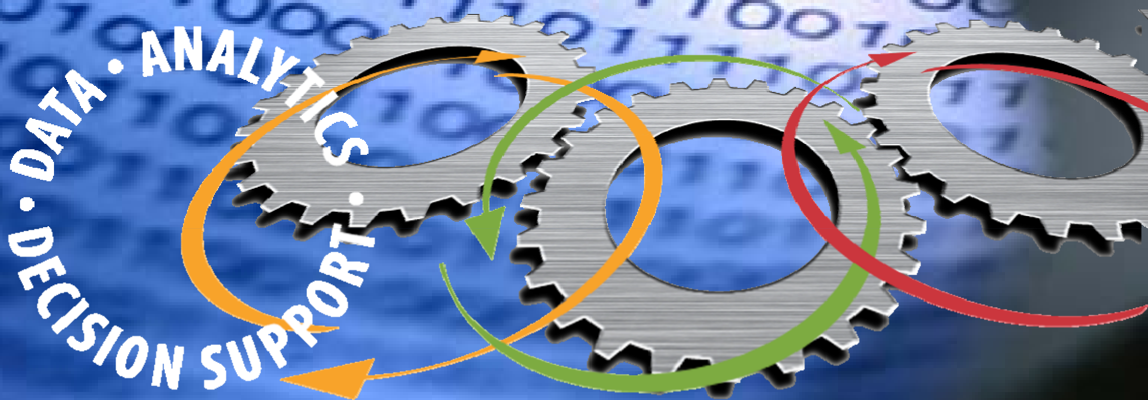
Innovations and Value Creation in Predictive Modeling

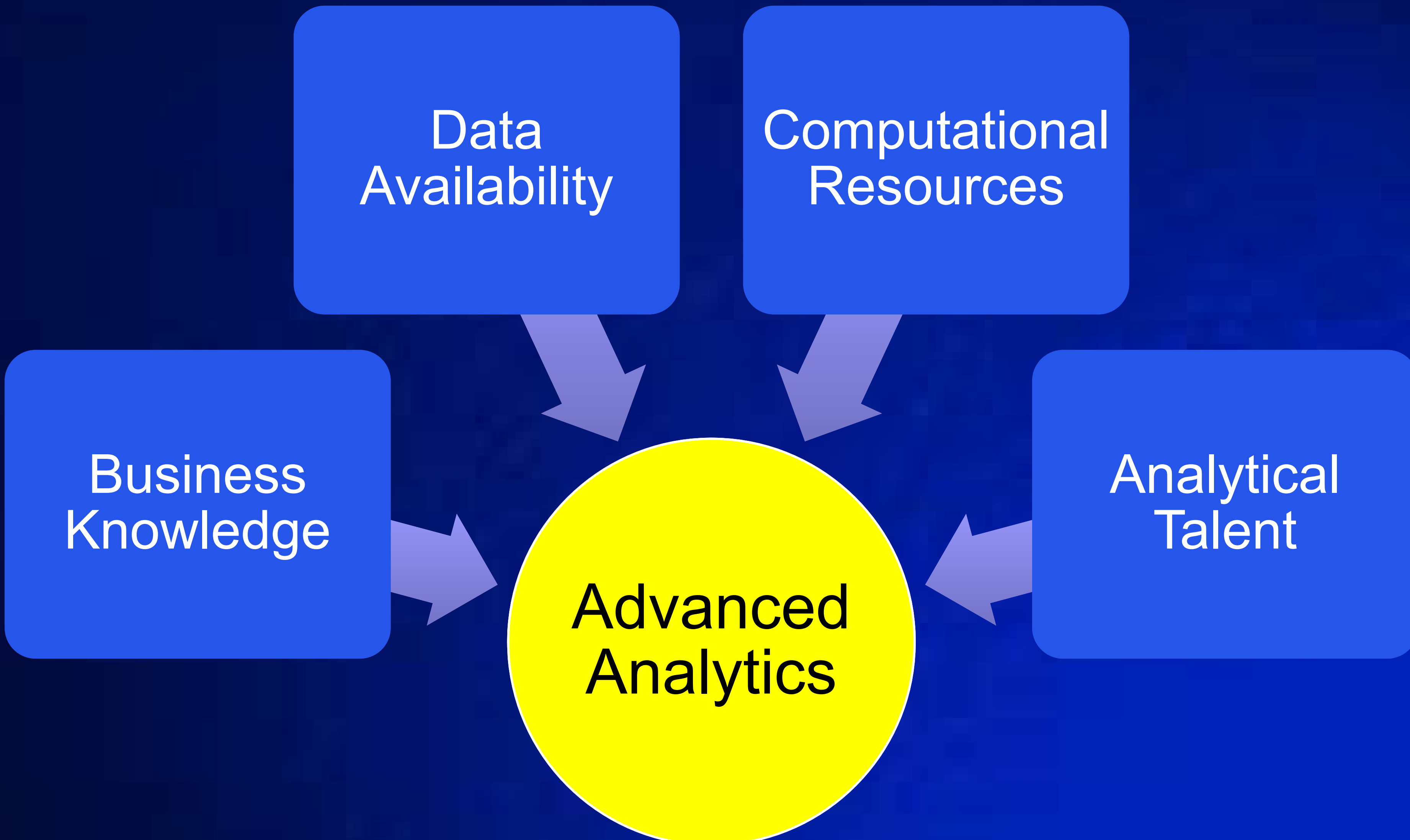
David Cummings
Vice President - Research
ISO Innovative Analytics



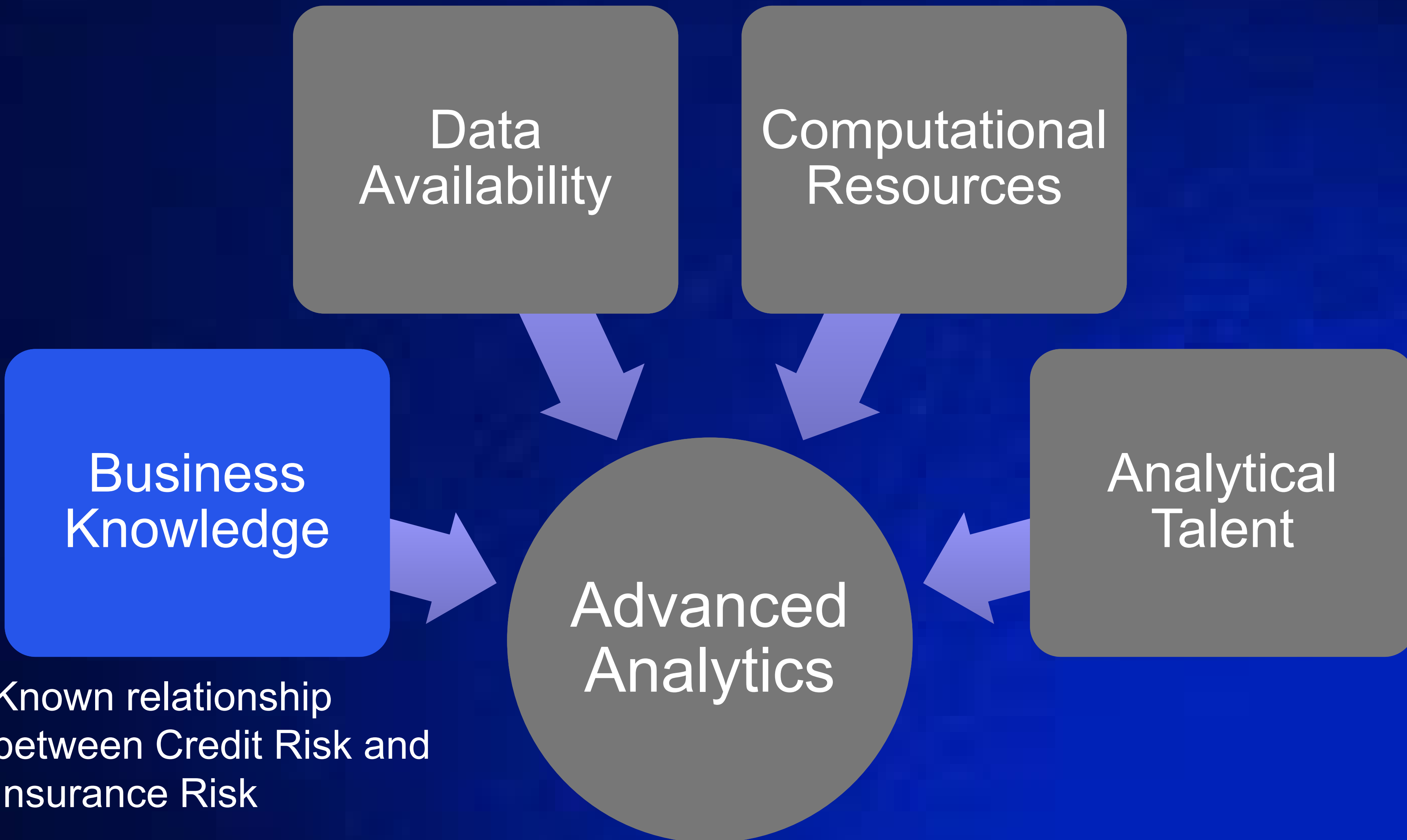
Innovations and Value Creation in Predictive Modeling

- A look back at the past decade of innovation in predictive analytics
- New innovations in predictive modeling in Auto and Homeowners Insurance
- Measuring the value of increased rate segmentation

The Recipe for Advanced Analytics

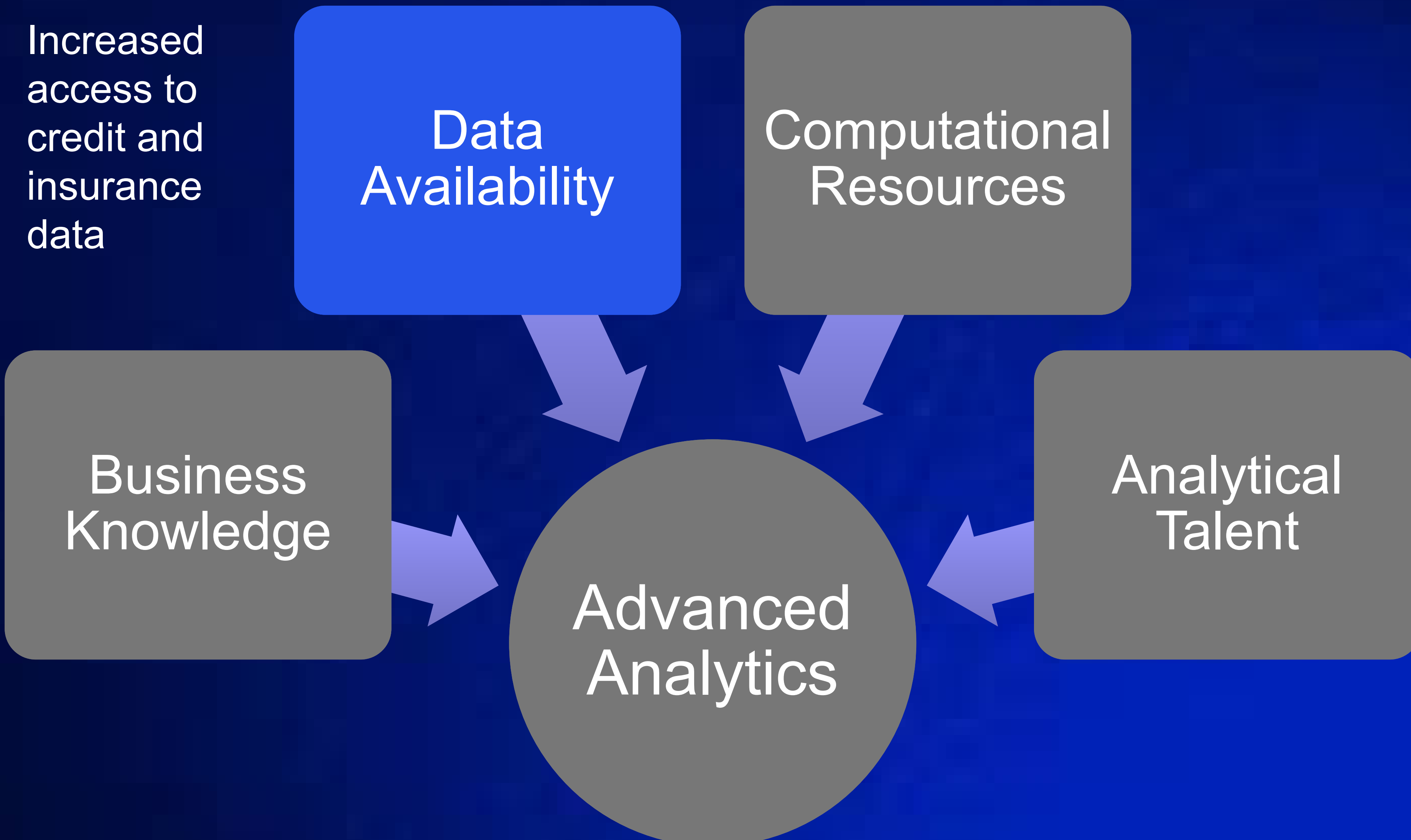


Example: Credit Scoring in Auto Insurance in the mid 90's

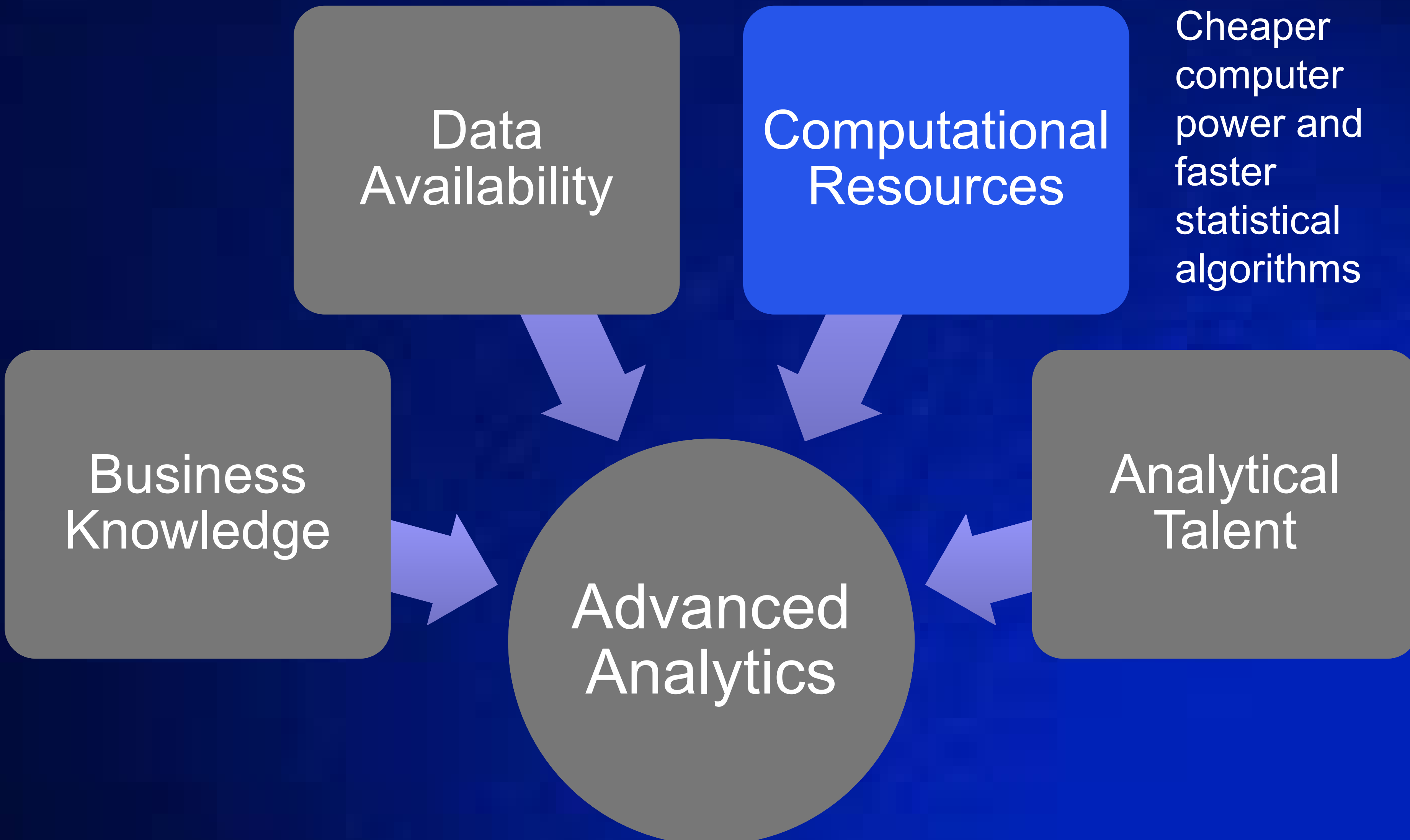


Known relationship between Credit Risk and Insurance Risk

Example: Credit Scoring in Auto Insurance in the mid 90's

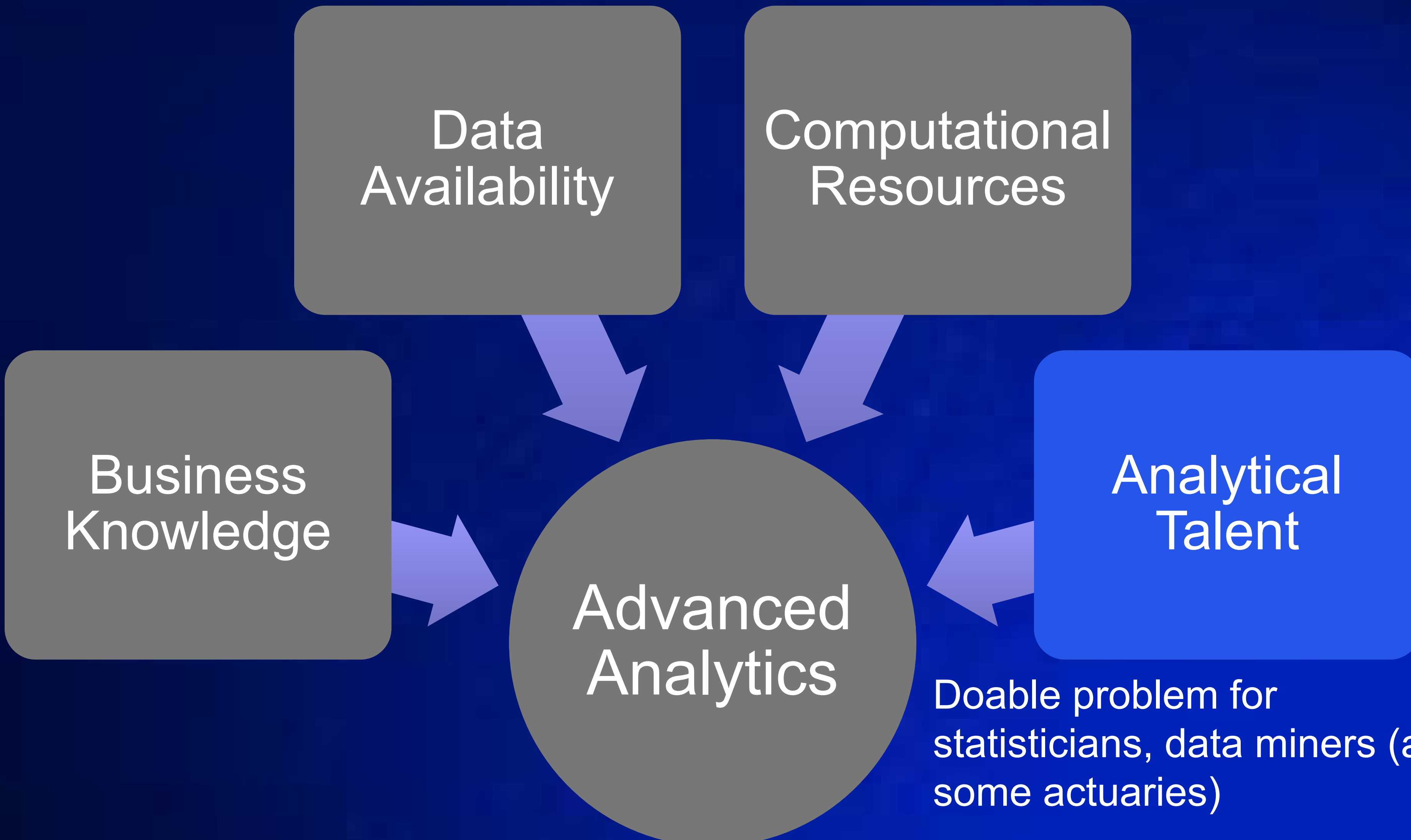


Example: Credit Scoring in Auto Insurance in the mid 90’s



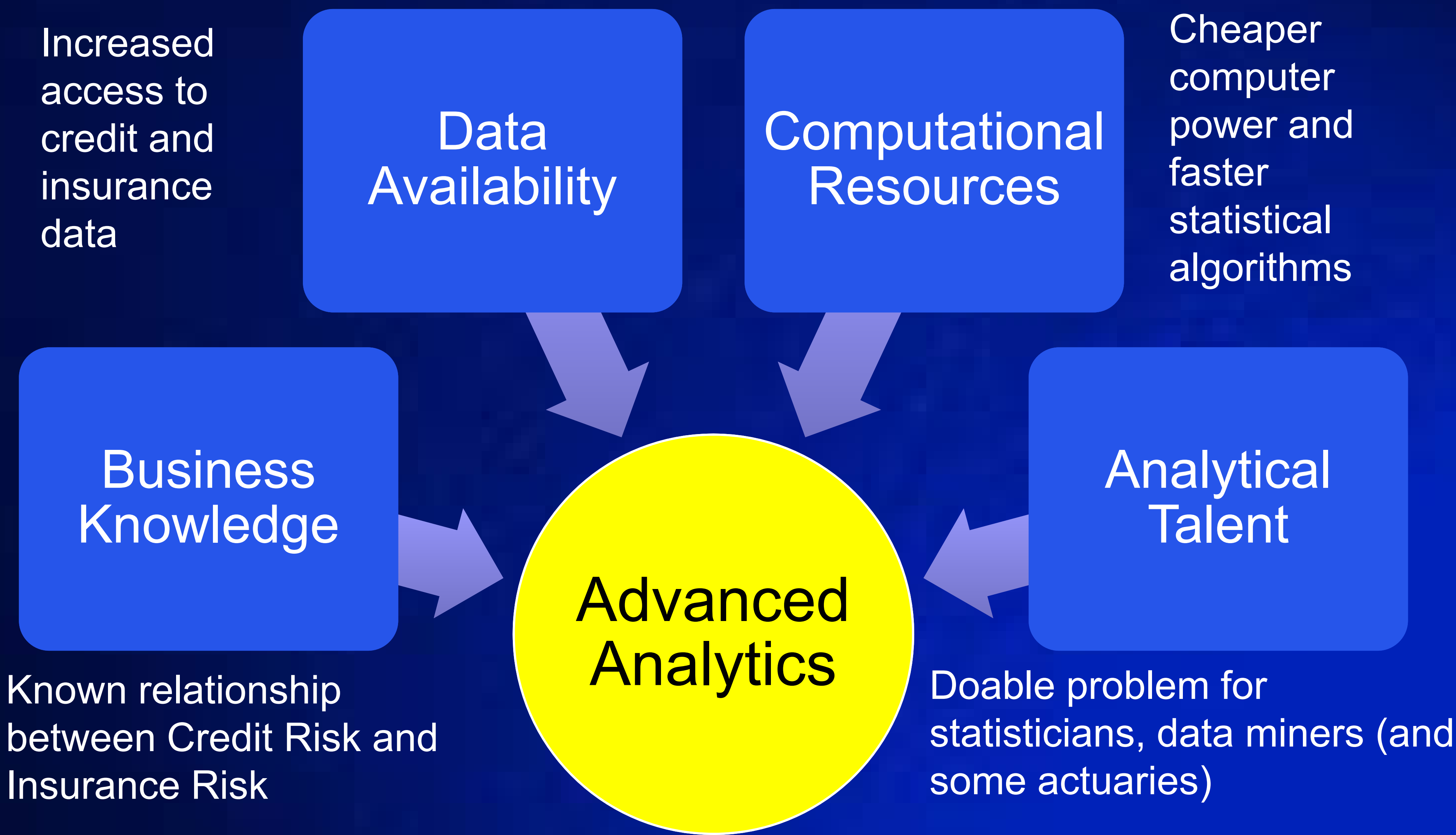
Cheaper computer power and faster statistical algorithms

Example: Credit Scoring in Auto Insurance in the mid 90's



Doable problem for statisticians, data miners (and some actuaries)

Example: Credit Scoring in Auto Insurance in the mid 90's

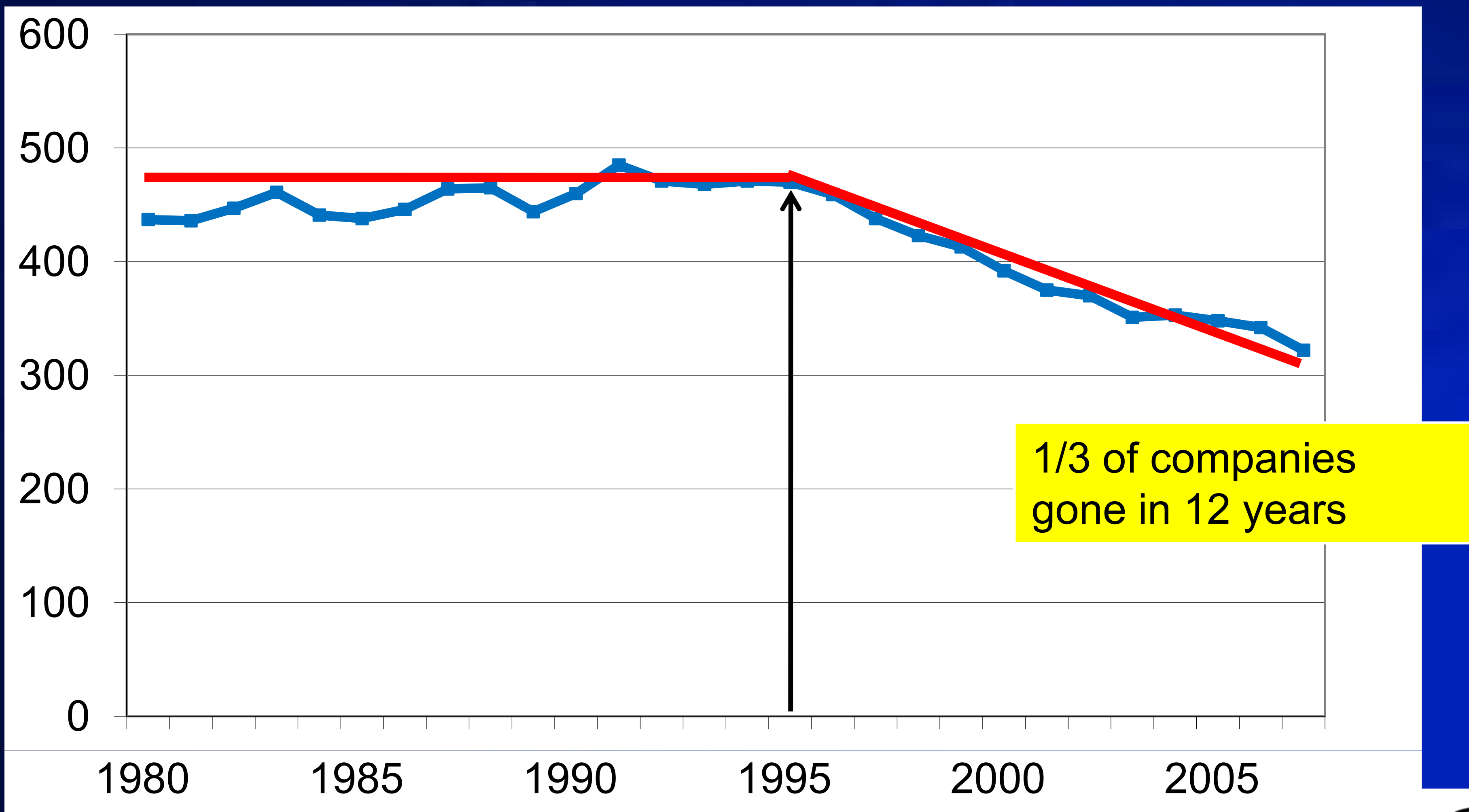


What has the impact been?

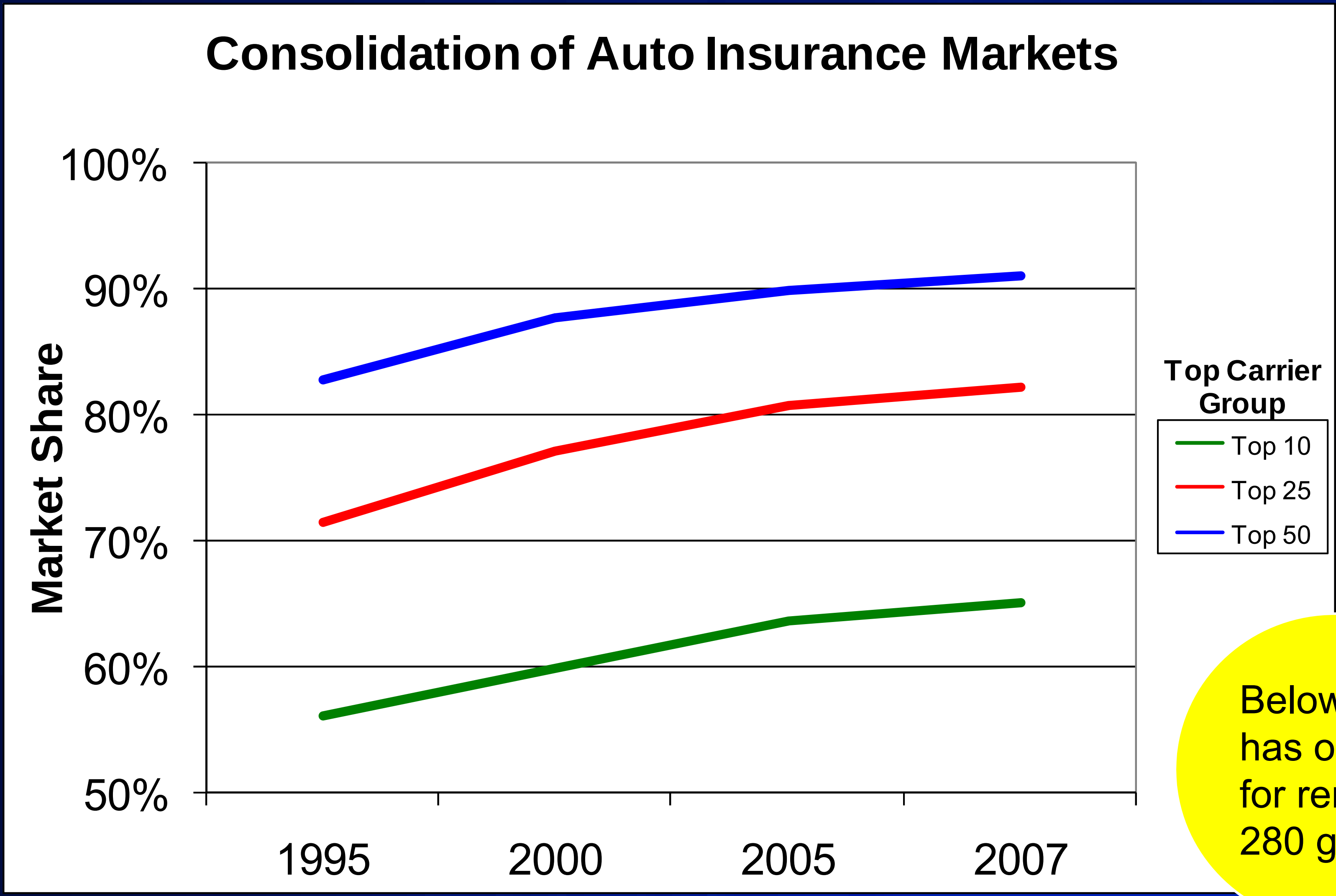
- Major innovations in an historically static rate plan
- Increased competition
- Profitable growth for adopters of advanced analytics
- Hunger for the next innovation

Indication of Increased Competition

Number of Companies writing Personal Auto Insurance in the US



Indication of Increased Competition

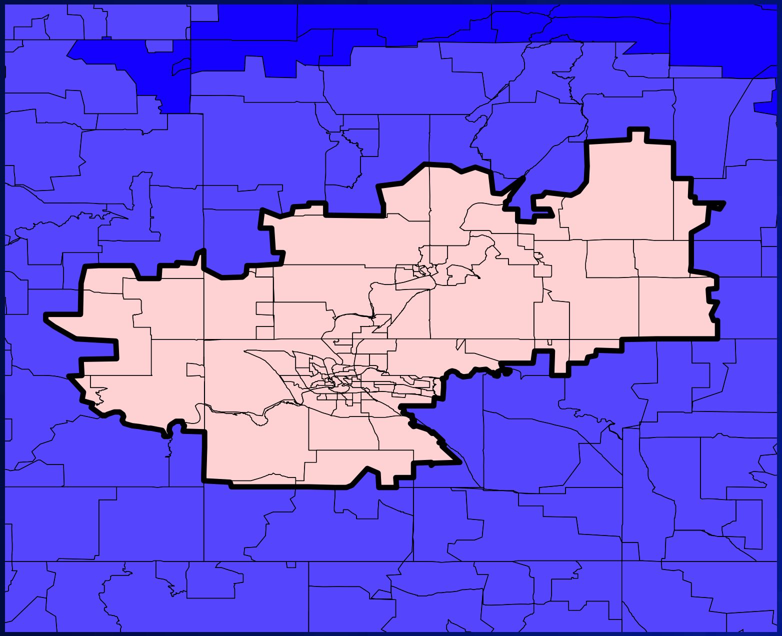


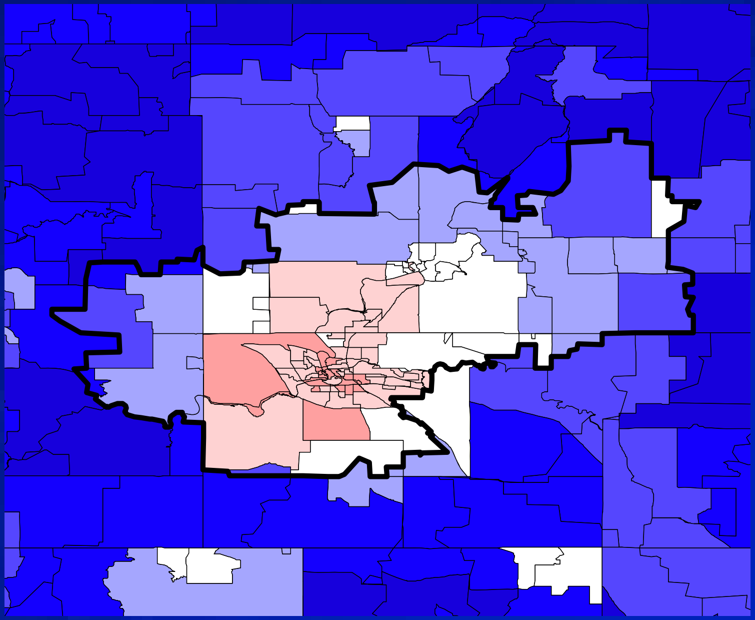
Below 50 now has only 9% for remaining 280 groups

Innovations in Predictive Modeling: Predictions at the Address Level

Territorial Conundrum

- **Territories should be big**
  - Have a sufficient volume of business to make credible estimates of the losses.





- **Territories should be small**
  - Conditions vary within territory.

View as Case Studies in Model Development

Data Driven Approach

- **Reduction in number of variables**
  - Necessary for small insurers
- **Special circumstances in fitting models to individual auto / home owners data.**
- **Diagnostics**
  - Graphics and Maps

Data Versus the Conundrum

Data

ISO Data

- Loss Cost
- Trend
- Location Products
  - Distance to Coast / major body of water
  - PPC

Development Partners

Obtained loss, exposure, policy features and address for individual policies from cooperating insurers

External Data

- Weather
- Census / Claritas
- Business Points
- Elevation

Some Environmental Features (Possibly) Related to Claims

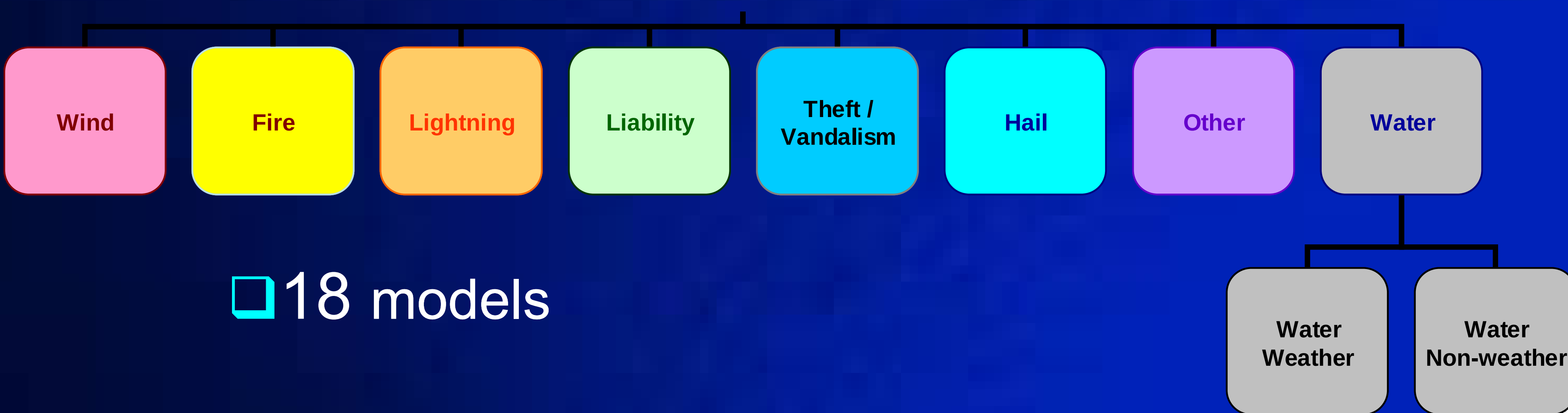
- **Proximity to Businesses and Attractions**
  - Workplaces, Shopping Centers, Contractors, etc.
- **Weather / Terrain: Wind, Temperature, Snowfall, Change in Elevation**
- **Population (Traffic) Density**
- **Others : Commuting Patterns, Coastal proximity, etc.**

Combining Environmental Variables at a Particular Address

- **Individually, the geographic variables have a predictable effect on claim rate and severity.**
- **Variables for a particular location could have a combination of positive and negative effects.**
- **ISO has built models to calculate the combined effect of all variables.**
  - Based on countrywide data – Actuarially credible

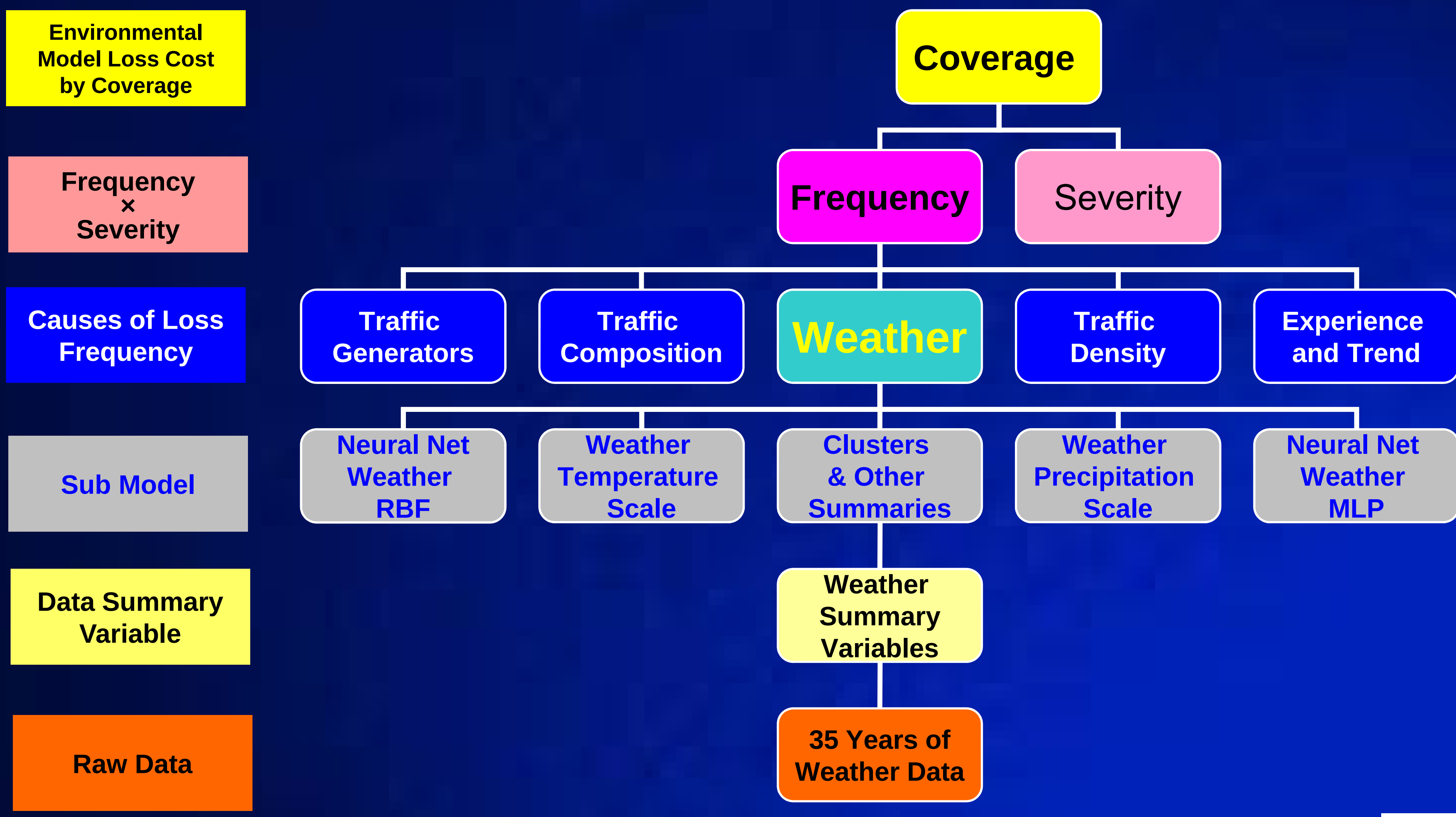
Variable Selection is Multiplied by the Number of Models

- **Frequency and Severity are modeled separately**
- **Models are at coverage / peril level**
  - Five auto coverages: BI, PD, PIP, Comp. & Coll.
    - 10 models
  - Nine home owners perils:



- 18 models

In Depth for Auto Weather Component



Environmental Model

Loss Cost = Pure Premium

= Frequency x Severity

$$\text{Frequency} = \frac{e^{\lambda}}{1+e^{\lambda}} \qquad \text{Severity} = e^{\mu}$$

λ = Intercept
+ Weather
+ Traffic Density
+ Traffic Generators
+ Traffic Composition
+ Experience and Trend

μ = Intercept
+ Weather
+ Traffic Density
+ Traffic Generators
+ Traffic Composition
+ Experience and Trend

Constructing the Components Frequency Model as Example

$\lambda =$ Intercept

$+\ \alpha_1 \cdot X_1 + \ldots + \alpha_{n_1} \cdot X_{n_1}$ = Weather

$+\ \alpha_{n_1+1} \cdot X_{n_1+1} + \ldots + \alpha_{n_2} \cdot X_{n_2}$ = Traffic Density

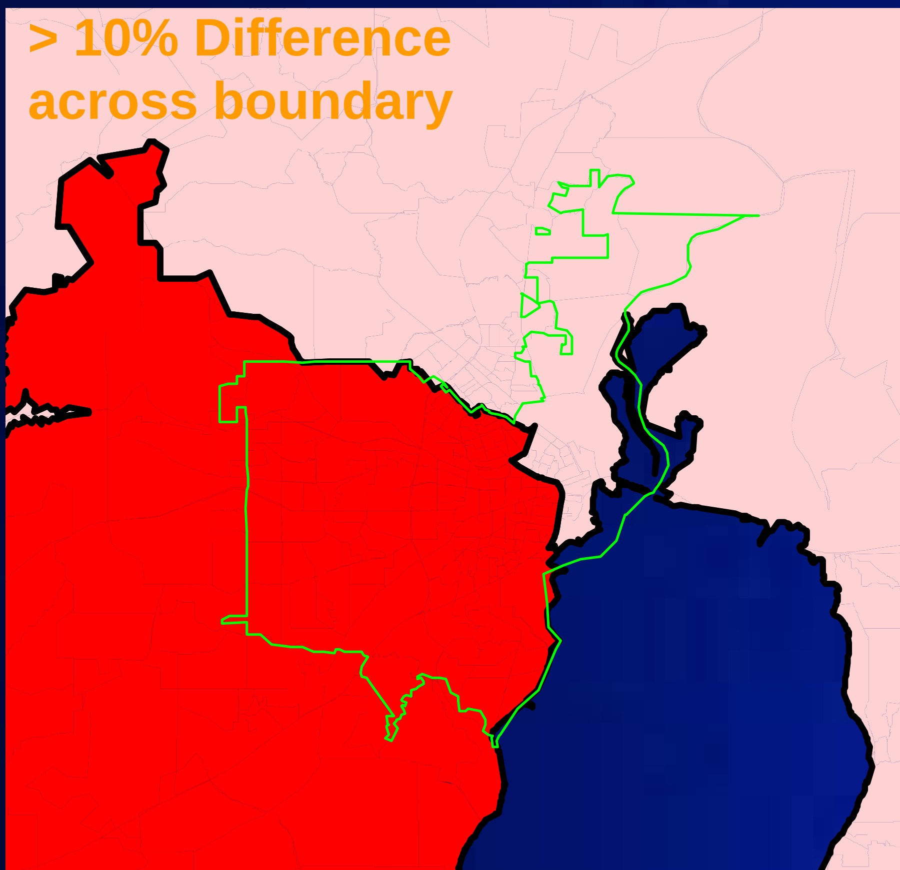
$+\ \alpha_{n_2+1} \cdot X_{n_2+1} + \ldots + \alpha_{n_3} \cdot X_{n_3}$ = Traffic Generators

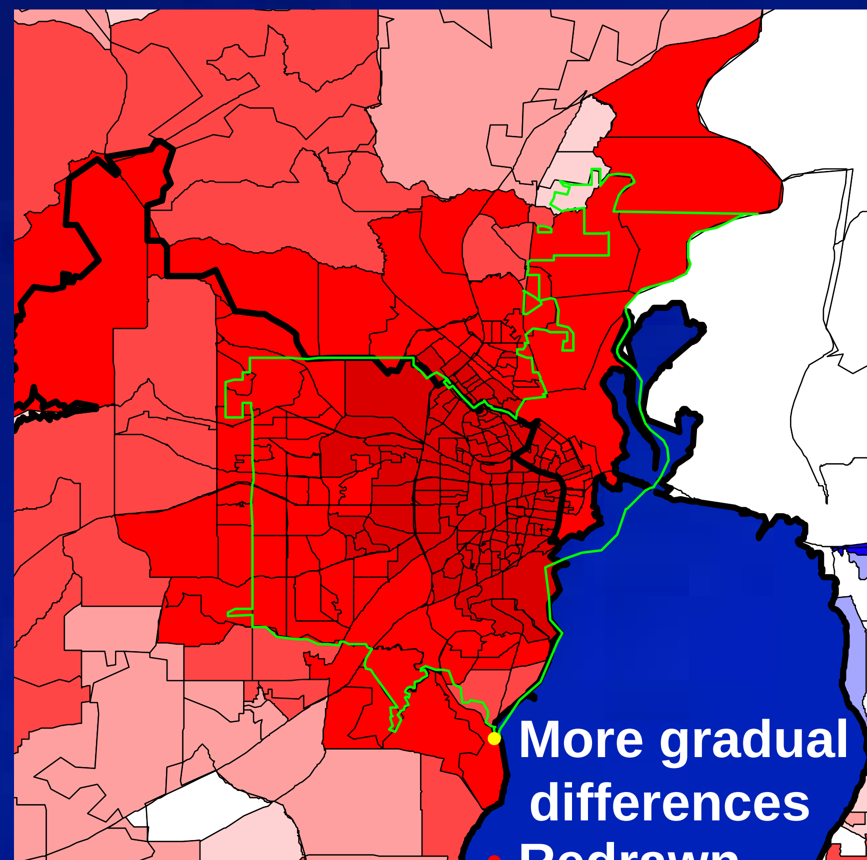
$+\ \alpha_{n_3+1} \cdot X_{n_3+1} + \ldots + \alpha_{n_4} \cdot X_{n_4}$ = Traffic Composition

$+\ \alpha_{n_4+1} \cdot X_{n_4+1} + \ldots + \alpha_{n_5} \cdot X_{n_5}$ = Experience & Trend

$+$ Other Classifiers

An Example on the Ground

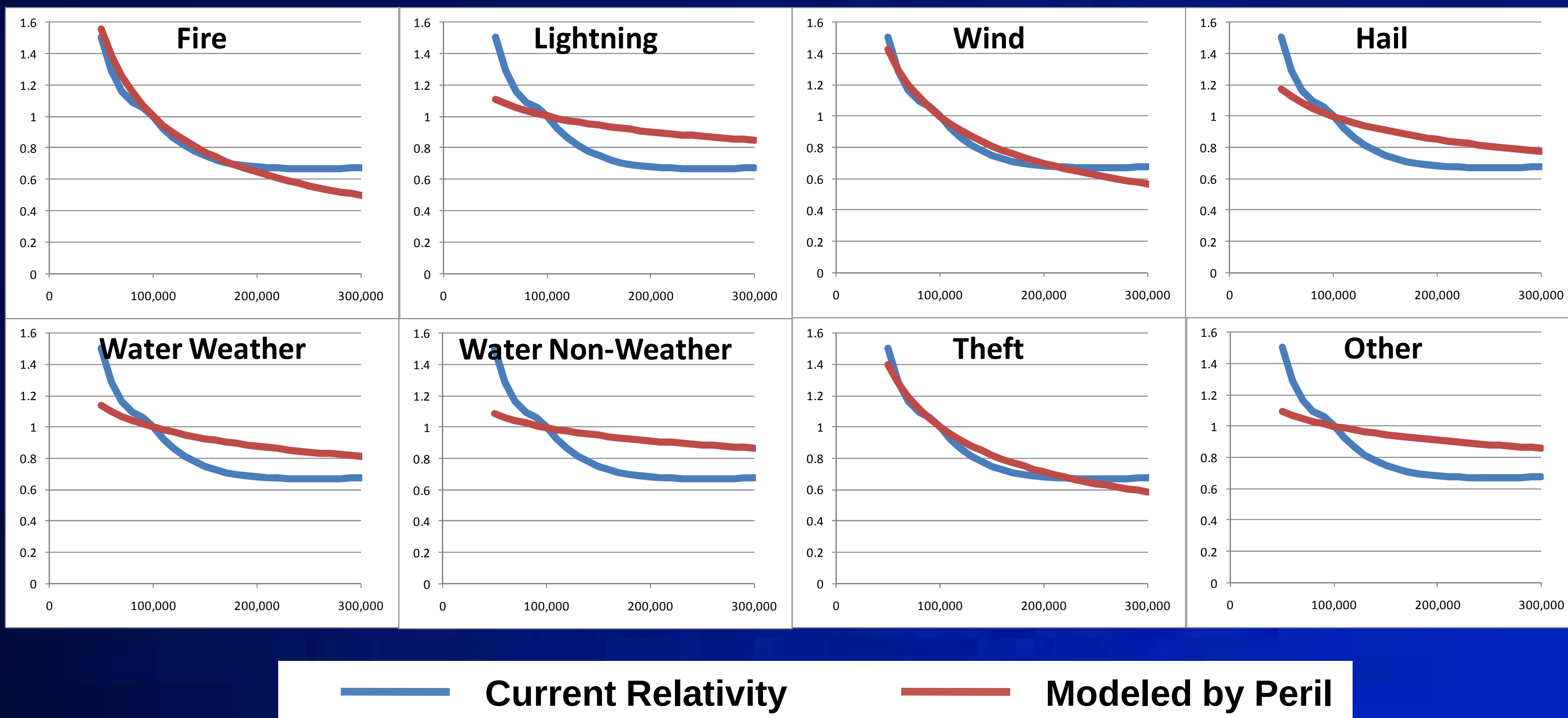




- More gradual differences
- Redrawn boundaries

Homeowners
Amount Relativities by Peril

Loss Cost per $1000 of Building Coverage



- **Significant variation by peril**

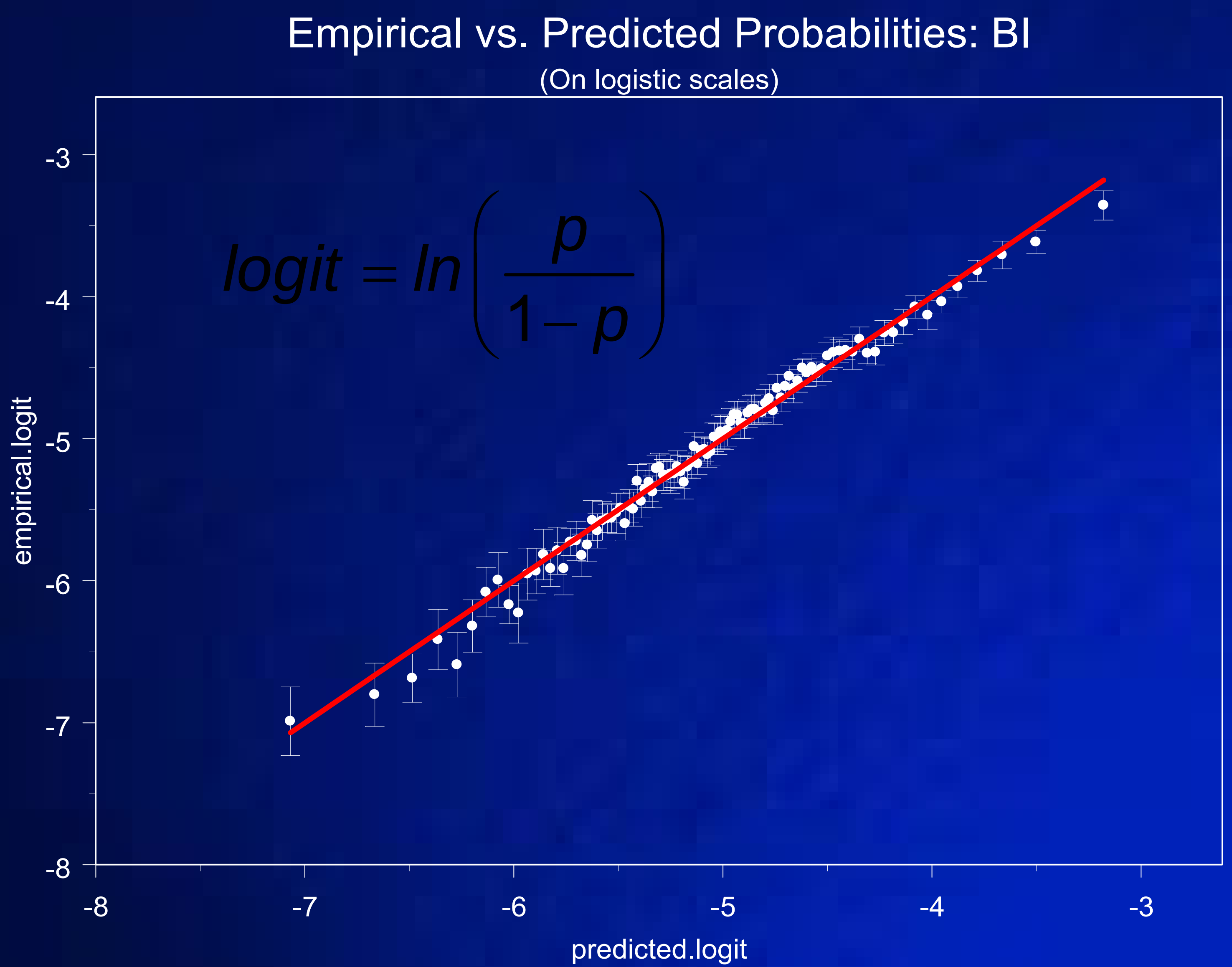
Homeowners Rating Factors by Peril

- Rating Factors that vary by peril provide lift
- Adds accuracy and complexity
  - All-peril relativities can be derived from peril-based relativities according to peril mix within the area
  - Local Prediction by peril may result in varying peril loss costs at the address level
- Effectively produces all-peril relativities that vary at the address level

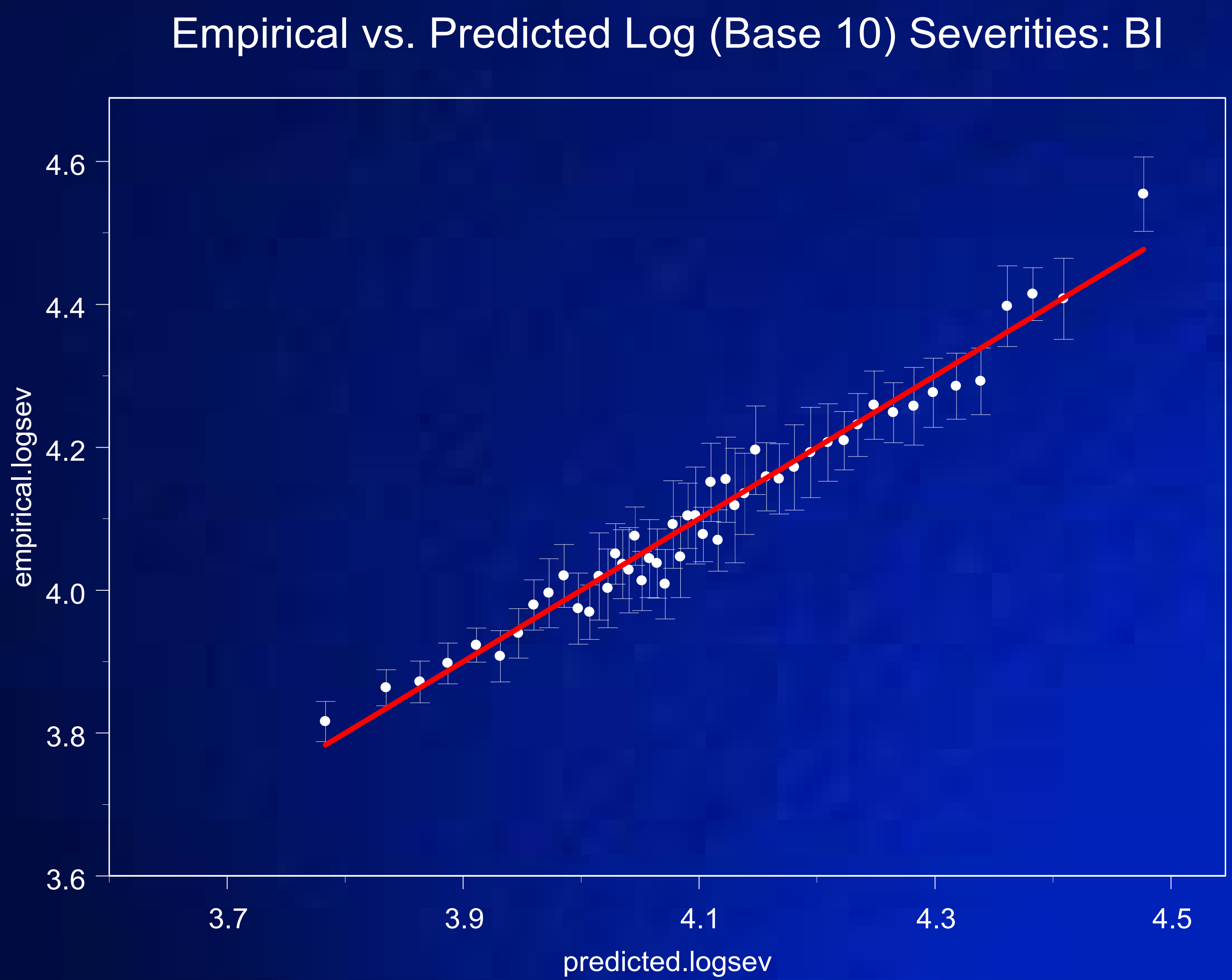
Overall Model Diagnostics

- **Sort in order of increasing prediction**
  - Frequency & Severity
- **Group observations in buckets**
  - 1/100th of record count for frequency
  - 1/50th of the record count for severity
- **Calculate bucket averages**
- **Apply the GLM link function for bucket averages and predicted value**
  - logit for frequency
  - log for severity
- **Plot predicted vs empirical**
  - With confidence bands

Overall Diagnostics - Frequency



Overall Diagnostics - Severity



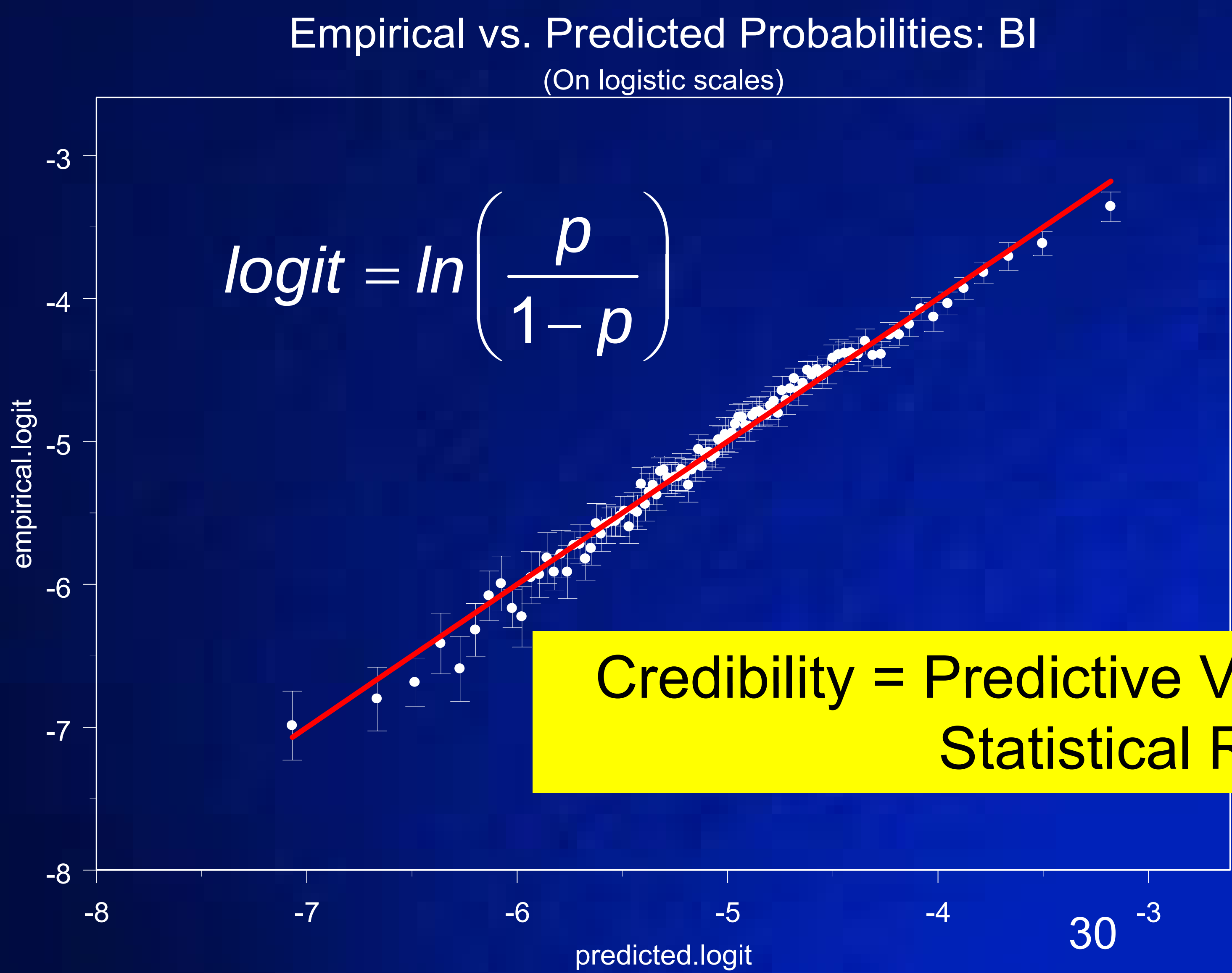
Credibility
Statement of Principles regarding P&C Insurance Ratemaking (adopted 1988)

Credibility is a measure of the predictive value that the actuary attaches to a particular body of data. Credibility is increased by making groupings more homogeneous or by increasing the size of the group analyzed. A group should be large enough to be statistically reliable. Obtaining homogeneous groupings requires refinement and partitioning of the data. There is a point at which partitioning divides data into groups too small to provide credible patterns. Each situation requires balancing homogeneity and the volume of data.

Credibility
Statement of Principles regarding P&C Insurance Ratemaking (adopted 1988)

Credibility is a measure of the predictive value that the actuary attaches to a particular body of data. Credibility is increased by making groupings more homogeneous or by increasing the size of the group analyzed. A group should be large enough to be statistically reliable. Obtaining homogeneous groupings requires refinement and partitioning of the data. There is a point at which partitioning divides data into groups too small to provide credible patterns. Each situation requires balancing homogeneity and the volume of data.

Overall Diagnostics - Frequency



Credibility = Predictive Value and Statistical Reliability

30

Credibility
Statement of Principles regarding P&C Insurance Ratemaking (adopted 1988)

Credibility is a measure of the predictive value that the actuary attaches to a particular body of data. Credibility is increased by making groupings more homogeneous or by increasing the size of the group analyzed. A group should be large enough to be statistically reliable. Obtaining homogeneous groupings requires refinement and partitioning of the data. There is a point at which partitioning divides data into groups too small to provide credible patterns. Each situation requires balancing homogeneity and the volume of data.

31

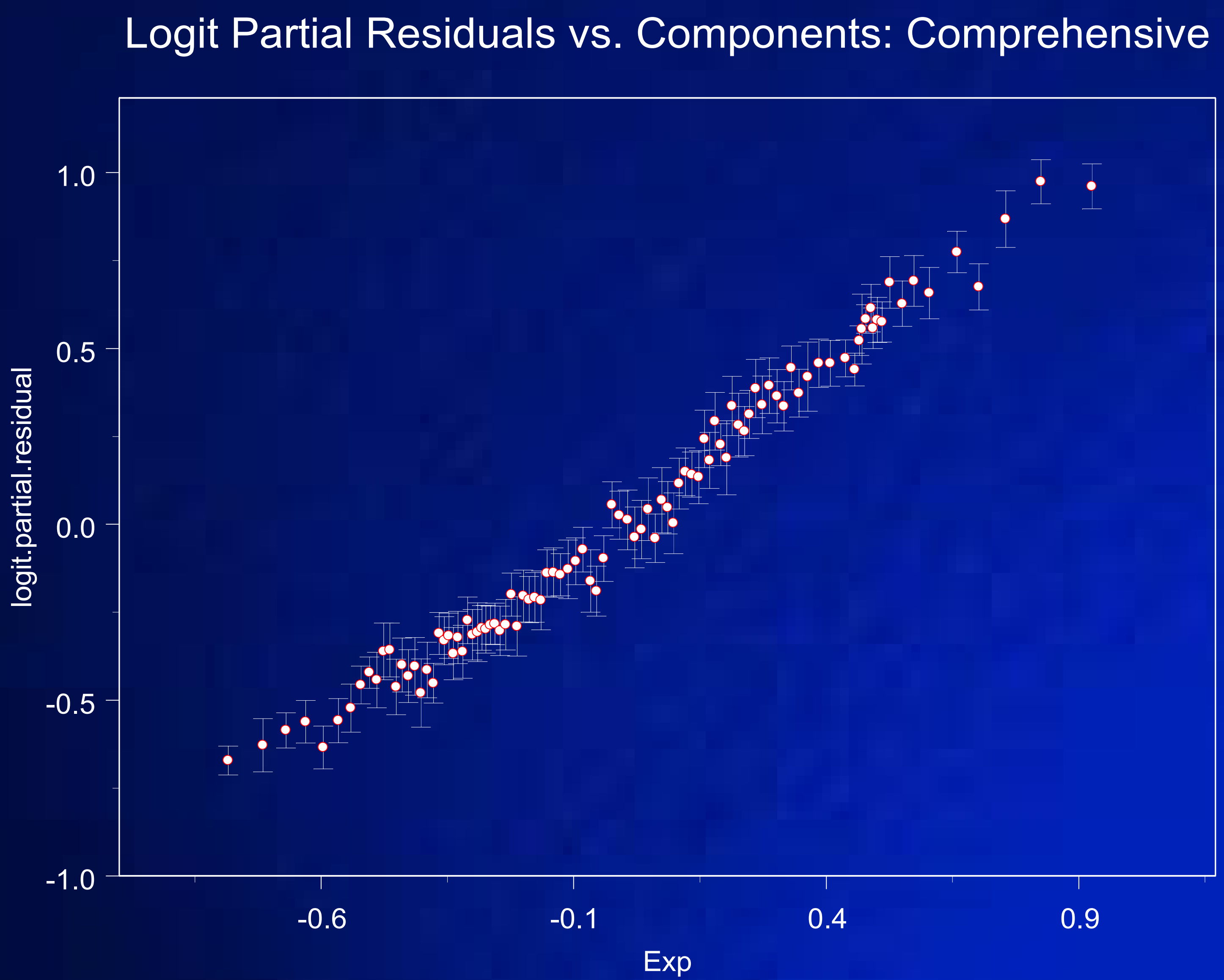
Component Diagnostics Frequency Example

- Sort observations in order of i^{th} component C_i
- Bucket as above and calculate
  - C_{ib} = Average C_i in bucket b
  - p_{ib} = Average p_i in bucket b
  - Partial Residuals

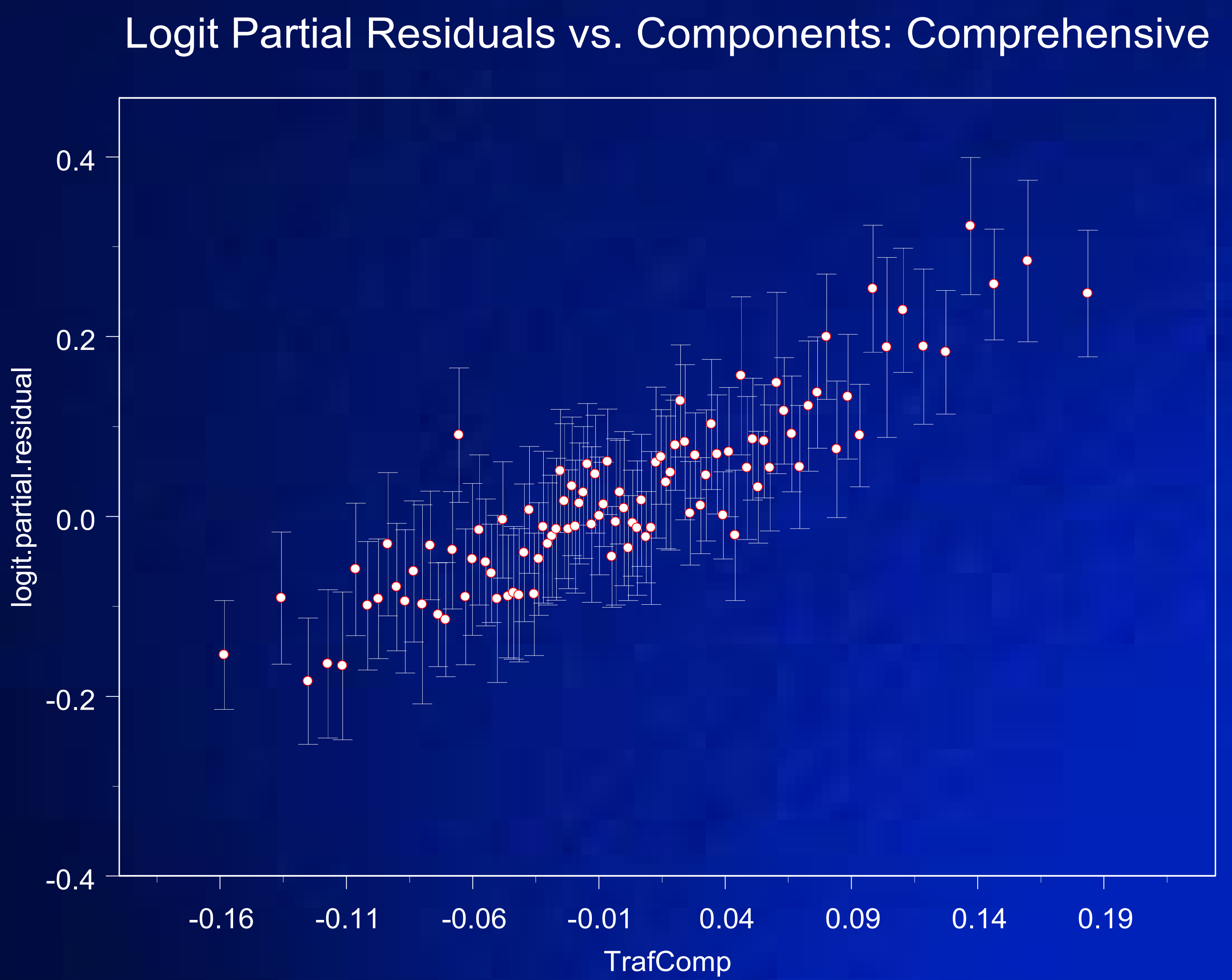
$$R_{ib} = ln\left(\frac{p_{ib}}{1-p_{ib}}\right) - \left(\lambda + \sum_{k \neq i} C_{kb}\right)$$

- Plot C_{ib} vs R_{ib} – Expect linear relationship

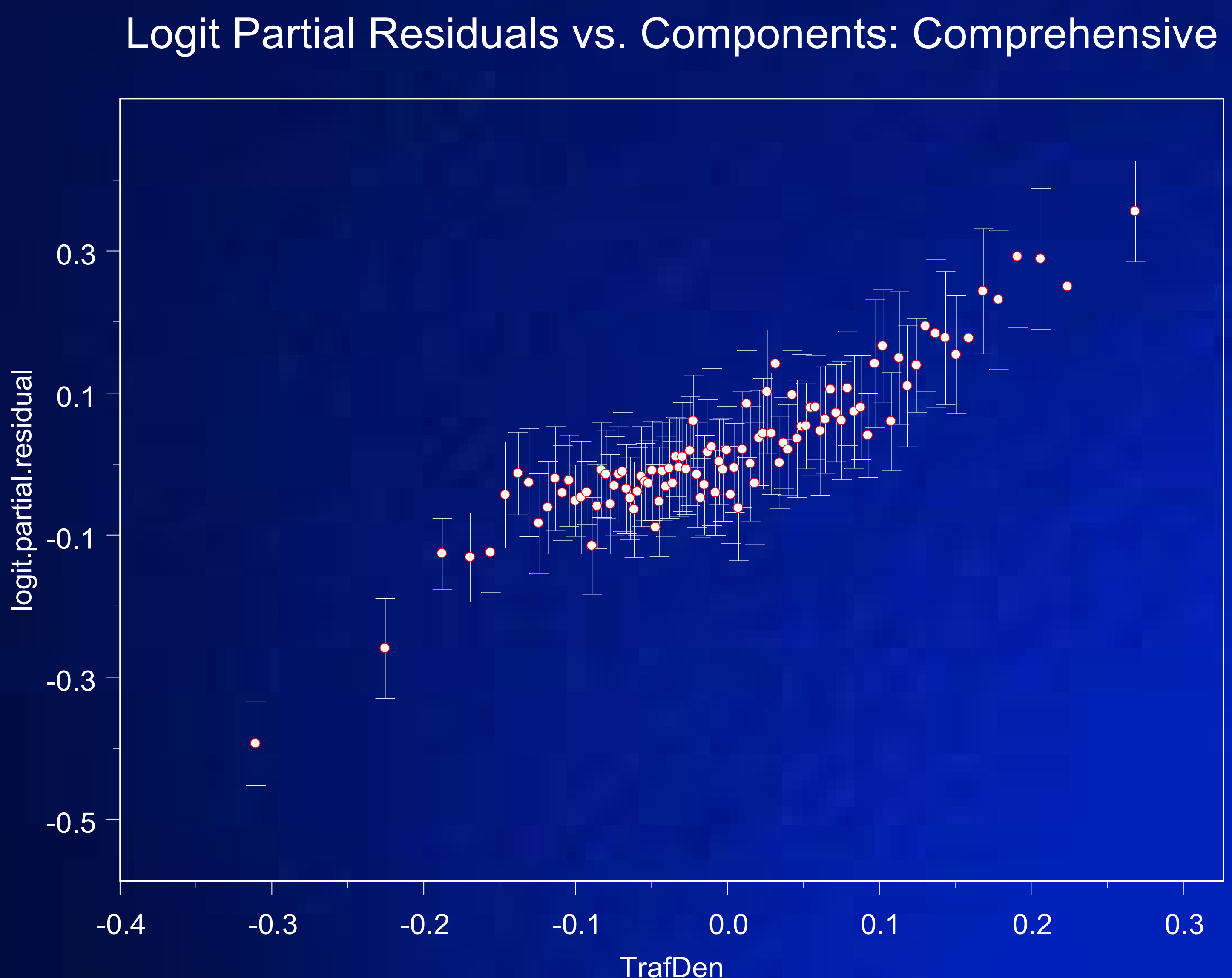
Component Diagnostics Experience and Trend



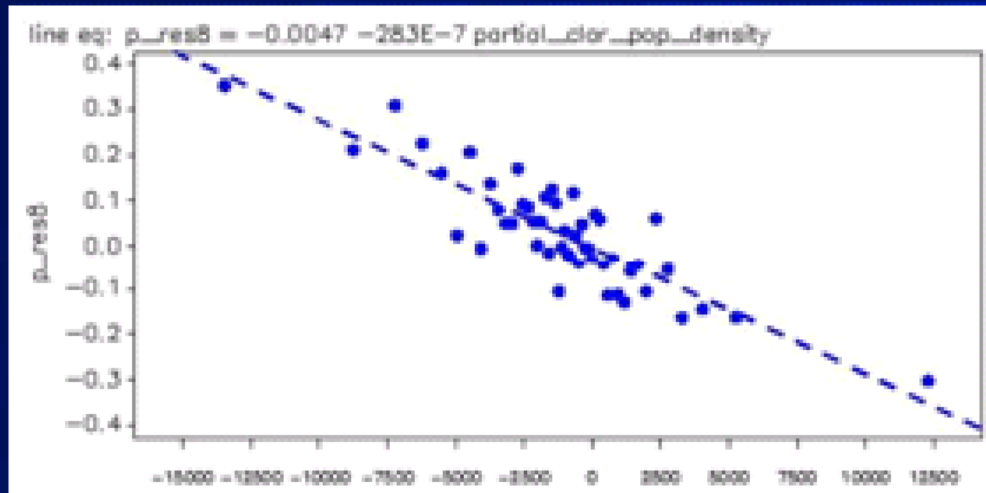
Component Diagnostics Traffic Composition

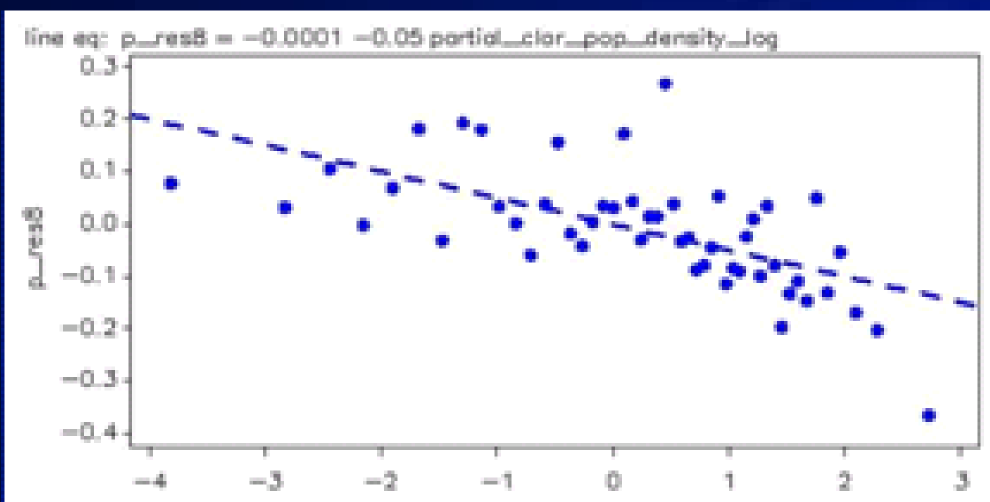


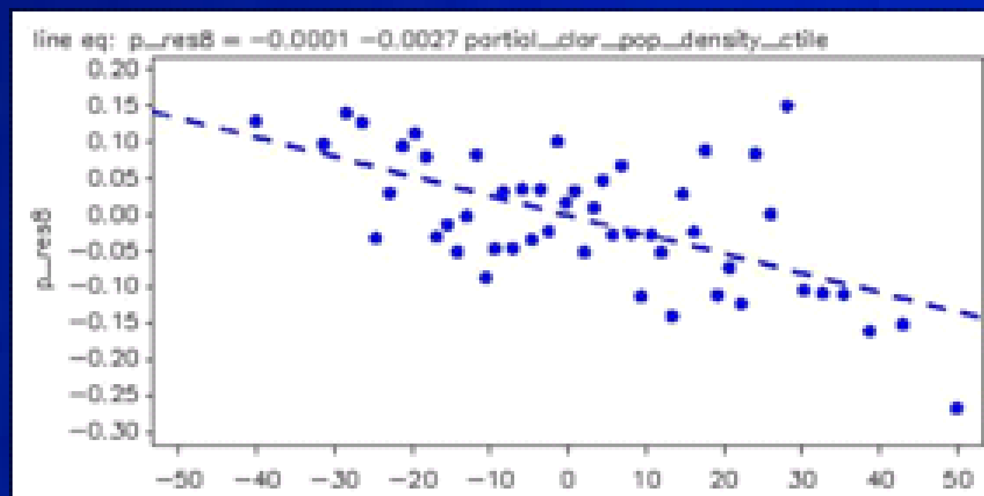
Component Diagnostics
Traffic Density



Partial Residual Plot : Finding Transformations







Example of Diagnostics Collinearity and Multicollinearity

- **Correlations Matrix: measures the correlations among each pair of variables in the models, but does not consider multicollinearity.**

- **Variance Inflation Factors (VIF): A measure of the multicollinearity among independent variables.**

Customized Model

Loss Cost = Pure Premium

= Frequency x Severity

$$\text{Frequency} = \frac{e^{\lambda}}{1+e^{\lambda}}$$

$$\begin{aligned}\lambda = \; & \alpha_0 \\ & + \alpha_1 \cdot \text{Weather} \\ & + \alpha_2 \cdot \text{Traffic Density} \\ & + \alpha_3 \cdot \text{Traffic Generators} \\ & + \alpha_4 \cdot \text{Traffic Composition} \\ & + \alpha_5 \cdot \text{Experience and Trend} \\ & + \text{Other Classifiers}\end{aligned}$$

$\alpha_1 \ldots \alpha_5 \equiv 1$
in industry model

Severity model
customized similarly

Predictions at the Address Level Summary

- Model estimates loss cost as a function of business, demographic and weather conditions associated with address.
- Preparing data for models based on geography is not a trivial exercise
- Showed fit assessment and model diagnostics
- Indicated how to customize the model

Measuring the Value of Rate Segmentation

Our Challenge

- Enhanced rate segmentation can add significant value

BUT

- Increased segmentation has a cost

- How do we evaluate the value vs. cost?
- How do we make the case to decision makers?

How Some Actuaries Make the Case to Increase Segmentation

We need to enhance our analytics in order to maintain our competitive pricing advantage!

I don't want to lose our pricing advantage. How much will it cost to implement an enhanced pricing strategy?

How Some Actuaries Make the Case to Increase Segmentation

It will take 100,000 IT man-hours costing $10 million to modify our underwriting and agency systems.

That’s a lot of money to spend! How much additional revenue will we bring in?

How Some Actuaries Make the Case to Increase Segmentation

We will implement the new rate structure so that it will be revenue neutral.

You want me to spend $10 million to get NO additional revenue? That doesn't make any sense!

How Some Actuaries Make the Case to Increase Segmentation

Why doesn't he understand how important this pricing strategy is to our business?

Where can I find an actuary with some business sense?

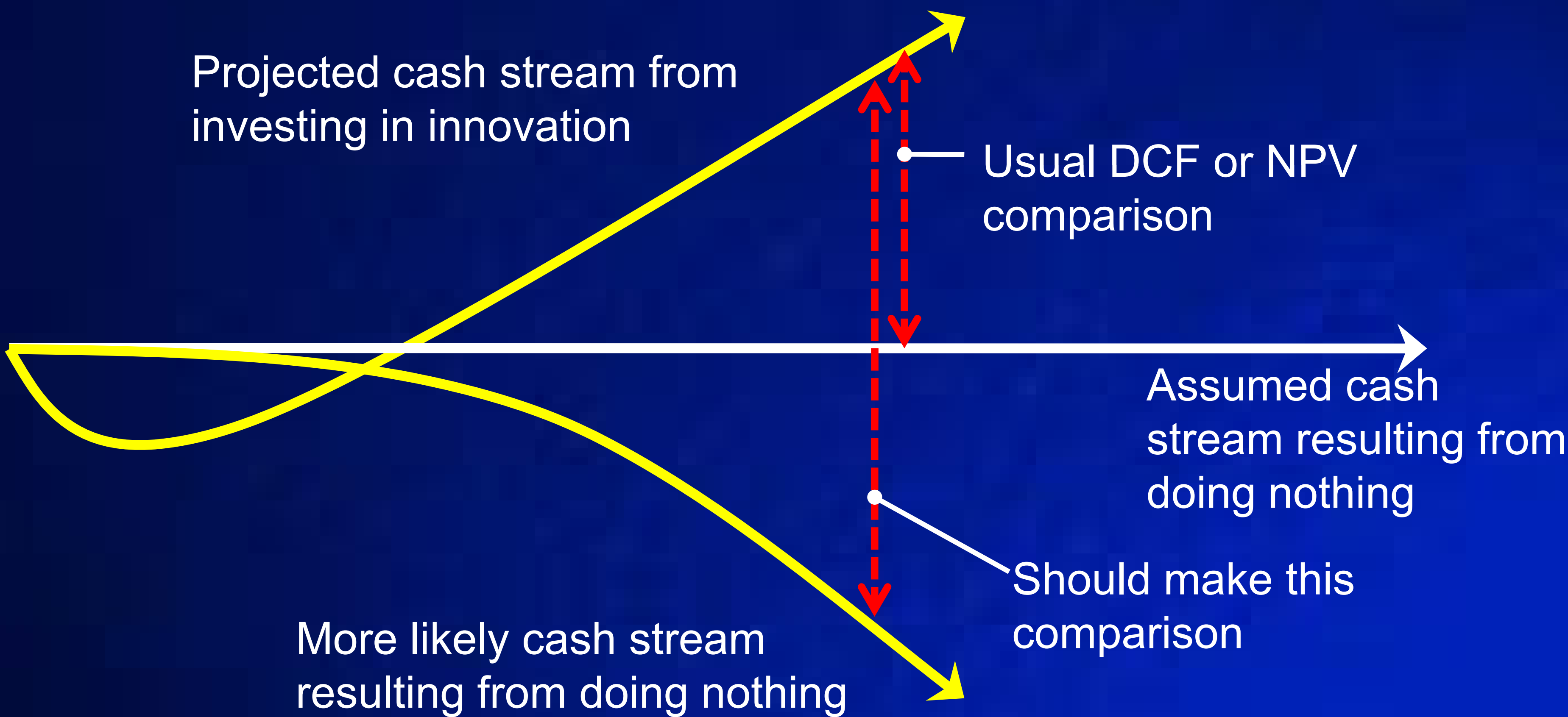
What's wrong with this dialog?

- **Focus only on implementation costs**
  - In a competitive marketplace, there is a cost to doing nothing
  - Lost business, lost revenue, and increasing cost of remaining policies
- **Short-term view of revenue impact**
  - "Revenue Neutral" applies only to average premiums on current book
  - There can be long-term revenue impacts

How to make the case better

- **Better projections of revenue and profit impacts**
  - Look beyond "Revenue Neutral" implementation
- **Better consideration of marketplace dynamics**
  - Includes customer retention and competitive effects
- **Demonstrate the value in monetary terms**

The Discounted Cash Flow Trap



Source: Christensen, Kaufmann, Shih, "Innovation Killers: How Financial Tools Destroy Your Capacity to Do New Things", Harvard Business Review, Jan 2008

Illustration

- **Insurer writes 3 policies**
- **All policies priced in the same class**
  - Expected Loss Ratio = 50%
  - Profit if Loss Ratio < 60%
- **More accurate segmentation is available in the marketplace**
  - Used by competitors
  - Places some policies at risk

Illustration – Base Case

Policy #	Premium	Insurer's Expected Loss	Break-Even Loss
1	60	30	36
2	60	30	36
3	60	30	36
Total	180	90	108
Ratio to Premium		50%	60%

Accurate Expected Loss	Insurer's Profit
20	16
30	6
40	-4
90	18
50%	10%

Illustration – Year 1

Policy #	Premium	Insurer's Expected Loss	Break-Even Loss
1	60	30	36
2	60	30	36
3	60	30	36
Total	180	90	108
Ratio to Premium		50%	60%

Accurate Expected Loss	Insurer's Profit	
20	~~16~~	0
30		6
40		-4
90	~~18~~	2
50%	~~10%~~	1%

Lost Profit = 16

Value of Lift (VoL)

- **Assume a competitor comes in and takes away the above average risks.**
- **Because of adverse selection, the new loss ratio will be higher than the current loss ratio.**
- ***What is the value of avoiding this fate?***
  - $16 in this illustration
  - Insurer could have spent additional $16 for segmentation and been no worse off
- **May express the VoL as a $ per car year.**
  - $5.33 per policy

Value of Lift – ISO Risk Analyzer Personal Auto Environmental Module

Coverage	Value of Lift
Bodily Injury	$4.99
Property Damage	$3.63
Collision	$1.61
Comprehensive	$4.85
Personal Injury (PIP)	$15.04
Combined	$13.29

Based on holdout sample of all coverages industry data (4.5 million records)

Illustration – Year 2

Policy #	Premium	Insurer's Expected Loss	Break-Even Loss
2	70	35	42
3	70	35	42
Total	140	70	84
Ratio to Premium		50%	60%

Accurate Expected Loss	Insurer's Profit
30	12
40	2
90	14
50%	10%

Illustration – Year 2

Policy #	Premium	Insurer's Expected Loss	Break-Even Loss
2	70	35	42
3	70	35	42
Total	140	70	84
Ratio to Premium		50%	60%

Accurate Expected Loss	Insurer's Profit	
30	~~12~~	0
40		2
90	~~14~~	2
50%	~~10%~~	1.4%

Lost Profit = 12

Illustration – Year 3

Policy #	Premium	Insurer's Expected Loss	Break-Even Loss
3	80	40	48
Total	80	80	48
Ratio to Premium		50%	60%

Accurate Expected Loss	Insurer's Profit
40	8
40	8
50%	10%

Illustration – Summary

No Enhanced Segmentation

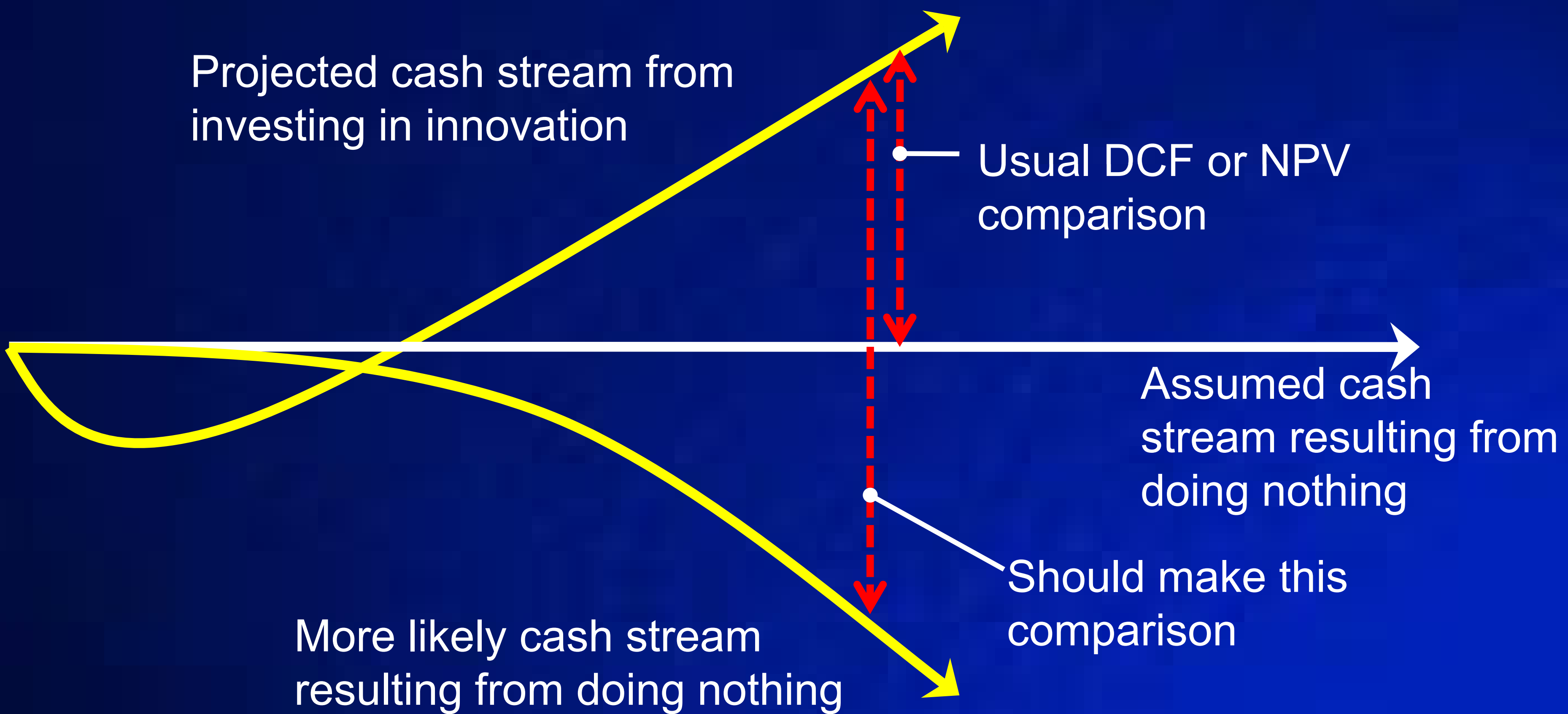
Year	Premium	Profit
0	180	18
1	120	2
2	70	2
3	80	8

NPV	25

- **Declining Revenue**
- **Declining Profit**

- **Calculate NPV**
  - Using 10% discount rate
- **Proper Basis of Comparison**

The Discounted Cash Flow Trap



Source: Christensen, Kaufmann, Shih, “Innovation Killers: How Financial Tools Destroy Your Capacity to Do New Things”, Harvard Business Review, Jan 2008

Alternative Scenario Enhanced Segmentation

Year	Premium	Profit excl Marginal Costs	Marginal Costs	Profit
0	180	18	10	8
1	180	18	3	15
2	180	18	3	15
3	180	18	3	15

NPV	41

- **Assume premium and policies are retained**
- **Directly consider implementation costs**
  – Higher first year expenses

Comparison

No Enhanced Segmentation

Year	Premium	Profit
0	180	18
1	120	2
2	70	2
3	80	8

NPV	25

Enhanced Segmentation

Year	Premium	Profit
0	180	8
1	180	15
2	180	15
3	180	15

NPV	41

- Greater NPV for Enhanced Segmentation

References

- Glenn Meyers, “Value of Lift”, Actuarial Review, May 2008

- David Cummings, “Value of Lift – A Net Present Value Framework”, Actuarial Review, Feb 2009

Extensions of this Approach

- **Refined considerations of retention and conversion effects**
- **Consider different premium scenarios**
- **Projections are inherently uncertain**
  - Use stochastic simulation to project future scenarios under uncertainty
  - Connection with Strategic Risk Management

Summary

- Predictive Modeling has had a profound impact on the insurance industry
- Significant innovations in progress for the next wave of advanced analytics
- Assessing the value of segmentation requires understanding of marketplace dynamics
- Profitability and market share are at risk for those who do nothing