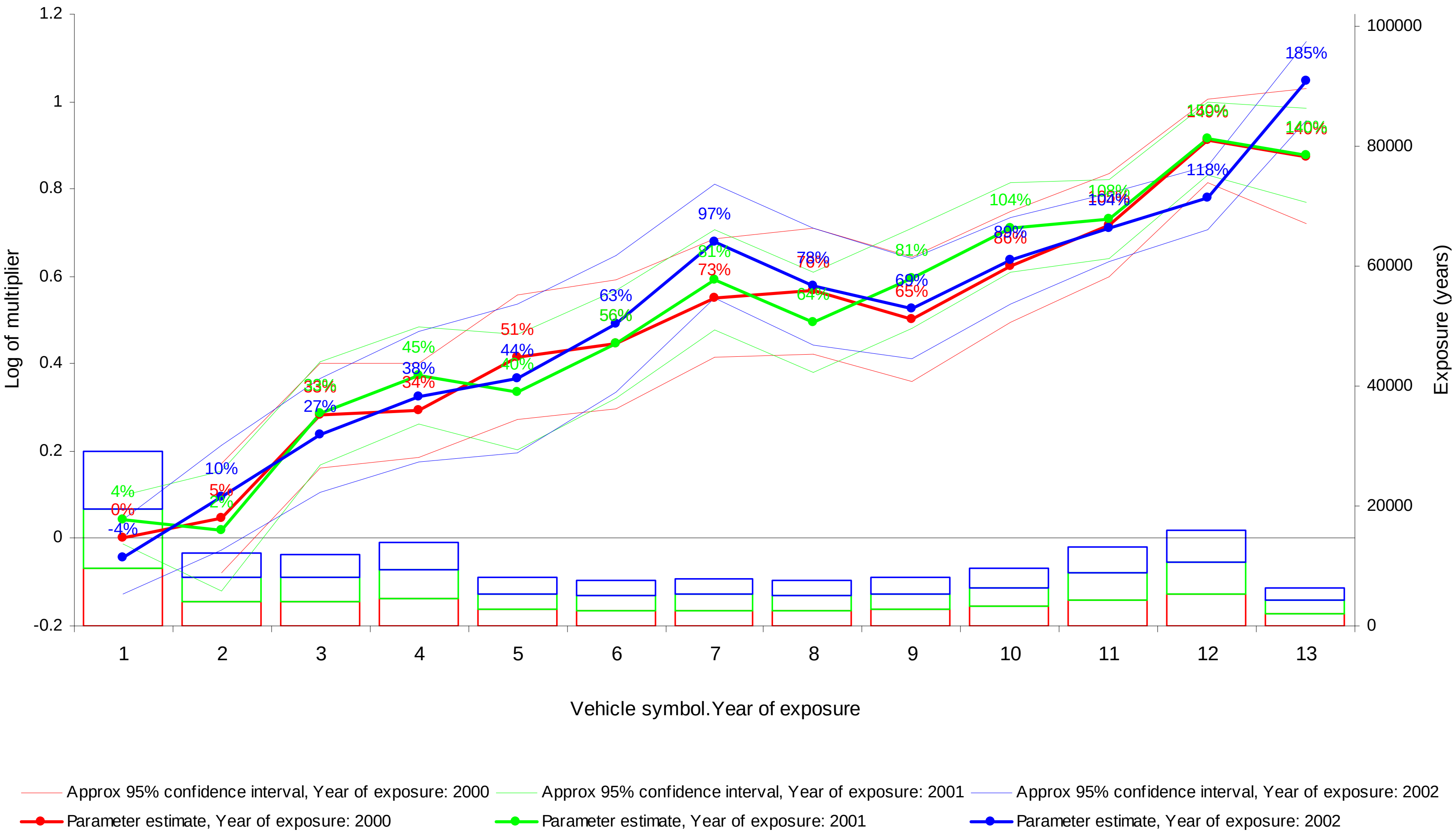
Deviances

- Single figure measure of goodness of fit
- Try model with & without a factor
- Statistical tests show the theoretical significance given the extra parameters

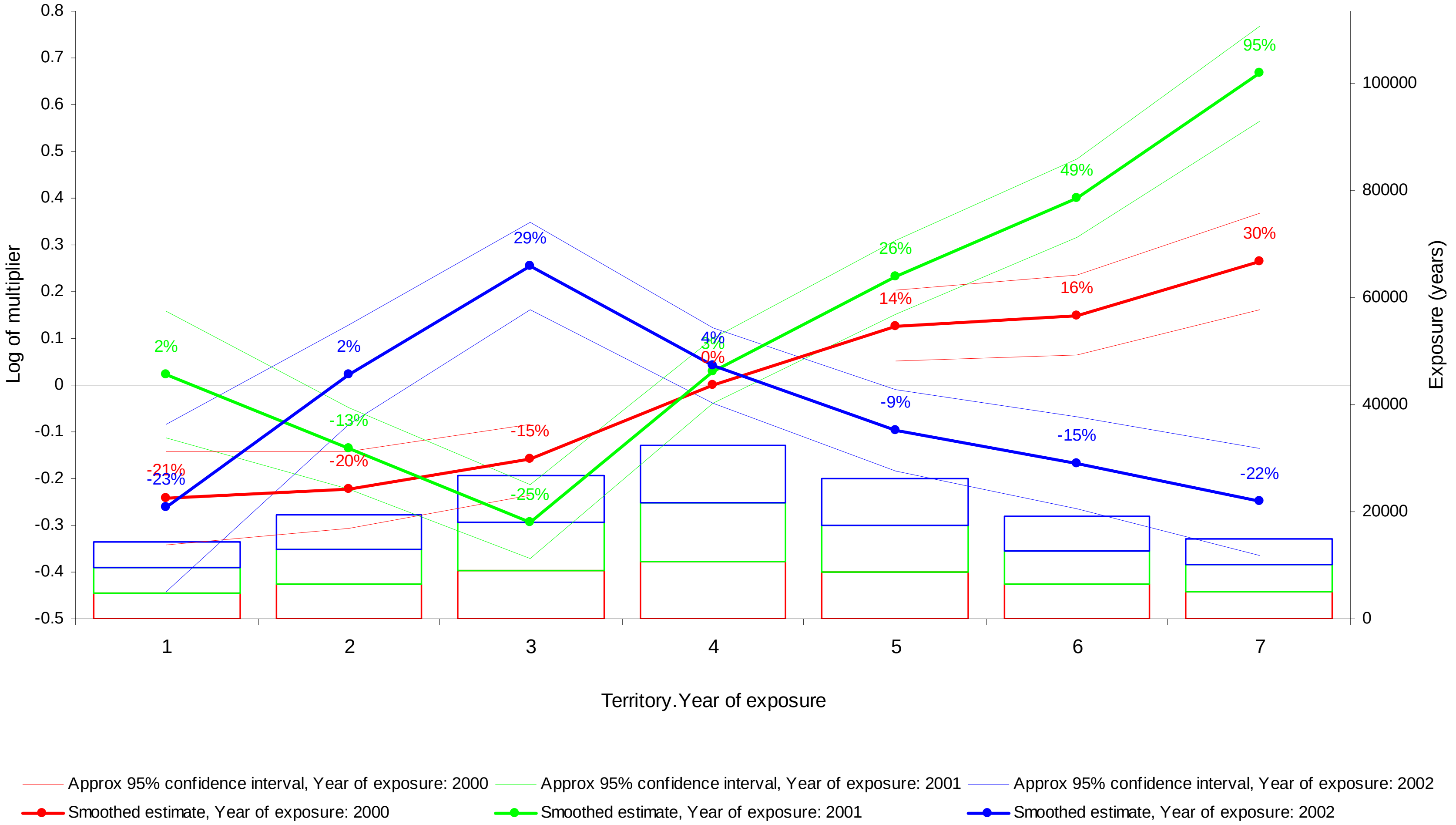
Deviances


Age
Sex
Vehicle
Zone
Mileage
Claims
Model A
Fitted value
Deviance = 9585
df = 109954
?
Age
Sex
Vehicle
Mileage
Claims
Model B
Fitted value
Deviance = 9604
df = 109965

Consistency over time

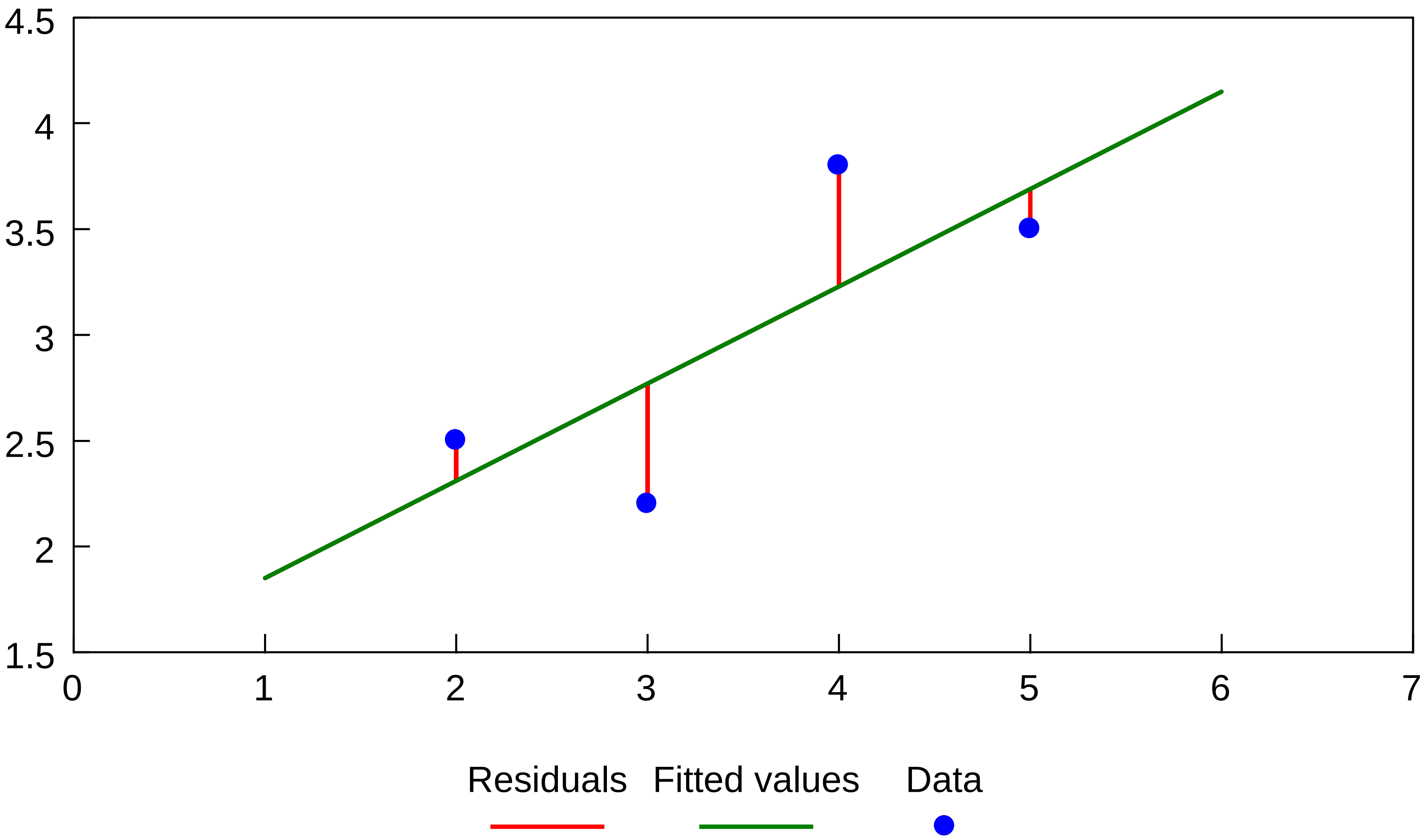


Consistency over time



Watson Wyatt Worldwide

Residuals



Residuals

- Several forms, eg
  - standardized deviance

$$\text{sign}\,(Y_u-\mu_u)\,/\,(\,\phi\,(1-h_u)\,)^{\frac{1}{2}}\sqrt{2\,\omega_u\int_{\mu_u}^{Y_u}(\,Y_u-\zeta\,)\,/\,V(\zeta)\,d\zeta}$$

  - standardized Pearson

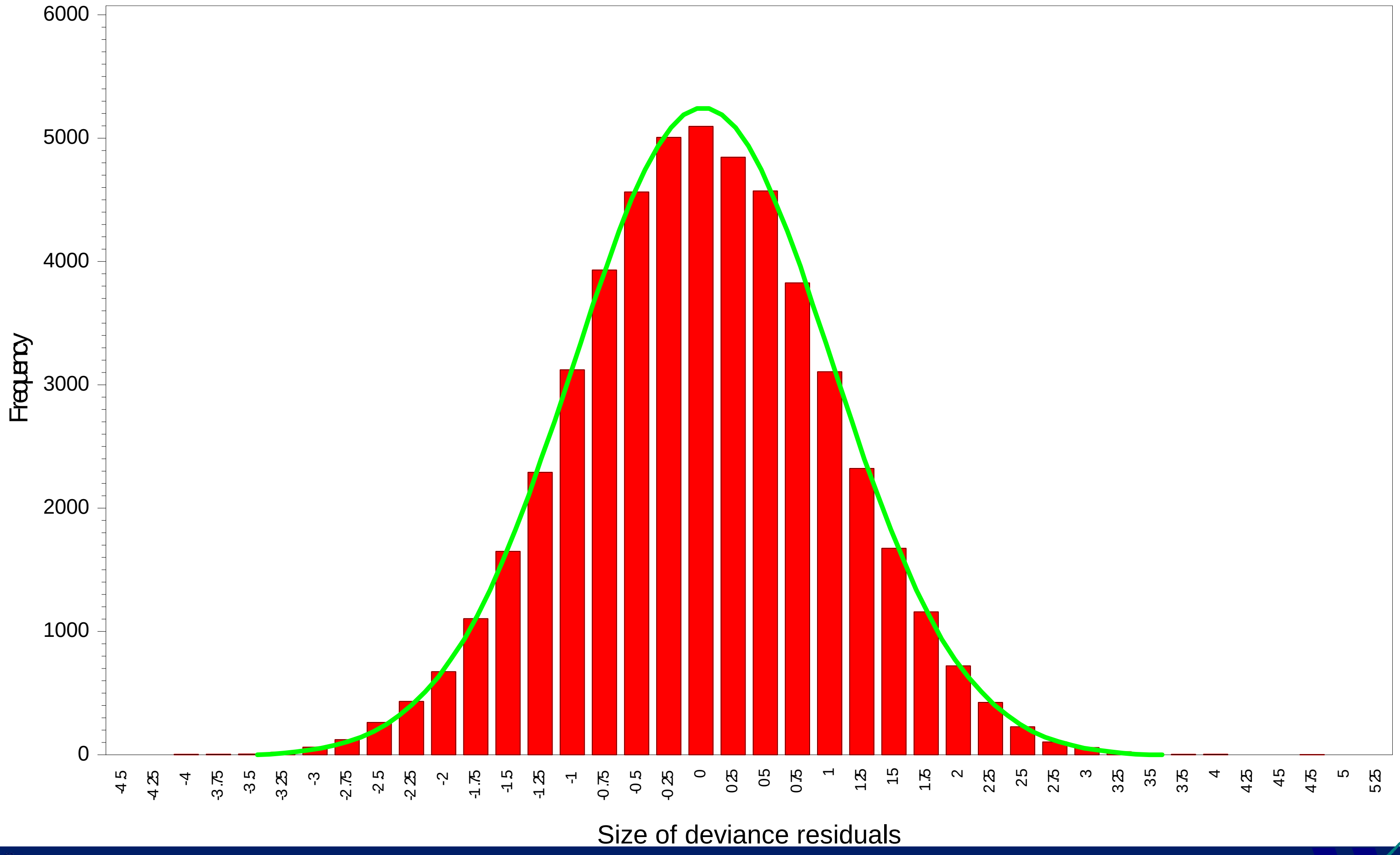
$$\frac{Y_u\ -\ \mu_u}{(\,\phi.V(\mu_u).(1-h_u)\,/\,\omega_u\,)^{\frac{1}{2}}}$$

- Standardized deviance - Normal (0,1)
- Numbers/frequency residuals problematical

Residuals

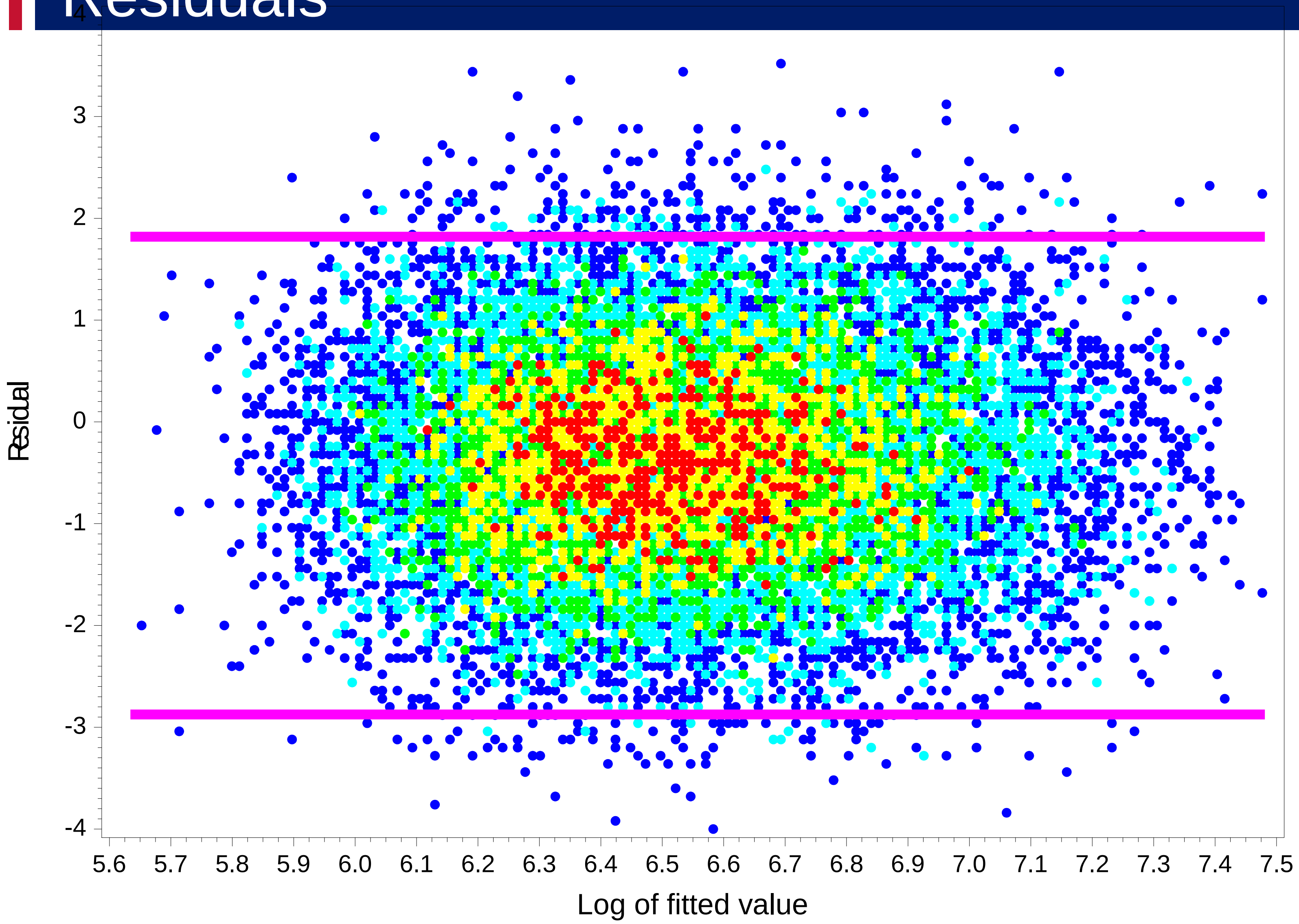
Histogram of Deviance Residuals

Run 12 (Final models with analysis) Model 8 (AD amounts)



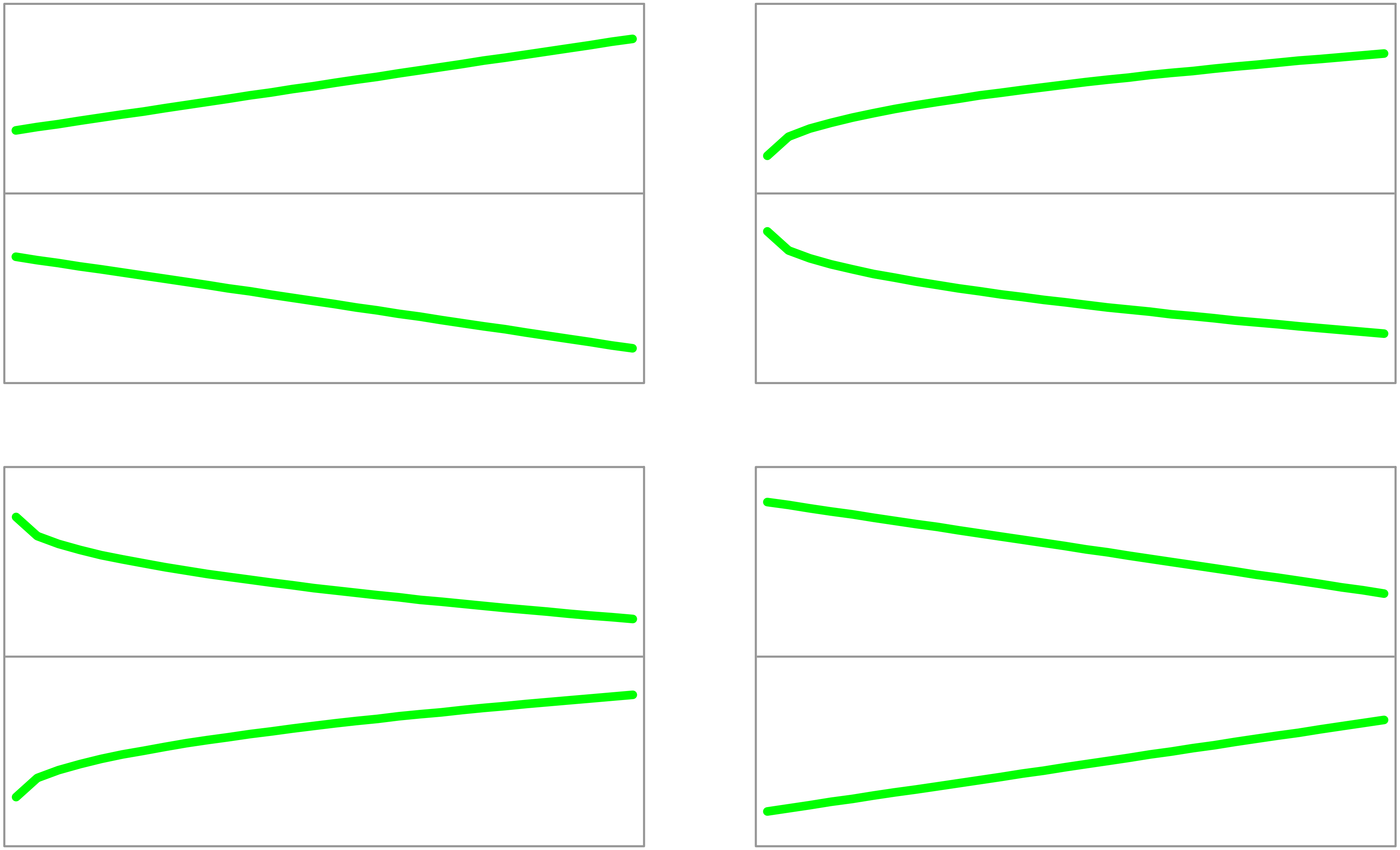
Pretium 08/01/2004 12:24

Residuals



Pretium 04/05/2004 11:03

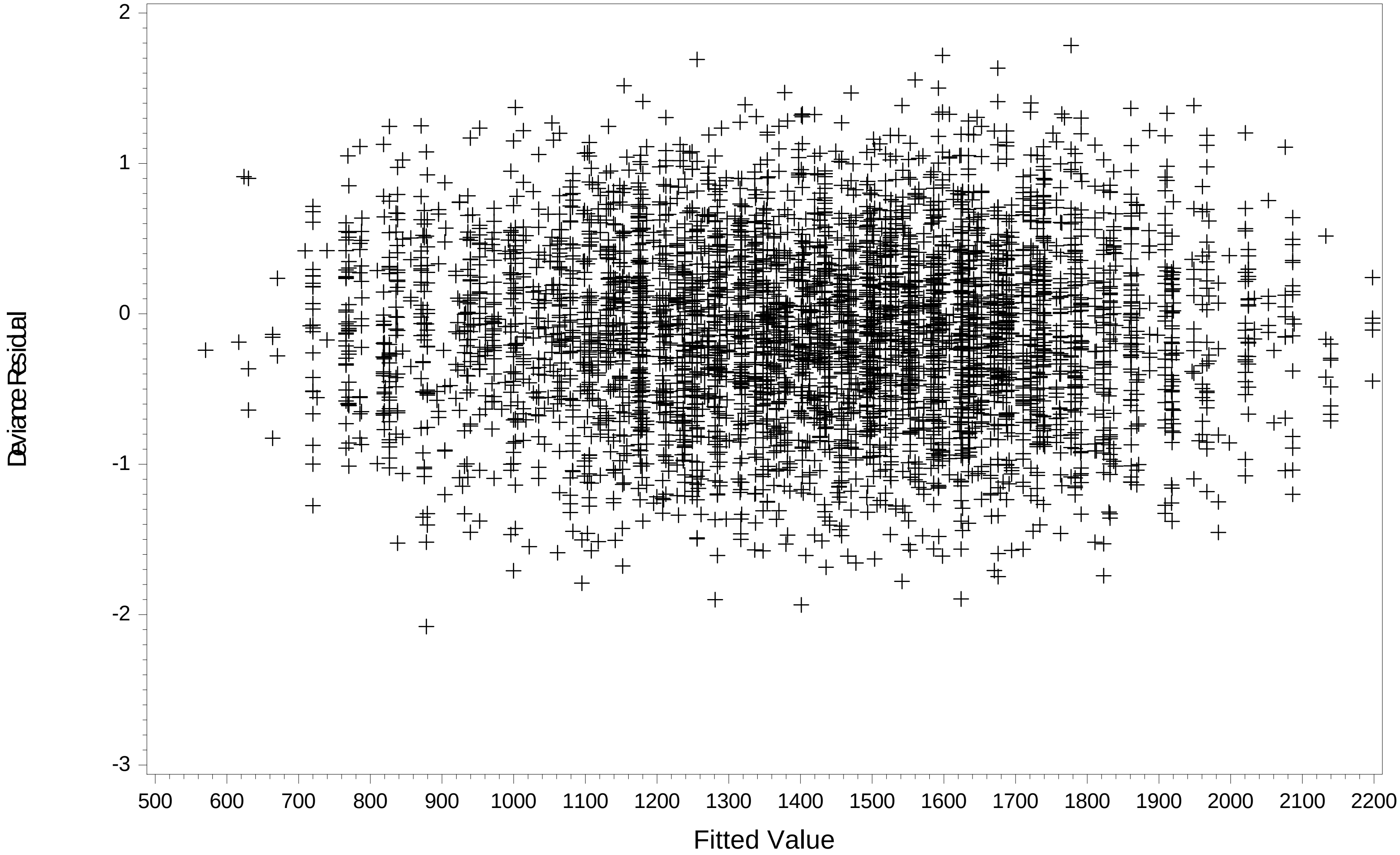
Residuals



Gamma data, Gamma error

Plot of deviance residual against fitted value

Run 12 (All claim types, final models, N&A) Model 6 (Own damage, Amounts)

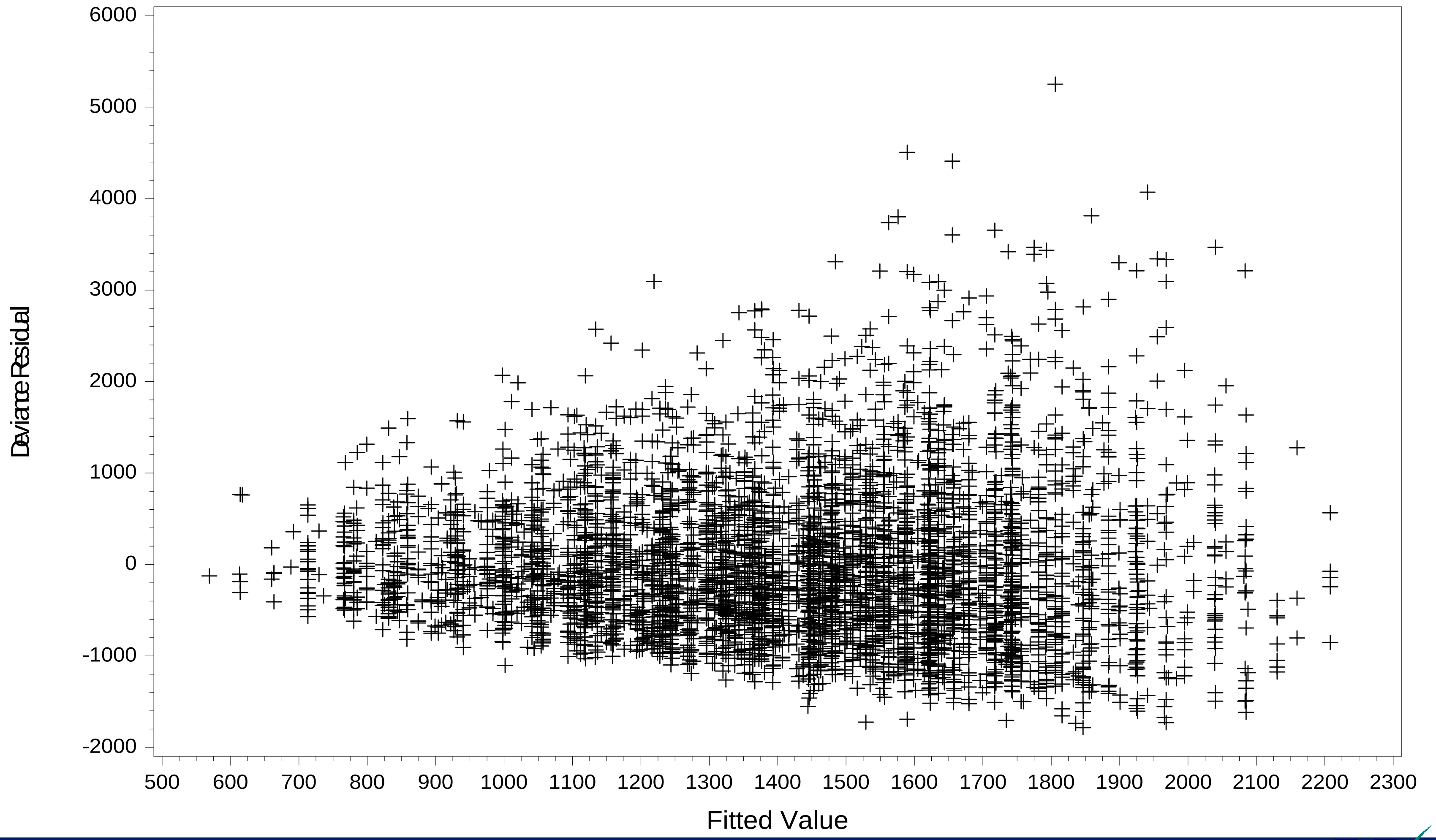


Pretium 08/01/2004 12:32

Gamma data, Normal error

Plot of deviance residual against fitted value

Run 12 (All claim types, final models, N&A) Model 7 (Own damage, Amounts)



Pretium 08/01/2004 12:32

Agenda

- GLM formulae
- Model testing
  - – use only variables that are predictive
  - – make sure model is reasonable
- Aliasing
- Model refinements

Aliasing and "near aliasing"

- Aliasing
  - the removal of unwanted redundant parameters
- Intrinsic aliasing
  - occurs by the design of the model
- Extrinsic aliasing
  - occurs "accidentally" as a result of the data

Intrinsic aliasing

$\mathbf{X}.\underline{\beta} = \alpha + \beta_1$ if $\underline{\text{age}}$ 20 - 29

~~$+ \beta_2$ if $\underline{\text{age}}$ 30 - 39~~

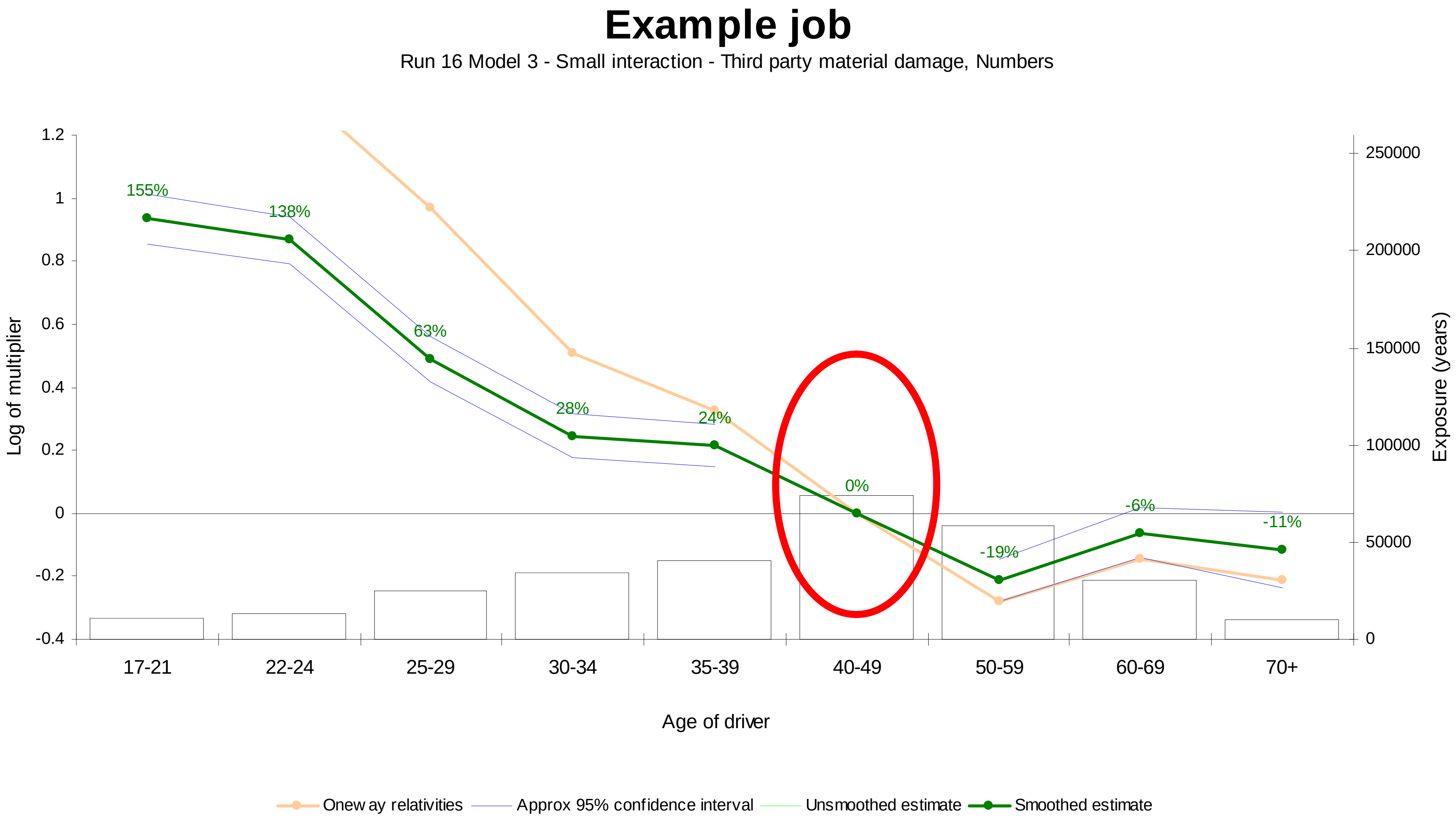
$+ \beta_3$ if $\underline{\text{age}}$ 40 +

~~$+ \gamma_1$ if $\underline{\text{sex}}$ male~~

$+ \gamma_2$ if $\underline{\text{sex}}$ female

- "Base levels"

Intrinsic aliasing



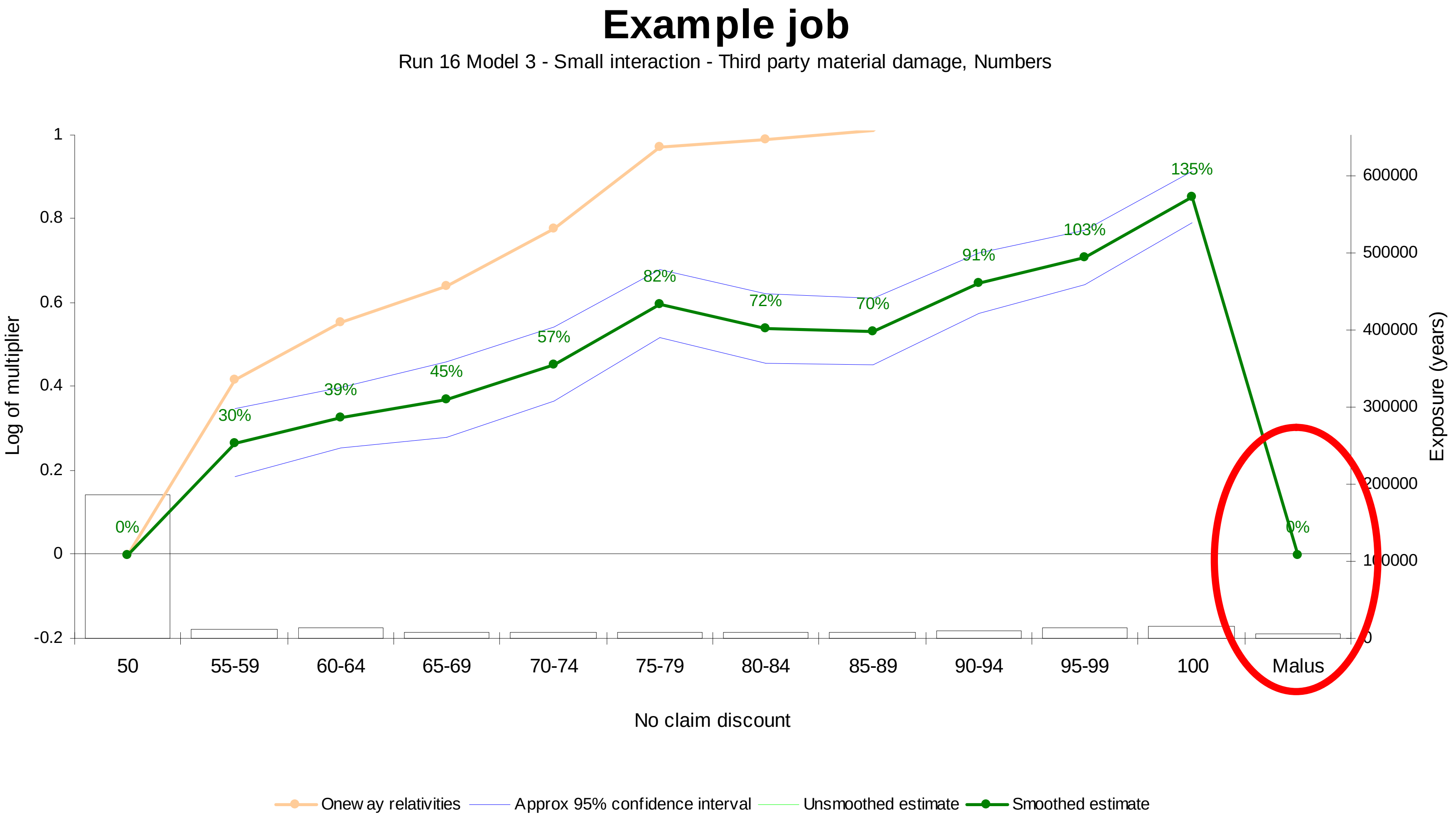
Extrinsic aliasing

- If a perfect correlation exists, one factor can alias levels of another
- Eg if doors declared first:

	Exposure: # Doors →	2	3	Selected base ~~4~~	5	Unknown
	Color ↓					
Selected base	~~Red~~	13,234	12,343	13,432	13,432	0
	Green	4,543	4,543	13,243	2,345	0
	Blue	6,544	5,443	15,654	4,565	0
	Black	4,643	1,235	14,565	4,545	0
Further aliasing	~~Unknown~~	0	0	0	0	3,242

- This is the only reason the order of declaration can matter (fitted values are unaffected)

Extrinsic aliasing



"Near aliasing"

- If two factors are almost perfectly aliased, convergence problems can result as a result of low exposures and/or results can become hard to interpret

Exposure: # Doors → Color ↓	2	3	Selected base ~~4~~	5	Unknown
Selected base ~~Red~~	13,234	12,343	13,432	13,432	0
Green	4,543	4,543	13,243	2,345	0
Blue	6,544	5,443	15,654	4,565	0
Black	4,643	1,235	14,565	4,545	2
Unknown	0	0	0	0	3,242

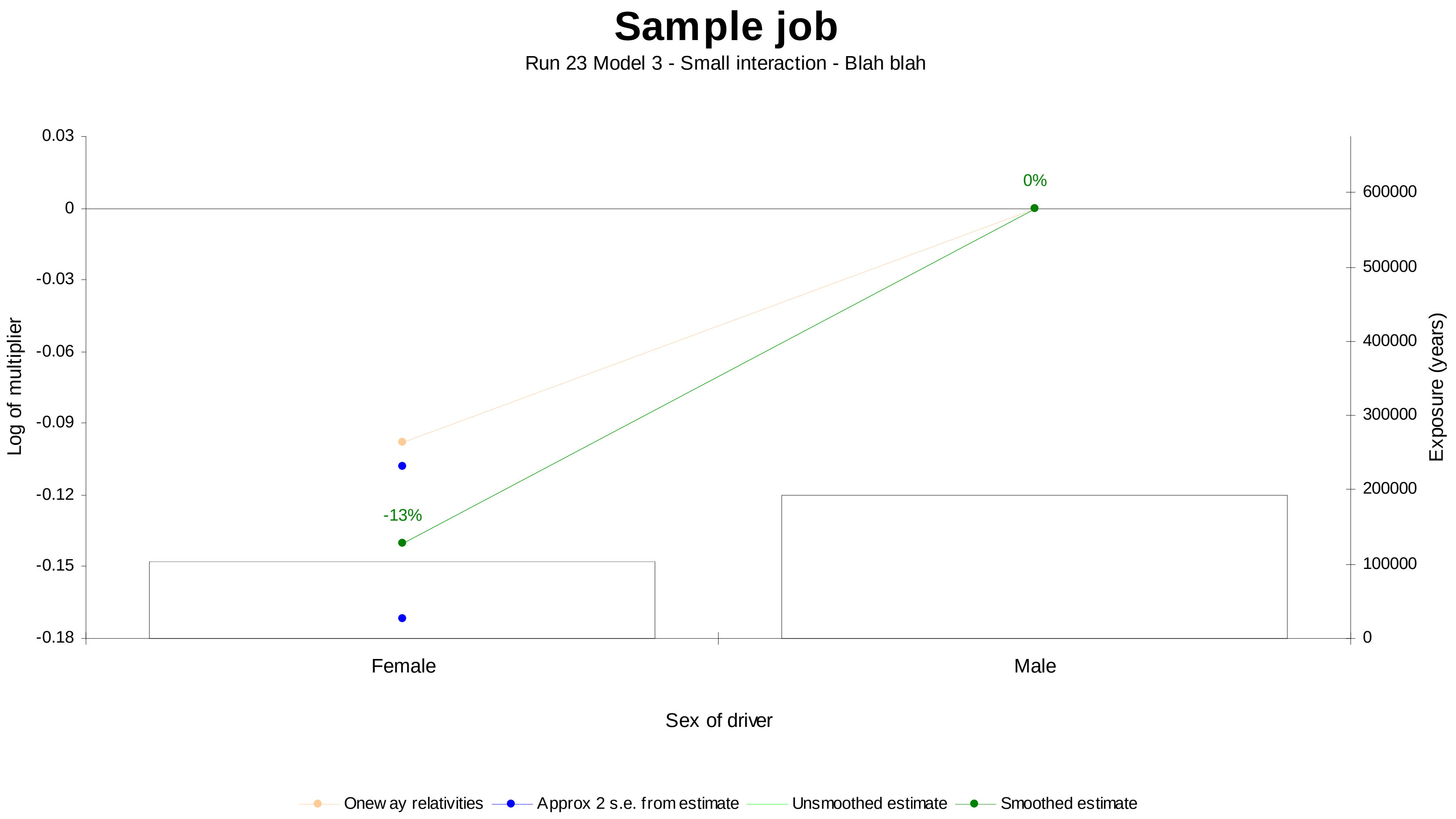
- Eg if the 2 black, unknown doors policies had no claims, GLM would try to estimate a very large negative number for unknown doors, and a very large positive number for unknown color

Watson Wyatt Worldwide

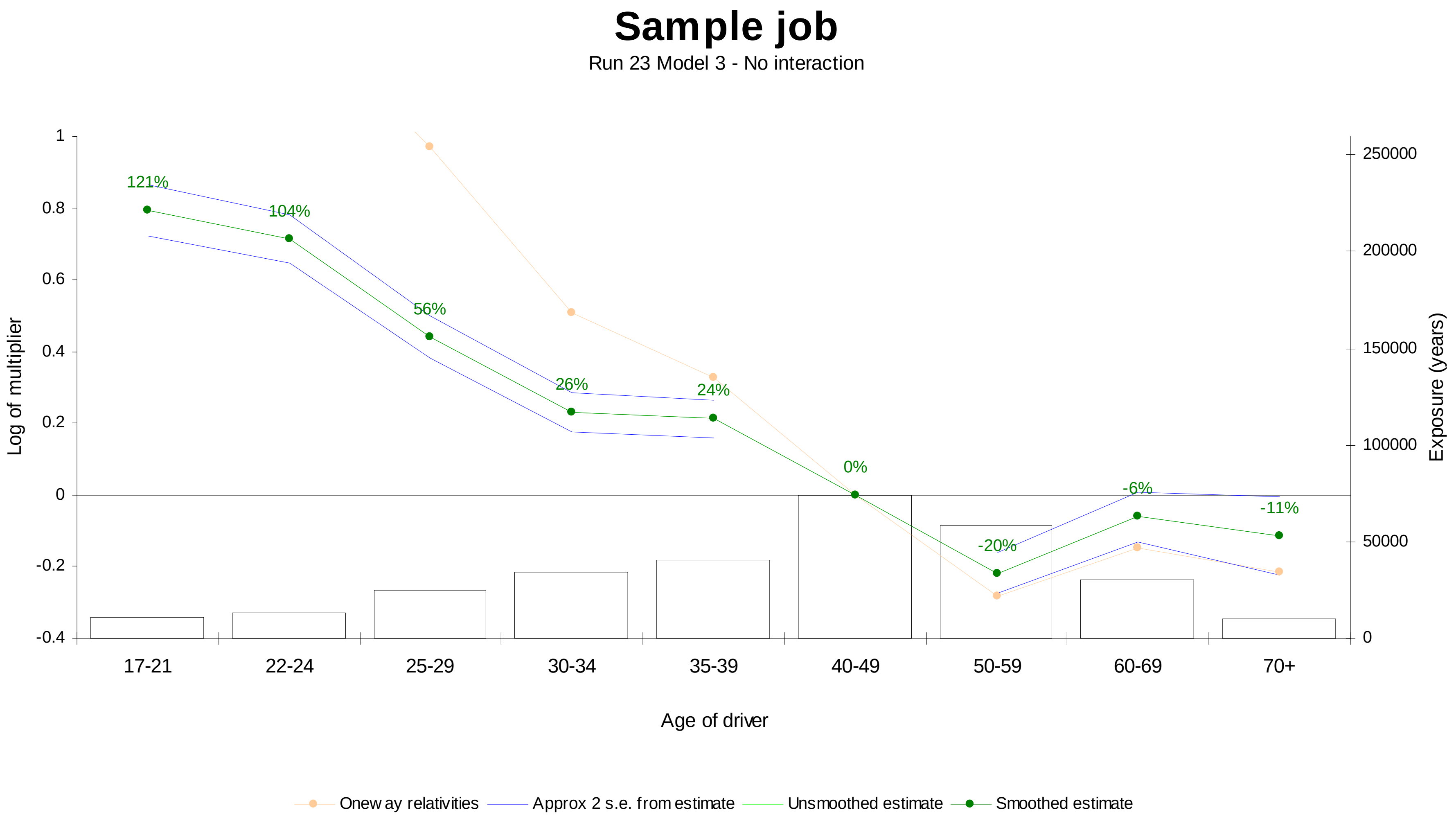
Agenda

- GLM formulae
- Model testing
  - use only variables that are predictive
  - make sure model is reasonable
- Aliasing
- Model refinements
  - Interactions
  - Splines
  - Restrictions

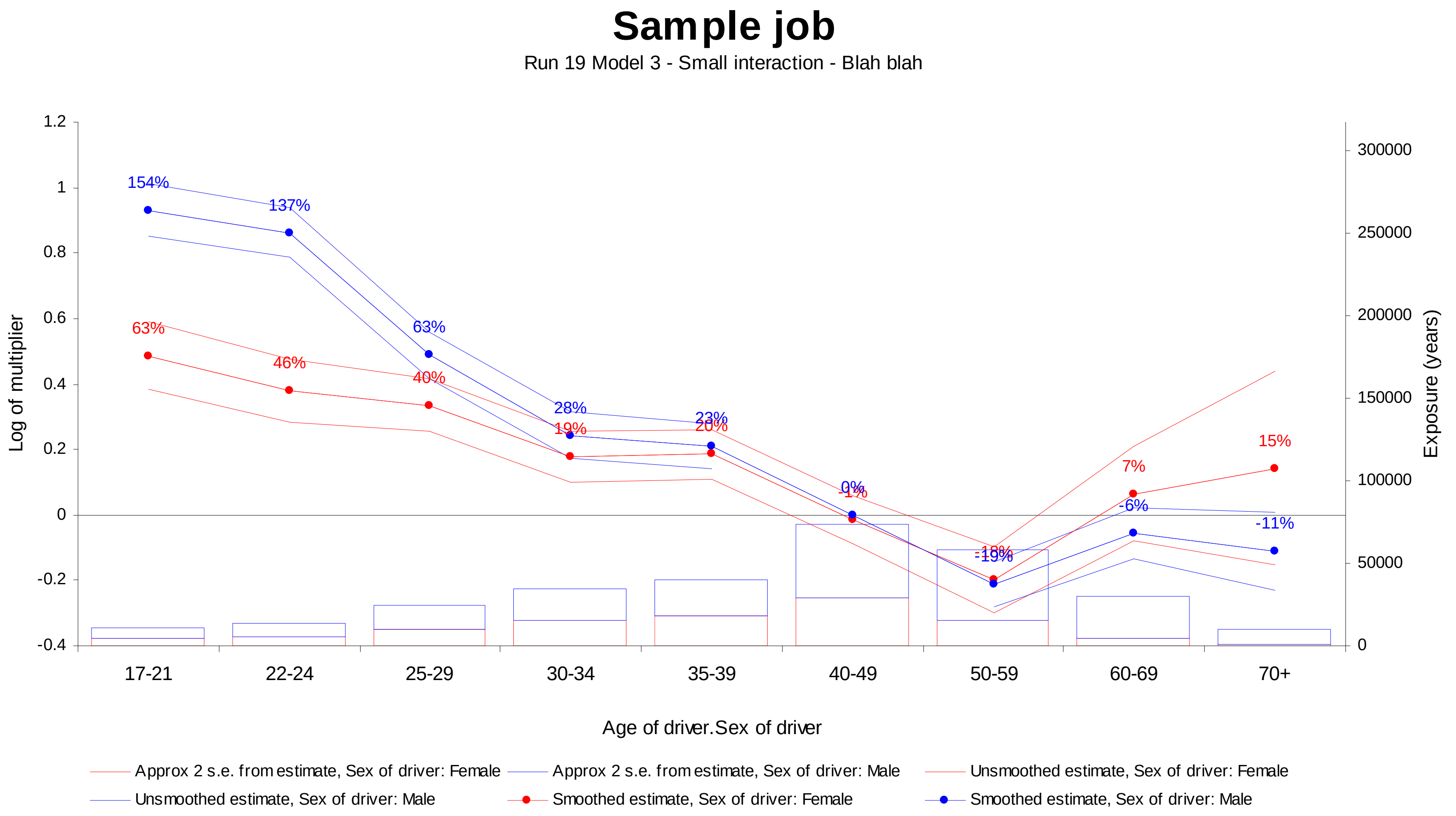
Interactions



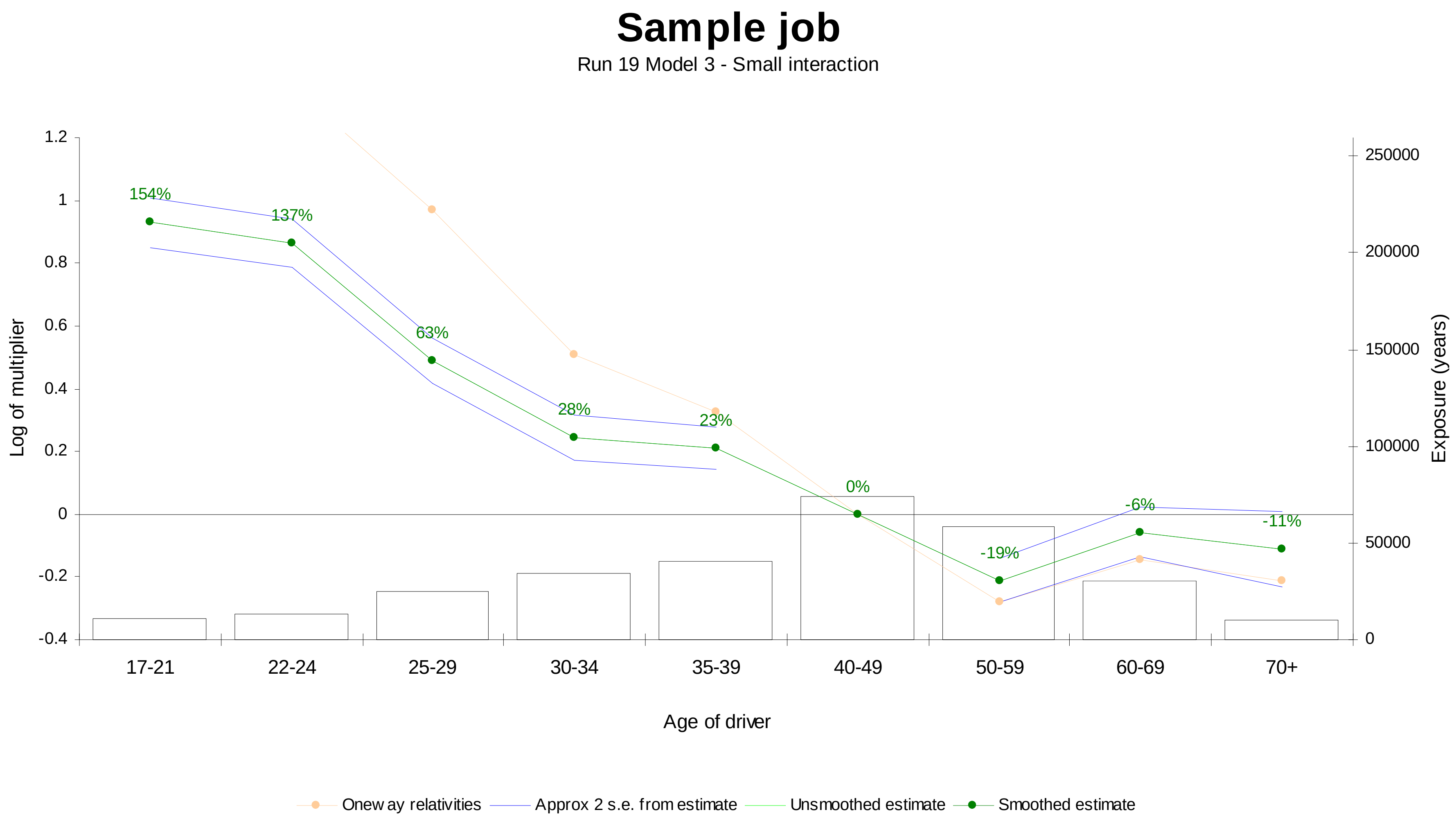
Interactions



Interactions

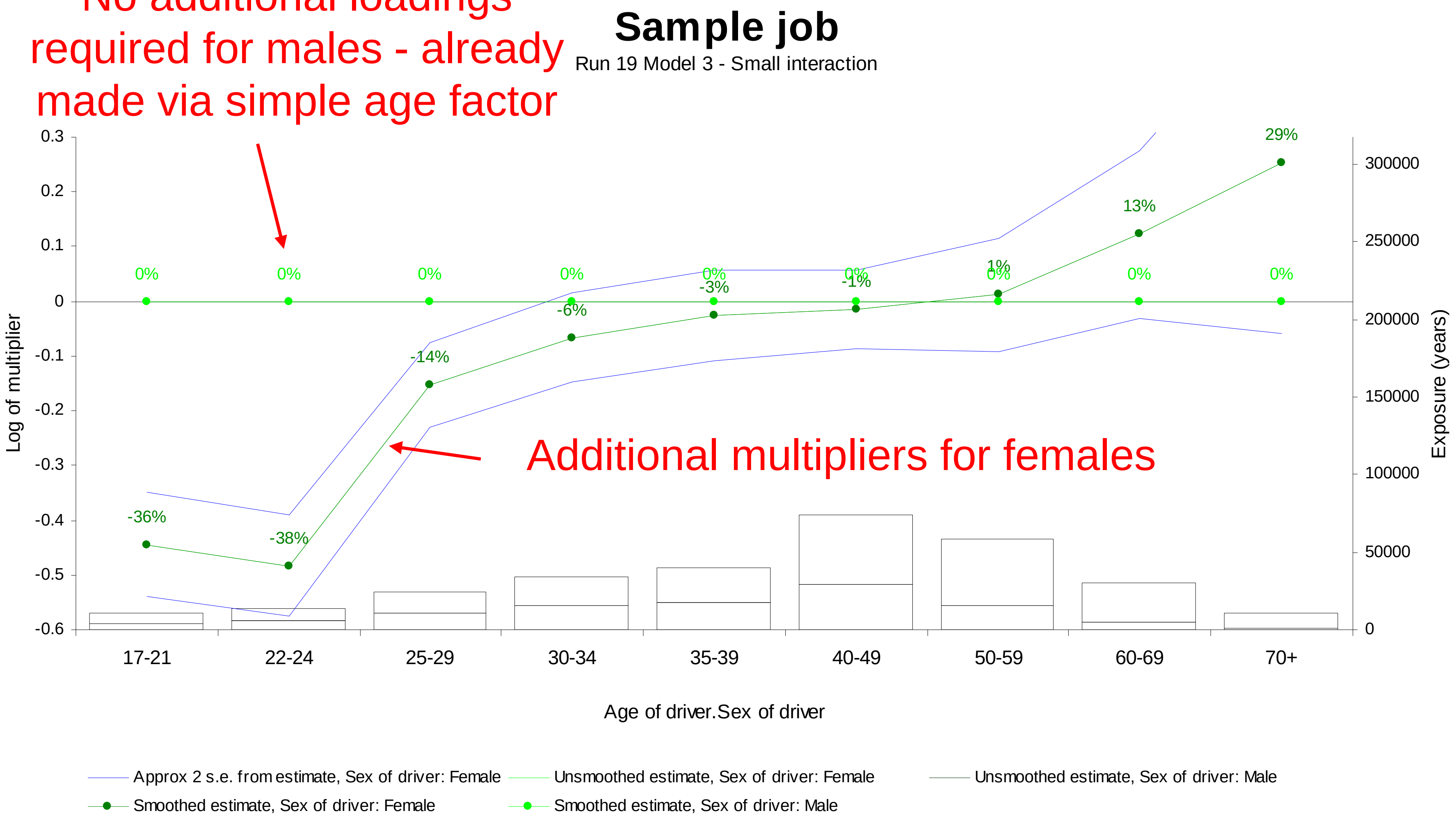


Marginal interaction:
Age effect



Marginal interaction: Age.Sex (ie additional female multipliers)

No additional loadings required for males - already made via simple age factor



Additional multipliers for females

Interactions

Group >	1	2	3	4	5	6	7	8	9	10	11	12	13
Age v													
17	1.36	1.64	1.79	2.09	2.27	2.42	2.56	2.65	3.27	3.71	4.08	4.36	4.84
18	1.12	1.31	1.47	1.76	1.84	2.00	2.11	2.19	2.43	2.97	3.29	3.55	3.90
19	1.08	1.30	1.46	1.63	1.82	1.91	2.02	2.11	2.53	2.88	3.30	3.35	3.63
20	0.98	1.18	1.36	1.54	1.68	1.79	1.83	1.97	2.19	2.66	3.02	3.20	3.38
21-23	0.96	1.13	1.24	1.51	1.65	1.64	1.80	1.85	2.04	2.26	2.55	2.53	2.89
24-26	0.82	0.99	1.10	1.31	1.43	1.52	1.51	1.64	1.81	1.93	2.13	2.22	2.47
27-30	0.78	0.90	1.07	1.19	1.32	1.39	1.41	1.51	1.65	1.77	1.91	2.01	2.24
31-35	0.63	0.78	0.86	0.99	1.09	1.17	1.22	1.32	1.42	1.54	1.66	1.71	1.88
36-40	0.55	0.64	0.71	0.85	0.91	0.93	0.99	1.07	1.18	1.29	1.40	1.41	1.53
41-45	0.51	0.61	0.66	0.79	0.88	0.88	0.94	0.99	1.09	1.15	1.29	1.31	1.42
46-50	0.46	0.55	0.61	0.70	0.76	0.81	0.84	0.92	1.02	1.07	1.12	1.18	1.31
51-60	0.40	0.49	0.56	0.64	0.68	0.71	0.78	0.82	0.90	0.99	1.02	1.12	1.20
60+	0.43	0.52	0.55	0.67	0.72	0.73	0.78	0.83	0.93	0.98	1.04	1.11	1.25

Interactions

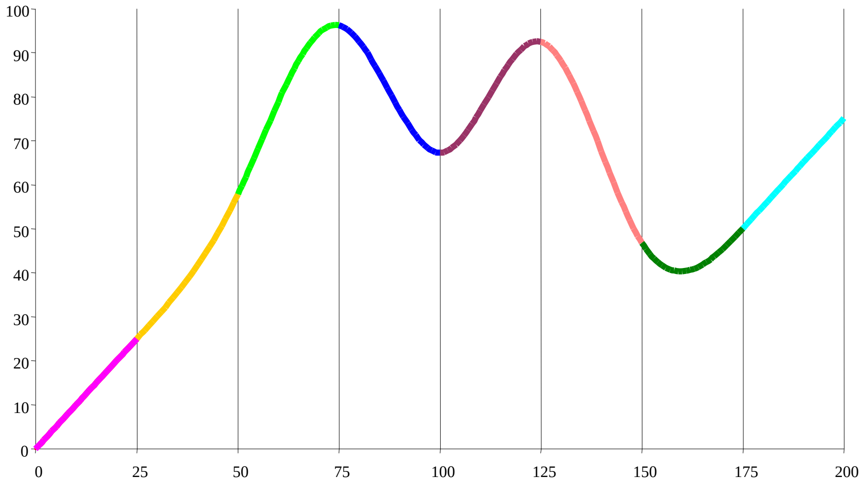
Group	1	2	3	4	5	6	7	8	9	10	11	12	13
Factor	0.54	0.65	0.73	0.85	0.92	0.96	1.00	1.08	1.19	1.26	1.36	1.43	1.56

Age	Factor
17	2.52
18	2.05
19	1.97
20	1.85
21-23	1.75
24-26	1.54
27-30	1.42
31-35	1.20
36-40	1.00
41-45	0.93
46-50	0.84
51-60	0.76
60+	0.78

1.00	1.17
1.00	1.00

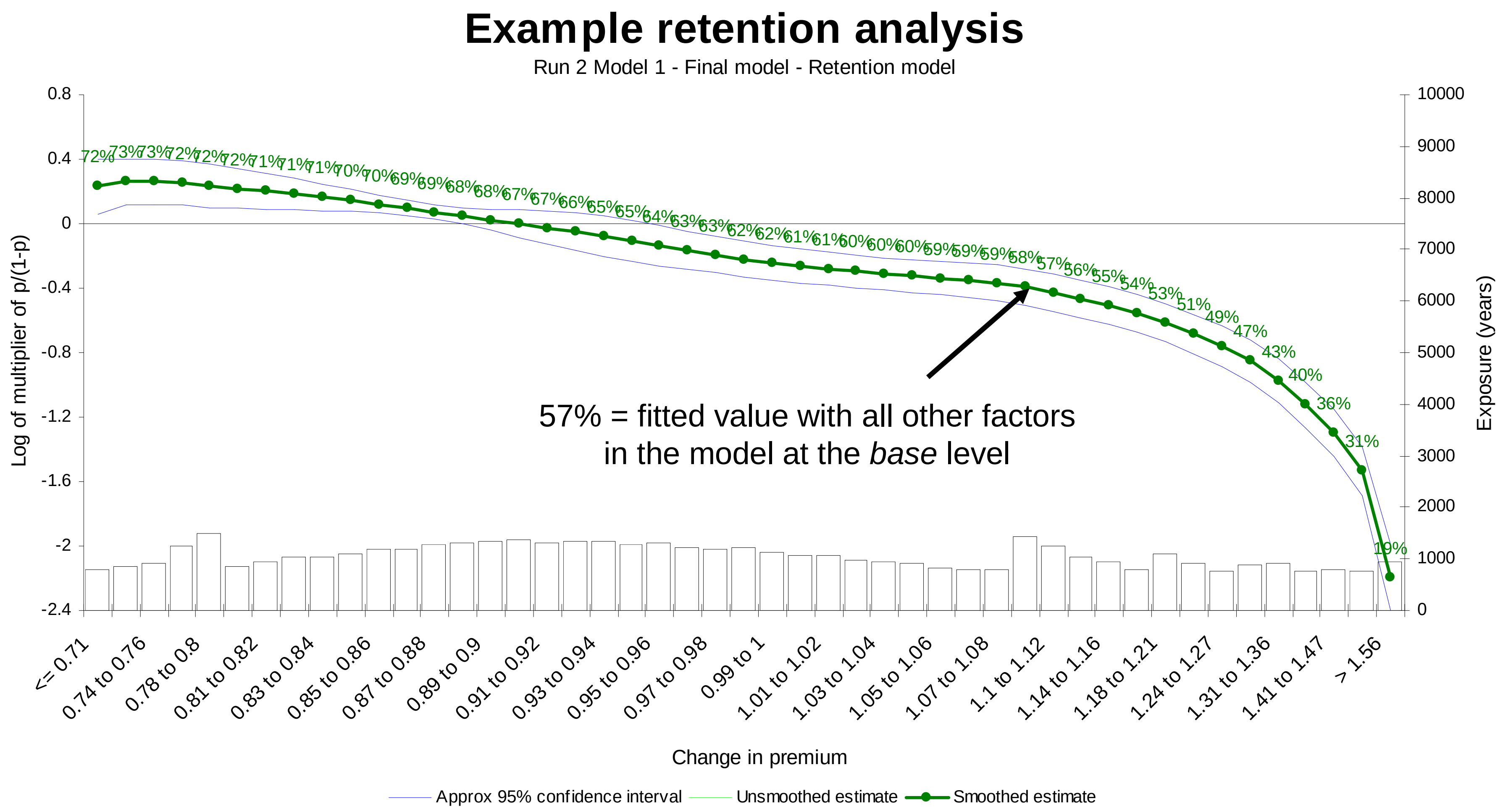
Spline definition

- A series of polynomial functions, with each function defined over a short interval



- Intervals are defined by k+2 knots
  - two exterior knots at extremes of data
  - variable number (k) of interior knots
- At each interior knot the two functions must join "smoothly“
- Regression splines are a form of generalized additive models

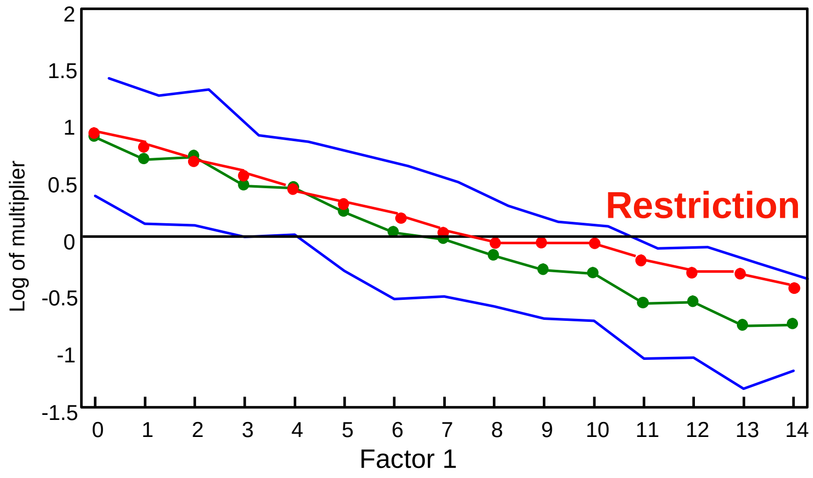
Example retention elasticity curve

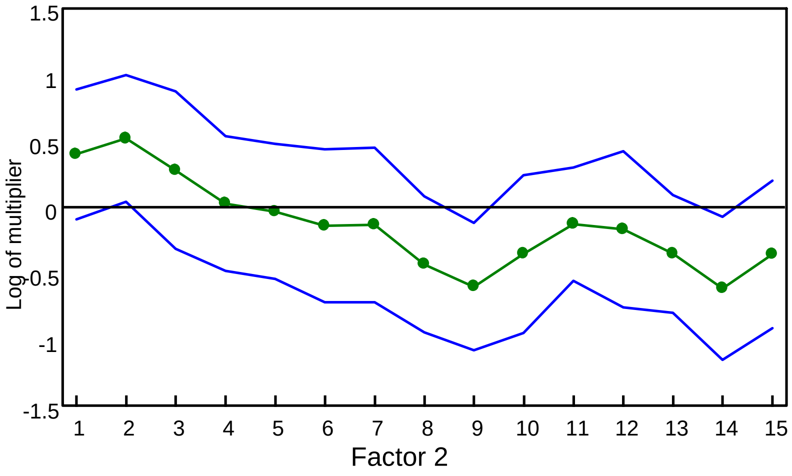


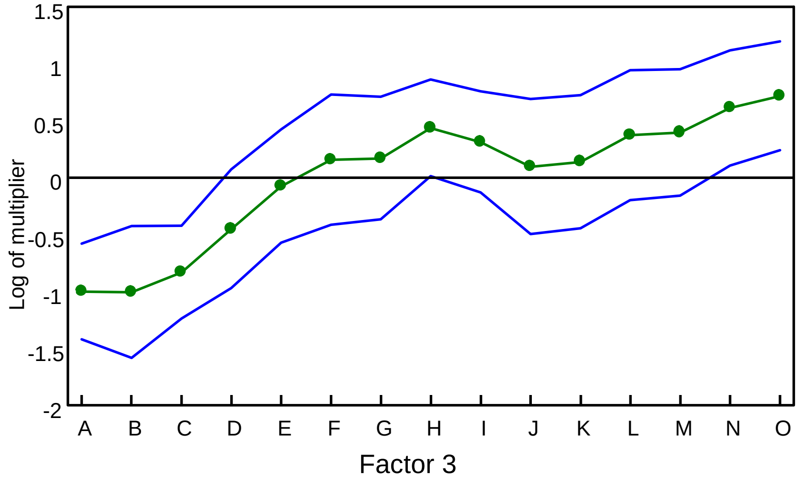
57% = fitted value with all other factors in the model at the *base* level

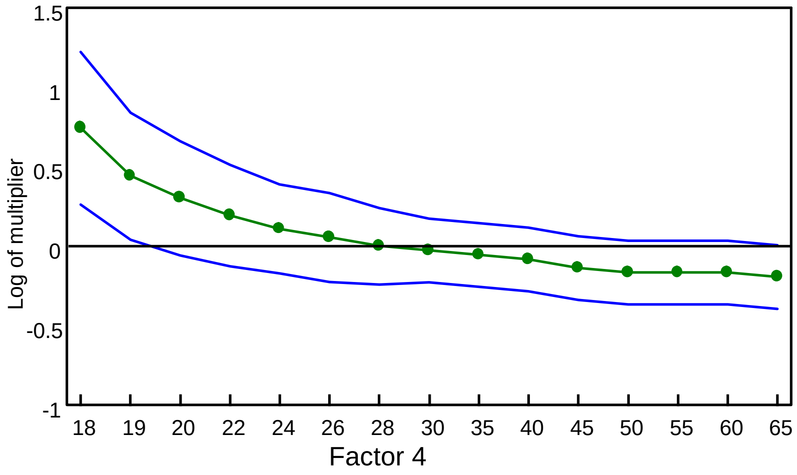
Watson Wyatt Worldwide

Restricted models

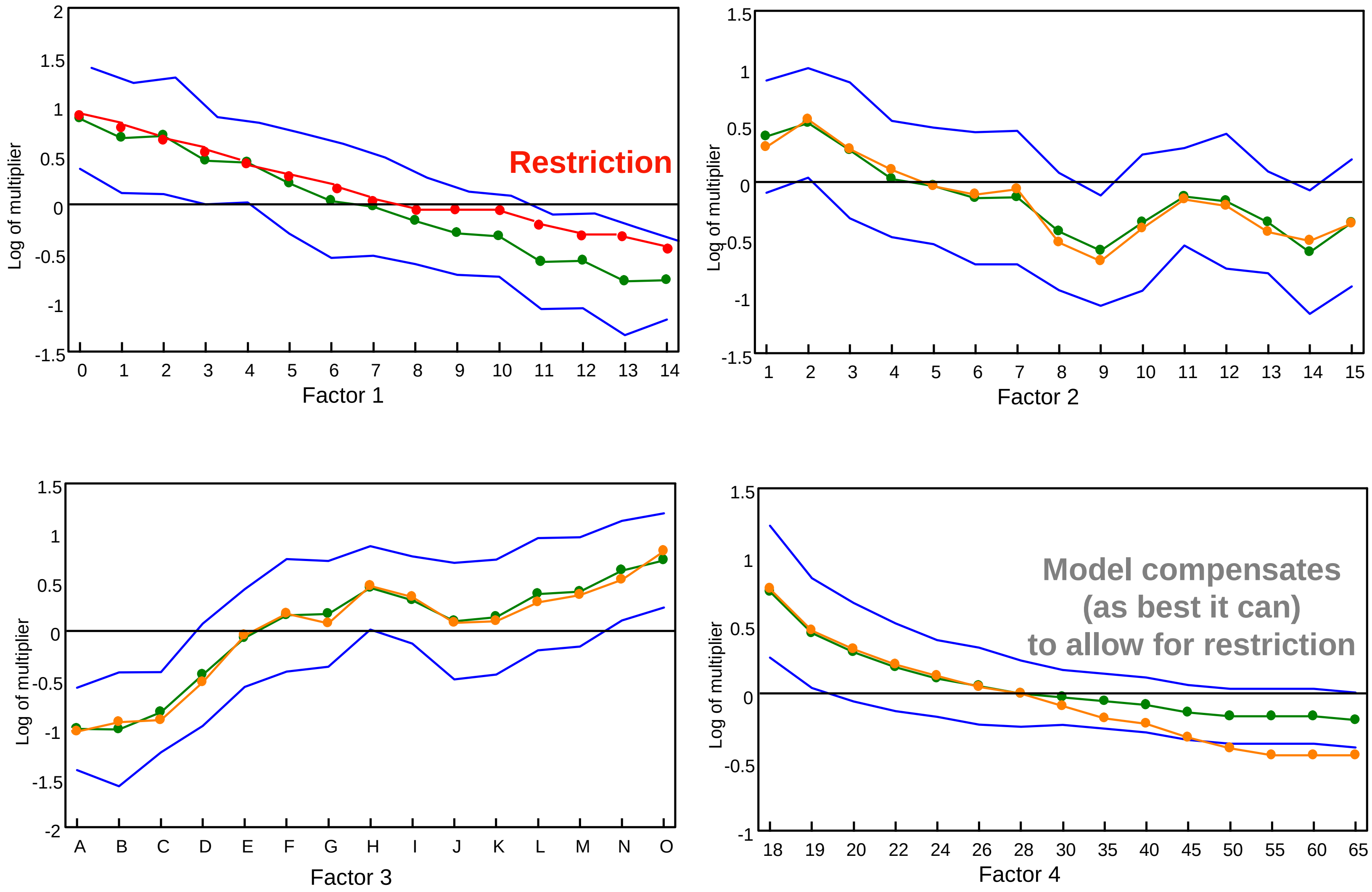








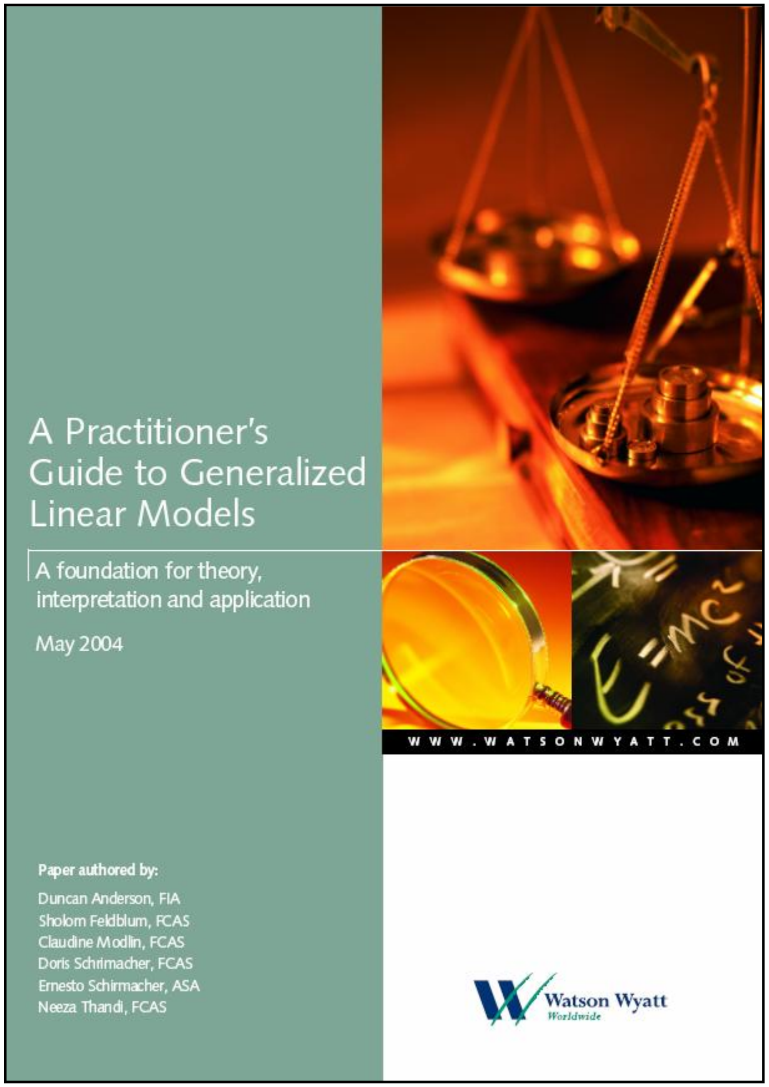
Restricted models



Restriction

Model compensates
(as best it can)
to allow for restriction

"A Practitioner's Guide to Generalized Linear Models"



A Practitioner's Guide to Generalized Linear Models

A foundation for theory, interpretation and application

May 2004

WWW.WATSONWYATT.COM

Paper authored by:

Duncan Anderson, FIA
Sholom Feldblum, FCAS
Claudine Modlin, FCAS
Doris Schirmacher, FCAS
Ernesto Schirmacher, ASA
Neeza Thandi, FCAS

Watson Wyatt
Worldwide

- CAS 2004 Discussion Paper Program
- CAS Exam 9 syllabus as of 2006
- Copies available at www.watsonwyatt.com/glm

watsonwyatt.com

PM-2
An Introduction to GLM Theory

CAS Seminar on Ratemaking
Boston, March 17, 2008

Claudine Modlin, FCAS, MAAA

Watson Wyatt
Worldwide