Motor Insurance Pricing - Using GLM

Sameer R Kakrambe

Disclaimer

- All the points covered in this presentation are my personal views and general guidelines of Motor Insurance Pricing using GLM. They must not be construed as methodology employed by my current and/or previous organization(s) in any shape or form

Agenda

- Pricing
- Linear Modelling vs. Generalized Linear Modelling
- Data Preparation
- Modelling in real-life
- Spatial Smoothing
- Other applications of GLM

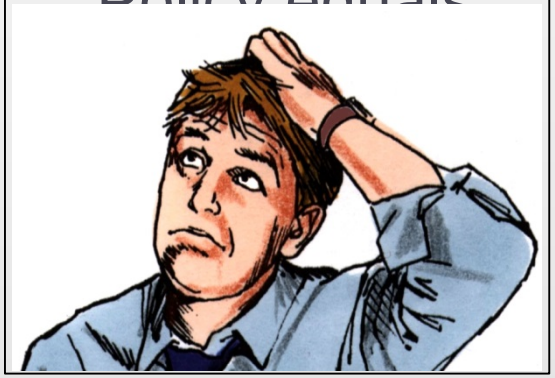
- Pricing

Pricing: Non-Insurance Products

- Price of a Pen equals:
  - Manufacturing cost of all components
  - Labor cost
  - Assembly Cost
  - Packaging Cost
  - Distribution Cost etc.

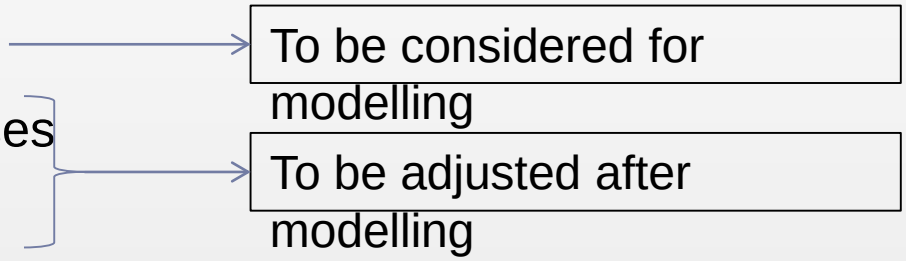
Pricing: Motor Insurance

- Price of a Motor Insurance Policy equals:



- What information is required to determine costing structure of a motor insurance policy like that of a pen?
  - Will there be a claim?
  - If yes, how many claims will there be?
  - If yes, what is the amount of claim paid?
  - If yes, when will a claim be lodged?
  - If no, does potential policyholder get insurance for free?
- Will this information be available at policy issuance?
  - Certainly NO!
- Insurer has to ***predict*** all above points in order to fairly price individual Motor Insurance Policy

- Motor Insurance Pricing has two aspects:
  - Risk Premium:
    - Pure Risk Rate, and
    - A loading for large losses
    - IBNR's



  - Office Premium:
    - Risk premium, plus
    - Trending
    - Loading the cost of reinsurance
    - Loading for expenses and commission
    - Charge to reflect cost of capital
    - Investment income
    - Taxes

Basic Ratemaking

- One-way or Two-way Analyses:
  - Summarized insurance statistics, claim frequency/severity/loss ratio for each value of each explanatory variable
  - Does NOT take into account effect of other variables

Zone	Cost per claim
North	20,000
South	7,000
East	13,500
West	11,000

	Gender	
Zone	Male	Female
North	13,000	7,000
South	3,000	4,000
East	9,000	4,500
West	6,000	5,000

- Limitations:
  - Can be distorted by correlation between factors
  - Do not consider interdependencies between factors
  - Do not exhibit true relative variation between factor levels as it double counts effects of underlying individual variables

Overcoming Limitations

- Where possible and statistically relevant, available data should be subdivided into homogeneous subsets
- Subdivision will help avoid cross-subsidies and profitability will not depend on particular cross section of risks
- Company will be less exposed to changes in business mix
- Company will be less exposed to anti-selection

- Linear Modelling vs.

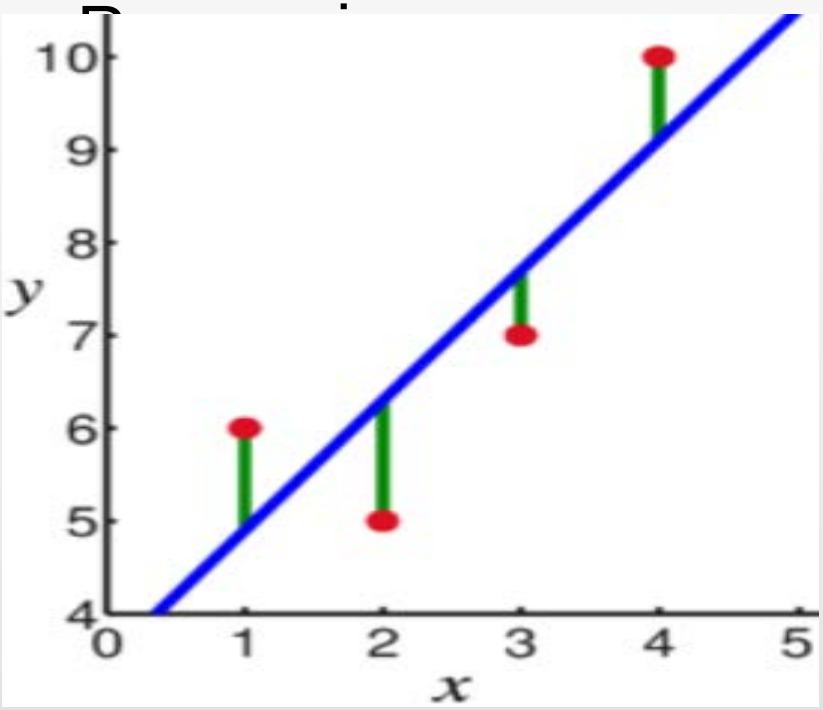
Generalized Linear Modelling

Modelling

- General Steps in Modelling
  - Goal: To explain how a variable of interest depends on some other variable(s).
  - Collect data with which models are to be built
  - Parameterize models from observed data
  - Evaluate if observed data follow or violate model assumptions
  - Evaluate model fit using appropriate statistical tests
    - Significance of explanatory factors
    - Predictive power of models
  - Validate the model
  - Use the model to predict future outcomes on similar (not necessarily identical) risks

Linear Models (LM)

- Simple (Classical)

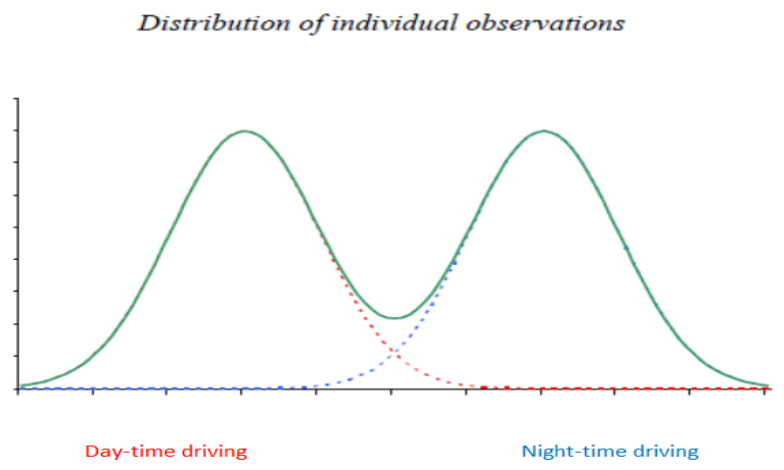


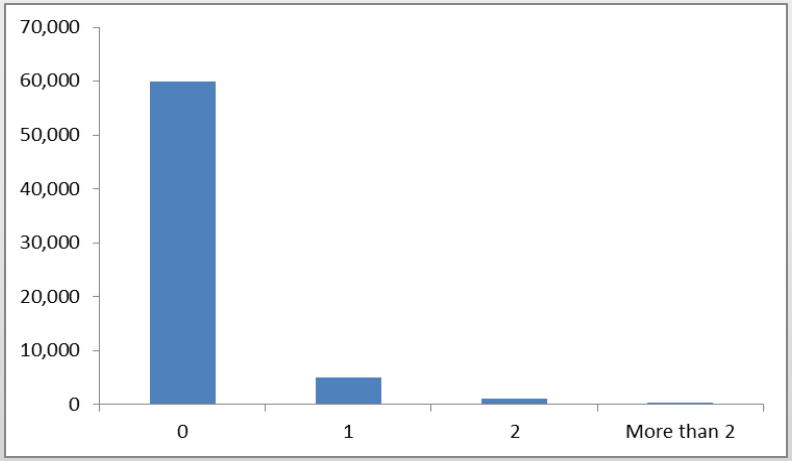
- Model: $Yi = b_0 + b_1 X_i + e_i$
- Assumptions:
  - There exists a linear relationship
  - Errors are independent
  - Variance of e_i is constant
  - $e_i \sim N(0, \sigma_e^2)$

- Parameters b_0, b_1 are calculated to minimise square error
- Can be easily extended to multiple explanatory variables

Limitations

- Limitations:
  - Linear models assume all observations are independent and each comes from a Normal distribution.
  - This assumption does not relate to the aggregate of the observed item, but to each observation individually.
  - Difficult to assert normality and constant variance
  - Values of response variables could be strictly positive
  - Many insurance risks tend to vary multiplicatively and not additively





Generalized Linear Models (GLM)

- Linear models are a special case of GLM's
- LM assumptions of Normality, constant variance & additivity effects are removed
- The effect of the covariates on the response variable is assumed to be additive on a transformed scale
- The response variable is assumed to be a member of the exponential family of distributions

$$f_i(y_i;\theta_i,\phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i,\phi)\right\}$$

- The distribution is completely specified in terms of its mean and variance

- The variance of Y_i is a function of its mean

$$Var(Y_i) = \frac{\phi . V(\mu_i)}{\omega_i}$$

- where V(x) = variance function, is a specified function
- φ = parameter that scales the variance
- ω_i = constant that assigns a weight, or credibility, to observation i.
- Familiar distributions belonging to exponential family and their respective variance functions:

Distribution	V (x)
Normal	1
Poisson	x
Gamma	x^2
Binomial	x (1 - x)
Inverse Gaussian	x^3

Comparison – LM vs GLM

- LM:
  - *Random Component* : Each component of $\underline{Y}$ is independent and normally distributed. The mean μ_i allowed to differ, but all Y_i have common variance σ^2_e
  - Systematic Component: The n covariates are combined to give the "linear predictor" $\underline{\eta} = \underline{\beta}\, X$
  - Link Function : The relationship between the random and systematic components is specified via a link function, that is identity function.
  - $E[\underline{Y}] = \underline{\mu} = \underline{\eta}$

- GLM:
  - *Random Component* : Each component of $\underline{Y}$ is independent and from one of the exponential family of distributions
  - Systematic Component: The n covariates are combined to give the "linear predictor" $\underline{\eta} = \underline{\beta}\, X$
  - Link Function : The relationship between the random and systematic components is specified via a link function ***g***, that is differentiable and monotonic
  - $E[\underline{Y}] = \underline{\mu} = g^{-1}(\underline{\eta})$

Common Link & Error Functions

Response Variable	Link Function	Error Structure	V (x)
Claim Frequency	Log	Poisson	x
Claim Severity	Log	Gamma	x^2
Burning Cost	Log	Tweedie	x^p (1<p<2)
Retention Rate	Logit [Ln (y/1-y)]	Binomial	x (1 - x)

Ways of Modelling

- Burning Cost:
  - Defined as "*Actual cost of claims during a past period of years expressed as an annual rate per unit of exposure*"
  - Advantages:
    - Simplicity
    - Quicker since needs less number of models to be modelled
    - Allows for experience of individual risk or portfolios
  - Disadvantages:
    - Harder to spot trends
    - Provides less understanding of changes impacting individual risks
- Frequency - Severity:
  - *Frequency = Claim Count per unit exposure*
  - *Severity = Claim amount per unit claim*
  - Advantages:
    - Mirrors the underlying process
    - Can use for complex insurance structures
    - Can gain additional insights into aggregate losses
    - Helps us identify trends
  - Disadvantages:
    - More onerous data requirements
    - Time-consuming
    - Requires more expertise

Perils in Motor Insurance

- Typical Perils modelled in Motor Insurance:
  - Accidental Damage
  - Third Party Liability
    - Property Damage
    - Death
    - Bodily Injury
  - Windscreen
  - Theft
  - Fire
  - Combination of above perils based on credible data availability

- Data Preparation

Data Preparation

- Typical datasets available with Insurers:

Data Preparation

- Contains all policy information like:
  - Policy number
  - Policy start date, end date
  - Policy endorsements
  - Vehicle information: Make, model, registration date, Price of the car etc.
  - Premium information
  - Coverage
  - Gender
  - Age of policyholder
  - No Claim Bonus (NCB) level
  - Geography
  - Fuel type
  - Voluntary Excesses
  - Occupation
  - Annual Mileage
  - Parking Location (covered or not)

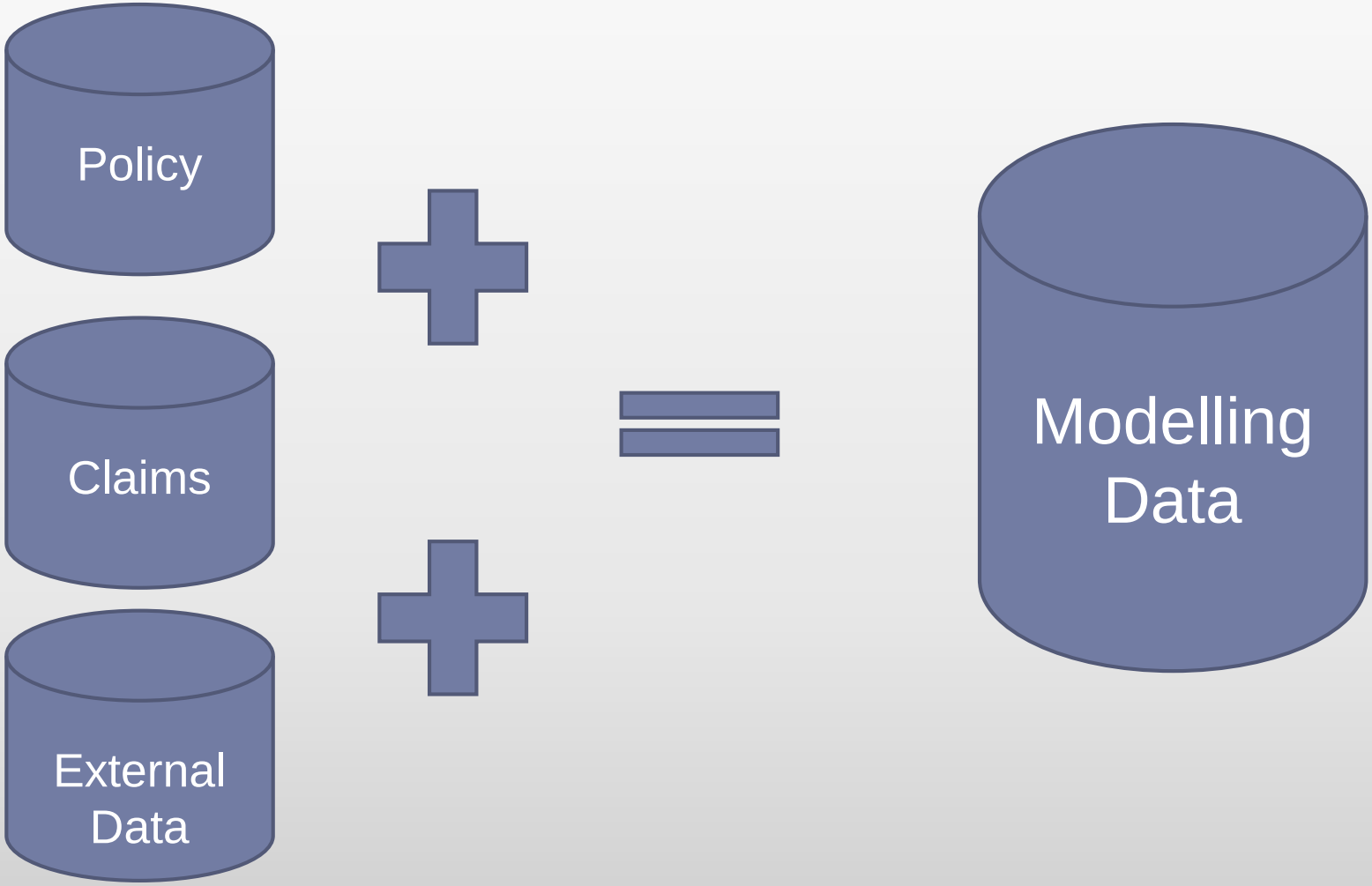
Data Preparation

- Contains claim related information like:
  - Claim number
  - Claim type
  - Loss date
  - Intimation date
  - Settlement date
  - Claim last updated date
  - Claims paid
  - Claims outstanding
  - Large loss threshold determination
  - Floored loss threshold determination
  - Orphan claims
  - Policy number

Data Preparation

- External data could include:
  - Credit score information
  - Vehicle additional details like
    - Dimensions
    - Gearbox type
    - Safety information
  - Criminal history of a policyholder
  - Theft indices of geographies
  - Flood scores of geographies
  - Other demographic information of policyholders
  - Claims history with previous insurers
- Linked to policy data using unique ID of policyholders

Data Preparation



Data Validation

- Policy Data:
  - Check missing values – missing value for zone, deductible, make etc.
  - Investigate factors are sensible
    - Min/Max/Average of factors
    - Absurd levels – Policyholder age < 18 years
    - Continuous variables banding – equal exposure banding or equal class width banding
    - Exposure at levels – Highest exposure for Renewal when it is known that New Business should have the highest exposure. (*Exposure is weight used in the model fit to attach an importance to each observation)*
  - Check if all details are available for modelling period:
    - For example, Modelling period is decided to be 2008 – 2012. Check all policy details are available for this duration

Data Validation

- Policy Data:
  - Exposure Calculation:
    - General exposure measure is “Vehicle-Years” because it’s easily quantifiable, easy to record, easy to administer
  - Case 1:

Policy No	Start Date (A)	End Date (B)
IAM007	01-Apr-12	31-Mar-13

    - Exposure = [(B) – (A)] / [(B) – (A)]
  - Case 2: Suppose policyholder moved from Mumbai to Delhi, effective 30-Sep-12. Data for modelling should look like

Policy No	Start Date (A)	End Date (B)	Endorsement Date (C)	City	Exp Start Date (D)	Exp End Date (E)
IAM007	01-Apr-12	31-Mar-13	30-Jun-12	Mumbai	01-Apr-12	30-Jun-12
IAM007	01-Apr-12	31-Mar-13	30-Jun-12	Delhi	01-Jul-12	31-Mar-13

    - Exposure for each row of policy = [(E) – (D)] / [(B) – (A)]

Data Validation

- Policy Data:
  - Split the policy at desired level of temporal granularity

Policy No	Start Date (A)	End Date (B)	Premium	AQ Start Date (D)	AQ End Date (E)
IAM007	01-Apr-12	31-Mar-13	12,000	01-Apr-12	30-Jun-12
IAM007	01-Apr-12	31-Mar-13	12,000	01-Jul-12	30-Sep-12
IAM007	01-Apr-12	31-Mar-13	12,000	01-Oct-12	31-Dec-12
IAM007	01-Apr-12	31-Mar-13	12,000	01-Jan-13	31-Mar-13

  - Usually all motor insurance policies are of 1 year duration
  - Policies are split monthly/quarterly/half-yearly to analyse seasonality of claims

Data Validation

- Claims Data:
  - Check values are sensible – min/max/average of claims registered
  - Investigate open claims:
    - Incurred claims = Paid claims + Outstanding Reserves
    - Modelling data period should be considered such that most of the claims are developed
    - For Outstanding reserves, latest position of a claim should be considered

Policy No	Claim No	Paid	Outstanding	Loss Date	Intimation Date	Paid Date
IAM007	CL00001	1,000	20,000	12-Sep-12	13-Sep-12	15-Sep-12
IAM007	CL00001	2,000	19,000	12-Sep-12	13-Sep-12	17-Sep-12
IAM007	CL00001	4,000	17,000	12-Sep-12	13-Sep-12	19-Sep-12
IAM007	CL00001	10,000	11,000	12-Sep-12	13-Sep-12	21-Sep-12
IAM007	**CL00001**	**15,000**	**6,000**	**12-Sep-12**	**13-Sep-12**	**10-Oct-12**

Data Validation

- Claims Data:
  - Negative, nil claims should be eliminated from modelling
  - Claim amount below a certain threshold should be removed from modelling because:
    - Such claims could be below deductible limit
    - Such claims could be so trivial that they do not represent reality
  - Large Losses:
    - Could distort genuine severity trends
    - Can be capped. Ways of capping:
      - Flat threshold amount across years
      - Indexation of threshold amount across years
      - Different threshold amounts for different vehicle segments – e.g. Separate large loss definitions for Alto and BMW 5-series
    - Remove large losses from the analysis altogether. *(in this case, remember to remove claim count from frequency model as well!!)*
  - All removed claim amounts and counts to be added back post modelling

Claims Merging

- Case 1:
  - Only 1 claim in a policy year
  - Input:

Policy No	Start Date (A)	End Date (B)	AQ Start Date (C)	AQ End Date (D)
IAM007	01-Apr-12	31-Mar-13	01-Apr-12	30-Jun-12
IAM007	01-Apr-12	31-Mar-13	01-Jul-12	30-Sep-12
IAM007	01-Apr-12	31-Mar-13	01-Oct-12	31-Dec-12
IAM007	01-Apr-12	31-Mar-13	01-Jan-13	31-Mar-13

Policy No	Claim No	Loss Date	Incurred Amount
IAM007	CL00001	12-Sep-12	21,000

  - Output:

Policy No	Start Date (A)	End Date (B)	AQ Start Date (D)	AQ End Date (E)	Claim Amount
IAM007	01-Apr-12	31-Mar-13	01-Apr-12	30-Jun-12	.
IAM007	01-Apr-12	31-Mar-13	01-Jul-12	30-Sep-12	21,000
IAM007	01-Apr-12	31-Mar-13	01-Oct-12	31-Dec-12	.
IAM007	01-Apr-12	31-Mar-13	01-Jan-13	31-Mar-13	.

Policy No	Start Date (A)	End Date (B)	AQ Start Date (D)	AQ End Date (E)	Claim Amount
IAM007	01-Apr-12	31-Mar-13	01-Apr-12	30-Jun-12	.
IAM007	01-Apr-12	31-Mar-13	01-Jul-12	12-Sep-12	21,000
IAM007	01-Apr-12	31-Mar-13	13-Sep-12	30-Sep-12	.
IAM007	01-Apr-12	31-Mar-13	01-Oct-12	31-Dec-12	.
IAM007	01-Apr-12	31-Mar-13	01-Jan-13	31-Mar-13	.

- Case 2:
  - 2 claims in a policy year
  - Input:

Policy No	Claim No	Loss Date	Incurred Amount
IAM007	CL00001	12-Sep-12	21,000
IAM007	CL00002	31-Dec-12	50,000

  - Output:

Policy No	Start Date (A)	End Date (B)	AQ Start Date (D)	AQ End Date (E)	Claim Amount	Claim Count
IAM007	01-Apr-12	31-Mar-13	01-Apr-12	30-Jun-12	.	0
IAM007	01-Apr-12	31-Mar-13	01-Jul-12	12-Sep-12	21,000	1
IAM007	01-Apr-12	31-Mar-13	13-Sep-12	30-Sep-12	.	0
IAM007	01-Apr-12	31-Mar-13	01-Oct-12	31-Dec-12	50,000	1
IAM007	01-Apr-12	31-Mar-13	01-Jan-13	31-Mar-13	.	0

- Case 3:
  - 2 claims in one quarter
  - Input:

Policy No	Claim No	Loss Date	Incurred Amount
IAM007	CL00001	12-Oct-12	30,000
IAM007	CL00002	01-Dec-12	6,000

  - Output:

Policy No	Start Date (A)	End Date (B)	AQ Start Date	nd Date (E)	Claim Amount	Claim Count
IAM007	01-Apr-12	31-Mar-13	01-Apr-12	30-Jun-12	.	0
IAM007	01-Apr-12	31-Mar-13	01-Jul-	30-Sep-12	.	0
IAM007	01-Apr-12	31-Mar-13	01-O	31-Dec-12	36,000	2
IAM007	01-Apr-12	31-Mar-13	01-Jan-13	31-Mar-13	.	0

Policy No	Start Date (A)	End Date (B)	AQ Start Date (D)	nd Date (E)	Claim Amount	Claim Count
IAM007	01-Apr-12	31-Mar-13	01-Apr-12	30-Jun-12	.	0
IAM007	01-Apr-12	31-Mar-13	01-Jul-12	30-Sep-12	.	0
IAM007	01-Apr-12	31-Mar-13	01-Oct-12	12-Oct-12	30,000	1
IAM007	01-Apr-12	31-Mar-13	13-Oc	01-Dec-12	6,000	1
IAM007	01-Apr-12	31-Mar-13	02-Dec-	31-Dec-12	.	0
IAM007	01-Apr-12	31-Mar-13	01-Jan-1	31-Mar-13	.	0

- Modelling in real life

Modelling

- The overall structure of a dataset for GLM claims analysis consists of
  - Explanatory variables – discrete or continuous
  - Temporal variable
  - Earned exposure fields – preferably by claim type if certain policy features are optional
  - Number of incurred claims associated with exposure in consideration
  - Incurred loss amounts fields
  - Premium fields – GWP, GEP even though premium columns are not used in modeling but are used in Loss Ratio calculations post modelling
  - Data extracts – entire dataset doesn't have to be used

Output of Modelling

Base 10,000

Analysis Period

Analysis Period	
2008	0.8100
2009	0.9000
2010	1.0000
2011	1.1000
2012	1.1500

Gender of Main Driver

Gender of Main Driver	
Male	1.0000
Female	0.9300

Rated Area

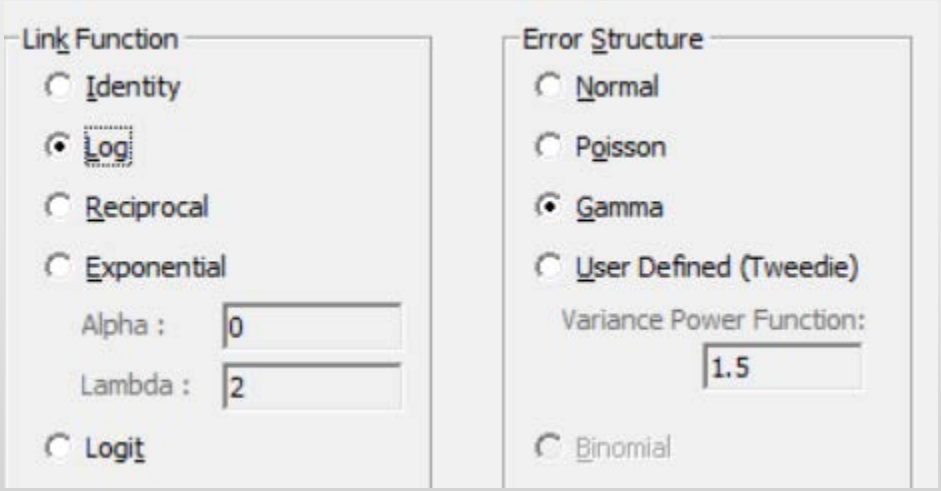
Rated Area	
1	0.4305
2	0.4783
3	0.5314
4	0.5905
5	0.6561
6	0.7290
7	0.8100
8	0.9000
9	1.0000
10	1.1000
11	1.2100
12	1.3310
13	1.4641
14	1.6105
15	1.7716
16	1.9487
17	2.1436
18	2.3579
19	2.5937
20	2.8531

- Output of modelling exercise is beta values at each level of factors included
- For ex. For a “Male” from “Rated area 15” in “2011”, severity would be
  - 10,000 * 1.0 * 1.7716 * 1.10 = 19,490

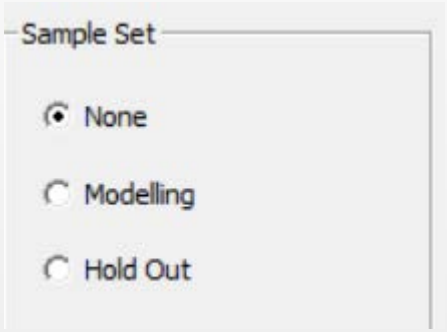
Modelling

- Preliminary analyses:
  - Check if range of response variable values makes sense

  - Select link and error functions appropriately



- Sampling

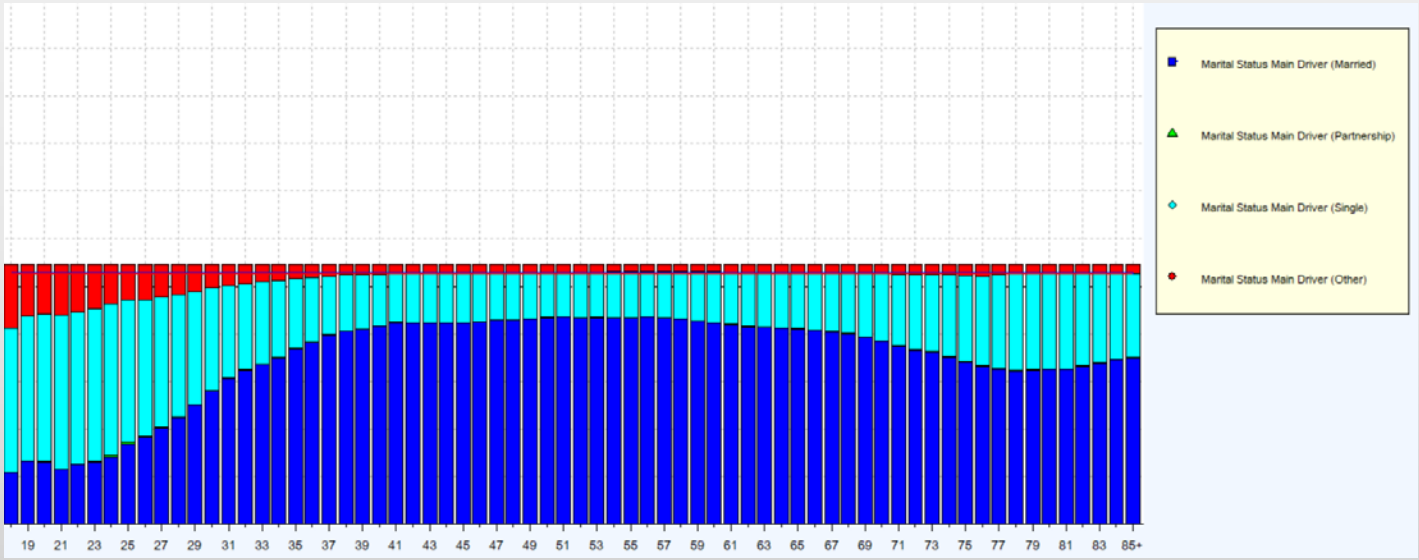


- Check base levels of every factor brought in GLM environment

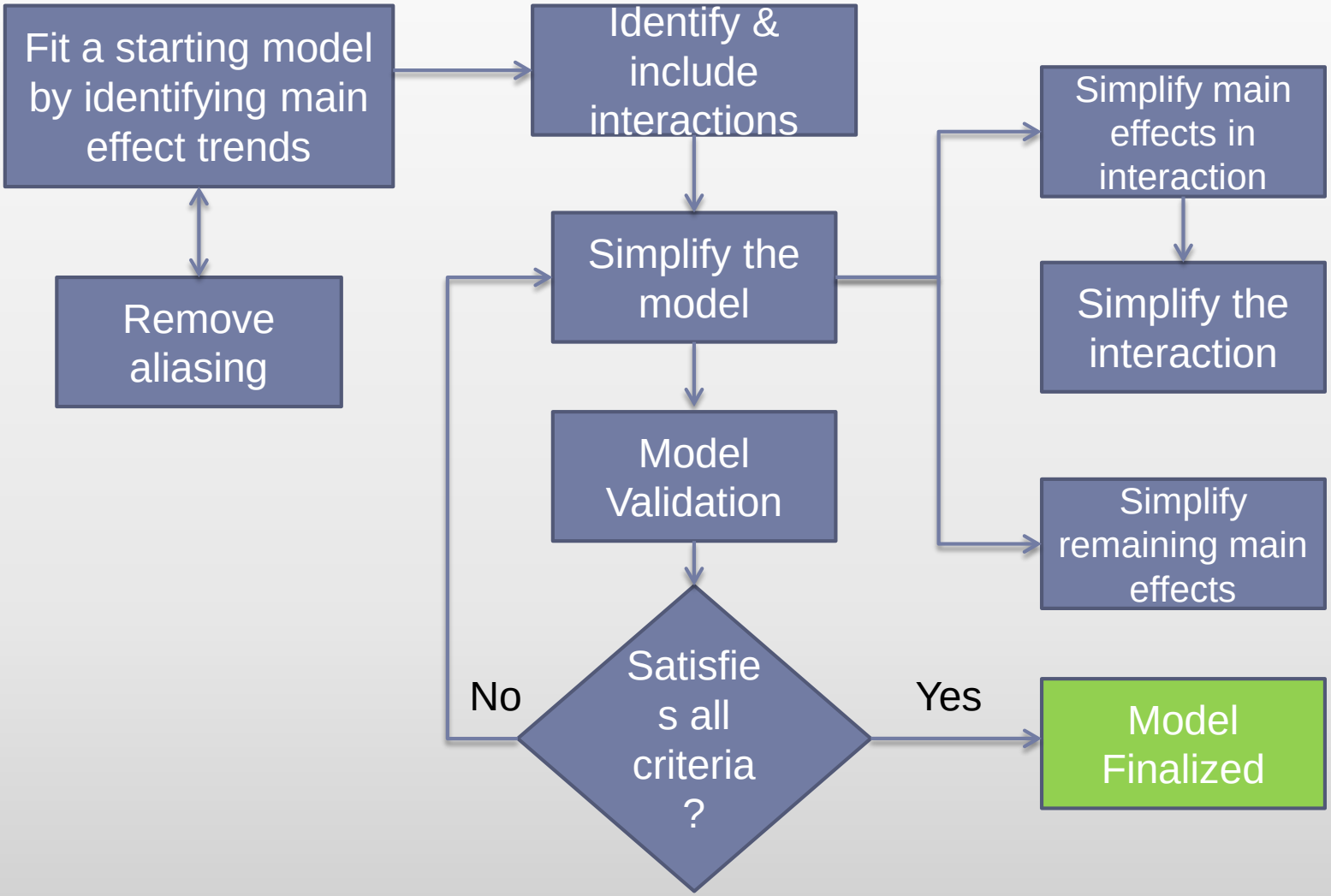
Vehicle Age	No. Observations	Weight
0 (1)	79	82
1 (2)	1,649	1,690
2 (3)	3,327	3,397
3 (4)	4,449	4,560
4 (5)	5,264	5,379
5 (6)	5,830	5,977
6 (7)	5,580	5,706
7 (8)	5,397	5,532
8 (9)	5,385	5,524
9 (10)	5,657	5,809
10 (11)	4,656	4,753

- Check Cramer's V Correlation Matrix
  - Correlations: Data is dependent between factors

Grid Results : Cramer's V Factor (#Levels)	Accident Year	Accident Quarter
Accident Year	0	0
Accident Quarter	1	0



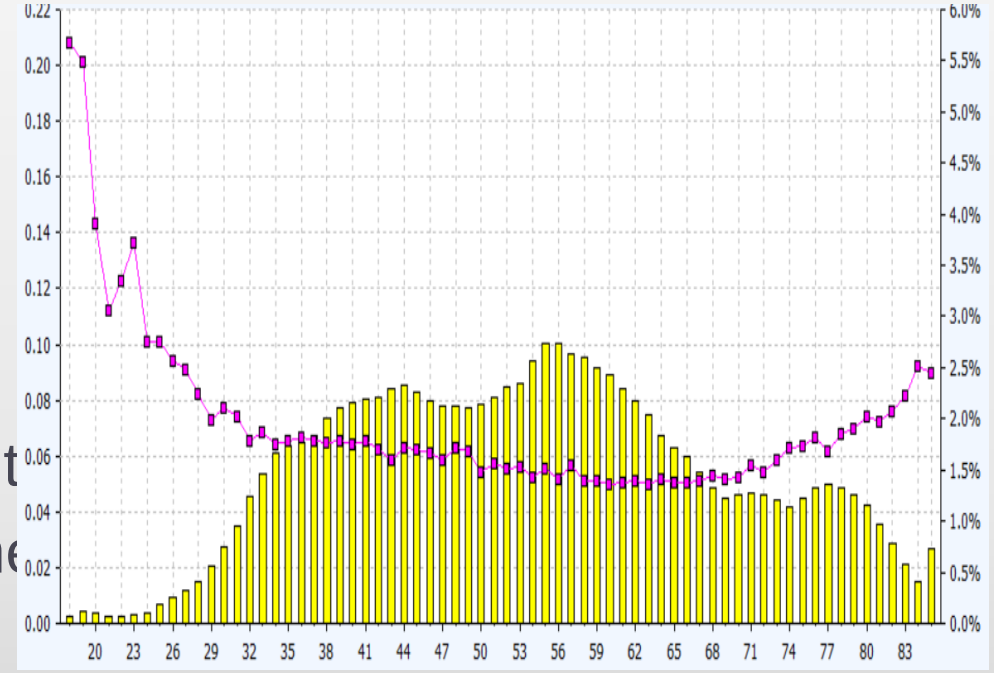
Modelling Process



Main Effect

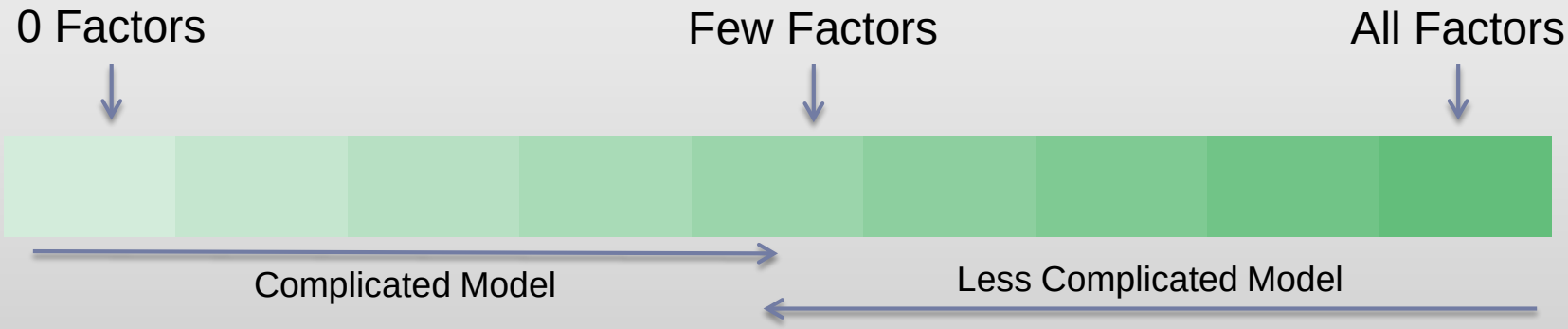
- Main effect: Raw variable being used in regression

- Trends identification:
  - What is the trend?
  - Is it significant?
  - Is it logical?
  - Is it statistically significant
  - Is it consistent across time
  - Is this trend reliable?

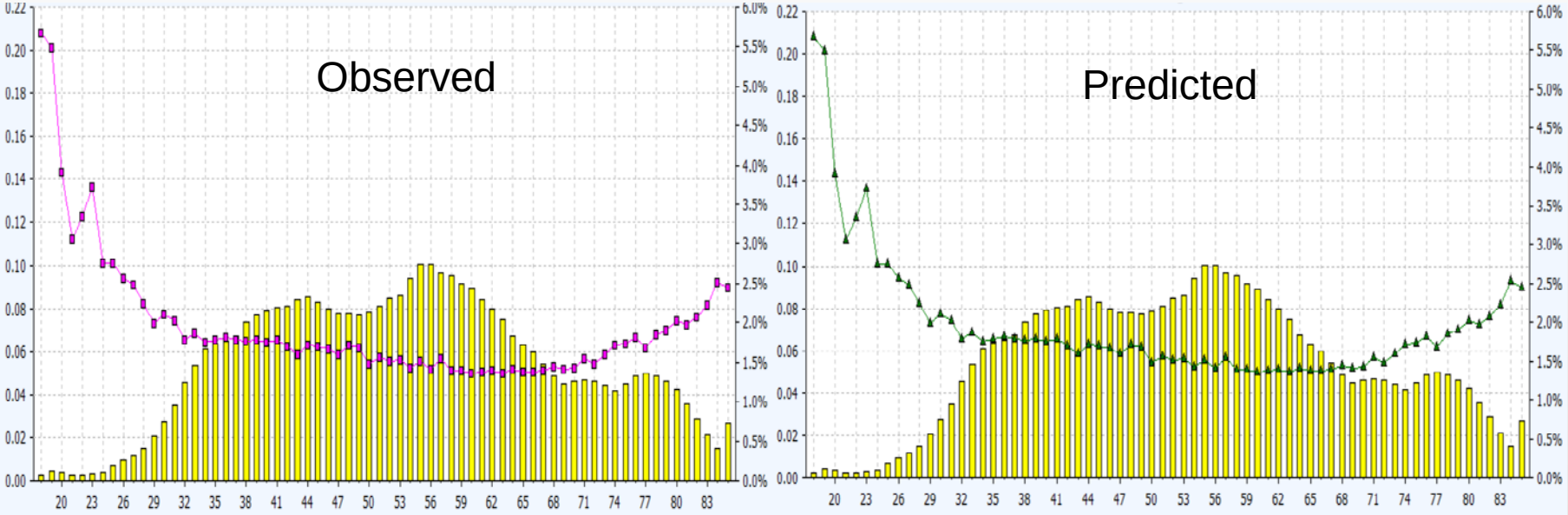


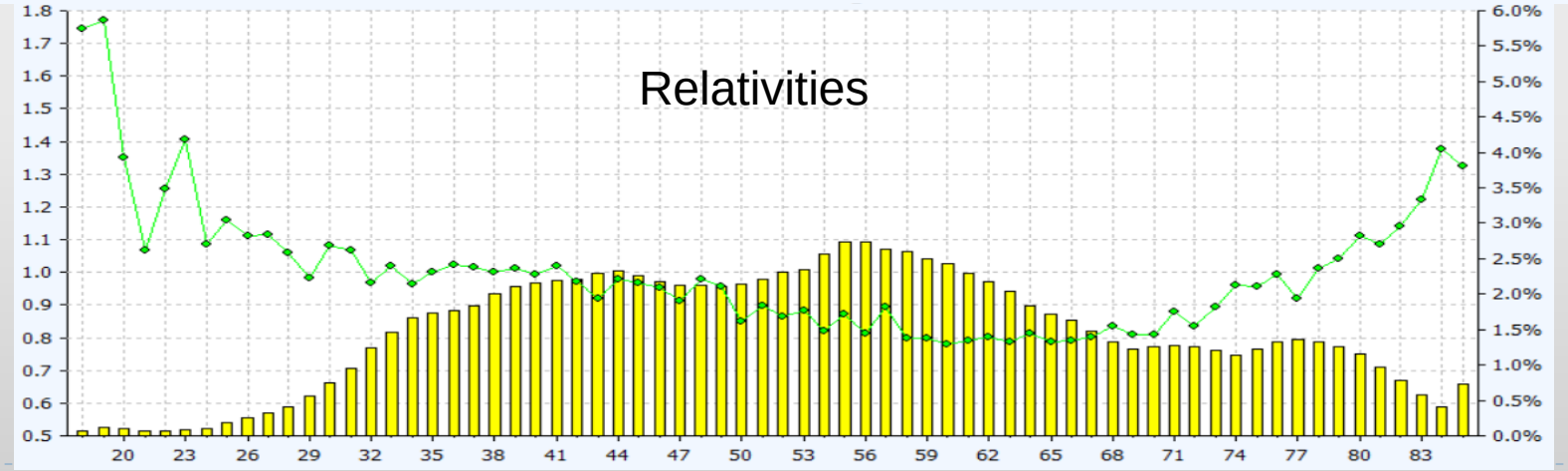
Building a start model

- Include factors which systematically affect response
- Criteria considered:
  - Standard Errors
  - Deviance tests
  - Time consistency
  - Random factor consistency
  - Common sense



Can you guess the variable?



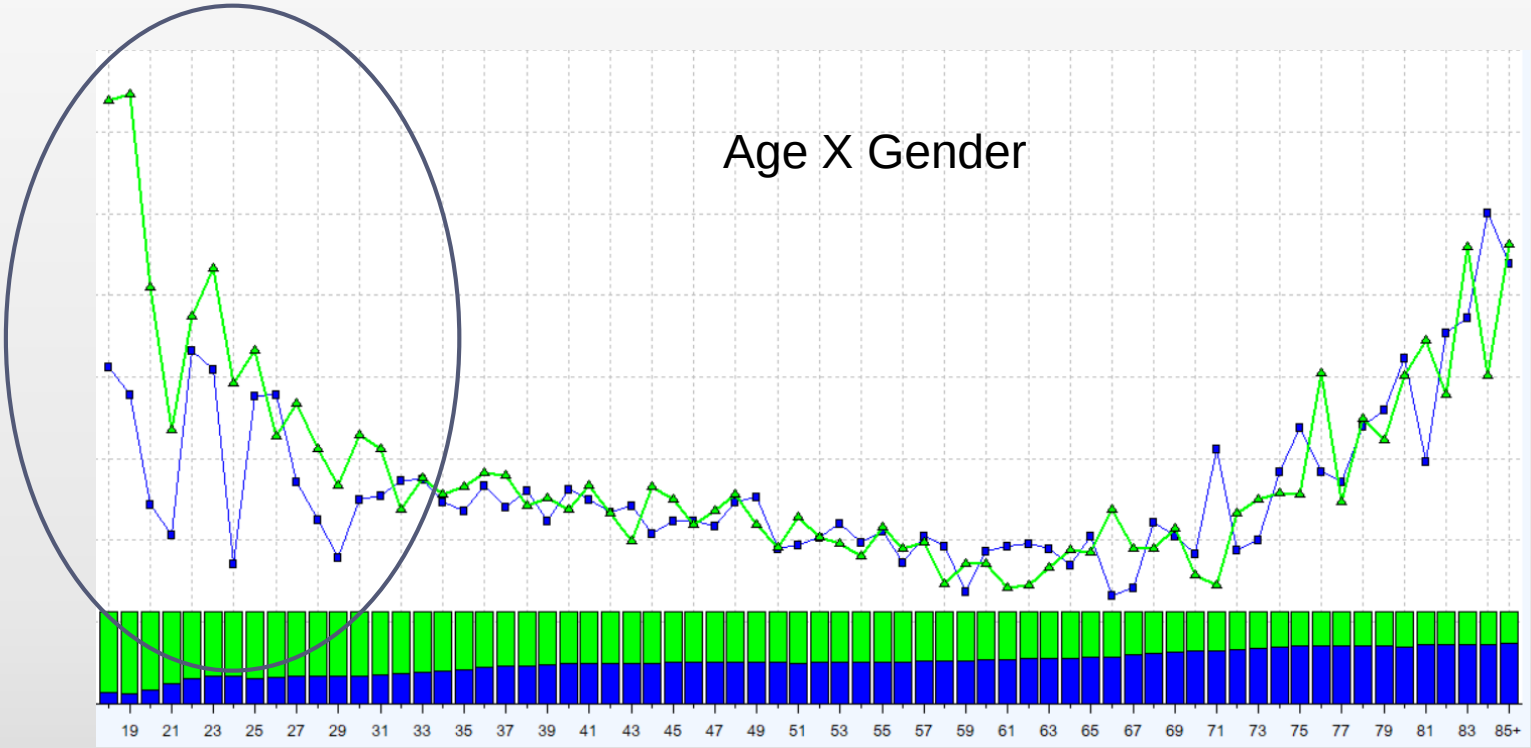


Aliasing

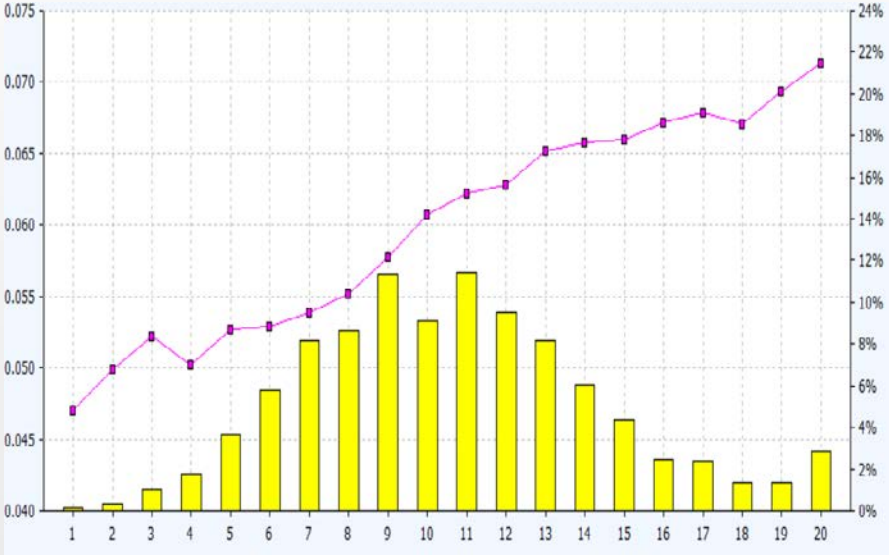
- Occurs when two factors or levels within factors are so highly correlated that it is impossible to tell from the data which factor is causing the underlying effect
- Types of aliasing:
  - Fixed aliasing: there are two perfectly correlated factor levels in the dataset
  - Complex aliasing: A combination of custom factors and variates is included in a model which makes certain factors correlated
- Effects:
  - Do not affect fitted values
  - Can slow down model fitting

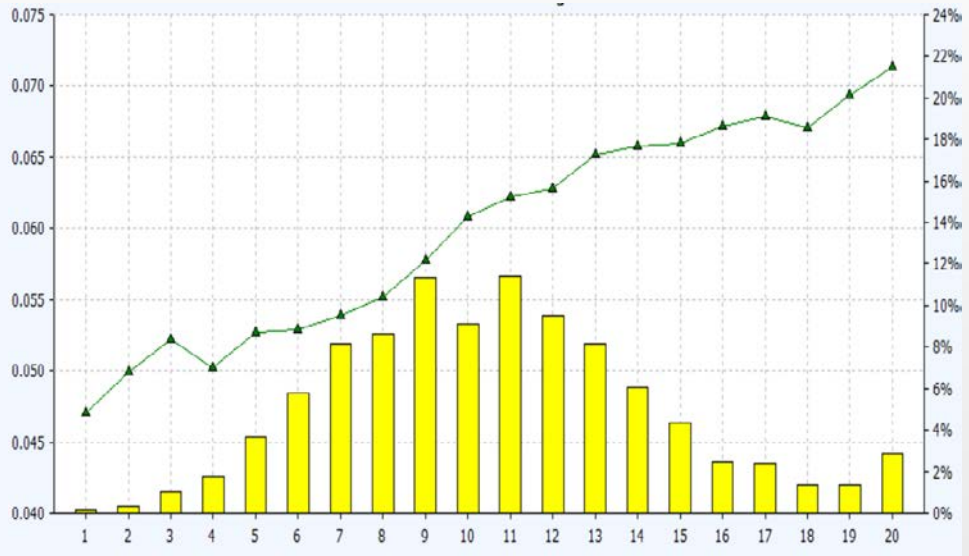
Interactions

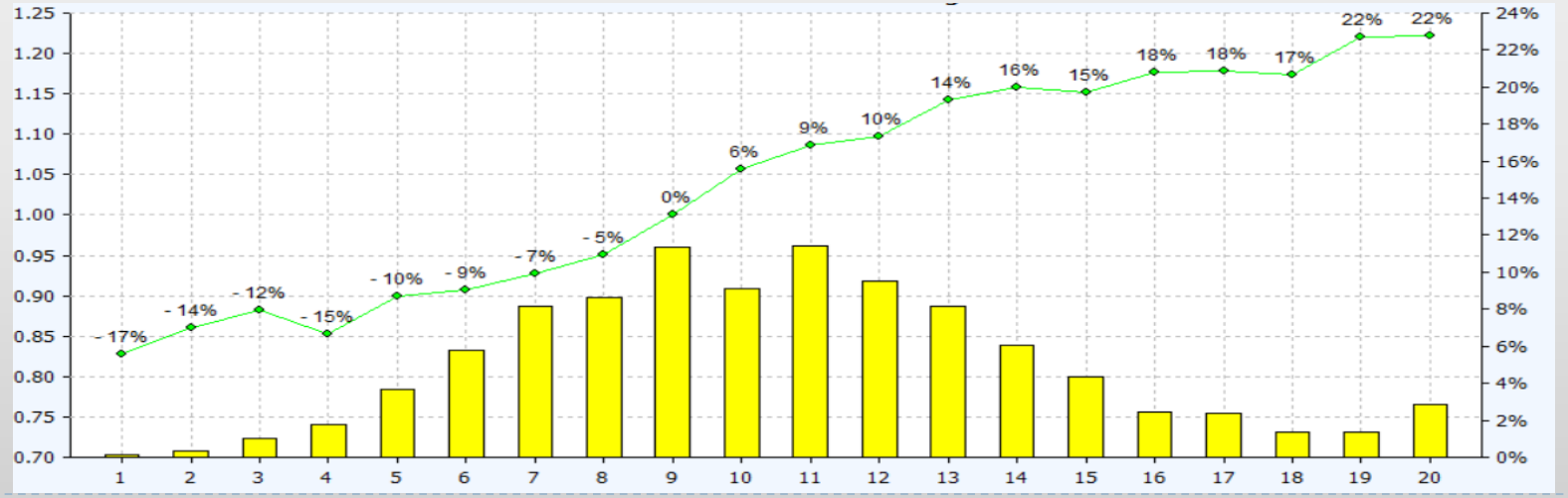
- The effect is dependent between factors



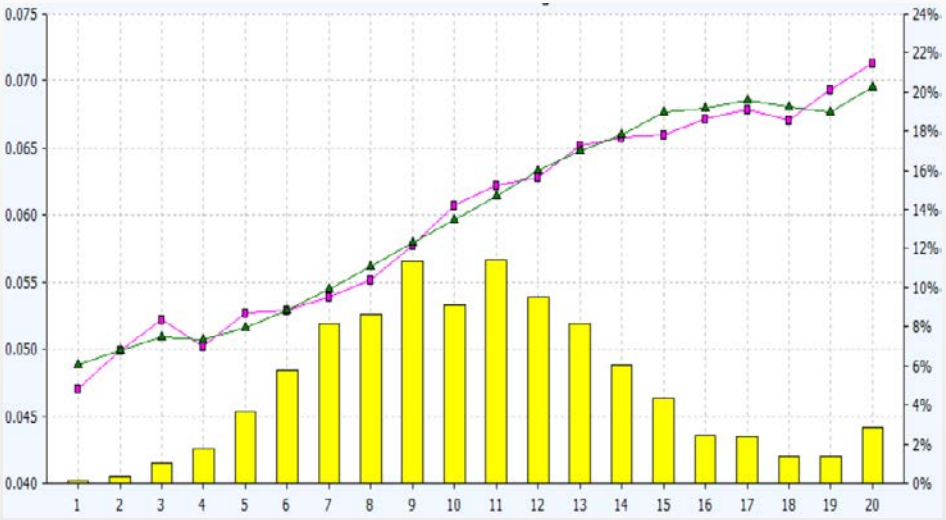
Factor Simplification

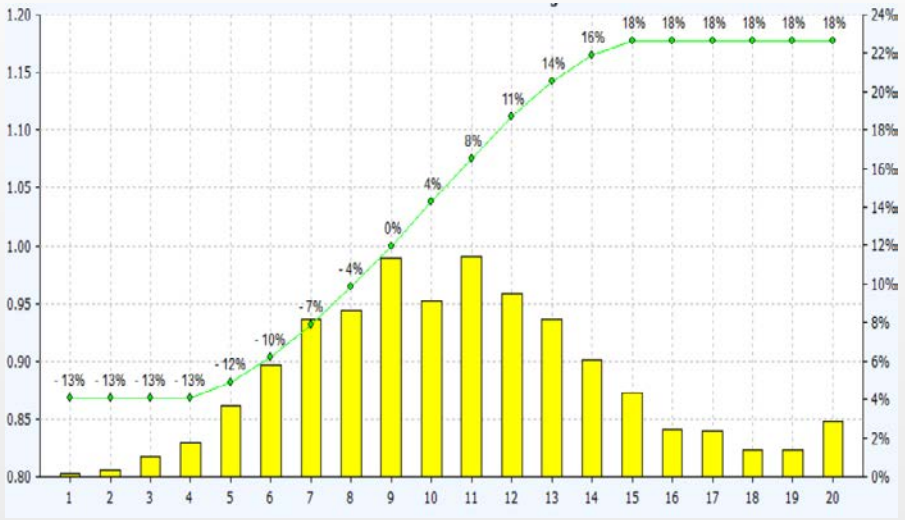




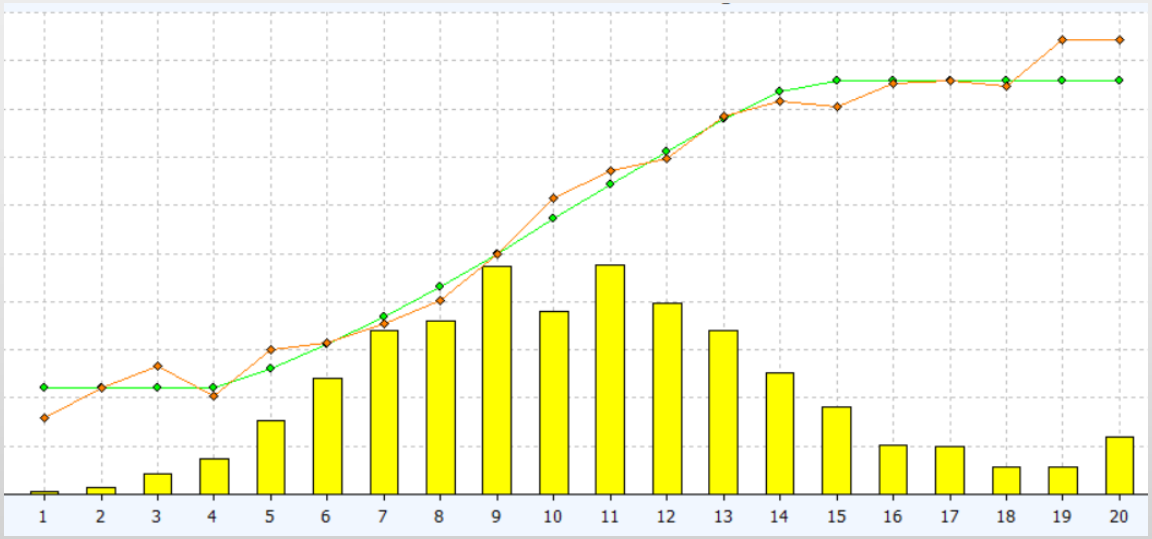


Factor Simplification - Variates

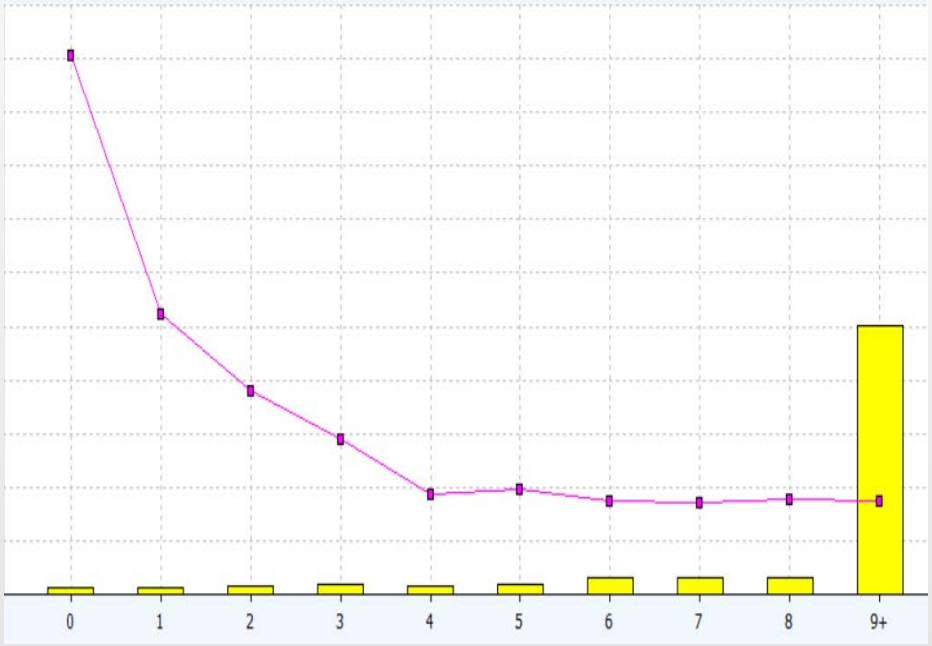


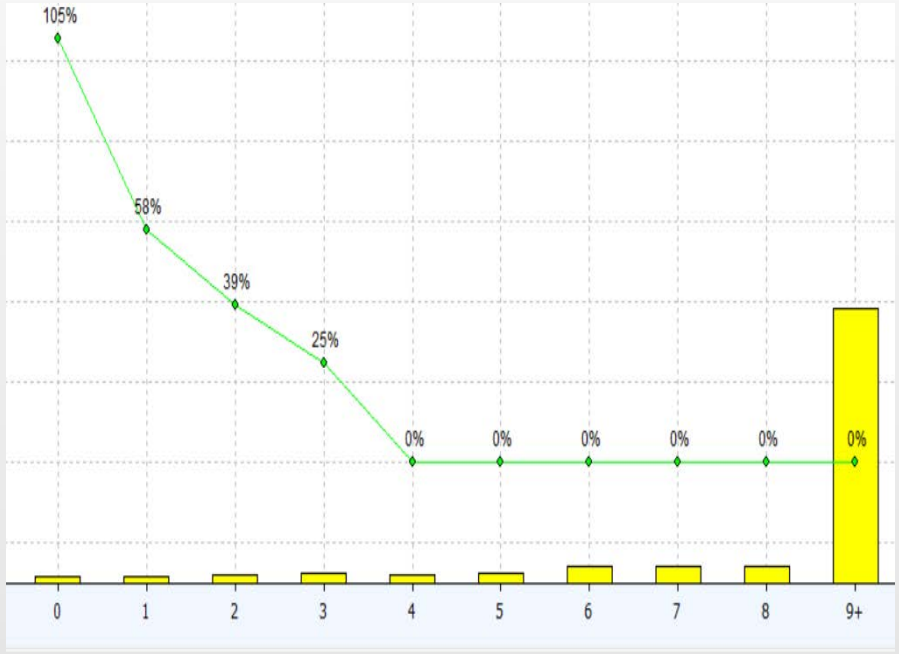


No. of Parameters fit after simplification = ??



Factor Simplification - Custom Factors



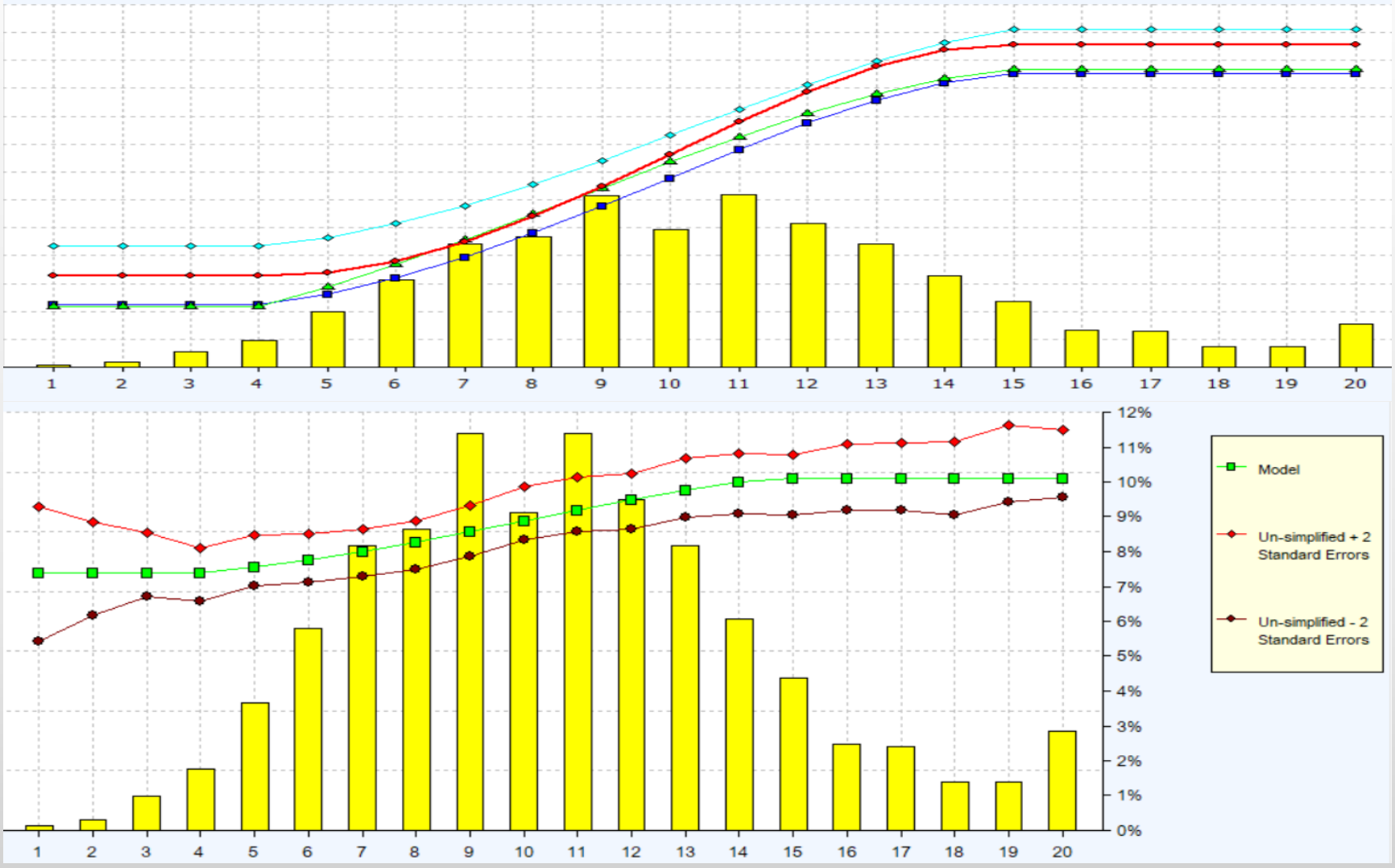


No. of Parameters fit after simplification = ??

Simplification tests

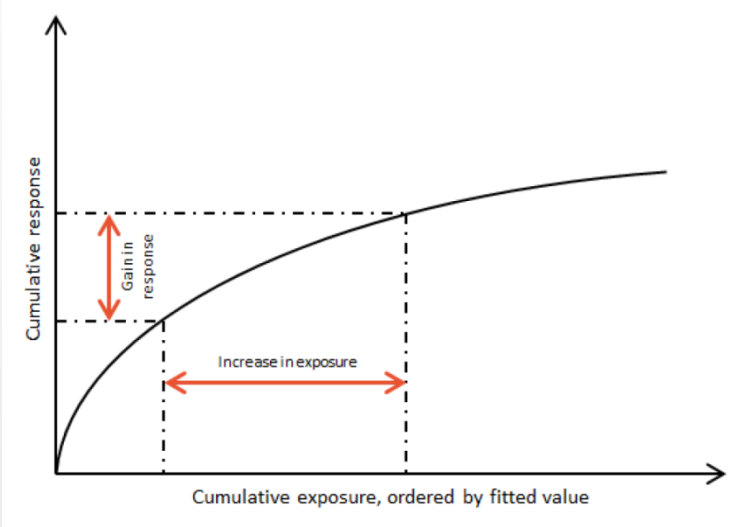
- Following factors need to be considered before finalizing simplification:
  - Significance – reduction in AIC, deviance, parameters
  - Main effect fit – how observed vs. predicted averages are after simplification
  - Comparison with Un-simplified
  - Un-simplified +/- 2% Standard Error interval
  - Time consistency
  - Random Factor consistency
  - Complex aliasing
  - Validation sample inconsistencies

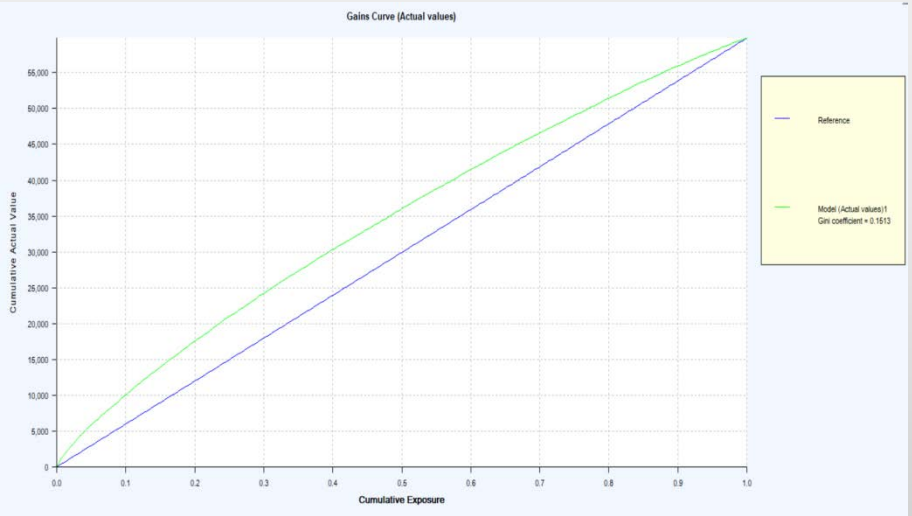
Simplification tests



Model Validation

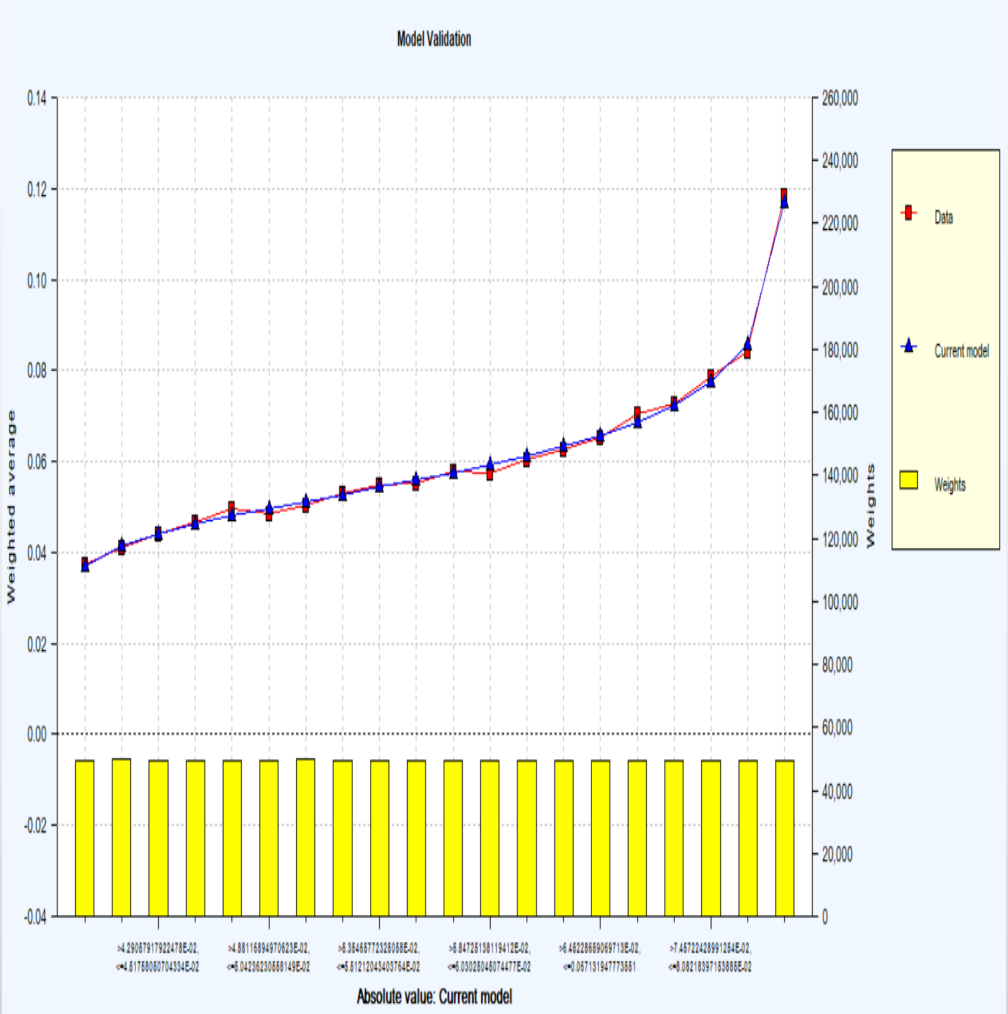
- Backward step-wise regression
  - Removes redundant factors
- Gains Curve
  - Measures predictive power of a model
  - If model is predictive, high fitted values should correspond to high observed values
  - Straight line is nothing but Mean Model
  - Gini coefficient is a measure of area under the curve for a model
  - Gini coefficient in the adjacent graph is difference between the fitted model and mean model





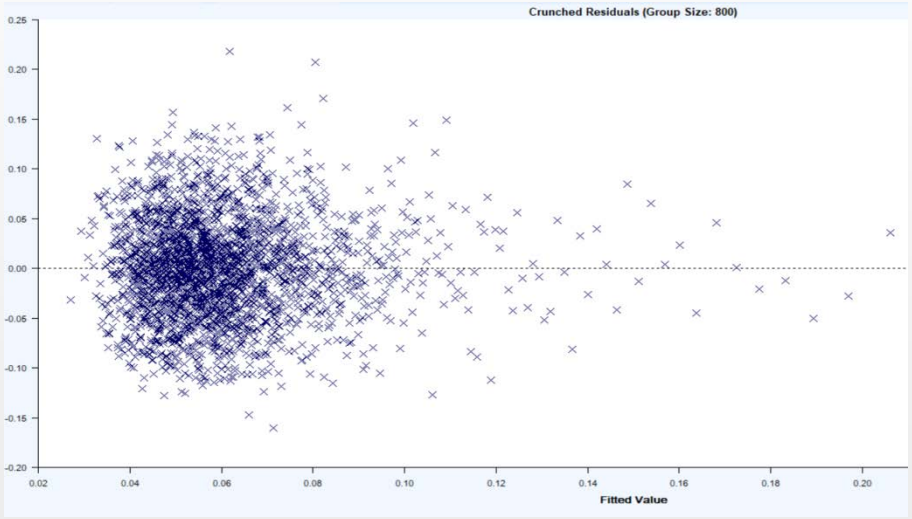
Model Validation

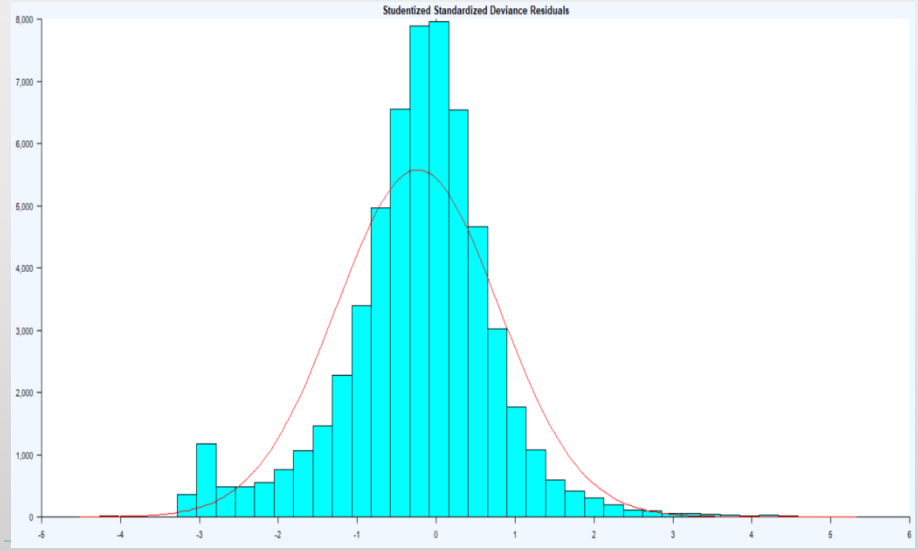
- Lift Curve
  - Graph displays points grouped in ascending absolute values from the fitted model
  - Within each group, the average value of the data and the average value of the comparison model predictions are calculated and plotted
  - For a predictive model, two lines should coincide considerably well
- Validation Sample trends
  - All beta trends should be similar to modelling sample
  - Observed vs Expected trend should be good for every factor



Model Validation

- Residuals
  - Deviance measure corrects skewness meaning deviance residuals should be more closely normally distributed
  - Checks error structure appropriateness
  - If correct, plot should have following characteristics:
    - Average residual will be zero
    - Pattern of residuals will be symmetrical about x-axis
    - Range of residual values will be fairly constant across the width
  - For discrete distributions individual residuals can be grouped.





Post-modelling Adjustments

- Floored losses adjustments
  - Losses removed due to flooring of claims to be added back
- Orphan losses adjustments
  - Orphan losses to be added back flat across all accident years
- Large Loss adjustments
  - Based on method adopted for large loss threshold determination, loading factor to be applied to add back curtailed amount
- IBNR adjustments
  - IBNYR – considering reporting delays, IBNYR factors to be applied, usually across accident years considered in modelling
  - IBNER – considering settlement delays, IBNER factors to be applied, usually across accident years considered in modelling

Post-modelling Adjustments

- Trending:
  - Investigate any trends in the base data that are likely to continue in future.
  - e.g. Improvement in Third Party Death Frequency owing to advancements in vehicle safety technology
- Inflation:
  - Inflating base values to the present day using broadly known inflation rates
  - Projecting from the present day to future using estimated inflation rates
  - For OD, we need to consider estimates of:
    - Motor spare parts' inflation
    - Wage inflation
    - Vehicle Price inflation
  - For liability,
    - Earning inflation
    - Court inflation where there is no fixed formula for compensation

Model Combining

- Risk Premium = Claim Frequency X Claim Severity
- Models can be combined at peril-level and/or portfolio level
- Specifically this can be done by
  - Selecting a dataset which most accurately reflects the likely future mix of business
  - Calculating an expected claim frequency and severity by claim type for each record in the data
  - Combining these fitted values, for each record, to derive the expected cost of claims (according to the individual GLMs) for each record
  - Fitting a further generalized linear model to this total expected cost of claims, containing the union of all factors along with simplifications and interactions, in all of the underlying models.

Policy No	Gender	Rated Area	AD_Freq	AD_Sev	AD_BC	Theft_freq	Theft_Sev	Theft_BC	Total_BC
IAM007	M	2	0.09	80,000	7200	0.0007	1,00,00,000	7,000	14,200
IAMVIR8	M	18	0.04	45,000	1800	0.001	50,00,000	5,000	6,800

Model Combining - Restrictions

- Artificial restrictions need to be imposed on certain factors like NCB.
- Although restrictions could be applied either to Frequency or Severity models, generally it is more appropriate to impose the restriction on the model at the risk premium stage.
- This allows a more complete and balanced compensation of relativities by other correlated factors.

Using Risk Premium results

- Compare with existing rating structures
- Compare shift in average premium if any
- Calculate Loss Ratios to
  - Identify profitable segments
  - To improve pricing in unprofitable segments

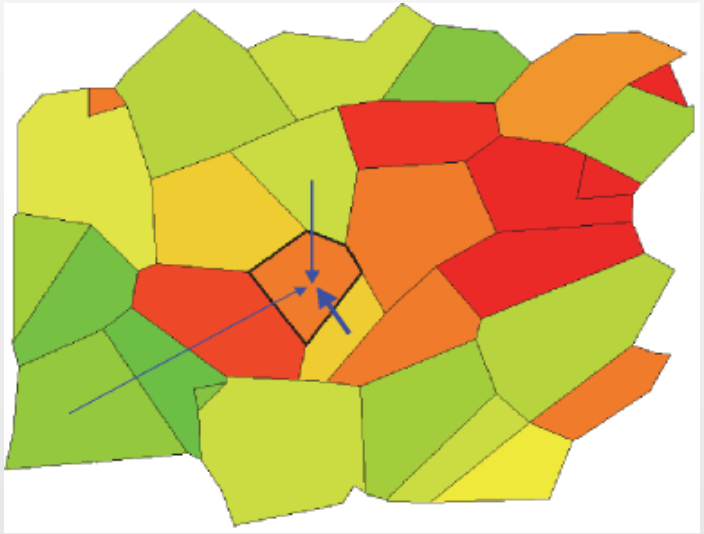
- Spatial Smoothing

Spatial Smoothing

- Purpose:
  - GLM models can be effectively used where number of levels of a factor is small
  - Cannot analyse individual postcodes, VIN's within GLM environments
  - Cannot directly use Classifications produced by industry standard bodies like ABI in UK, since organization's own experience could be different
  - GLM's fail to produce credible results for low exposure areas
- Underlying assumption is neighbouring entities are likely to have similar risk experience
- Two main forms of Spatial Smoothing:
  - Distance-based spatial smoothing
  - Adjacency-based spatial smoothing
- Degree of smoothing:
  - Employing too high or too low degree of smoothing destroys true underlying residual variation

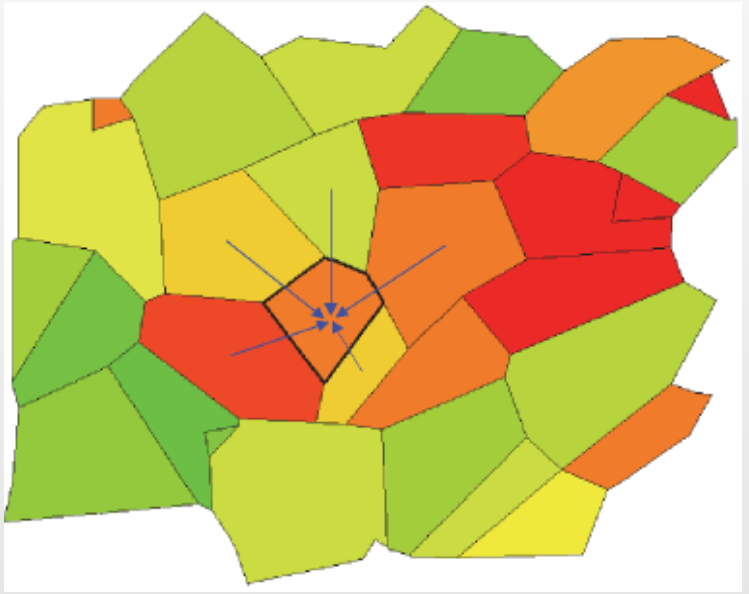
Distance based spatial smoothing

- Incorporates information about nearby location codes based on distance between them
- More the distance, lesser the influence
- Often used for weather related perils
- Easy to implement
- No distributional assumptions required



Adjacency based spatial smoothing

- Incorporates information about directly neighbouring location codes iteratively
- Prior knowledge of claims processes can be incorporated
- Urban-rural differences are explained more appropriately
- Complex to implement
- Used in creating zoning structures as well as vehicle classification



- Other applications of GLM

Other applications of GLM's

- Other applications which could assist in Motor Pricing:
  - Deriving a scoring algorithm for calculating Insurance Scores
  - Retention/conversion analysis