Territory Analysis with Mixed Models and Clustering

PRESENTED BY

ERIC J. WEIBEL AND J. PAUL WALSH

2008 CAS PREDICTIVE MODELING SEMINAR

SAN DIEGO, CA
OCTOBER 6 - 7, 2008

Agenda

- Introduction
- Motivation: Risk Classification Challenges
- Mixed Model Approach
- Mixed Model Results
- Constrained Cluster Analysis
- Final Results
- Future Research
- Conclusions

Agenda

- **<u>Introduction</u>**
- Motivation: Risk Classification Challenges
- Mixed Model Approach
- Mixed Model Results
- Constrained Cluster Analysis
- Final Results
- Future Research
- Conclusions

Introduction

- Views of the speakers are not necessarily identical to the views of the cosponsors of this program nor the employers or clients of the speakers

Introduction

- Broad Objective: Address Risk Classification Challenges to Territory Analysis
- Technical Objective: Assign zip codes to the appropriate territory for frequency or severity (Frequency or Severity Band)
- Two stage technical solution
- Stage 1: Mixed model – credibility weight three indications: 1) Raw indication for zip code, 2) Indication generated by an arithmetic model of auxiliary data, 3) Indication generated by surrounding zip codes (proximity complement)
- Stage 2: Constrained cluster analysis of stage 1 results using Nonlinear Programming

Agenda

- Introduction
- **Motivation: Risk Classification Challenges**
- Mixed Model Approach
- Mixed Model Results
- Constrained Cluster Analysis
- Final Results
- Future Research
- Conclusions

Motivation: Risk Classification Challenges

- Non-Actuarial
  - Causality
  - Controllability
  - Loss Control/Incentive Value
  - Objectivity
  - Affordability
- Actuarial
  - Integration
  - **Homogeneity vs. Credibility**

Motivation: Risk Classification Challenges

- Homogeneity vs. Credibility
  - Loss Cost Gradient (LCG) Dominance
    - Solutions with other variables
    - Why no simple solution in territory analysis?
  - Traditional Resolution in Territory Analysis
    - Without Auxiliary Data: 1) McDonald Approach, 2)Proximity Complement Approach, 3) Spline & Graduation Approaches
    - Subjective Resolution with Auxiliary Data
    - Objective Resolution with Auxiliary Data: Riegel's Approach
  - Possible New Methods of Resolution
    - Arithmetic Model of Causal Geographical Variables
    - Our Approach: Mixed Model of Zip Code Indication, Arithmetic Model, and Proximity Complement

Motivation: Risk Classification Challenges

- Objective Groupings
  - Once we have Mixed Model indications for each zip code, it is not a trivial matter to *objectively* select which zip codes should be placed into the ten frequency or severity bands
  - Furthermore, what if we want to simultaneously introduce non-actuarial risk classification criterion into the procedure?
  - Constrained Cluster Analysis provides the means

Agenda

- Introduction
- Motivation: Risk Classification Challenges
- **Mixed Model Approach**
- Mixed Model Results
- Constrained Cluster Analysis
- Final Results
- Future Research
- Conclusions

Mixed Model Approach

- Early Actuarial Discussion of Mixed Models
  - Introduction by Bishop, Fienberg and Holland (1975) [25]
  - Discussion in general terms as combination of cellular result and arithmetic model: Chang & Fairley (1978) [27], Venter (1990) [36], Mildenhall (1999) [33]
- Our Mixed Model
  - Three Estimators for Each Zip Code
    - Raw Indication for Zip Code
    - Arithmetic Model Indication for Zip Code
    - Raw Indication for nearby Zip Codes ("Proximity Complement")
  - Credibililty Weighting Formula

Mixed Model Approach: Raw Zip Code Indication

- Easy.

Mixed Model Approach: Arithmetic Model

- Multiple Linear Regression Elected for Simplicity
- Factor Selection
  - Select *Causal* Factors where possible
  - Select *Quantitative* Factors where possible
  - Select Acceptable Factors
- Variable Construction
  - Data Sources
    - 1990 Decennial Census
    - 2005 Survey of Economic Conditions
  - Spatial Interraction
    - Within Zip Code, and Mutually Exclusive Radii: 10, 25 and 50 miles

Mixed Model Approach: Arithmetic Model: Factor Selection

- Factors Included in our Model
  - Traffic Density
  - Legal Environment
  - Population Density
  - Binary Geographical Variable
- Factors Discussed in our Paper
  - Nature of Population
  - Traffic Enforcement
  - Weather
- Other Factors
  - Topography
  - Roads
  - Regulation
  - Driver Education
  - Medical Costs
  - Repair Costs

Mixed Model Approach: Arithmetic Model: Variable Construction

- Traffic Density
  - Numerator: Number of Minutes Spent Commuting One Way
    - Source 1990 Decennial Census
  - Denominator: Land Area
    - Source 1990 Decennial Census
- Legal Climate
  - Elected Variable = Lawyer Density: Feldblum, Conner & Feldblum
  - Numerator: Number of Persons Employed in Legal Offices
    - Source 2005 Survey of Economic Conditions
  - Denominator: Population
    - Source 1990 Decennial Census

Mixed Model Approach: Arithmetic Model: Variable Construction

- Population Density
  - Numerator: Population
    - Source 1990 Decennial Census
  - Denominator: Land Area
    - Source 1990 Decennial Census
- Binary Geographical Variables
  - If simply constructed, replicate current practice: Not Causal
  - Some argument for causality if correspond to legal jurisdiction
  - Central Los Angeles (90001-90077), Remainder Los Angeles, San Francisco

Mixed Model Approach: Proximity Complement

- Hunstad Method
  - Use Local Assigned Risk Territory Definition
- Tang Method
  - Use immediately contiguous zip codes as 1st complement
  - If necessary use Local Assigned Risk Territory as 2nd Complement
- Hunstad Suggestions
  - Weight each zip code by distance
  - Add individual zip codes until full credibility reached
- Our Approach – 10 mile radius

Mixed Model Approach: Credibility Weighting Formula

- Zip code indication credibility, z, determined by the 1,082 claim rule.
- Proximity complement credibility:

$$z_p = \frac{\left(\sqrt{\frac{c}{1082}}\right)(1-z)}{\left(\sqrt{\frac{c}{1082}} + R^2\right)}$$

- Arithmetic model credibility:

$$z_m = \frac{(R^2)\,(1-z)}{\left(\sqrt{\frac{c}{1082}} + R^2\right)}$$

- R^2 is the corresponding arithmetic model statistic, c is the number of claims in the proximity complement

Mixed Model Approach: Credibility Weighting Formula

- Our goal was to introduce the concept, rather than implement the best possible means of combining mixed model elements.
- As a result we did not devote much effort to arriving at a credibility weighting scheme.
- We leave it to future researchers to arrive at optimal credibility weighting scheme, which ideally would incorporate the relative local fit of the arithmetic model and proximity complement.
- Or, perhaps a more formal mixed model could be arrived at.
- Because of the rudimentary nature of our implementation, we were willing to intervene in the credibility weighting process in the event the local performance of the arithmetic model or proximity complement was too poor.

Agenda

- Introduction
- Motivation: Risk Classification Challenges
- Mixed Model Approach
- **Mixed Model Results**
- Constrained Cluster Analysis
- Final Results
- Future Research
- Conclusions

Mixed Model Results

- Plots of actual values, model predicted values and residuals, and proximity complements are presented in Appendix A.

Mixed Model Results: Regression Model Conclusions

- Ideally, simple binary variables will not need to be introduced, and other continuous causal variables could be introduced that would reflect these differences.
- Failing that, should try to define boundaries of geographical binary variables that correspond with court jurisdiction groupings

Mixed Model Results: Proximity Complement Conclusions

- Proximity Complement Conclusions
  - A dynamically determined radius would dramatically improve performance.
    - Information Sparseness = increase radius
    - Information Density = decrease radius
    - Steep LCG = decrease radius
    - Flat LCG = increase radius
  - In Appendix C of the paper, we compare our proximity complement performance with Hunstad for BI frequency, using mean absolute deviation for each CAARP territory.

Agenda

- Introduction
- Motivation: Risk Classification Challenges
- Mixed Model Approach
- Mixed Model Results
- **Constrained Cluster Analysis**
- Final Results
- Future Research
- Conclusions

Our Objectives in Creating Groupings

- The use of professional judgment in creating territorial groupings is a frequent source of criticism: Barber (1929) [1], Casey et al. (1976) [26], Phase I (1978) [19], Shayer (1978) [34].
- Our goal is to objectively group zip codes into bands that accurately reflect their expected relative frequency and severity rates.
- Using grouping to generate credible results is less of a concern for us because we already incorporated arithmetic model and proximity complement.

Desired Features of Groupings

- Binary decision variables arrayed in a matrix of 10 columns and 1,502 rows.
  - columns corresponding to frequency or severity bands
  - rows corresponding to a zip code.
- Only one column in each row can take on a value of "1", meaning that the zip code belongs to that band.
- This is set up as:
  - $x_{ij} \in [0,1] \in \mathbb{N}$ (2.3)
  - $\sum_{i} x_{ij} = 1$ (2.4)

Desired Features of Groupings

- Desirable L_2 or L_1 objective functions might be:

$$\min \sum_i \sum_j \left[\left(R_i - \frac{\sum_b x_{bj} R_b E_b}{\sum_c E_c x_{cj}} \right) x_{ij} E_i \right]^2 \qquad (2.5)$$

$$\min \sum_i \sum_j \left[abs \left(R_i - \frac{\sum_b x_{bj} R_b E_b}{\sum_c E_c x_{cj}} \right) x_{ij} E_i \right] \qquad (2.6)$$

- R_i is the computed mixed model relativity.
- E_i is the number of exposures in zip code i.

Desired Features of Groupings

- Social and regulatory acceptability constraints on the grouping process:
  - No band can consist of a land area of less than 20 square miles
  - We may wish to impose a minimum exposure or claim count for each band for credibility purposes
  - Factor weight constraints
- The 20 square mile constraint could be setup as follows, with Li representing the land area for zip code i.

$$\sum_i L_i x_{ij} \geq 20 \qquad (2.7)$$

Cluster Analysis Review

- Cluster Analysis literature is vast, diverse and somewhat unorganized. It developed somewhat independently under the auspices of different academic disciplines.
- The two standard texts are:
  - Kaufmann and Rouseauww (KR in sequel) (1990) [46]
  - Everitt, Landau and Leese (2001) [43]
- Han, Kamber and Tung (HKT in sequel) (2001) [45] also provide a remarkably brief introduction.

Cluster Analysis Review

- One major divide in Cluster Analysis techniques is the distinction between Hierarchical and Partitioning (KR) / Optimization (Everitt et al.).
- KR claim that Partitioning / Optimization methods will tend to arrive at the best groupings for a fixed number of clusters.

Cluster Analysis Review

- Imposition of constraints is a very new topic in cluster analysis.
  - KR does not even mention it.
  - Everitt et al. focus on proximity/contiguity constraints and certain constraints related to hierarchy.
  - HKT have a broader discussion of pioneering work being done. In particular they refer to Tung et al (2001) [52].

Cluster Analysis Review

- Tung et al divide constraints into six types.
- Summation constraint involves the sum of some quantity tied to the units being grouped. In our case land area would be an example.
- Factor weight constraints or minimum claim or exposure counts (for credibility purposes) are similar.
- Unfortunately, Tung et al do not provide a method of solution, and furthermore discuss the difficult nature of the problem.

Cluster Analysis Review

- Since HKT and Tung et al. both discuss how difficult *summation* constraints will be to solve, this leaves us in a bit of a pinch with respect to the cluster analysis literature.
- Cluster Analysis literature provides no answers for summation constraints at this time.
- However...

Cluster Analysis Review

- Teboulle et al. (2006) [51] draws relationship between *k-means* cluster analysis and nonlinear programming gradient-type method.
- Given that a relationship between partitioning / optimization cluster analysis and nonlinear programming has been made, it would seem we should look to nonlinear programming to see if it offers a solution.

Nonlinear Programming Approach to Constrained Clustering

- A review of our objective function and initial constraints reveals that it can be considered a nonlinear programming problem from operations research. See Hillier and Lieberman (1995) [60].
  - Non-convex objective function
  - Binary decision variables
  - Linear / binary type constraints
- Choice of a Solver Application
  - R
  - Excel Solver
  - **Frontline Systems, Inc., Premium Solver**

Frontline KNITRO^{tm}

- KNITRO^{tm} appeared to be the best solver engine to use for our problem.
- All integer programming type problems employ branch & bound.
- KNITRO^{tm} uses one of three methods each time it conducts a minimization step
  - Interior Point Algorithms (Barrier Methods): Byrd, Gilbert and Nocedal (2000) [56], Byrd, Nocedal and Waltz (2003) [58].
    - Conjugate Gradient
    - Direct
  - Active Set (Sequential Linear Quadratic Programming): Byrd, Gould, Nocedal and Waltz (2004) [57].

Problem Setup

- As originally configured, our problem is too large to be solved in a reasonable amount of time.
- It is the modeler's task to creatively specify the model in a manner that makes maximum use of the structure present, increasing chances of success and decreasing computational demands.
- The size of the problem can be significantly reduced, and its structure made more clear with a few steps.

Problem Setup

- Sort the zip codes by increasing mixed model indication.
- Pre-assign the zip codes to frequency or severity bands by mixed model indication deciles, e.g lowest 10% assigned to band 1.
- Prune irrelevant decision variables.
  - For example, the rightmost rows are clearly irrelevant for zip codes with low mixed model indications – an optimal solution will never assign those zip codes to one of the high bands.
- Non-Decreasing Band Assignment Constraint.

$$0 \le \sum_{j=1}^{10} j\left[x_{(i+1),j} - x_{i,j}\right] \le 1 \text{ for } i \text{ from } 1 \text{ to } 1{,}501 \qquad (3.5)$$

- Restrict problem in terms of the number of zip codes considered at one time.

$$x_{ij} \text{ for } i \le 148, j \le 2 \text{ and for } 149 \le i \le 296\ j \le 3, \text{ and } 297 \le i \le 444, 2 \le j \le 4$$

Final Model Formulation

- We elected to use the L_1 objective function (2.6), which converted to the range specified above is:

$$\min\left[\begin{array}{l} \sum_{i=1}^{148}\sum_{j=1}^{2}\left[abs\left(R_i-\frac{\sum_b x_{bj}R_bE_b}{\sum_c E_c x_{cj}}\right)x_{ij}E_i\right] \\ +\sum_{i=149}^{296}\sum_{j=1}^{3}\left[abs\left(R_i-\frac{\sum_b x_{bj}R_bE_b}{\sum_c E_c x_{cj}}\right)x_{ij}E_i\right] \\ +\sum_{i=297}^{444}\sum_{j=2}^{4}\left[abs\left(R_i-\frac{\sum_b x_{bj}R_bE_b}{\sum_c E_c x_{cj}}\right)x_{ij}E_i\right] \end{array}\right] \qquad (3.7)$$

- In our initial attempts we decided to ignore the minimum land area constraint (2.7).

Sequential Solution Procedure

- The sequential solution procedure essentially involves moving downward and to the right through our original range of decision variables.

	i range	FB1	FB2	FB3	FB4	FB5	FB6	FB7	FB8	FB9	FB10
Setup1	1 to 148	1	0								
	149 to 296	0	1	0							
	297 to 444		0	1	0						

	i range	FB1	FB2	FB3	FB4	FB5	FB6	FB7	FB8	FB9	FB10
Setup1	1 to 148	1	0								
	149 to 296	0	1	0							
	297 to 444		0	1	0						
Solution1	1 to 116	1									
	117 to 275		1								
	276 to 444			1							

	i range	FB1	FB2	FB3	FB4	FB5	FB6	FB7	FB8	FB9	FB10
Solution1	1 to 116	1									
	117 to 275		1								
	276 to 444			1							
Setup2	117 to 275		1	0							
	276 to 444		0	1	0						

	i range	FB1	FB2	FB3	FB4	FB5	FB6	FB7	FB8	FB9	FB10
Solution1	1 to 116	1									
	117 to 275		1								
	276 to 444			1							
Setup2	117 to 275		1	0							
	276 to 444		0	1	0						
	445 to 592			0	1	0					

	i range	FB1	FB2	FB3	FB4	FB5	FB6	FB7	FB8	FB9	FB10
Solution1	1 to 116	1									
	117 to 275		1								
	276 to 444			1							
Setup2	117 to 275		1	0							
	276 to 444		0	1	0						
	445 to 592			0	1	0					
Solution2	117 to 276		1								
	277 to 453			1							
	454 to 592				1						

	i range	FB1	FB2	FB3	FB4	FB5	FB6	FB7	FB8	FB9	FB10
Solution2	117 to 276		1								
	277 to 453			1							
	454 to 592				1						
Setup3	277 to 453			1	0						
	454 to 592			0	1	0					

	i range	FB1	FB2	FB3	FB4	FB5	FB6	FB7	FB8	FB9	FB10
Solution2	117 to 276		1								
	277 to 453			1							
	454 to 592				1						
Setup3	277 to 453			1	0						
	454 to 592			0	1	0					
	593 to 740				0	1	0				

	i range	FB1	FB2	FB3	FB4	FB5	FB6	FB7	FB8	FB9	FB10
Solution2	117 to 276		1								
	277 to 453			1							
	454 to 592				1						
Setup3	277 to 453			1	0						
	454 to 592			0	1	0					
	593 to 740				0	1	0				
Solution3	277 to 474			1							
	475 to 628				1						
	629 to 740					1					

Setup4	475 to 628	1	0				
	629 to 740	0	1	0			
	741 to 888		0	1	0		
Solution4	475 to 637	1					
	638 to 766		1				
	767 to 888			1			
Setup5	638 to 766		1	0			
	767 to 888		0	1	0		
	889 to 1036			0	1	0	
Solution5	638 to 794		1				
	795 to 927			1			
	928 to 1036				1		
Setup6	795 to 927			1	0		
	928 to 1036			0	1	0	
	1037 to 1184				0	1	0
Solution6	795 to 928			1			
	929 to 1067				1		
	1068 to 1184					1	

Setup7	929 to 1067	1	0			
	1068 to 1184	0	1		0	
	1185 to 1332		0		1	0
Solution7	929 to 1084	1				
	1085 to 1220		1			
	1221 to 1332				1	
Setup8	1085 to 1220		1		0	
	1221 to 1332		0		1	0
	1333 to 1485			0		1
Solution8	1085 to 1223		1			
	1224 to 1339				1	
	1340 to 1485					1

Elected KNITRO^{tm} Solver Parameters

- Global Optimization of non-convex problems
  - Finding a global optimum is not usually guaranteed.
    - (Sometimes it can be guaranteed in integer programming problems, but usually it would take to long to arrive at a guaranteed solution.)
  - As a result, additional measures should be taken to make it likely that a good solution near the global optimum is arrived at:
    - Multi-Start Search
    - Topographic Search

Elected KNITRO^{tm} Solver Parameters

- Automatic Scaling - Yes
- Derivatives – Forward 1st Order
- Sparse Optimization - Yes
- Integer Tolerance – 0.05
- Other Parameters – Defaults Used

Agenda

- Introduction
- Risk Classification Challenges to Territory Analysis
- Homogeneity vs. Credibility
- Mixed Model Approach
- Mixed Model Component 2: Arithmetic Model
- Mixed Model Component 3: Proximity Complement
- Mixed Model Results
- Constrained Cluster Analysis
- **Final Results**
- Future Research
- Conclusions

For BI frequency the Hunstad assignments modestly outperform mixed models with clustering. The mixed model outperforms the Hunstad result for bands 1 and 10, with results for the first band significantly better.

<u>BI Frequency</u>

	FB1	FB2	FB3	FB4	FB5	FB6	FB7	FB8	FB9	FB10
					Relatvities					
Mixed Model	0.5438	0.6180	0.6730	0.7253	0.7866	0.8602	0.9870	1.1386	1.3374	1.7544
Actual	0.4895	0.5775	0.6589	0.7232	0.7882	0.8619	0.9940	1.1488	1.3472	1.7708
Hunstad	0.5334	0.6715	0.7456	0.8037	0.8767	0.9795	1.0752	1.1856	1.3425	1.7393
					MAD					
New Cell	0.00105	0.00092	0.00047	0.00039	0.00037	0.00045	0.00058	0.00071	0.00109	0.00315
Hunstad Cell	0.00121	0.00041	0.00034	0.00029	0.00048	0.00035	0.00052	0.00052	0.00086	0.00319
New Total					0.00087					
Hunstad Total					0.00083					

For PD frequency, mixed models with clustering moderately outperformed the Hunstad result. Our approach again outperformed for bands 1 and 10.

PD Frequency

	FB1	FB2	FB3	FB4	FB5	FB6	FB7	FB8	FB9	FB10
					Relatvities					
Mixed Model	0.6548	0.7265	0.7853	0.8423	0.9171	0.9663	1.0127	1.0598	1.1247	1.3036
Actual	0.6132	0.7137	0.7827	0.8423	0.9173	0.9671	1.0140	1.0613	1.1271	1.3102
Hunstad	0.7301	0.8634	0.9297	0.9642	0.9965	1.0219	1.0492	1.0740	1.1117	1.2430
					MAD					
New Cell	0.00223	0.00094	0.00081	0.00074	0.00067	0.00049	0.00047	0.00044	0.00114	0.00299
Hunstad Cell	0.00261	0.00129	0.00048	0.00042	0.00030	0.00027	0.00027	0.00029	0.00060	0.00318
New Total					0.00082					
Hunstad Total					0.00097					

For BI severity, our approach significantly outperformed the Hunstad result, and again outperformed in bands 1 and 10.

BI Severity

	SB1	SB2	SB3	SB4	SB5	SB6	SB7	SB8	SB9	SB10
					Relatvities					
Mixed Model	0.8297	0.8777	0.9026	0.9267	0.9499	0.9805	1.0136	1.0422	1.0761	1.1268
Actual	0.8224	0.8728	0.8985	0.9253	0.9508	0.9833	1.0154	1.0427	1.0765	1.1293
Hunstad	0.8380	0.8902	0.9202	0.9525	0.9792	1.0049	1.0232	1.0445	1.0675	1.1156
					MAD					
New Cell	207.61	129.62	91.92	87.93	87.16	124.18	90.86	92.81	100.82	206.48
Hunstad Cell	229.64	100.22	113.12	158.01	210.82	171.97	139.16	144.30	145.46	243.90
New Total					117.85					
Hunstad Total					168.71					

For PD severity, our approach moderately outperformed the Hunstad result, and again outperformed for bands 1 and 10.

PD Severity

	SB1	SB2	SB3	SB4	SB5	SB6	SB7	SB8	SB9	SB10
					Relatvities					
Mixed Model	0.8387	0.8770	0.9078	0.9346	0.9615	0.9905	1.0181	1.0423	1.0803	1.1487
Actual	0.8355	0.8755	0.9076	0.9349	0.9625	0.9909	1.0181	1.0421	1.0807	1.1503
Hunstad	0.8505	0.8989	0.9406	0.9771	0.9983	1.0155	1.0283	1.0449	1.0700	1.1303
					MAD					
New Cell	28.94	11.79	11.40	12.11	13.73	12.68	8.33	9.46	19.58	35.53
Hunstad Cell	29.83	18.28	20.34	12.95	10.06	5.18	7.54	8.00	14.25	42.84
New Total					14.67					
Hunstad Total					17.01					

Agenda

- Introduction
- Motivation: Risk Classification Challenges
- Mixed Model Approach
- Mixed Model Results
- Constrained Cluster Analysis
- Final Results
- **Future Research**
- Conclusions

Directions for Future Research

- Within the existing framework of Territory Analysis:
  - Refine Arithmetic Model, Proximity Complement, and Credibility Weighting Scheme
  - Refinement & Automation of Constrained Cluster Analysis
- Development of new Territory Analysis framework:
  - Introduce New Geographical Rating Variables
  - Integrate with classification plan ratemaking
- Refinements to California Personal Auto Ratemaking

Refinements to California Personal Auto Ratemaking

- Central argument against territorial rating by Prop 103's precursors
  - Not a causal variable
  - Subjective / arbitrary procedures in grouping
  - Intellectual Underpinnings of Prop 103
    - Casey *et al.* (1976) [26]
    - Shayer (1978) [34]
    - Ferreira (1978a) [28]
    - Ferreira (1978b) [29]
    - Chang & Fairley (1978) [27]
    - Stone (1978) [35]
    - Phase I (1978) [19]
    - Phase II (1979) [20]

Refine Ca. Personal Auto Ratemaking: Supplant Fq/Sv Bands w/ Causal Geog. Variables

- New Causal Geographical Rating Variables in California
- Insurance Commissioner can introduce new rating variables demonstrated to have a "substantial relationship to the risk of loss."
- Currently, two such geographical rating variables exist – relative claims frequency and relative claims severity
- As causal geographical variables are introduced, the more "undesirable" geographical variation in frequency and severity, with no known cause, would be captured in the relative frequency and severity bands.
- Scope of relative frequency and severity could be reduced as new causal geographical variables are introduced

Refine Ca. Personal Auto Ratemaking: Supplant Fq/Sv Bands w/ Causal Geog. Variables

- Traffic Density
- Traffic Enforcement
  - CDI itself investigated in Phase II - enforcement ratio
  - That study could be improved upon
  - Powerful loss prevention argument
  - Enforcement ratio already incorporates spatial interaction
  - Assign enforcement ratio for each zip code every year or so
- Medical and Repair Cost Indices

Refine California Pers. Auto Ratemaking: Const. Cluster Analysis vs. Pump/Temper

- Subsequent Court Criticism of Pumping and Tempering as "Arbitrary"
- Introduce factor weight as a constraint in the Cluster Analysis procedure
  - An investigational attempt to implement this form of constraint would be of interest

$$\frac{\sum_i \sum_j \left[abs\left(\frac{\sum_b x_{bj} R_b E_b}{\sum_c E_c x_{cj}} - 1 \right) x_{ij} E_i \right]}{\sum_d E_d} \leq M \qquad (3.9)$$

Agenda

- Introduction
- Motivation: Risk Classification Challenges
- Homogeneity vs. Credibility
- Mixed Model Approach
- Mixed Model Component 2: Arithmetic Model
- Mixed Model Component 3: Proximity Complement
- Mixed Model Results
- Constrained Cluster Analysis
- Final Results
- Future Research
- **Conclusions**

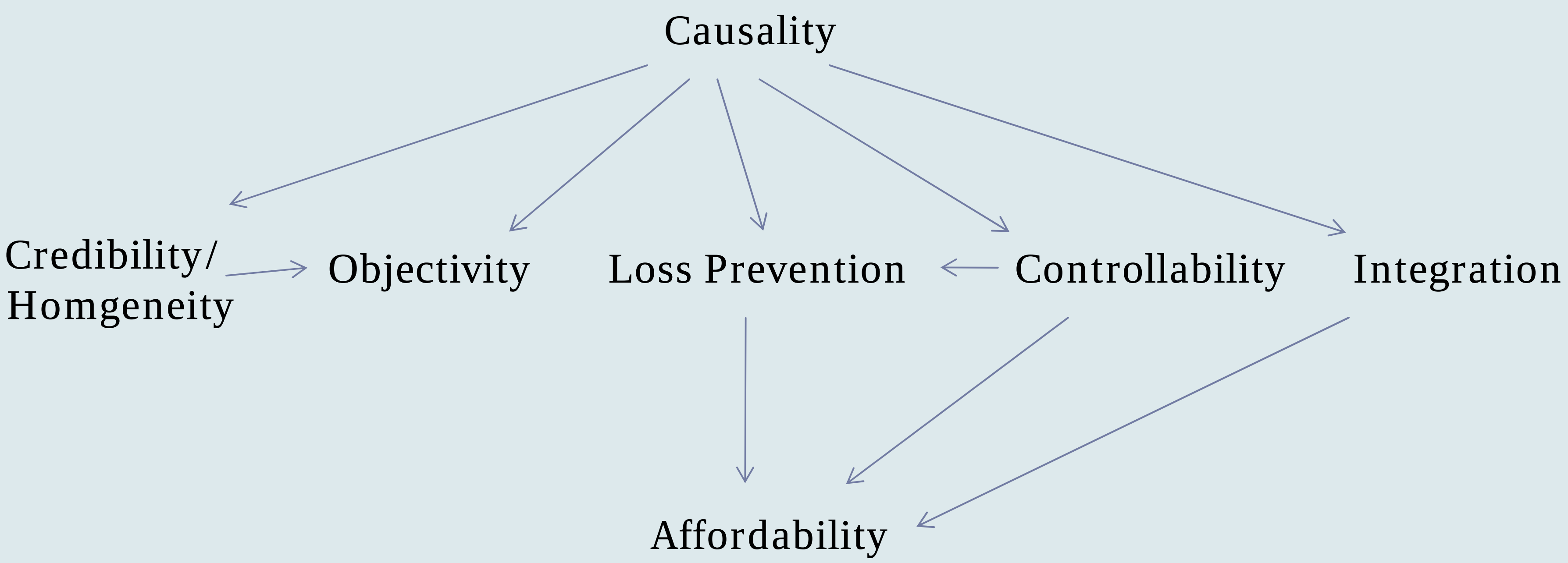
Conclusions

- Our approach is objective.
- Outperformed existing Manual despite the fact that the implementation of our concept was rudimentary.
- Significant further work can be done on improving each of the elements of the mixed model
- Sequential Cluster Analysis Procedure can be completely automated after it is perfected
- Causal analysis of geographical variation in loss costs associated with our approach could drive the development of new geographical rating variables

Conclusions

- New Causal Geographical Variables:
  - Eliminate criticisms regarding causality
  - Potentially invigorate local loss prevention initiatives
  - These largely continuous variables could be incorporated in the same predictive model as all the other rating variables

Conclusions



- Questions?

- Thank You