CAS RPM SEMINAR

How Predictive Analytics Can Change Your Market Footprint

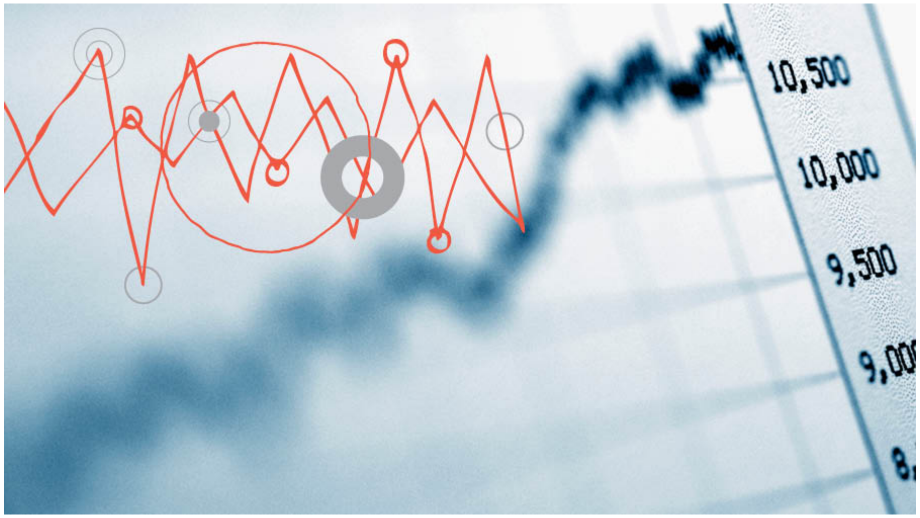
Serhat Guven, FCAS, MAAA

March, 2015

TOWERS WATSON

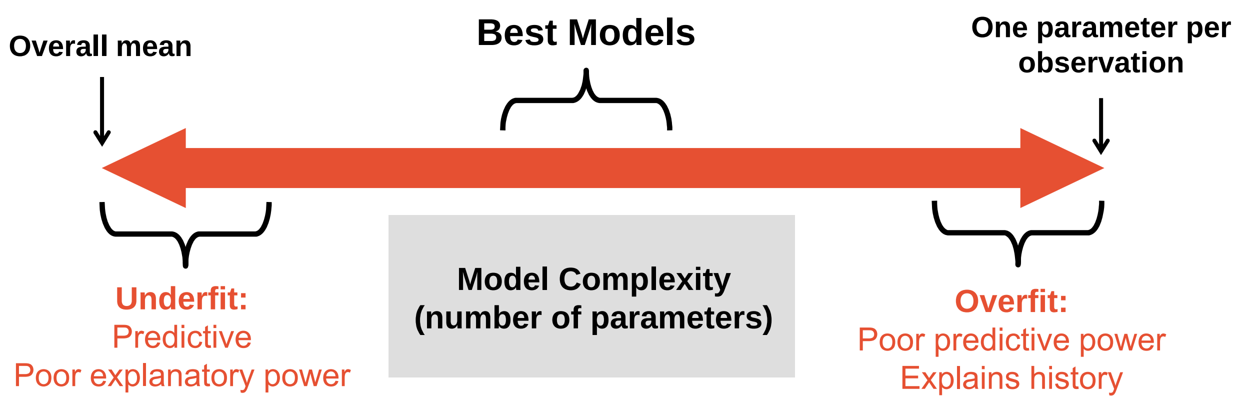
Overview

- Background
- Predictive Analytics and Modeling
- Competitive Analytics and Simulation
- Summary



Background

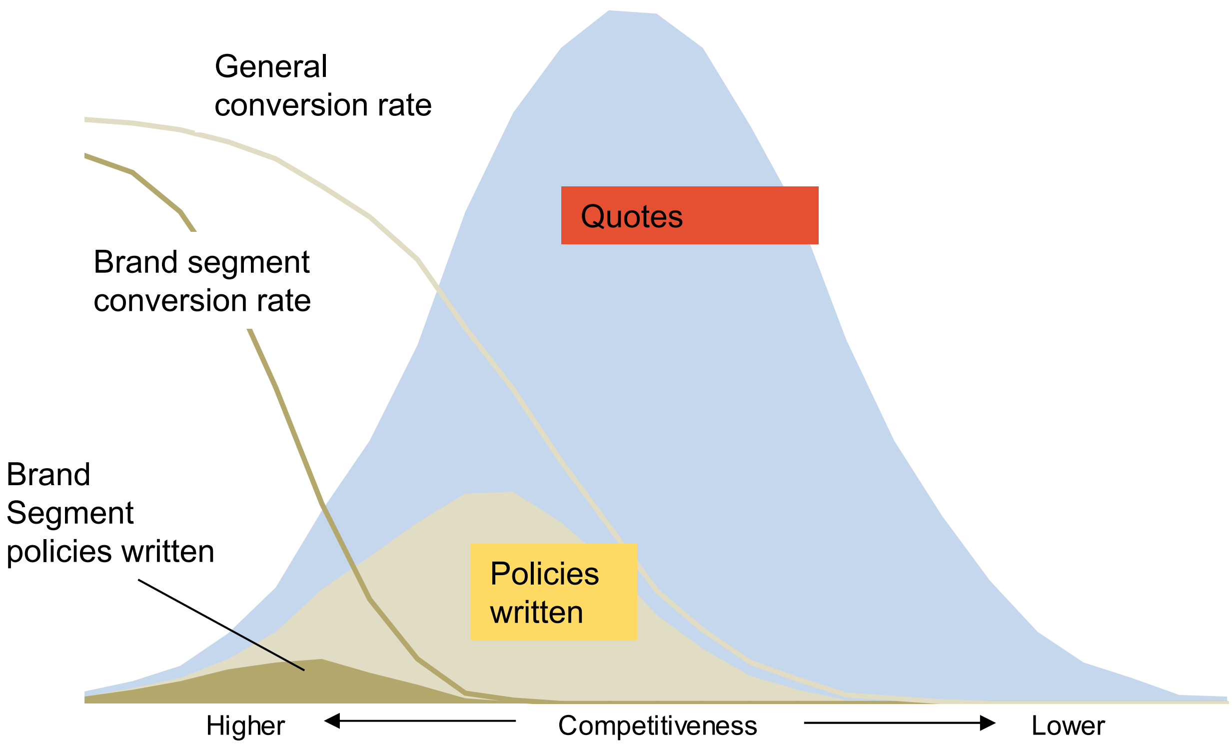
- Statistical data modeling is involves a trade off between predictive and explanatory powers



- Dilemma of over-fitting vs. danger of anti-selection
  - 100 + factors
  - Geographic spatial analysis
  - Many interactions
  - Multi-dimensional effects via scores

Background

- Policies written are more skewed to competitive segments



- Unintended competitiveness through under pricing degrades profitability

Background

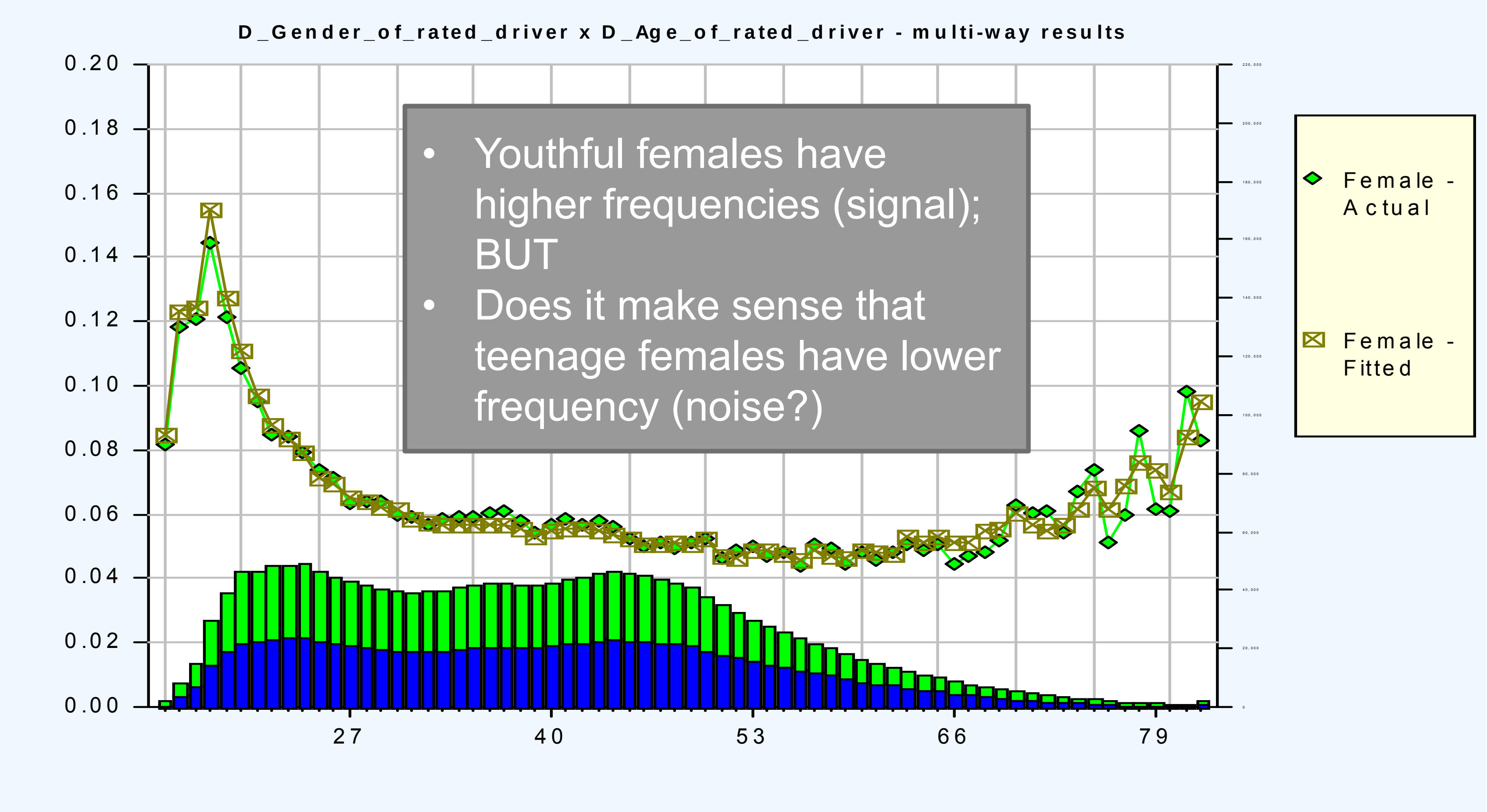
- What techniques are used to minimize the risk of overfitting?

- How is competitive pricing used to enter into new markets?

Predictive Analytics and Modeling

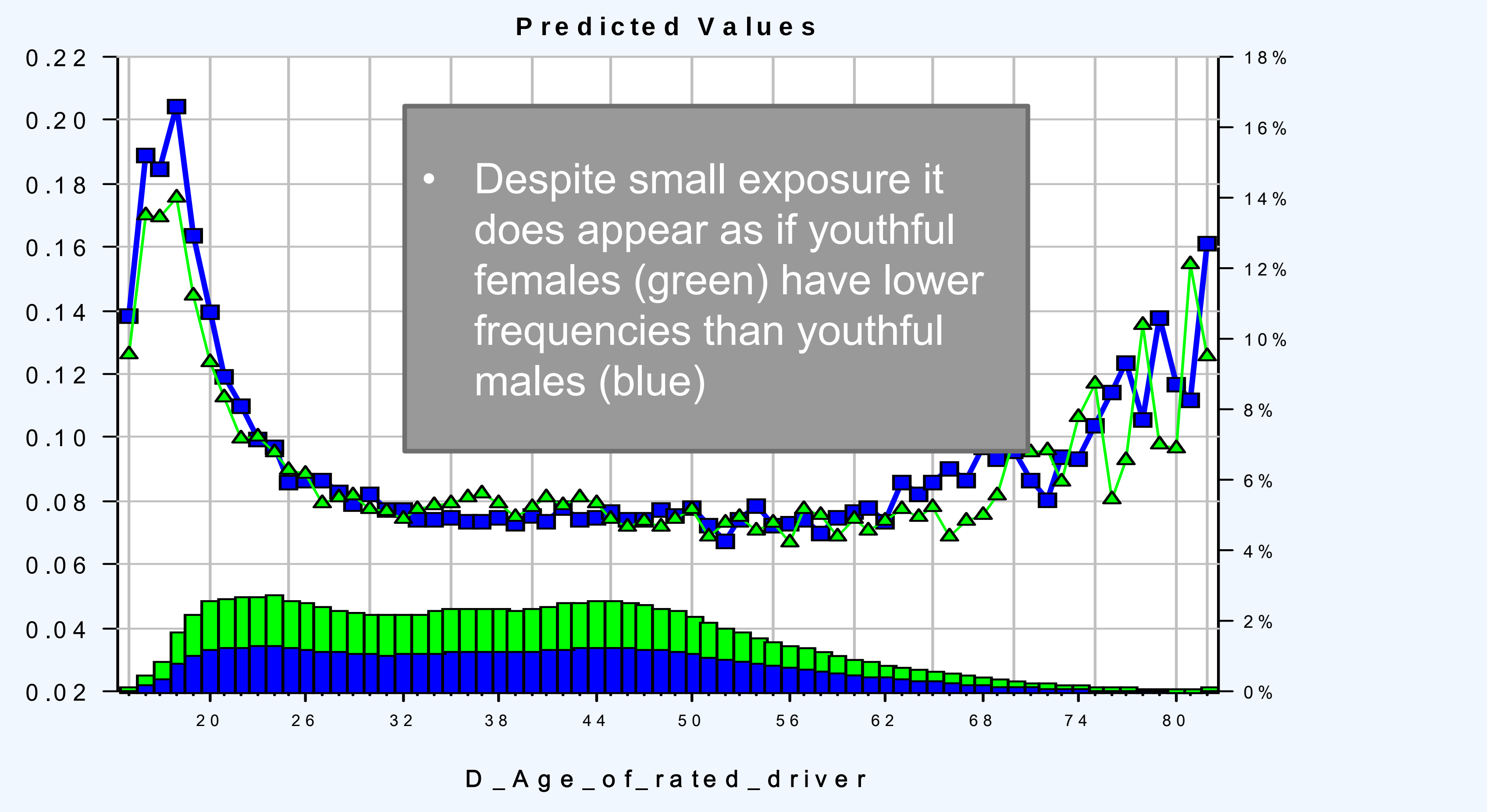
Motivations in Predictive Models

- Goal of a good model is to find the pattern and ignore the noise



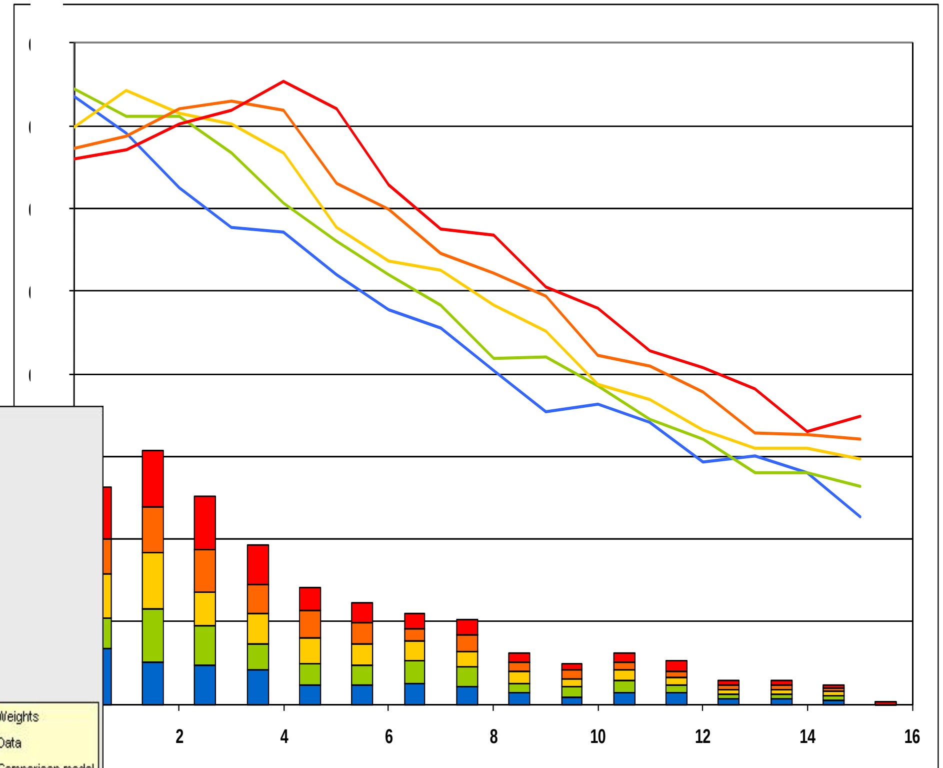
Motivations in Predictive Models

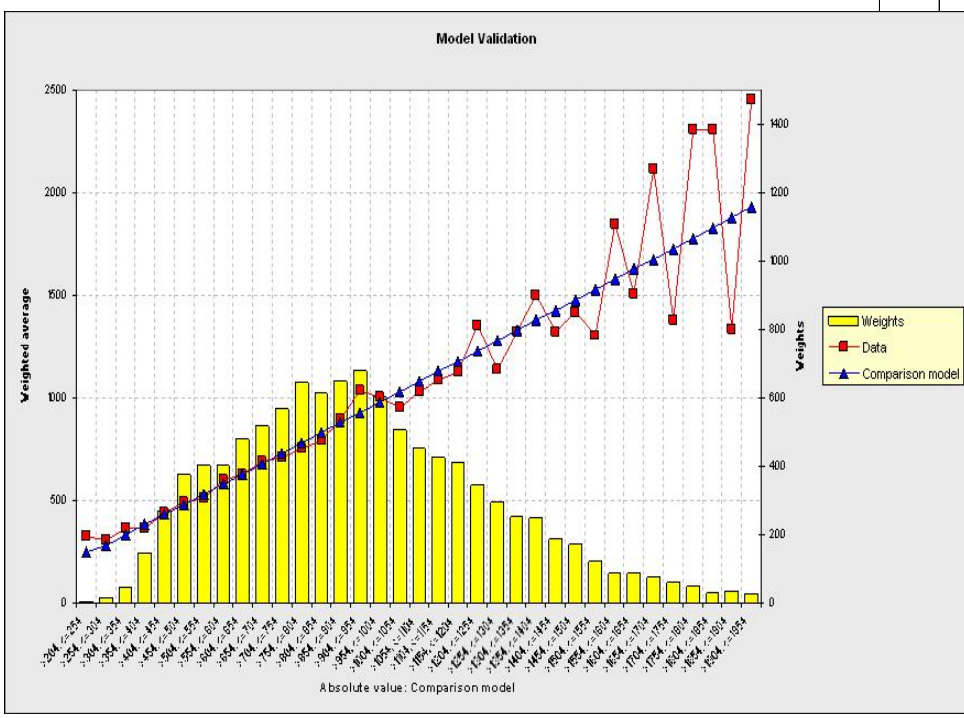
- Goal of a good model is to find the pattern and ignore the noise



Motivations in Predictive Models

- Variety of tests are applied in modeling in practice
- Models are rarely built blindly

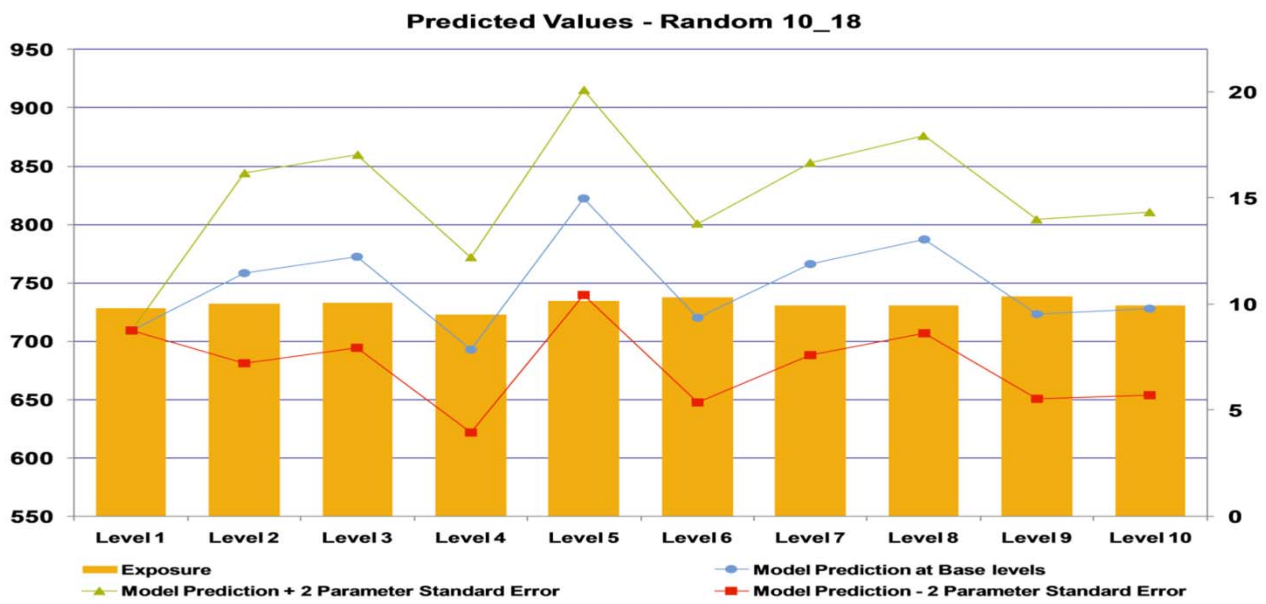




- Chi-squared & F-tests
- Wald p-values
- Akaike information criteria
- Consistency tests

Motivations in Predictive Models

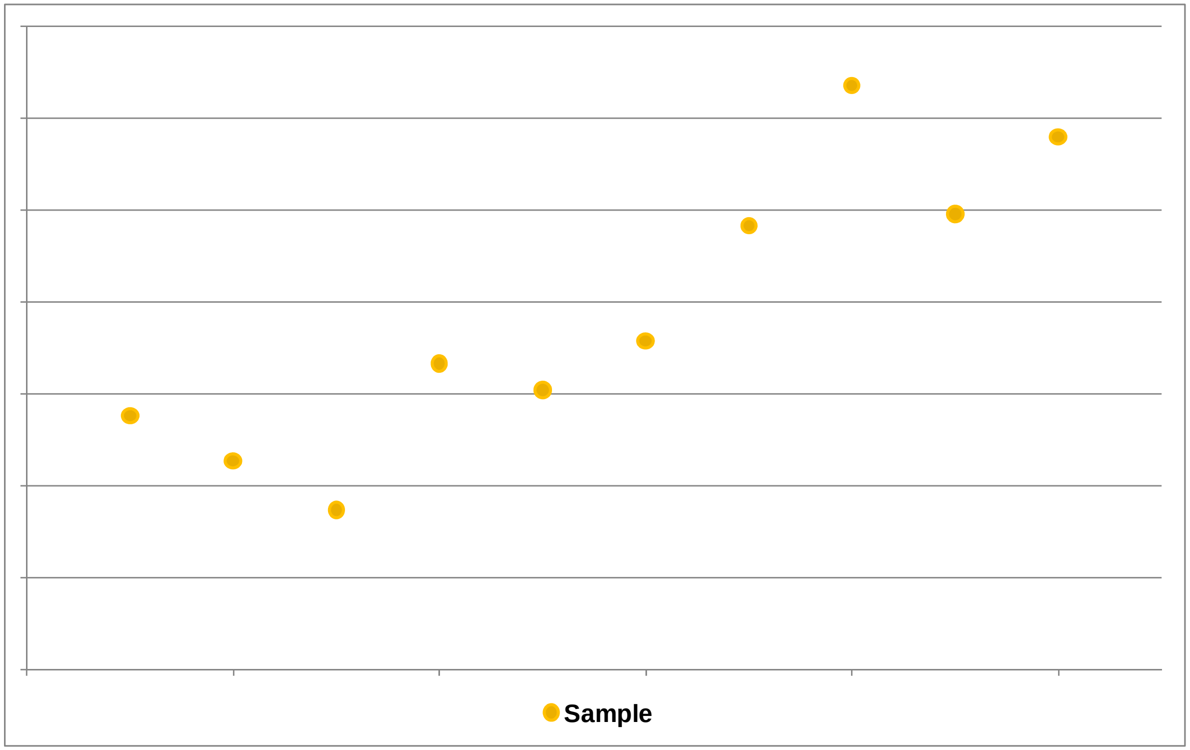
- Consistency example to the extreme: adding random factors that may appear more than 2 SD away from the null model



- Robot model maker might accept these parameters as significant
- Deviance measure decreases as more parameters added can mislead

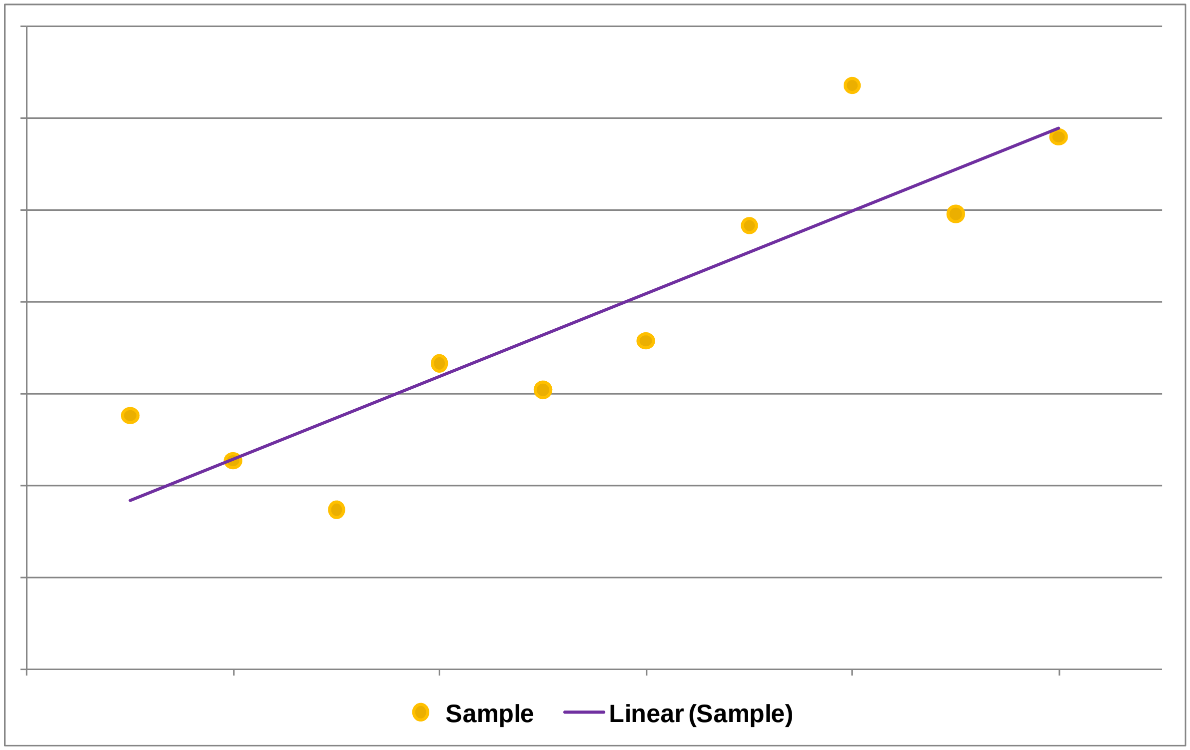
Motivations in Predictive Models

- Simple dataset with one factor:



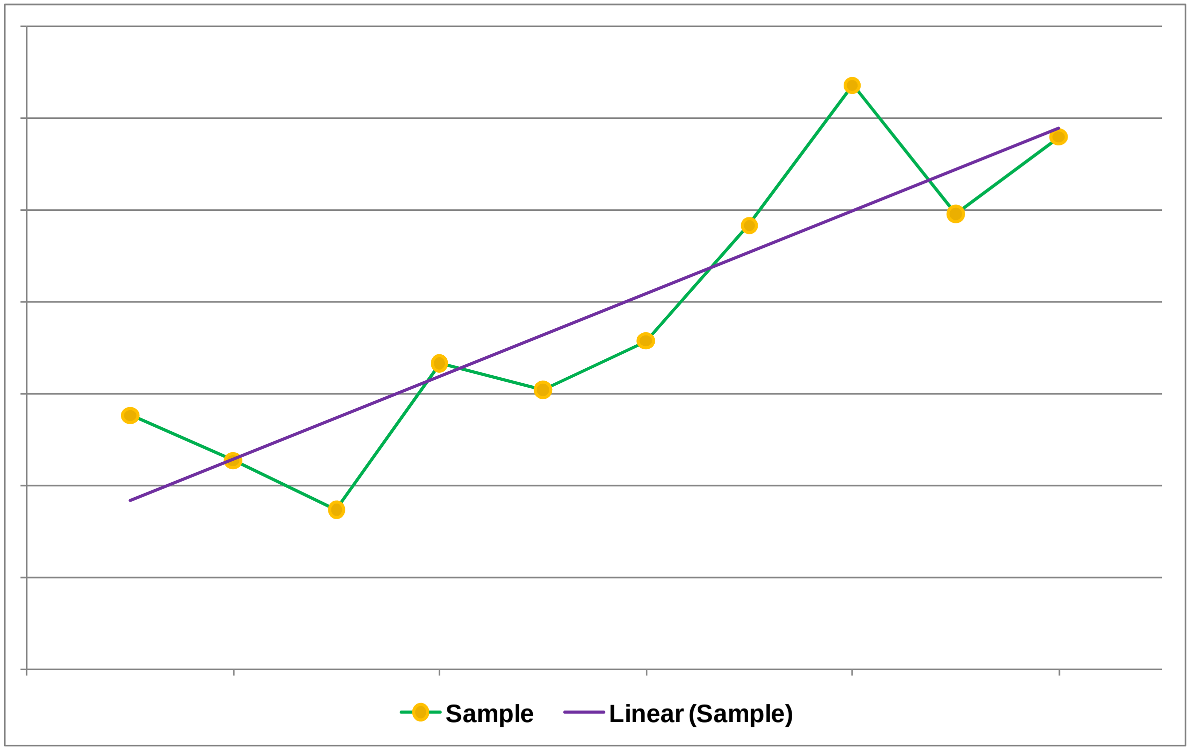
Motivations in Predictive Models

- Upward sloping trendline could be used as a factor



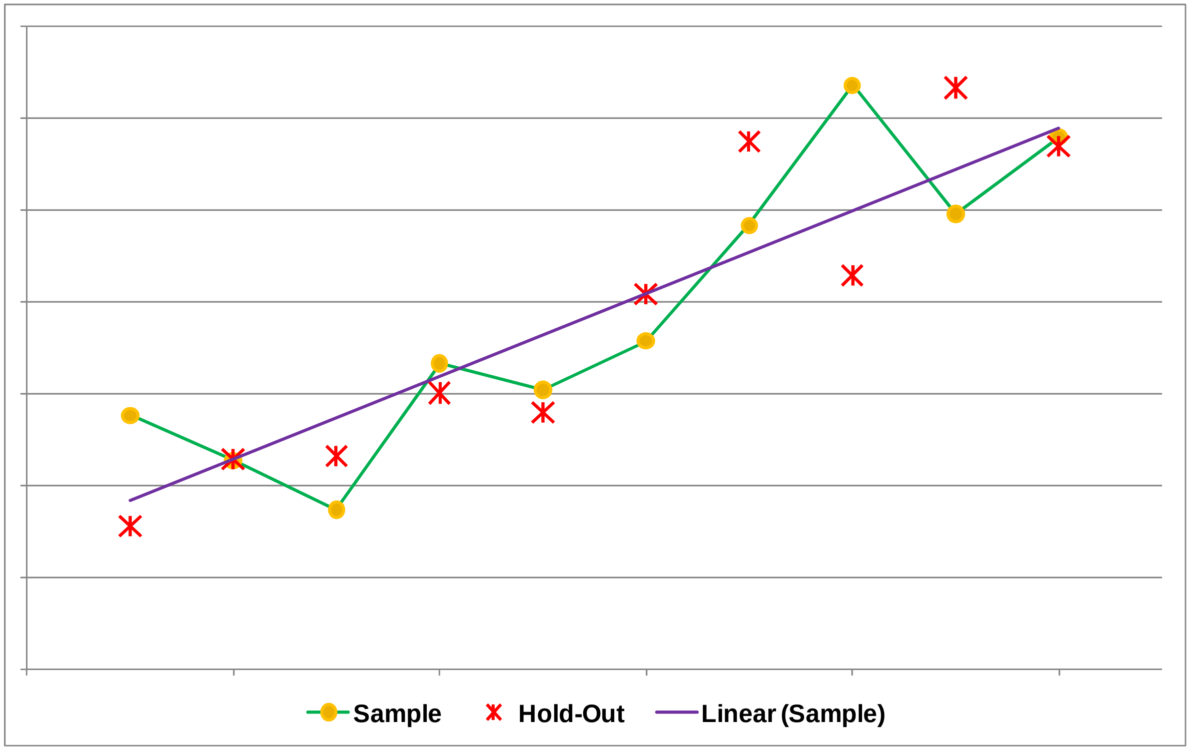
Motivations in Predictive Models

- Model could slavishly follow the data (green line) as the deviance is reduced:



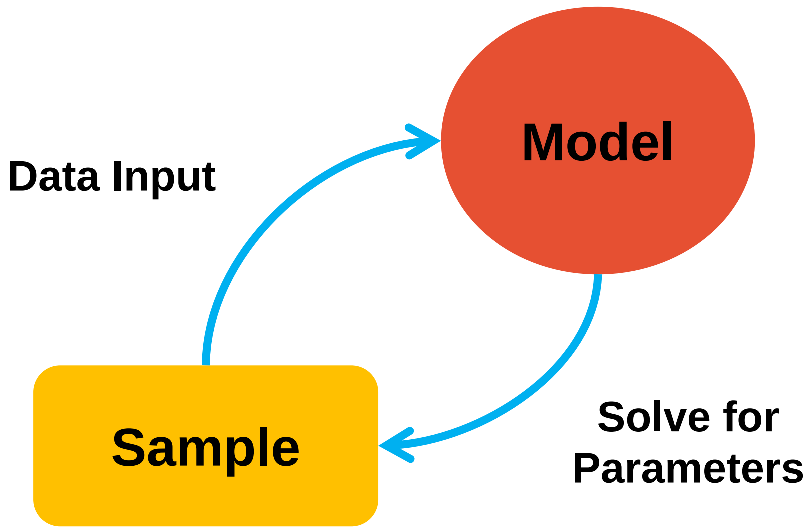
Motivations in Predictive Models

- However the model will not perform better on the hold out



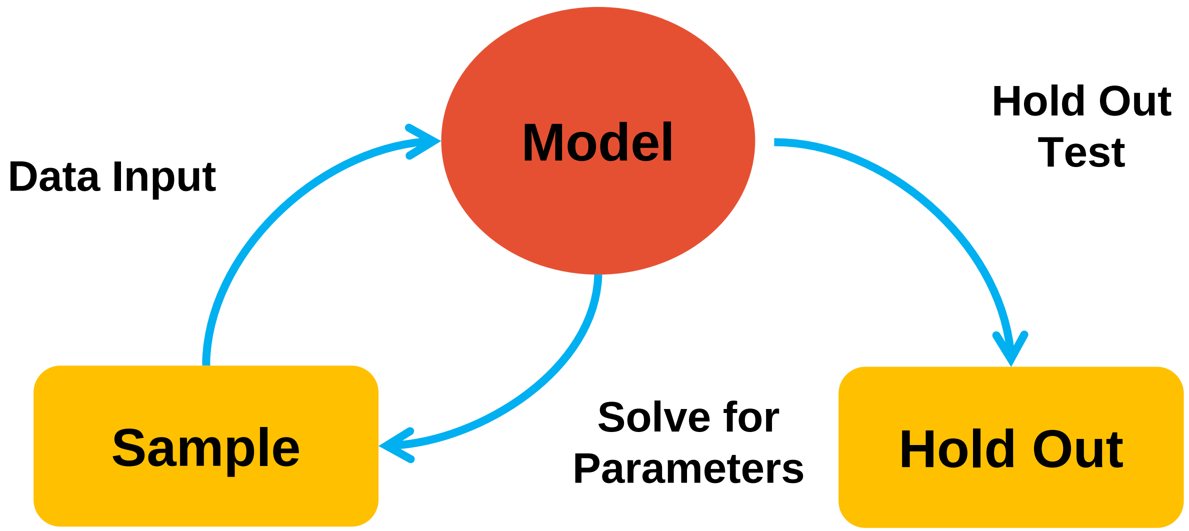
Motivations in Predictive Models

- Current practice uses the data to solve for the parameters



Motivations in Predictive Models

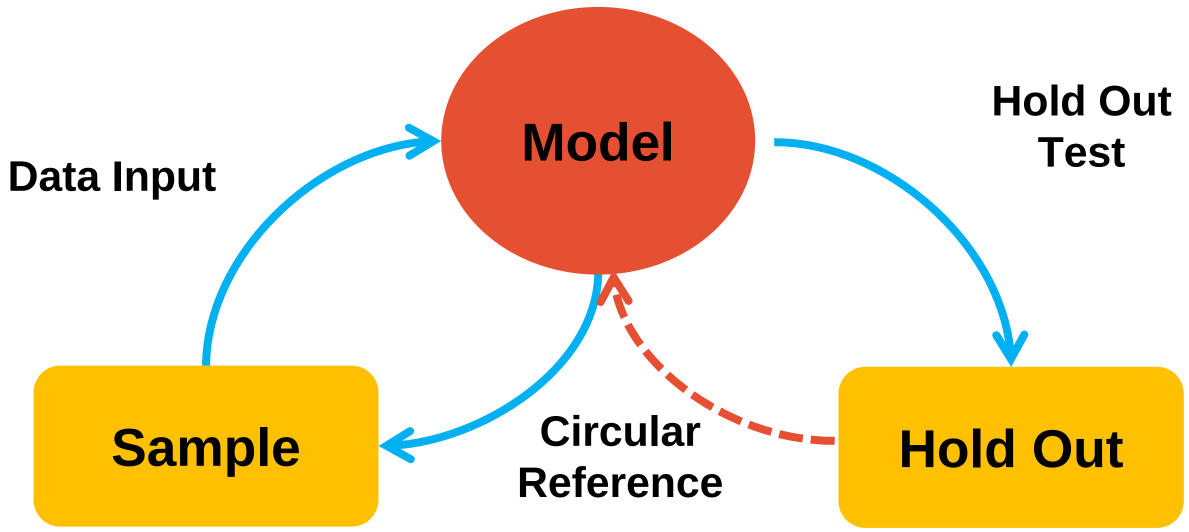
- This is then tested against a hold out sample



- Issues:
  - What is a good fit
  - What do adjust when there is a poor fit

Motivations in Predictive Models

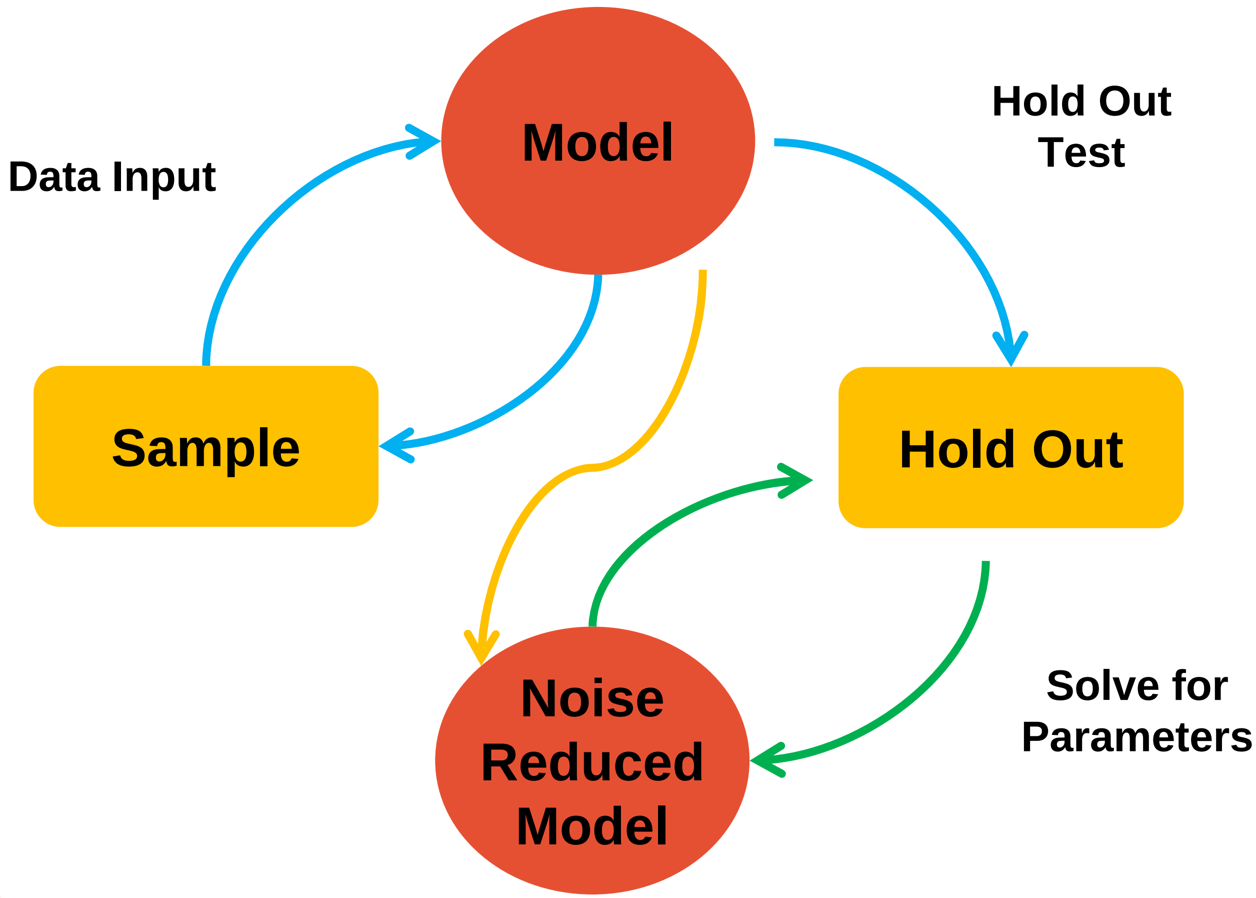
- Hold out sample could be used to adjust parameters



- This creates a circular reference

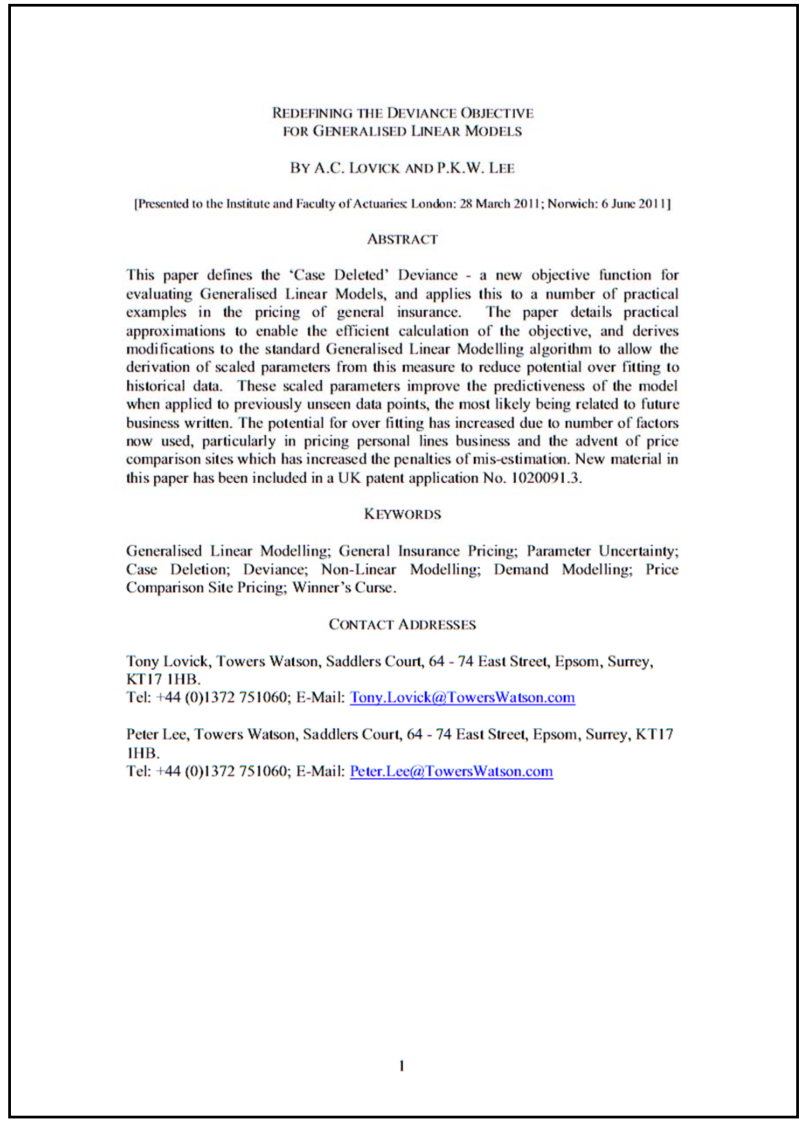
Motivations in Predictive Models

- The trick is to optimize parameters against the hold out without using it



"Case Deleted Deviance"

- Tony Lovick & Peter Lee
- Sessional Meeting of the Institute and Faculty of Actuaries 28 March 2011
- www.actuaries.org.uk



REDEFINING THE DEVIANCE OBJECTIVE
FOR GENERALISED LINEAR MODELS

BY A.C. LOVICK AND P.K.W. LEE

[Presented to the Institute and Faculty of Actuaries: London: 28 March 2011; Norwich: 6 June 2011]

ABSTRACT

This paper defines the 'Case Deleted' Deviance - a new objective function for evaluating Generalised Linear Models, and applies this to a number of practical examples in the pricing of general insurance. The paper details practical approximations to enable the efficient calculation of the objective, and derives modifications to the standard Generalised Linear Modelling algorithm to allow the derivation of scaled parameters from this measure to reduce potential over fitting to historical data. These scaled parameters improve the predictiveness of the model when applied to previously unseen data points, the most likely being related to future business written. The potential for over fitting has increased due to number of factors now used, particularly in pricing personal lines business and the advent of price comparison sites which has increased the penalties of mis-estimation. New material in this paper has been included in a UK patent application No. 1020091.3.

KEYWORDS

Generalised Linear Modelling; General Insurance Pricing; Parameter Uncertainty; Case Deletion; Deviance; Non-Linear Modelling; Demand Modelling; Price Comparison Site Pricing; Winner's Curse.

CONTACT ADDRESSES

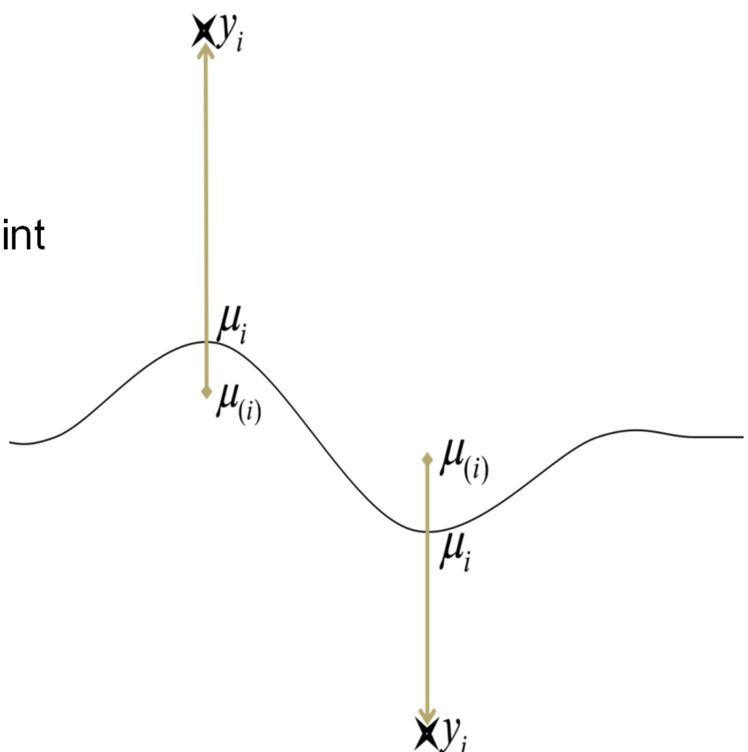
Tony Lovick, Towers Watson, Saddlers Court, 64 - 74 East Street, Epsom, Surrey, KT17 1HB.
Tel: +44 (0)1372 751060; E-Mail: Tony.Lovick@TowersWatson.com

Peter Lee, Towers Watson, Saddlers Court, 64 - 74 East Street, Epsom, Surrey, KT17 1HB.
Tel: +44 (0)1372 751060; E-Mail: Peter.Lee@TowersWatson.com

1

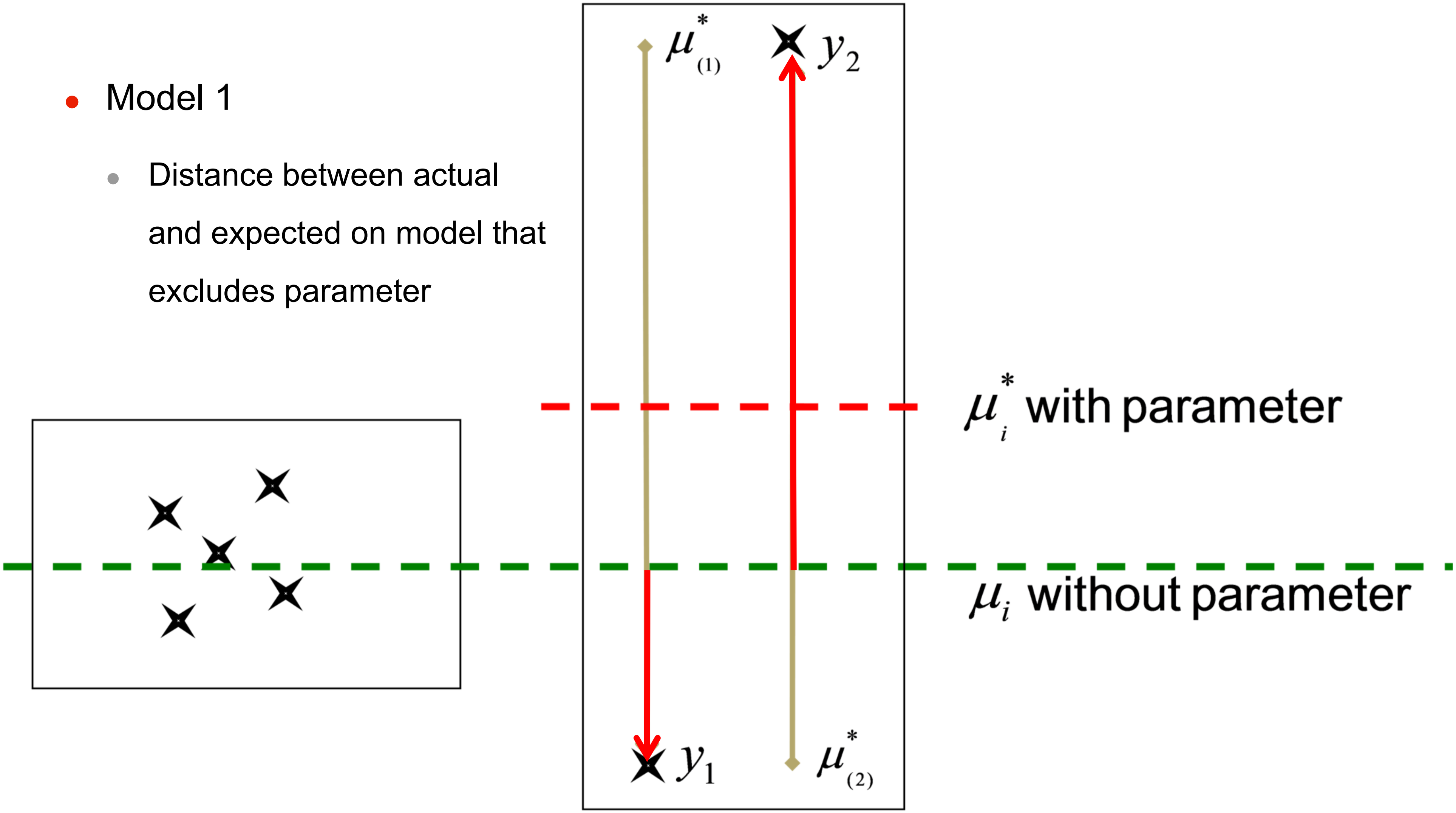
Concepts behind the Case Deleted Deviance

- Current practice is to parameterize a model with all data points and compare the fitted values (μ_i) with the observation (y_i)
  - Standard Deviance = SD (y_i, μ_i)
- A better approach is to physically refit the model with n-1 datapoints by excluding y_i to yield a new fitted value ($\mu_{(i)}$)
  - Case Deleted Deviance = CDD (y_i, $\mu_{(i)}$)
  - Case Deleted Deviance is independent of the point it relates to



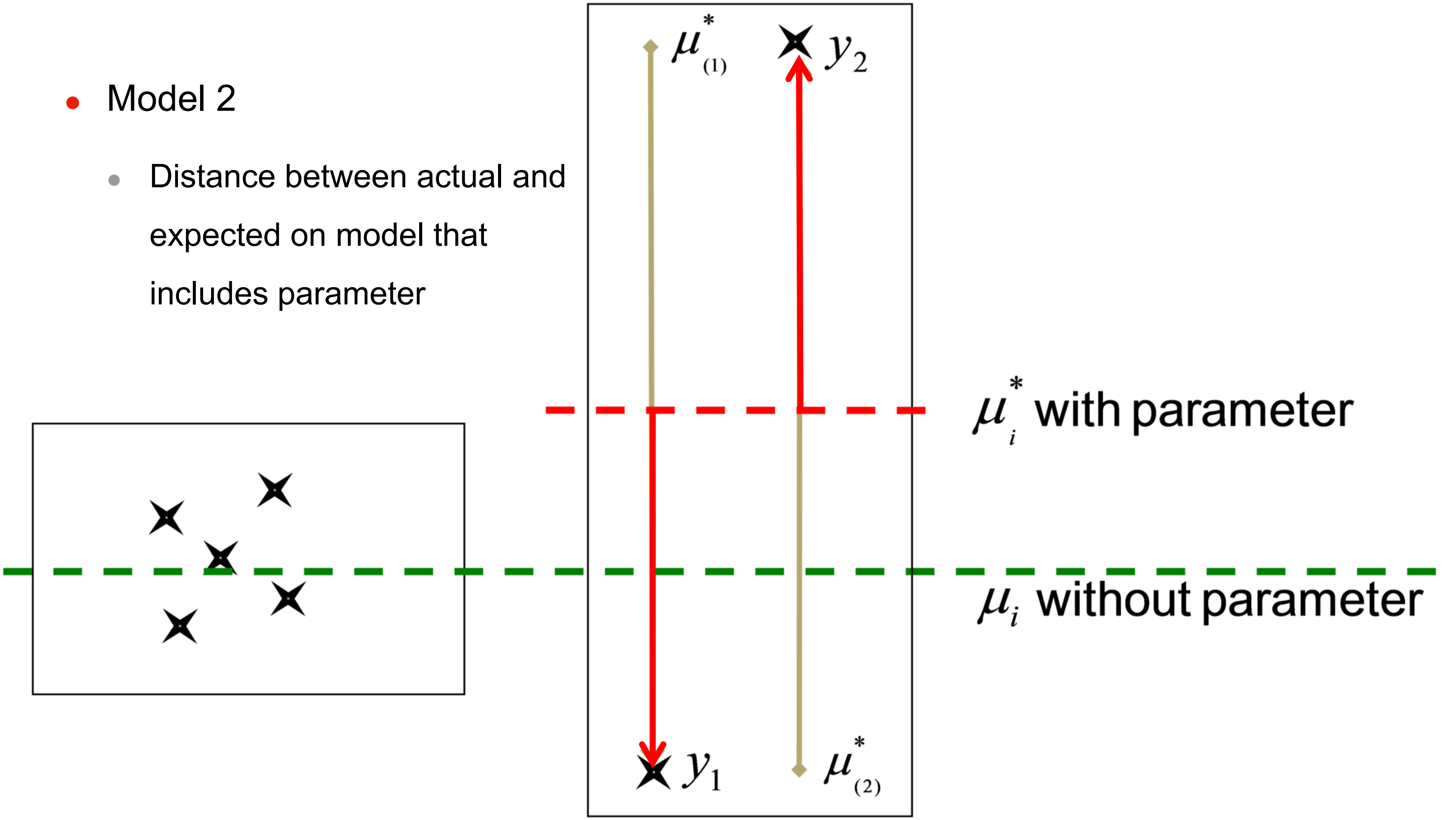
Concepts behind Noise Reduction

- Model 1
  - Distance between actual and expected on model that excludes parameter



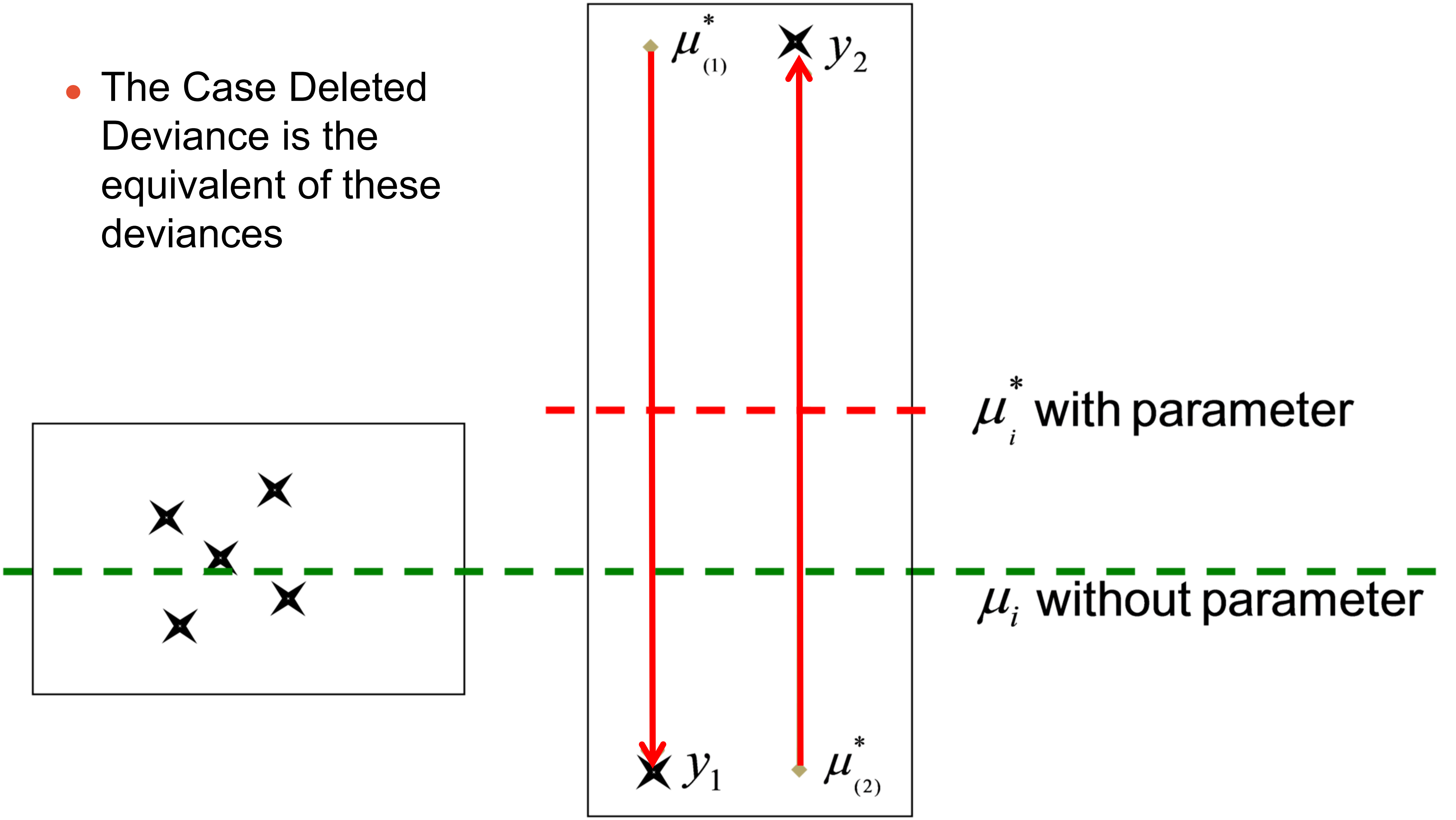
Concepts behind Noise Reduction

- Model 2
  - Distance between actual and expected on model that includes parameter



Concepts behind Noise Reduction

- The Case Deleted Deviance is the equivalent of these deviances

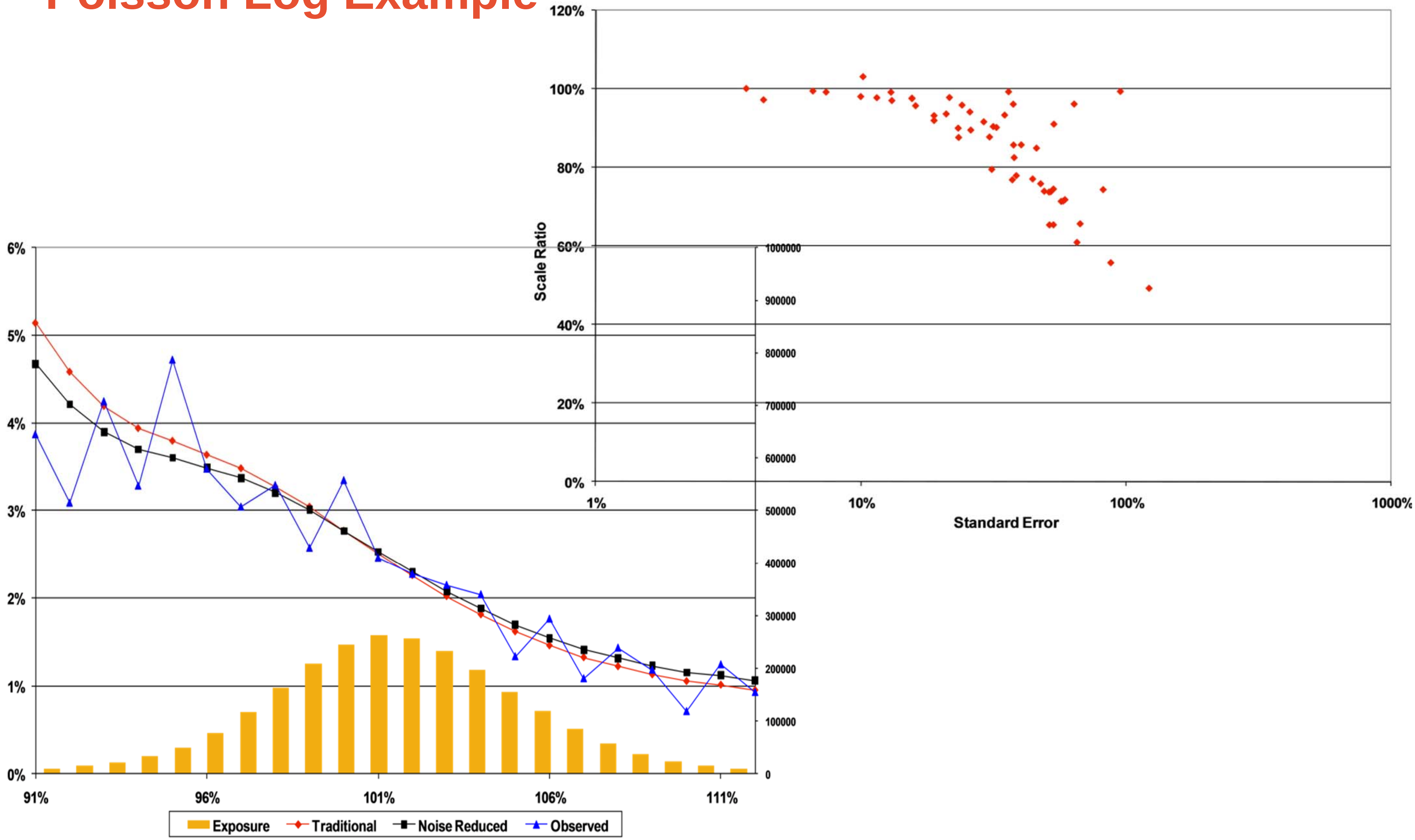


Concepts behind Noise Reduction

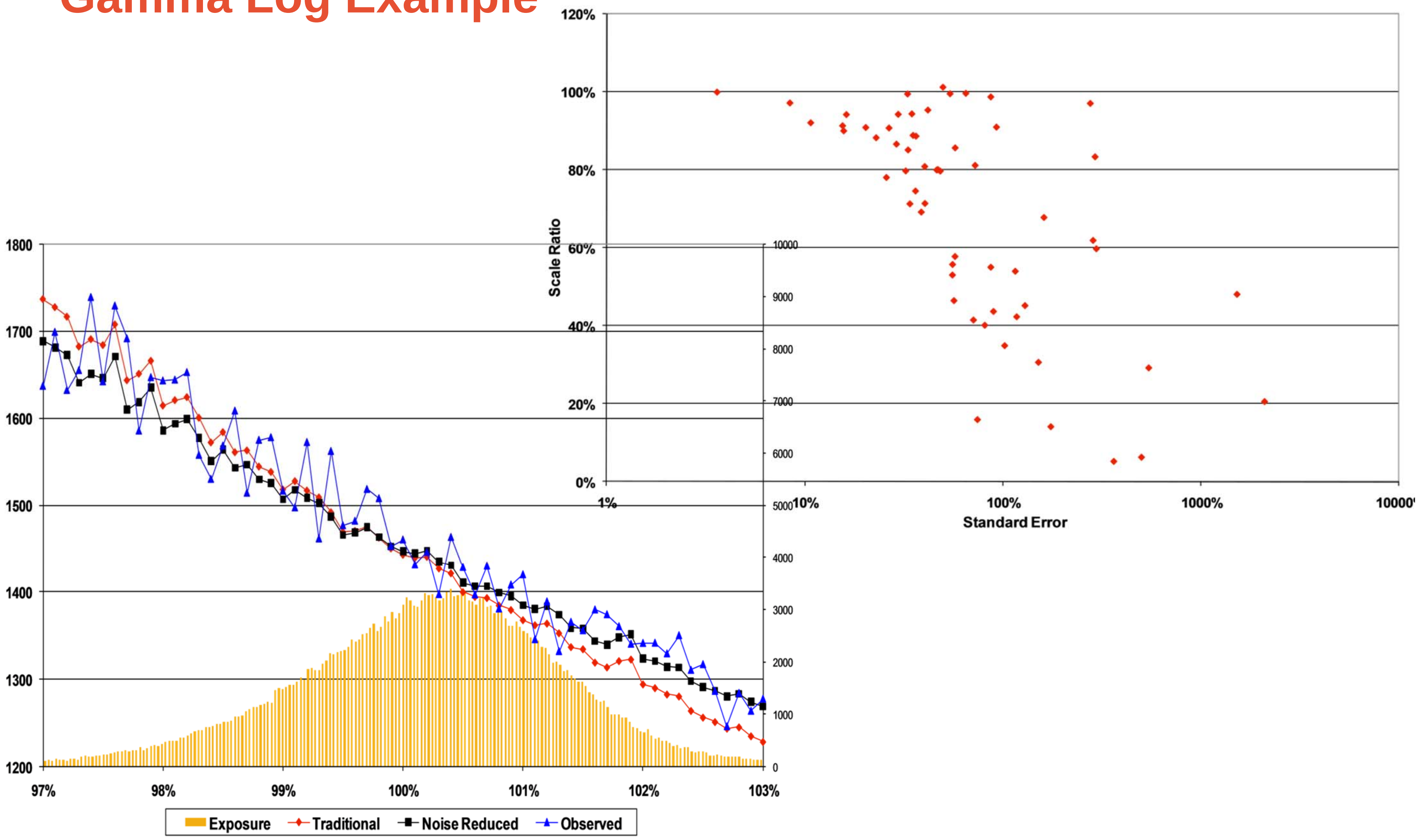
- Find the “best” scalars in the “Case Deleted Deviance” sense
- Higher Variance parameters get scaled back most
- Take account of parameter correlations

		Value	Standard Error	Standard Error (%)	Alias Indicator (%)	Weight	Weight (%)	Exp(Value)
		-2.743	0.031	1.1		236,207	100.0	0.064
		-0.060	0.020	34.1		100,320	42.5	0.942
						135,888	57.5	
						200,845	85.0	
		0.114	0.027	23.6		35,362	15.0	1.120
		0.117	0.010	8.9		236,207	100.0	1.124
		-0.265	0.016	6.0		219,928	93.1	0.768
		-0.076	0.017	22.5		219,928	93.1	0.926
16	PA Curve 1 spline 1 (OPoly(1))	0.212	0.039	18.5		8,194	3.5	1.236
17	PA Curve 1 spline 3 (OPoly(1))	0.041	0.009	23.1		229,373	97.1	1.042
18	PA Curve 1 spline 4 (OPoly(1))	-0.064	0.009	13.9		229,373	97.1	0.938
19	YADA Curve 1 (OPoly(1))	-0.176	0.014	8.2		236,207	100.0	0.838
20	YADA Curve 1 (OPoly(2))	0.062	0.012	19.0		236,207	100.0	1.064
21	VG Curve 1 spline 1 (OPoly(1))	-0.242	0.119	49.4		203,278	86.1	0.785
22	VG Curve 1 spline 2 (OPoly(1))	-0.116	0.070	60.0		235,863	99.9	0.890
23	VG Curve 1 spline 3 (OPoly(1))	-0.050	0.055	109.6		235,863	99.9	0.951
24	VG Curve 1 spline 4 (OPoly(1))	-0.177	0.092	52.3		235,863	99.9	0.838

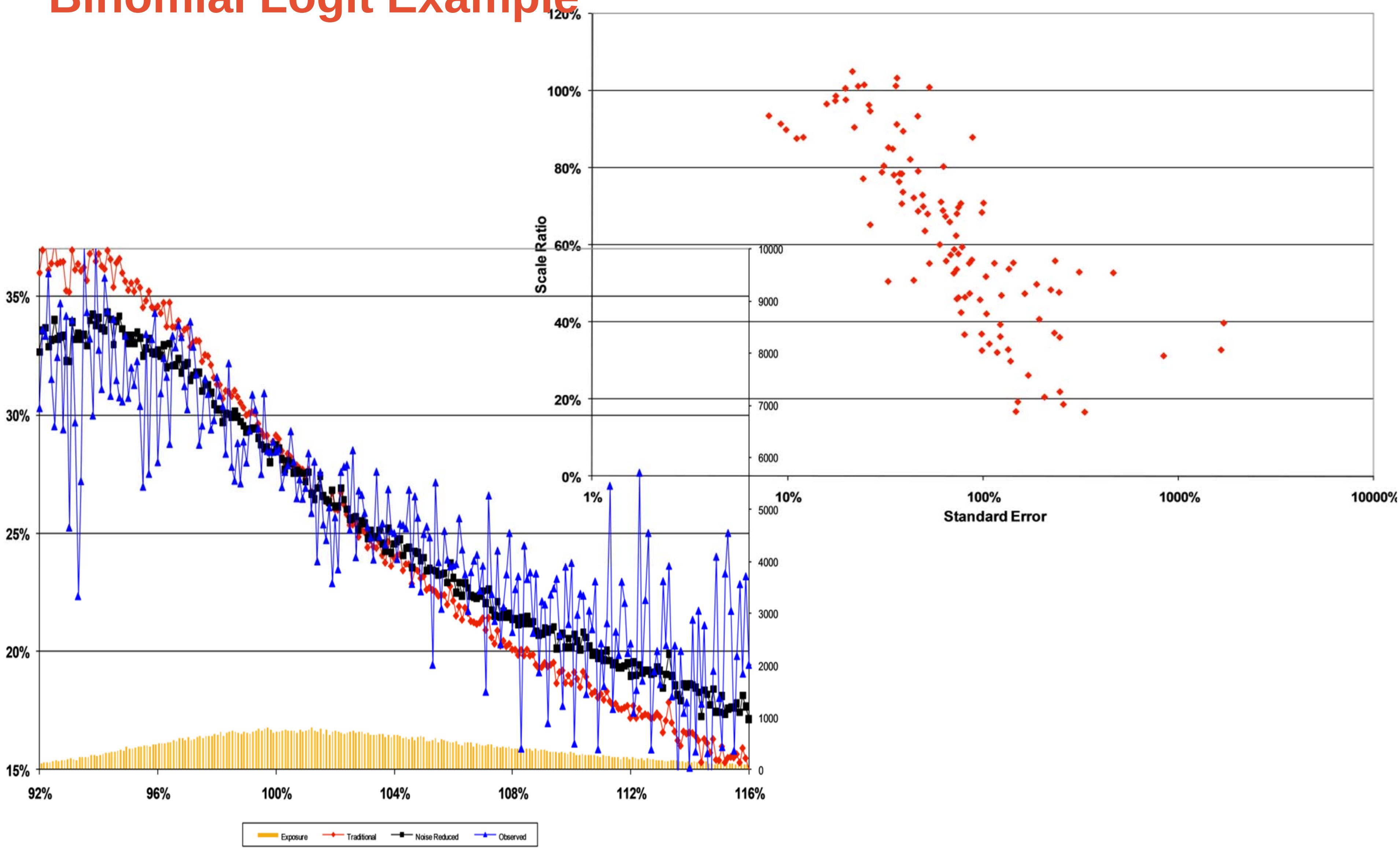
Poisson Log Example



Gamma Log Example

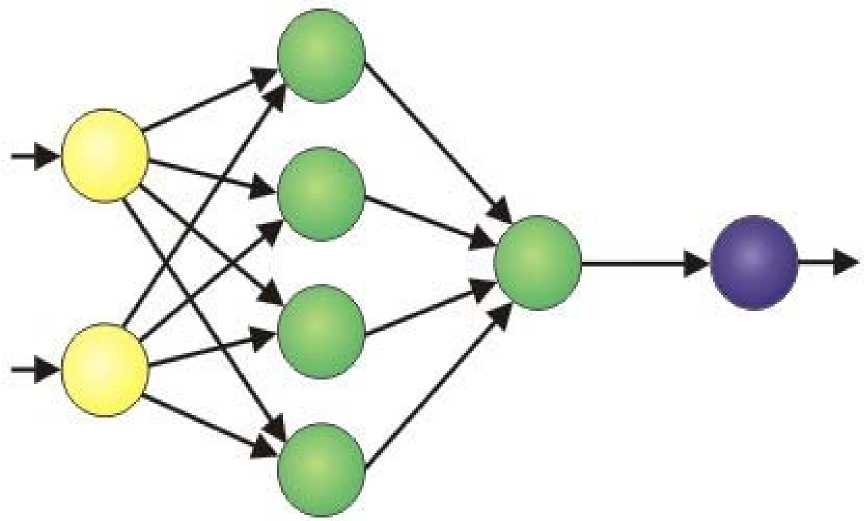


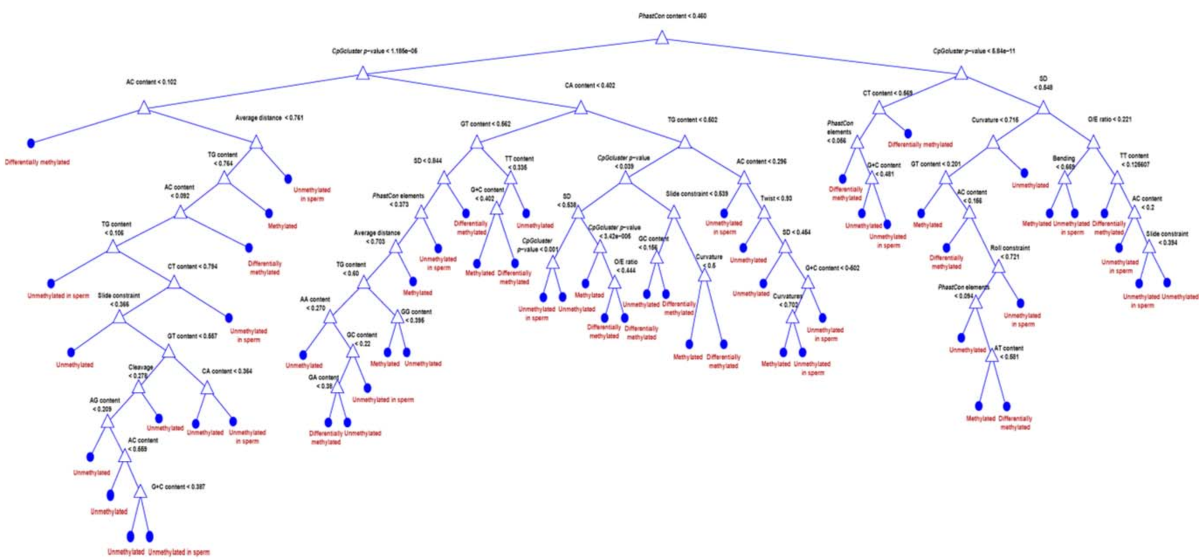
Binomial Logit Example

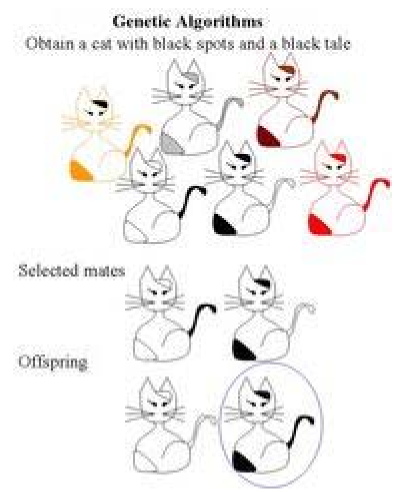


GLM Alternatives

- Noise Reduction can be applied to
  - Neural Networks
  - Genetic Algorithms
  - Decision Trees, etc







Competitive Analytics and Simulation

Competitive Analytics and Simulation

- Motivation: how to develop a rate product for a new market segment
  - Geographic expansion
  - New affinities,
  - Alternate products, etc

Competitive Pricing Analysis for New Products

- Rating Plan Development

Step 1

Plan project and launch

- Select state
- Select competitors and writing companies
- Identify exposure data needed to generate competitor premiums

Step 2

Collect competitive information

- Gather rating manuals and related filings for target competitors, including competitor or external information for:
  - Credit-based insurance scoring algorithms
  - Rating variables not used by the client
- Determine baseline coverage/product for comparison across carriers

Step 3

Develop insurance scoring or tier assignment

- Program each competitor's credit-based insurance scoring algorithms
- Validate results with competitor filing information (as available)
- Prepare comparative rating software to use the modeled credit-based insurance score instead of the default alignment option

Step 4

Create input database

- Create a database of in-force policies
- Populate variables that competitors use for which the client does not collect data
- Generate credit-based insurance score for each competitor by policy based on results from Step 3

Step 5

Generate competitor premiums

- Use batch-rating software to generate premiums
  - Generate new and renewal premiums for each risk in the input database
- Validate generated premiums against the filed competitor rule/rate manuals and other external publications (as available)

Step 6

Select Price and Develop Rate

- Identify target price based on market segment analysis
- Reverse engineer rate algorithm
- Prospectively assess performance

Collect Competitive Information

Identify Competitors

- Competitor profiling - vary from state to state
  - Identify how different competitors are approaching different market segments
  - Assess which competitors are profitable; which are growing; and which are growing profitably

- The goal is to identify key competitors for the target markets you wish to attract

Competitive Intelligence

Data Source(s)

- Competitor manuals and filings
- Comparative rating tool or spreadsheet software
- In-force book of business (biased sample)
- Business quoted but not written (optional; data may not be available or data may be incomplete)

Advantages

- Provides “on-the-street” premiums at a policy level for competitive analysis
- Provides a complete picture of the effectiveness of the current pricing structure, down to each individual rating segment
- Provides additional direction for internal pricing analysis
- Can be used to develop a new company rating plan
- Provides foundation for optimization

Disadvantages

- Time-intensive to generate competitor premiums if not already using a comparative rater
- Competitor information may not be readily available (especially tier/credit score)
- Easy to misinterpret information in collection/compilation of rating plan filings
  - Manual exchange programs typically not up to date
  - For groups with multiple writing companies, the full spectrum of tiers and rates may not be used in practice

Create Input Database

TOWERS WATSON

Field Types

- Build the new market basket using a combination of actual data with simulation
- Assess relevant fields and identify actions:

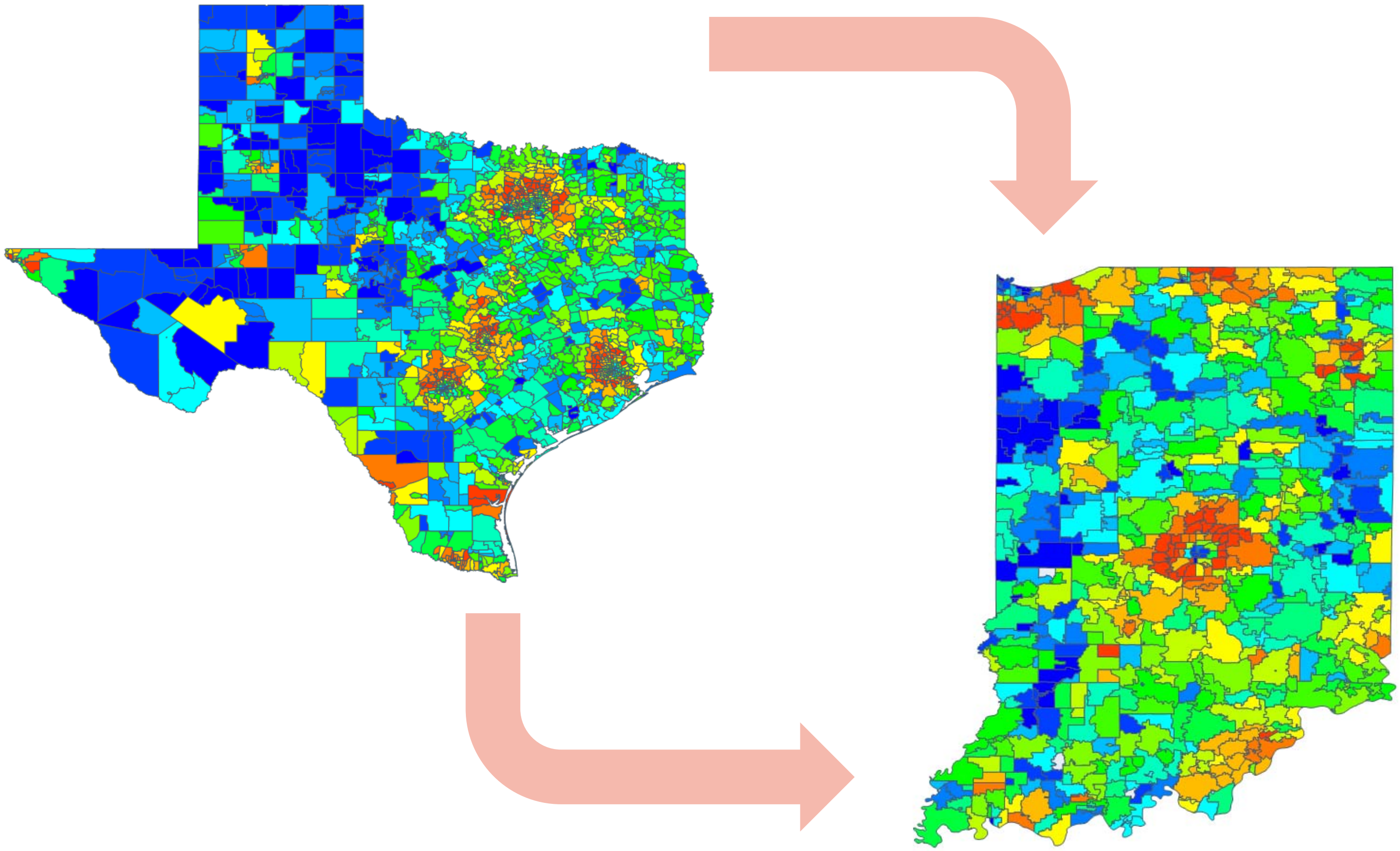
Policy Fields	Location Fields	Building Fields	Coverage Fields
• Policy identifier • Form • Primary residence	• Address • Latitude/Longitude • Distance to Coast • Protection Class • BCEG	• Year Built • Square footage • Coverage A • Construction • Wind mitigation features	• Wind coverage • Sinkhole coverage • Deductibles • Replacement cost on contents

Supplement Data with Other Sources

- The goal is to use as much ‘real’ data as possible

Geodemographic Extrapolation to New States

- Using census data to map distributions of existing data into new geographic units



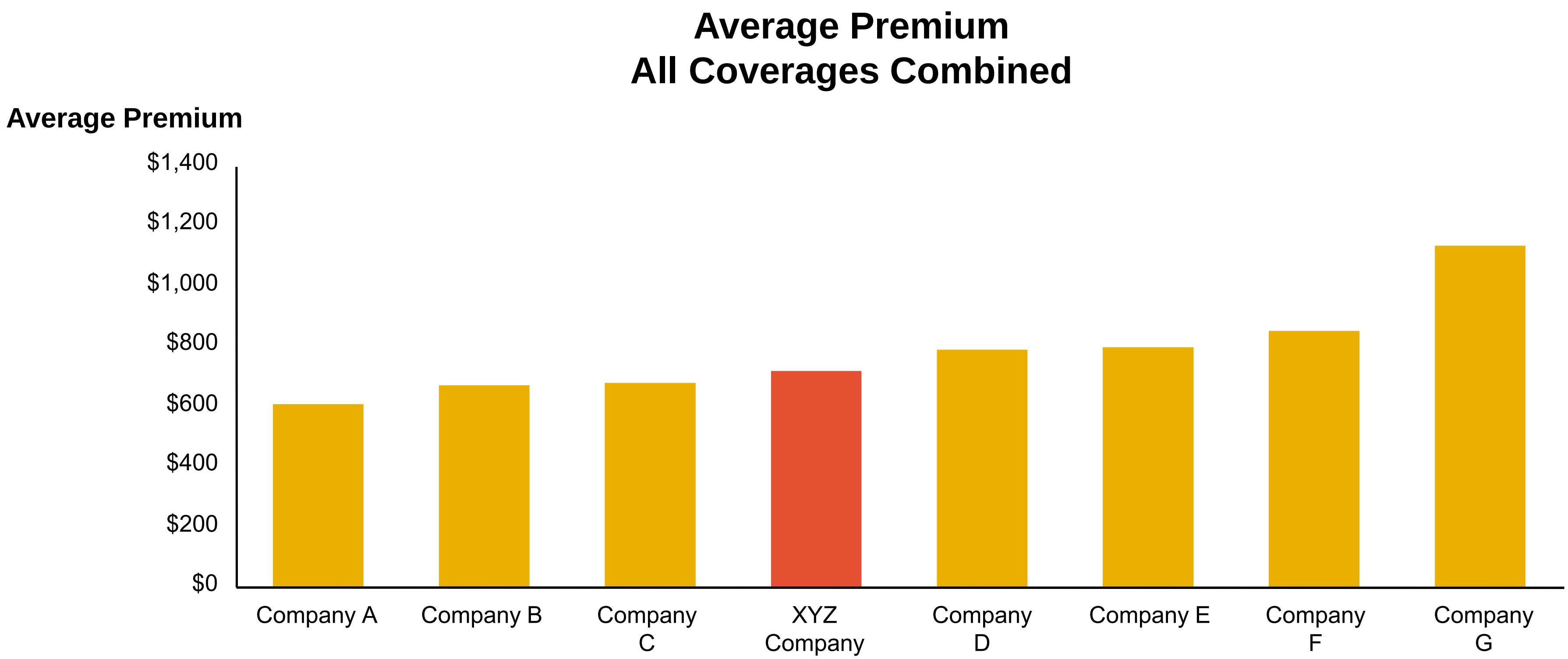
Simulate Missing Data

- Simple simulation
  - Need to specify the desired distribution of policies across the factor AND the correlation between that factor and other factors
- Location based simulation
  - Ties the simulation to specific segments within the book
- Stratified
  - Ties the correlation to scores (e.g. premiums)

Generate Competitor Premiums

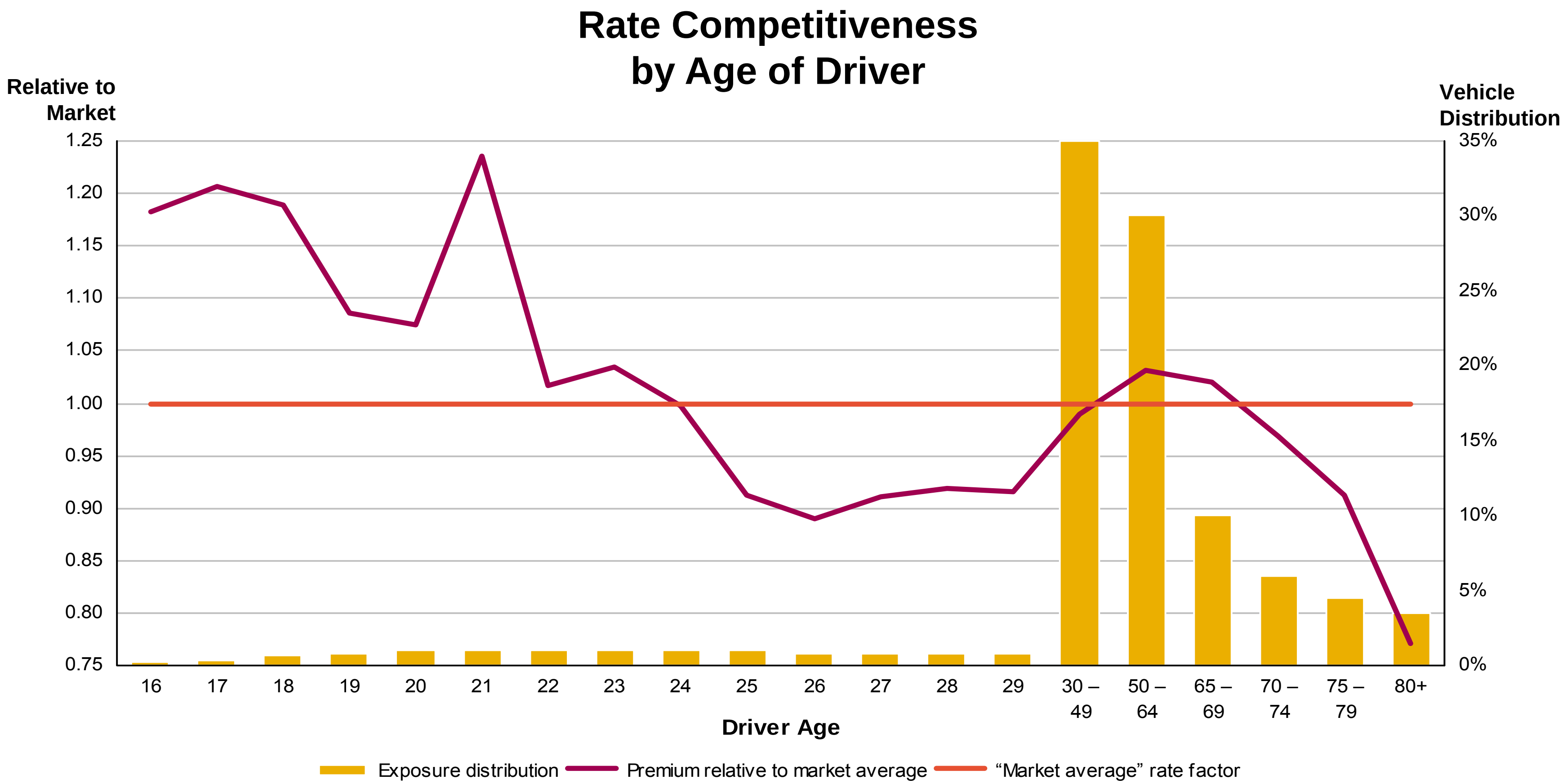
Batch rate the Market Basket

- Quantitative analysis starts with a comparison of algorithms on the market basket



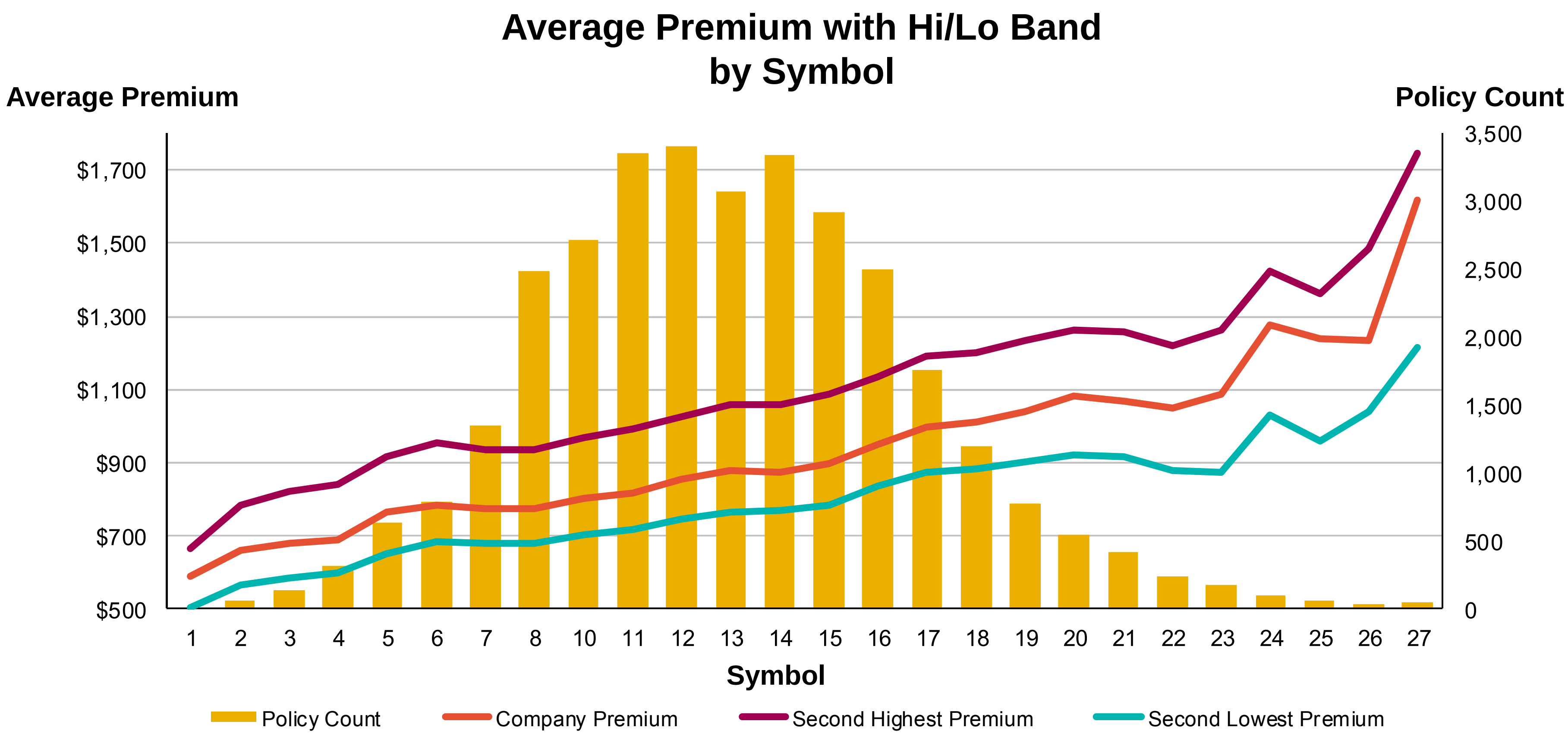
Beware of potential inherent bias in using mix of business

Batch rates are validated across segments



Use of a market average premium requires a distribution to be selected across competitors — simple average or weighted average

Alternate Metrics add Further Insight



The target market position should be identified and then metrics can be developed to monitor competitive position relative to target

Comparison by rating factor/segment

Auto Variables

- Driver-related variables
  - Driver age
  - Gender
  - Marital status
  - Education level
  - Employment status
  - Military status
  - Occupation
  - Driving record
  - Months licensed
  - Accident prevention discount
  - Advanced training discount
  - Good student discount
- Prior insurance
  - Length of time with prior carrier
  - Prior limits
  - Type of insurer
  - Lapse in coverage
- Household-related variables
  - Years at residence
  - Location
  - Policy tenure
  - Insurance score
  - Tier/insurance score for client and each competitor
  - Advanced shopper
  - Paid-in-full
  - EFT
  - Paperless documents
  - Multiple line discounts
  - Length of vehicle ownership
  - Household composition
  - Homeownership
  - Residence type
- Geography
  - Territory
  - Zip code
- Vehicle-related variables
  - Model year
  - Vehicle make
  - Cylinders
  - Performance
  - Symbol
    - — Liability and medical symbol
    - — Comprehensive and collision symbol
  - Annual mileage
  - Vehicle use
  - Miles driven to work
  - Location
  - Airbags
  - Disabling device
  - Anti-lock brakes
- Coverage-related variables
  - Limits (BI, PD, medical payment)
  - Deductibles (comprehensive, collision)

Comparison by rating factor/segment

Homeowners Variables

- Home-related variables
  - Construction type
  - Built with fire-resistive material
  - Year built
  - Presence of a basement
  - Presence of a burglar alarm
  - Presence of a sensaphone
  - Presence of a fire alarm
  - Presence of a sprinkler system
  - Presence of a pool
  - Distance to fire station
  - Distance to fire hydrant
  - Floor area
  - Type of garage
  - Home renovations
    - — Age of heating and cooling systems
    - — Age of plumbing
    - — Age of wiring
    - — Age of roof
  - Type of roof
  - Prior losses/claims
- Home-related variables (cont'd)
  - Number of family units
  - Number of bathrooms
  - Number of levels
  - Protection class
  - Town house
- Prior insurance
  - Length of time with prior carrier
- Geography
  - Territory
  - Zip code
- Coverage-related variables
  - Coverage A — dwelling amount of insurance
  - Coverage C — contents coverage
  - Coverage E — liability
  - Deductible
- Resident-related variables
  - Owner age
  - Marital status
  - Retired
  - Months owned
  - Presence of a mortgage
  - Number of occupants
  - Number of smokers
  - Policy tenure
  - Tier/insurance score for client and each competitor
  - Multiple line discount
    - — Auto
    - — Life
    - — Umbrella
  - Attendance at a safety seminar

Select Price & Develop Rate

Price Selection

- Prices can be selected using a wide array of approaches from extremely simple to very sophisticated:
  - Follow Progressive
  - Cheapest of n competitors
  - Market average
  - Clustering on relative position and intensity

- Analysis usually done on a policy basis
  - Cost models are used to allocate policy decisions to individual risks

Target price is based on individual competitors for clusters of risks

vs. Competitor A

- More vehicles than drivers, ages 40 – 65, in tiers 10+
- Drivers aged <30 or above 65, more vehicles than drivers

vs. Competitor B

- Drivers below age 20, one driver on the policy
- Drivers below 23, with three or more drivers on the policy

Percent of risks in State X where price is $50 or more below competitor

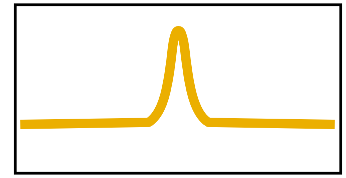
Percent of risks in State X where price is $50 or more above the competitor

Note: Text bullets show representative types of risks.

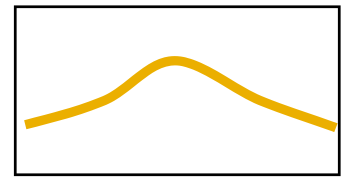
Competitive segmented in a cluster analysis

Market Competitive Intensity Index → Measure of price dispersion In the market

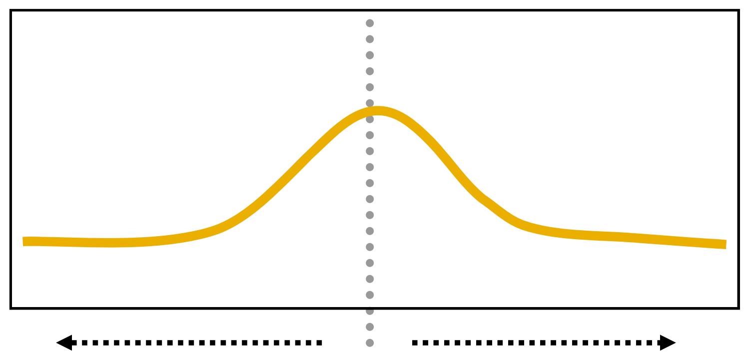
Low price dispersion/ high competitive intensity



High price dispersion/ Low competitive intensity



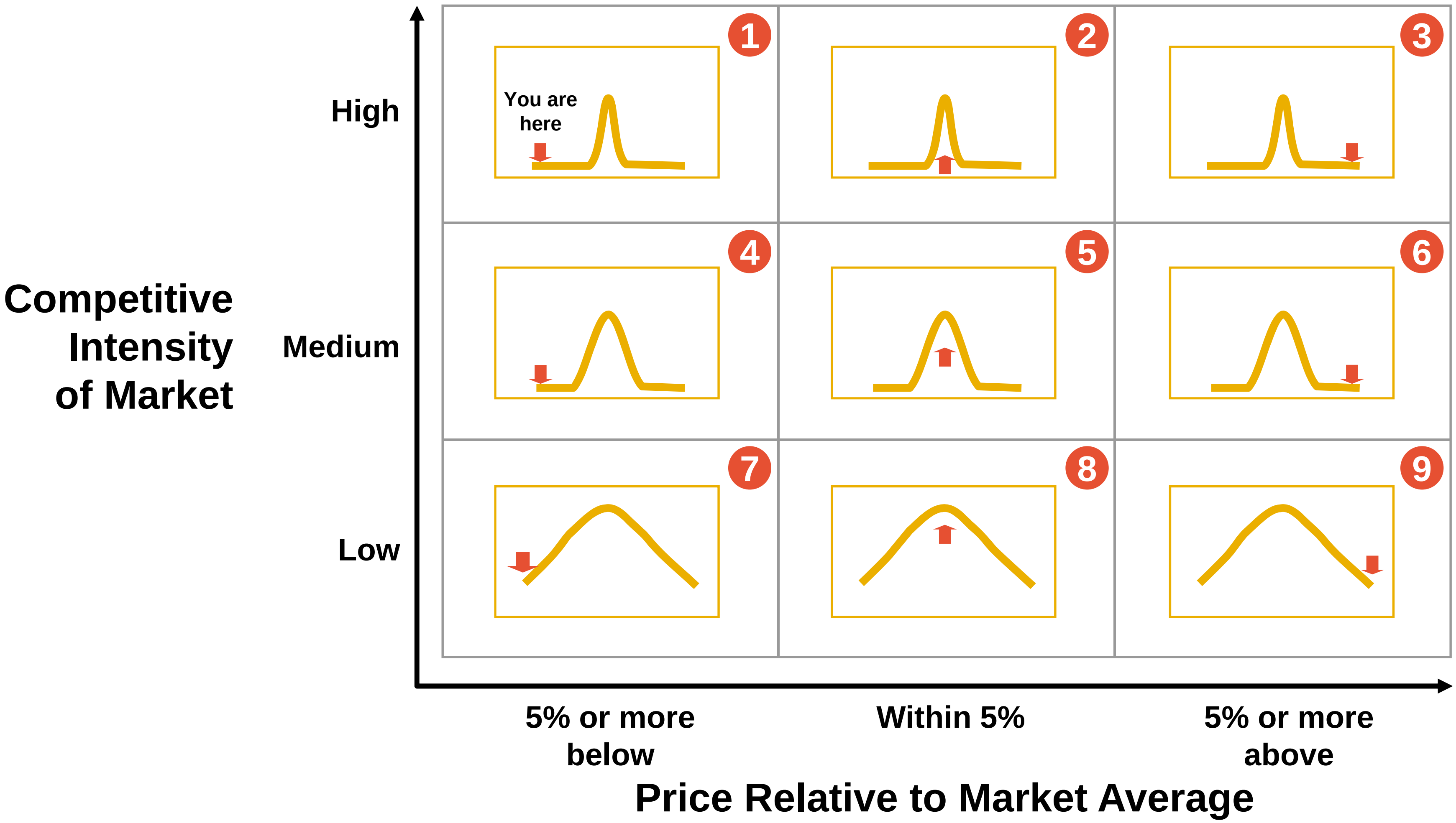
Price → Proposed price compared to market →



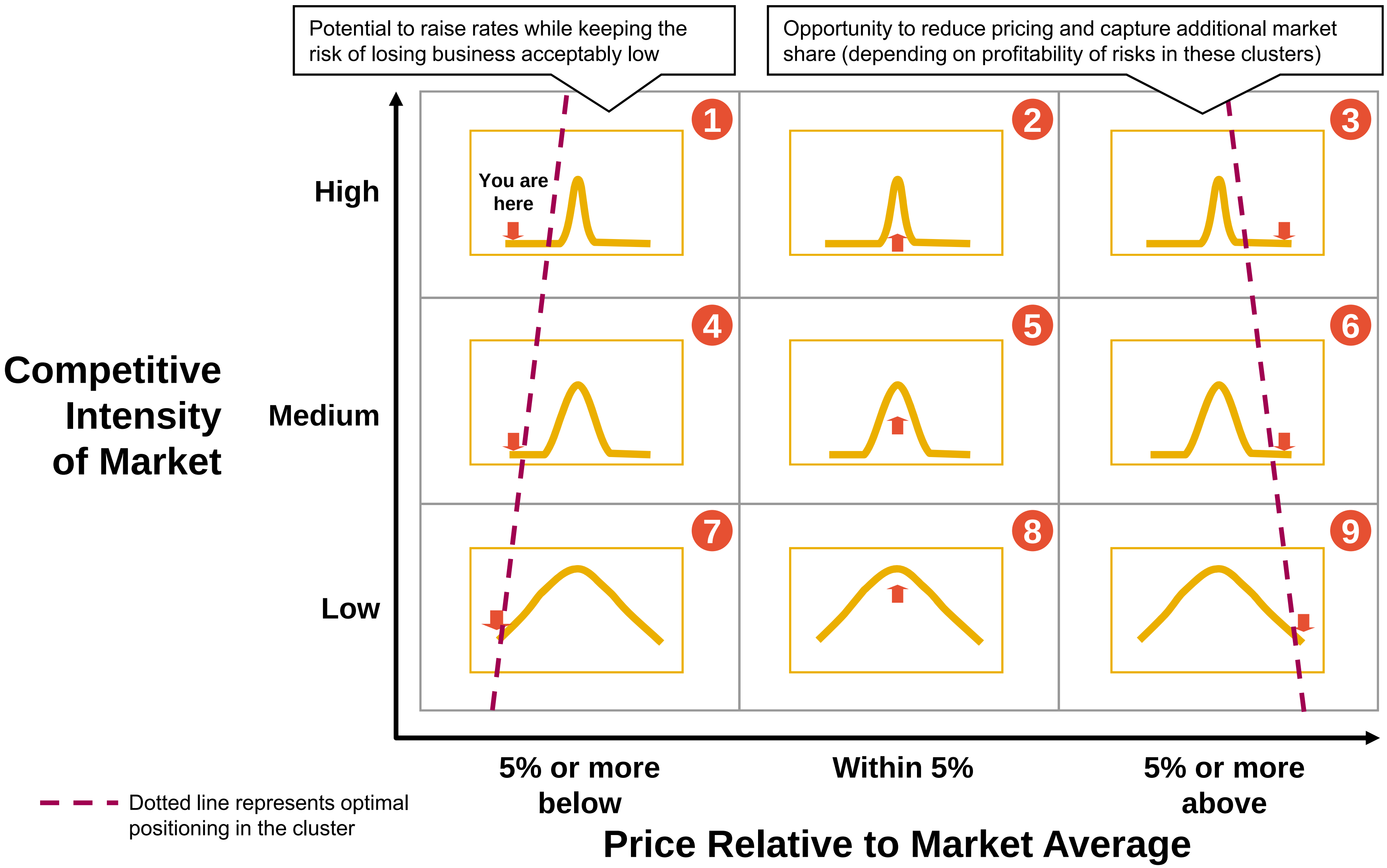
Price below market | *Price above market*

Market average

Final price is selected from cluster groups



The clusters suggest potential pricing strategies

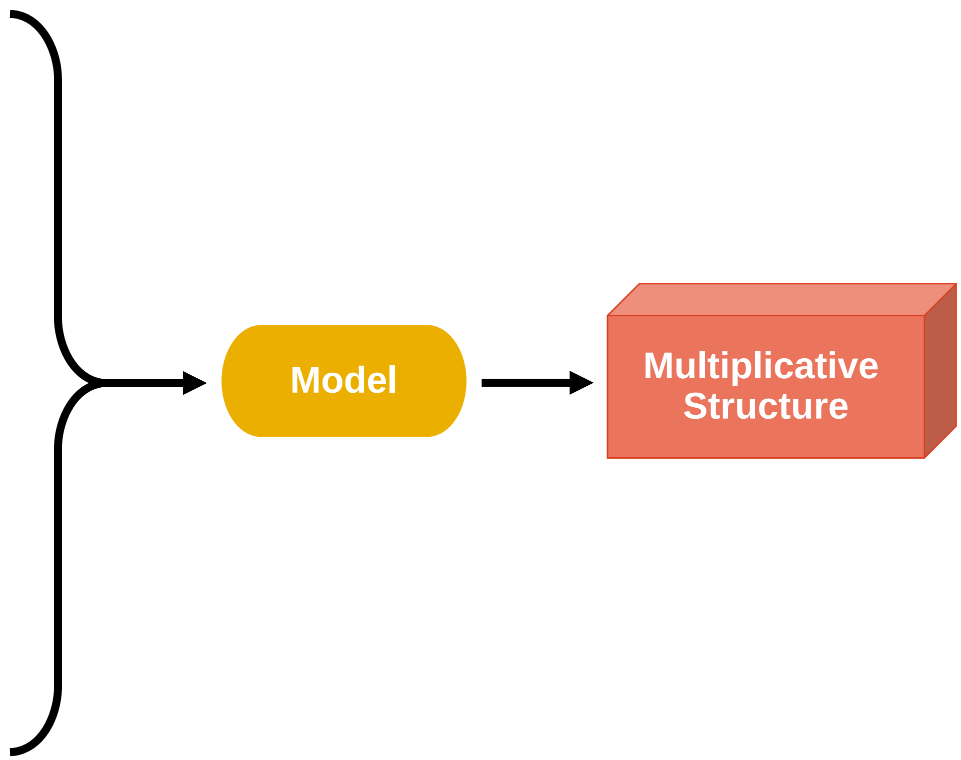


Reverse Engineer the Price into a Rate Algorithm

Individual Prices

Policy no.	Premium
PEL009759458	327
UQJ408808153	555
KZH964999642	261
DDU700866747	349
VUQ391058119	334
YUM718736198	331
GBQ270981530	279
CSR303293030	188
XTB008693907	175
TJJ330632016	319
MFD704472553	349
ZVI955030095	277
ZJY528736252	372
VRF026498810	647
BIN297260627	555
SXT608697514	203
JAE716278042	163
XUS991829954	633
IVN822320056	641
FOD690200573	232
DCI071346826	325
SEL511154881	538

Can fit model to results to yield multiplicative structure using standard rating factors

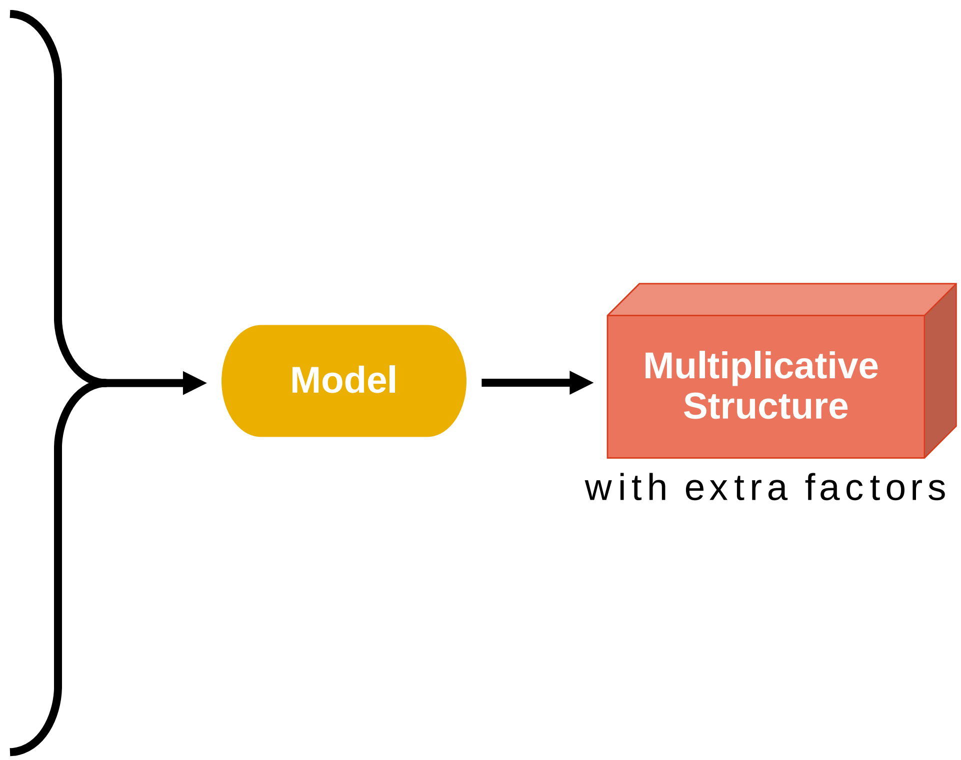


Reverse Engineer the Price into a Rate Algorithm

Individual Prices

Policy no.	**Premium**
PEL009759458	327
UQJ408808153	555
KZH964999642	261
DDU700866747	349
VUQ391058119	334
YUM718736198	331
GBQ270981530	279
CSR303293030	188
XTB008693907	175
TJJ330632016	319
MFD704472553	349
ZVI955030095	277
ZJY528736252	372
VRF026498810	647
BIN297260627	555
SXT608697514	203
JAE716278042	163
XUS991829954	633
IVN822320056	641
FOD690200573	232
DCI071346826	325
SEL511154881	538

Can fit model to results to yield multiplicative structure using standard rating factors plus alternative factors

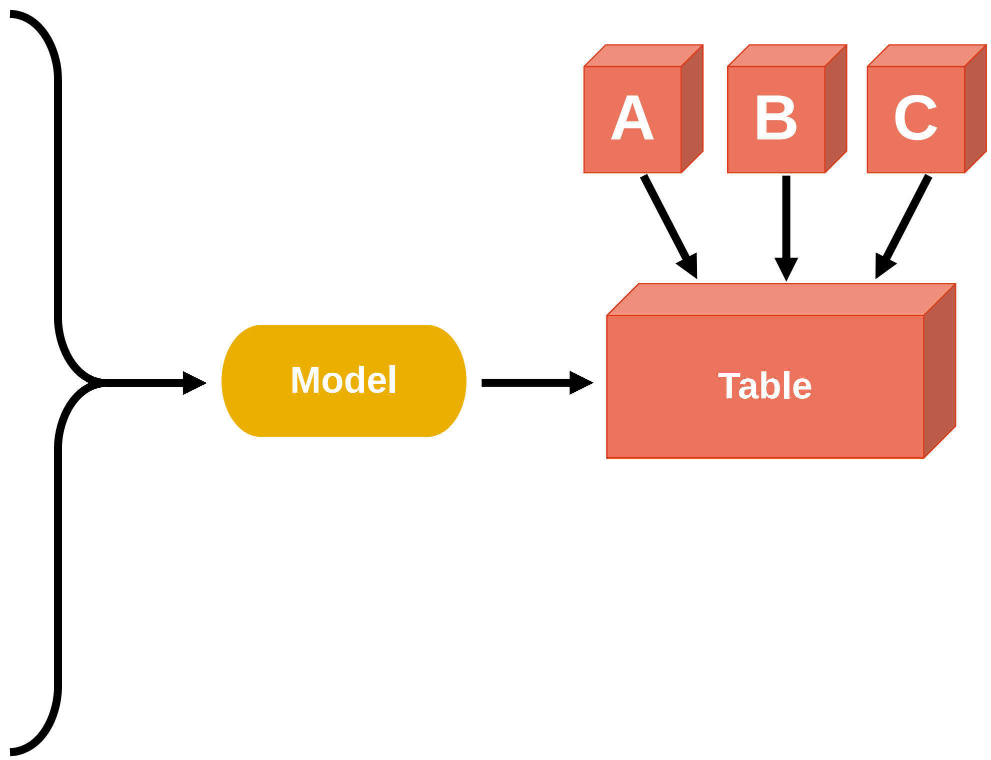


Reverse Engineer the Price into a Rate Algorithm

Individual Prices

Policy no.	Premium
PEL009759458	327
UQJ408808153	555
KZH964999642	261
DDU700866747	349
VUQ391058119	334
YUM718736198	331
GBQ270981530	279
CSR303293030	188
XTB008693907	175
TJJ330632016	319
MFD704472553	349
ZVI955030095	277
ZJY528736252	372
VRF026498810	647
BIN297260627	555
SXT608697514	203
JAE716278042	163
XUS991829954	633
IVN822320056	641
FOD690200573	232
DCI071346826	325
SEL511154881	538

Can use factors in combination, or secondary models, to derive score factors that feed into traditional table form

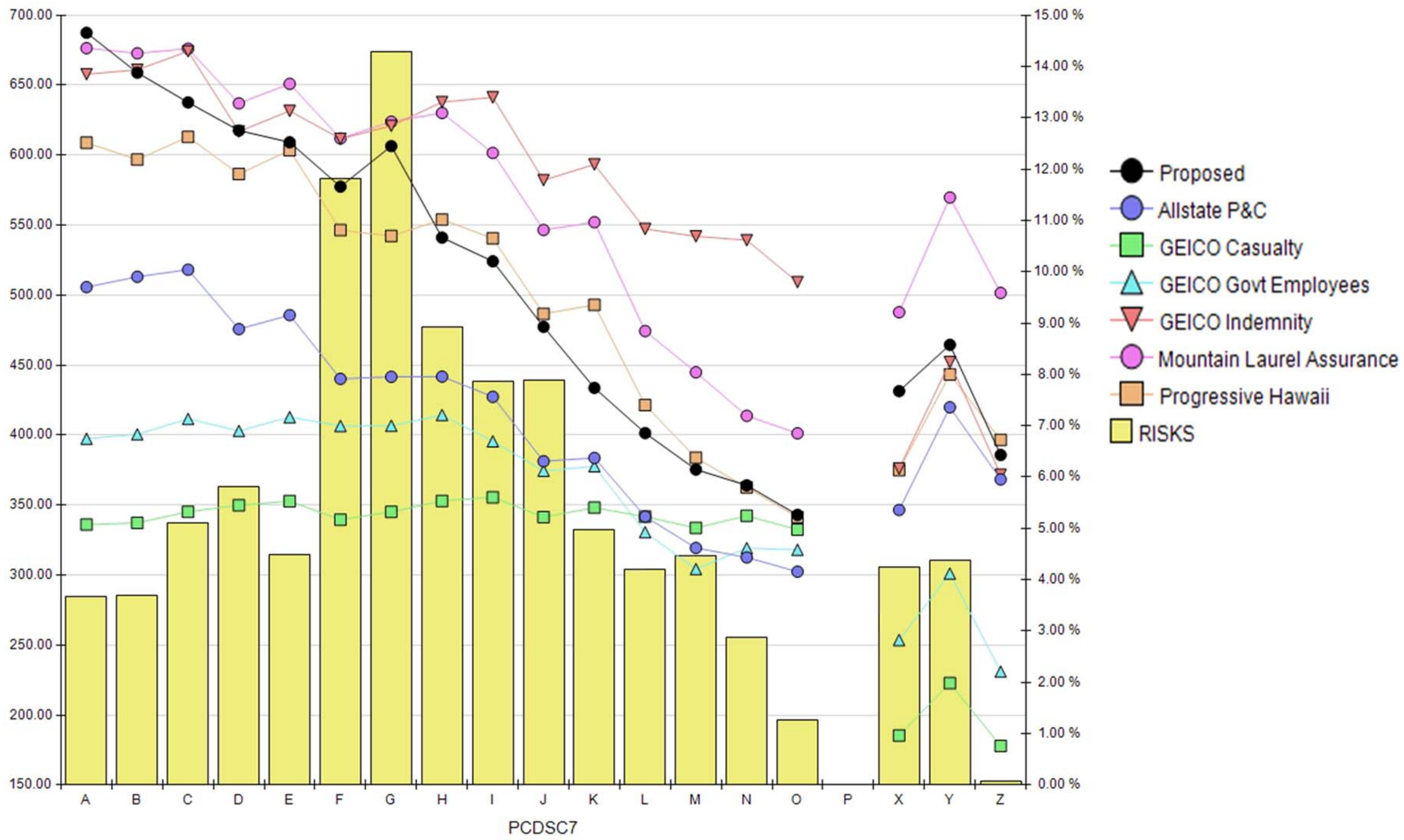


towerswatson.com

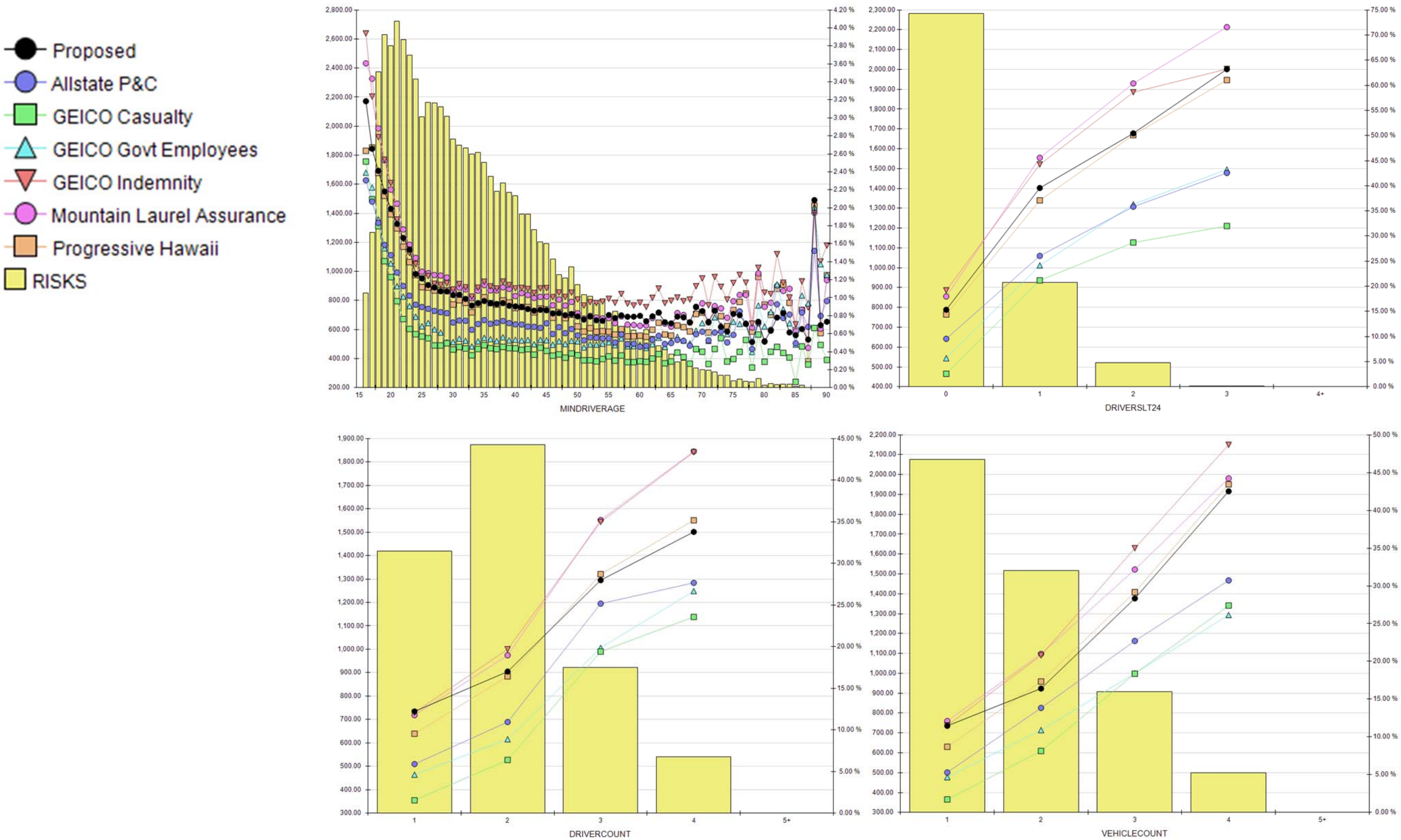
Prospective Assessment

TOWERS WATSON

Average Competitor & Proposed Vehicle Premium by Credit Score



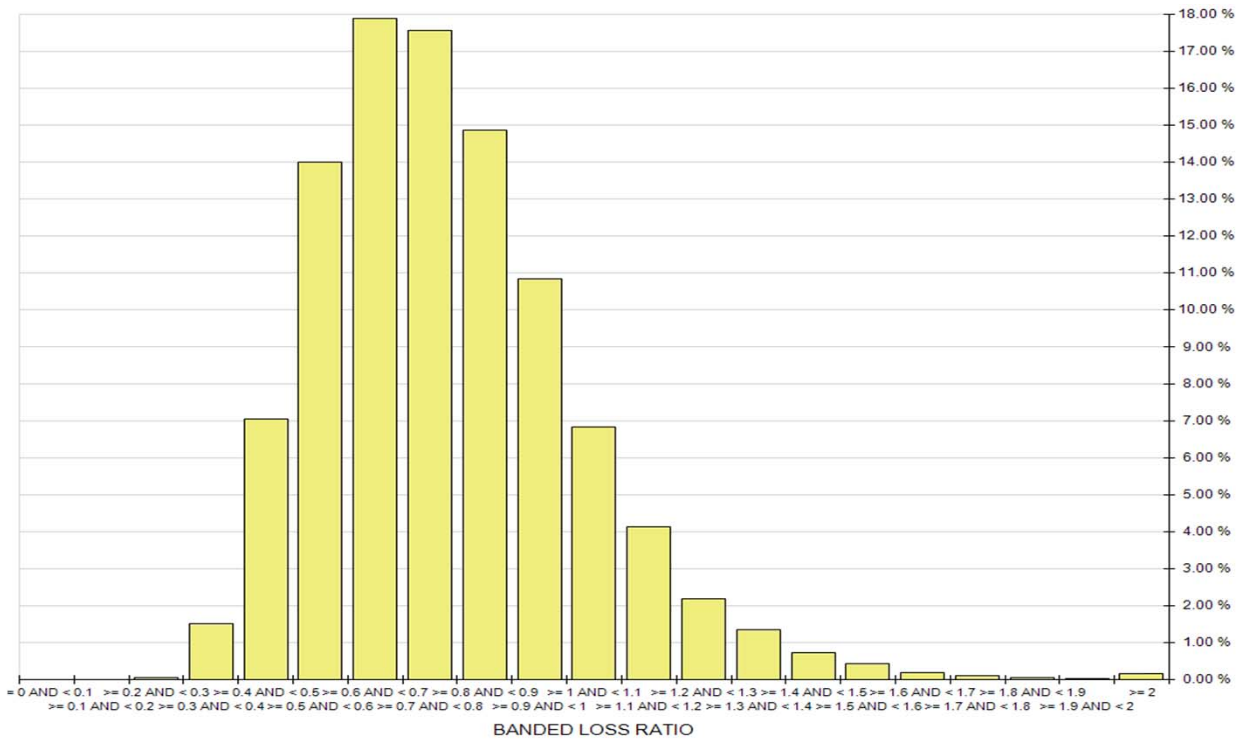
Policy Average Competitor & Proposed Premium Views



towerswatson.com

Proposed Loss Ratio Analysis

- Compare the new price to the existing cost model and identify potential profitability issues



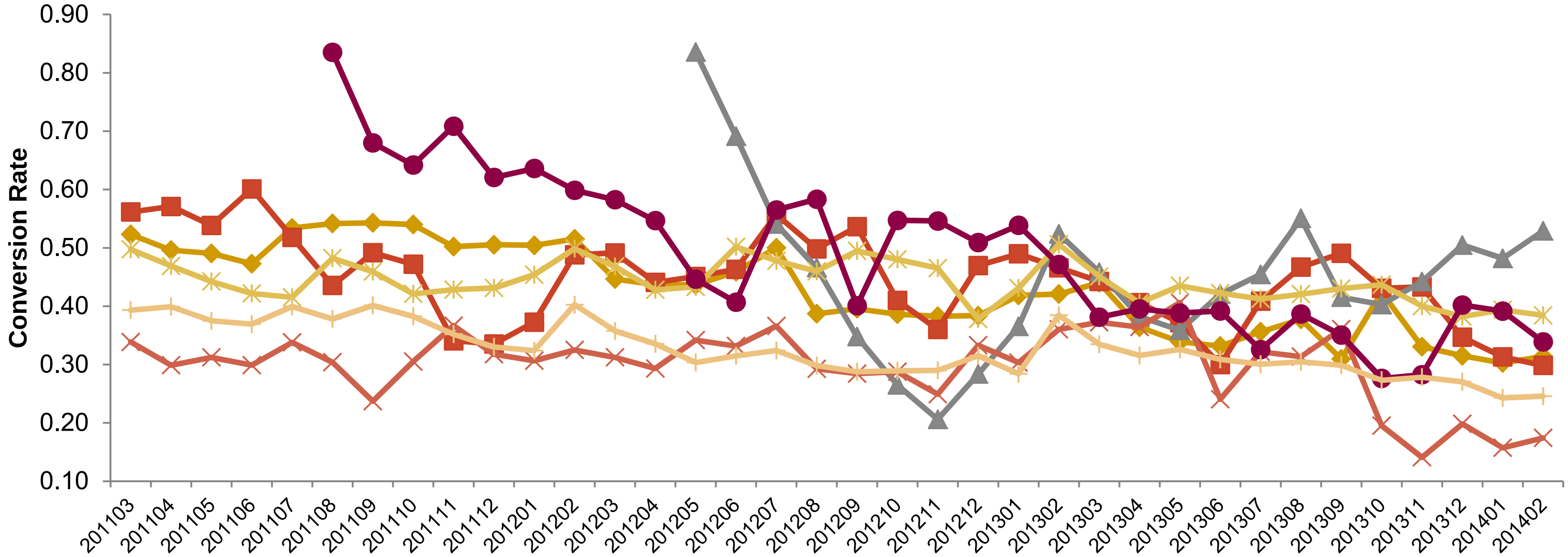
Monitoring

State Overview – Conversion

STATE	6 MONTH GROWTH RATE
State 1	-2.771%
State 2	-9.936%
State 3	5.519%
State 4	-10.665%
State 5	-2.597%
State 6	3.583%
State 7	-3.802%

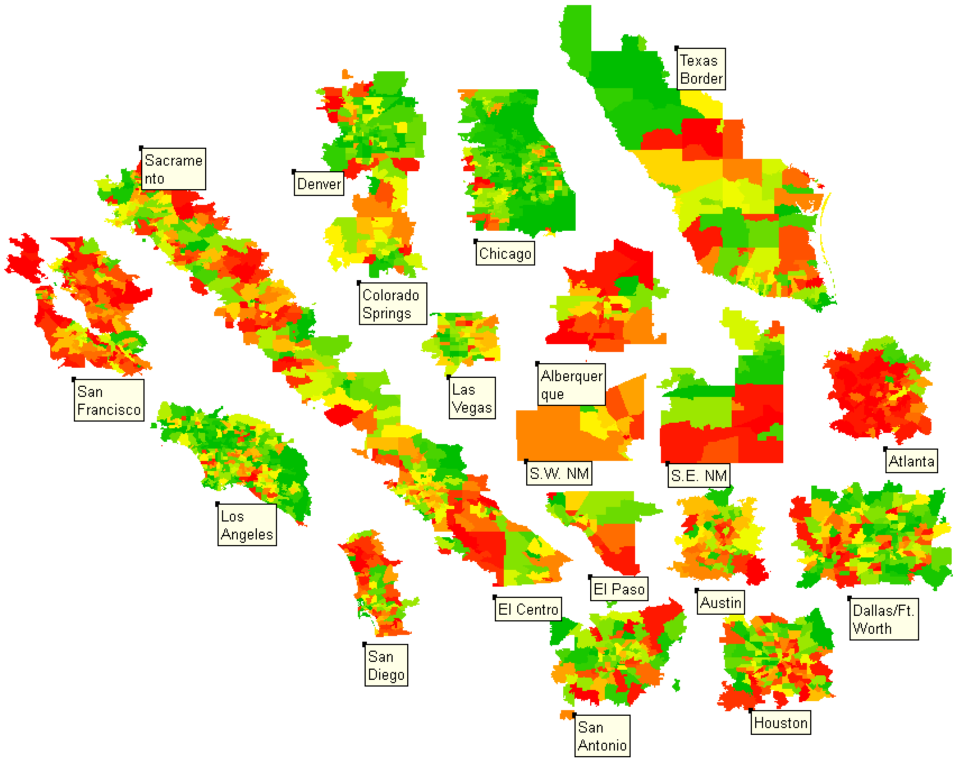
Commentary:

- State 2 and 4 appear to be overpriced.
- State 3 could be underpriced



Change in Conversion is studied within geographic areas

- Red areas show improvement in conversion
- Green areas show weakening in conversion rates



Summary

Sophistication is an Integral Part of Market Growth

Growing the book	The challenge in analytics is that my cost models represent the experience I have rather than the experience I want. It is imperative to apply alternate approaches to both predictive analytics and competitive analysis to change the market footprint
Predictive Analytics	Loss costs models are enhanced to minimize the effect of overfitting. Thus final algorithms are more responsive to the signal in the data rather than noise. By properly reflecting the signal you can be more confident in the extrapolation beyond the existing data set
Competitive Analytics	Collect competitive data is a time consuming process that is rife with the potential for error. However, once built this can provide valuable insight on how to price to alternate markets. We built upon the idea of the market basket to reflect the universe of shoppers we wish to attract. We then selected a price and built a rate algorithm based on the brand we want.

TOWERS WATSON