Institute and Faculty of Actuaries

General Insurance Claims Modelling with Factor Collapsing and Bayesian Model Averaging

Sen Hu, Adrian O'Hagan, T. Brendan Murphy

School of Mathematics and Statistics
Insight Centre for Data Analytics
University College Dublin, Ireland

18 September 2017

Motivating issues:

A variable appears to be borderline significant, included or excluded?

- Model uncertainty about variable selection
- How confident we should be about the final model.

Some categorical variables may have too many levels to be suitable for a standard GLM structure

Model parsimony and interpretability issues

- Lack of sufficient number of observations
- Insignificant levels should be merged (too many parameters).

Three motivating questions:

1

- Should a categorical predictor be included in modelling?

2

- When included, should certain levels be merged together?

3

- When included and with certain levels merged, how much confidence should be placed on this clustering of levels and this model?

A review of the current literature:

Collapse categories

- Pairwise multiple comparison with general linear hypothesis (Hothorn, Bretz and Westfall, 2008)
- Regularization methods such as lasso, OSCAR (Gertheiss and Tutz, 2010)
- Model-based clustering with Bayesian MCMC framework (Malsiner-Walli, Pauger and Wagner, 2017)
- "BMA" package in R (Raftery et al., 2015)

Keep all categories

- Generalized linear mixed models
- Combining GLM and credibility theory (Ohlsson and Johansson, 2008)

Method outline:

- Factor collapsing (FC) assesses which categories differ from one another with respect to the response
- There is uncertainty about the optimal combination
- Bayesian model averaging (BMA) takes such uncertainty into consideration

Variable selection → Categorical level selection

Two data sets are considered:

Illustrative set:

- Third party motor insurance claims data in Sweden in 1977 from Faraway (2016)
- 1797 observations
- 4 factors: Kilometres, Zone, Bonus, Make

Real industry set:

- Insurance claims history data from an Irish GI company from January 2013 to June 2014
- 452,266 policies
- Accidental damage (within comprehensive cover) is analyzed only.

Example: a question from the Sweden data set:

- Standard GLM output in R, for "Make" predictor in the frequency model

```
            Estimate Std. Error  z value Pr(>|z|)
(Intercept) -1.812178   0.013758 -131.721  < 2e-16 ***
...
Make2        0.086384   0.021240    4.067 4.76e-05 ***
Make3       -0.226013   0.025098   -9.005  < 2e-16 ***
Make4       -0.640736   0.024196  -26.481  < 2e-16 ***
Make5        0.161510   0.020235    7.982 1.44e-15 ***
Make6       -0.331235   0.017375  -19.063  < 2e-16 ***
Make7       -0.044705   0.023344   -1.915   0.0555 .
Make8       -0.008300   0.031606   -0.263   0.7929
Make9       -0.069596   0.009956   -6.990 2.74e-12 ***
```

- Standard GLM output in R, for "Kilometres" predictor in the severity model

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.39456    0.02335 359.474  < 2e-16 ***
...
Kilometres2  0.02455    0.01290   1.903  0.05718 .
Kilometres3  0.02124    0.01487   1.429  0.15327
Kilometres4  0.04306    0.02073   2.077  0.03793 *
Kilometres5  0.03945    0.02205   1.789  0.07374 .
```

Bayesian model averaging (BMA)

- Use BMA to average the best models (where possible)

$$\Pr(\Delta|D) = \sum_{k=1}^{K} \Pr(\Delta|M_k, D) \Pr(M_k|D)$$

$$\Pr(M_k|D) \approx \frac{\exp(-0.5\, BIC_k) \Pr(M_k)}{\sum_{r=1}^{K} \exp(-0.5\, BIC_r) \Pr(M_k)}$$

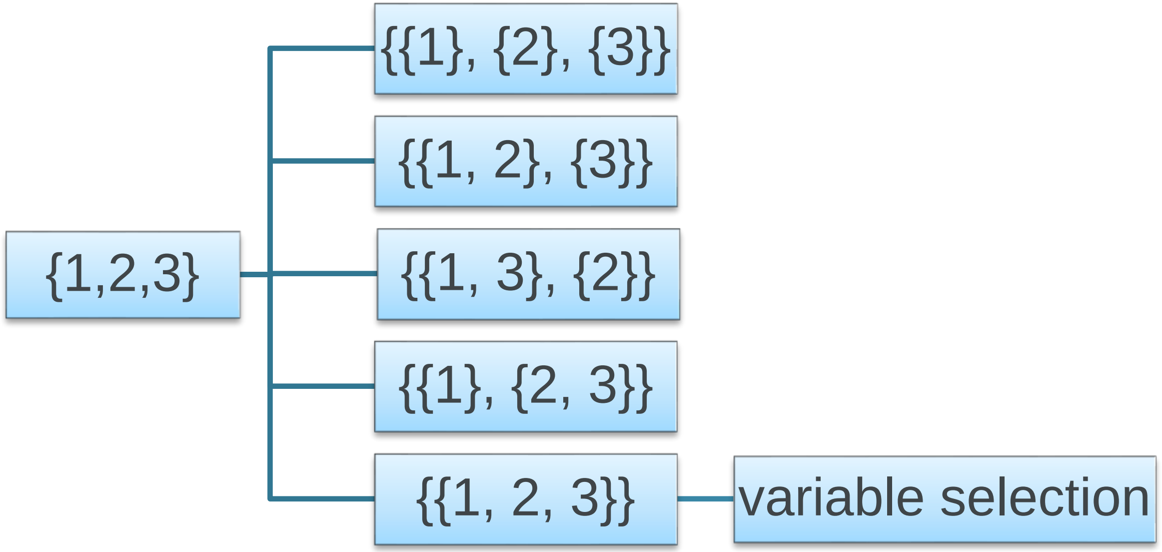
- $\Pr(M_{k)}$ is the prior for each model
- Average over model predictions
- Average over model coefficients.

Factor collapsing (FC)

Set partitions:

grouping elements within a set into non-empty subsets, in such a way that every element is included in one and only one subsets (“partitions” R package [Hankin, 2006])

Partitioning a 3-element set:



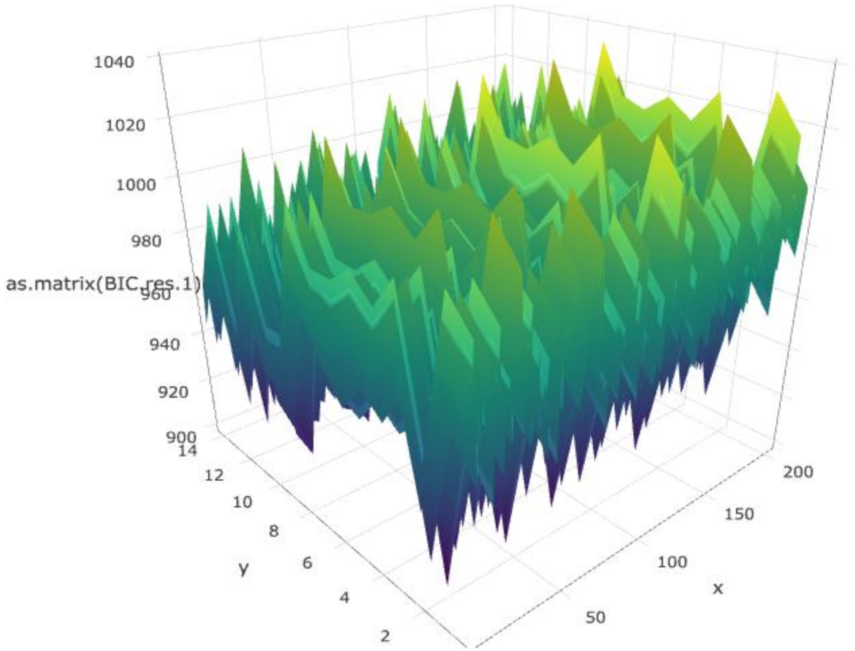
Fit each (combination of) partition into a pre-specified model

Bell number increases super exponentially

Stochastic search

Number of set partitions increases super exponentially

- It becomes computationally very intensive
- It becomes an optimization problem with a goal of finding the global minimum across model space



Optimization

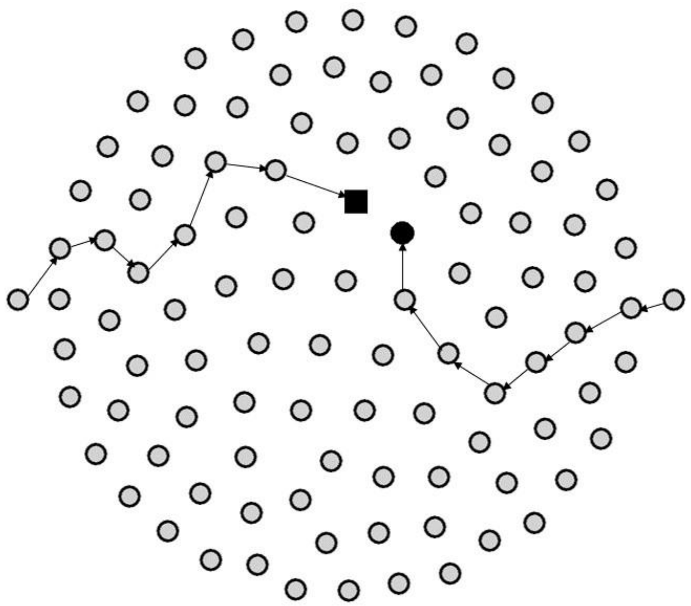
Many stochastic optimization (metaheuristic) methods work for this non-linear non-differential objective function

Simulated annealing:

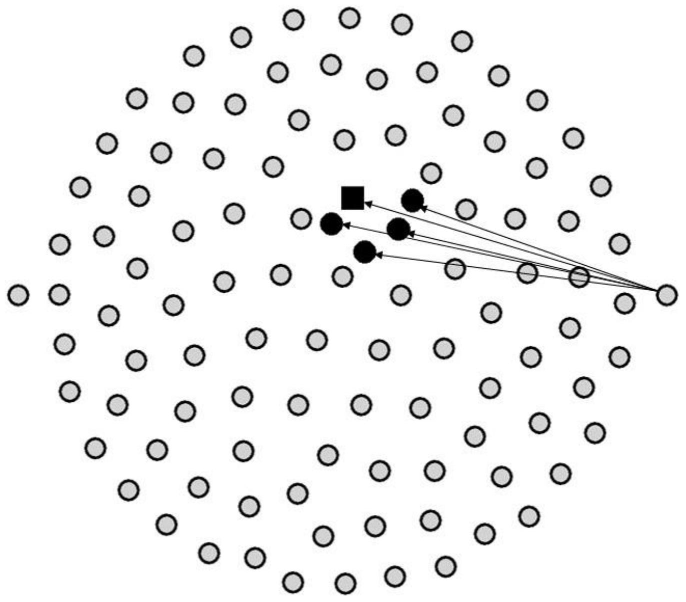
- Optimization technique based on Monte Carlo method
- Starting from a random state
- Make random state changes, accepting worse moves with probability determined by temperature
- Reduce temperature after reaching (close-to) equilibrium
- Stop once temperature gets very small

FC-BMA illustration

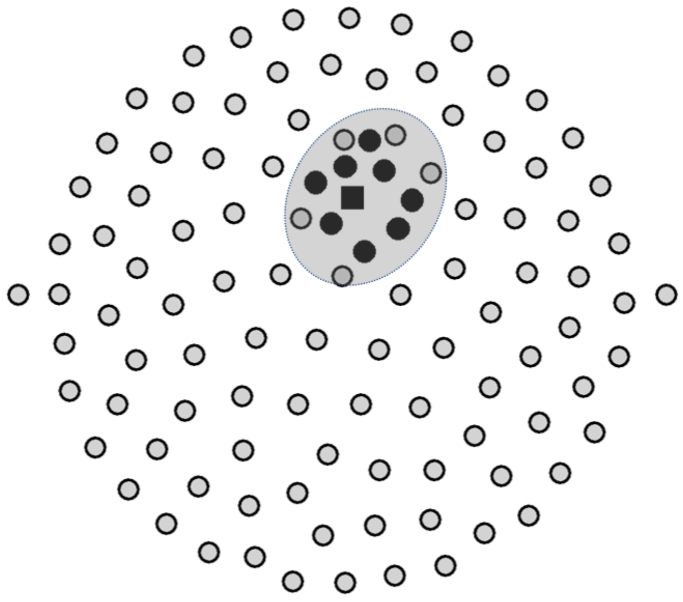
Comparing FC-BMA with stepwise selection using BIC or AIC



Forward and backward selection



FC-BMA



Optimal model region

Follow up example…

Predictor “Make” in the frequency model: results of multiple comparison using R package “multcomp” (Hothorn et al., 2016)

Hypothesis	Adjusted p-value
7 - 9 == 0 : coefficients of levels 7 and 9 equivalent	0.9583
8 - 9 == 0 : coefficients of levels 8 and 9 equivalent	0.4909
7 - 1 == 0 : coefficients of levels 7 and 1 equivalent	0.5600
8 - 1 == 0 : coefficients of levels 8 and 1 equivalent	1.0000
5 - 2 == 0 : coefficients of levels 5 and 2 equivalent	0.0839
8 - 2 == 0 : coefficients of levels 8 and 2 equivalent	0.1429
8 - 7 == 0 : coefficients of levels 8 and 7 equivalent	0.9837

Follow up example…

Predictor “Make” in the frequency model: results of multiple comparison using R package “multcomp” (Hothorn et al., 2016)

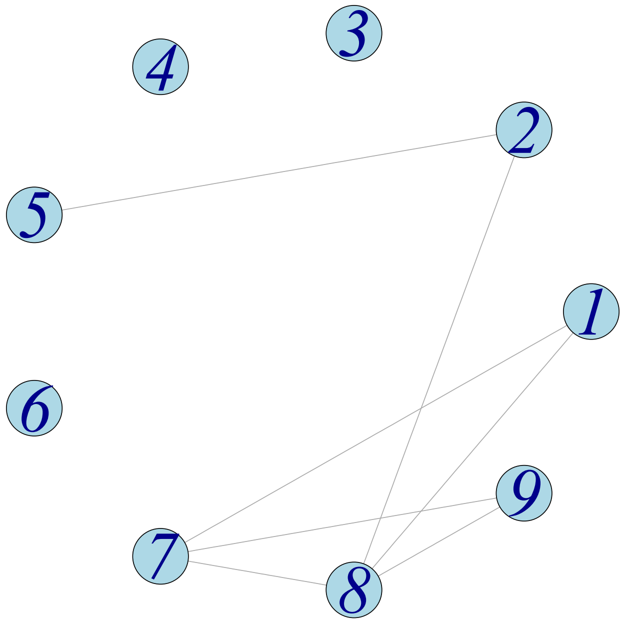


Illustration in R

Follow up example…

Results for collapsing “Make” factor only in the frequency model. Here only the best 5 models (based on their BIC values) are shown.

Partitions	BIC	BMA weights
(1,8)(2)(3)(4)(5)(6)(7,9)	10,301.11	0.3458
(1,8)(2,5)(3)(4)(6)(7,9)	10,301.81	0.2426
(1,7,8)(2)(3)(4)(5)(6)(9)	10,303.44	0.1076
(1,7,8)(2,5)(3)(4)(6)(9)	10,304.15	0.0754
(1)(2)(3)(4)(5)(6)(7,8,9)	10,304.92	0.0514

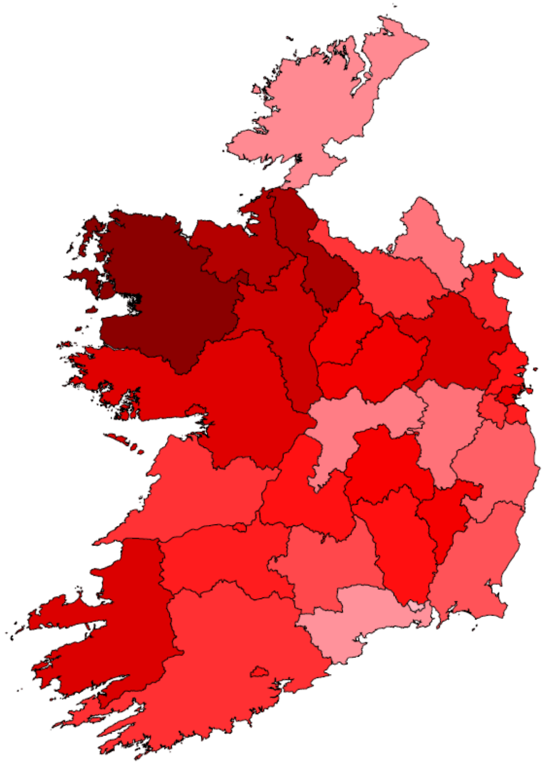
Follow up example…

Results for collapsing “Kilometres” factor only in the severity model. Here only the best 5 models (based on their BIC values) are shown.

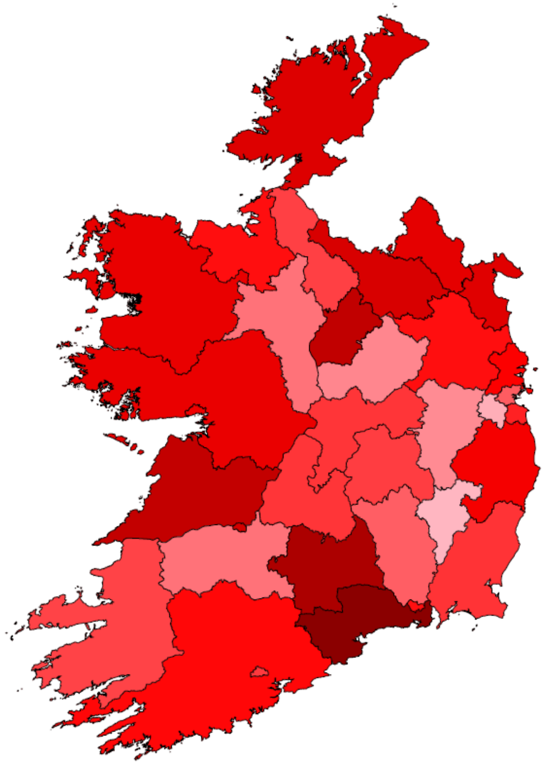
Partitions	BIC	BMA weights
(1)(23)(45)	1,878,161	0.8124
(1)(2)(3)(45)	1,878,164	0.1430
(1)(23)(4)(5)	1,878,167	0.0379
(1)(2)(3)(4)(5)	1,878,170	0.0067
(1)(25)(3)(4)	1,878,198	0.0000

Irish GI insurer data: counties

Coefficients of Irish counties from the standard GLM over an Irish map:



frequency



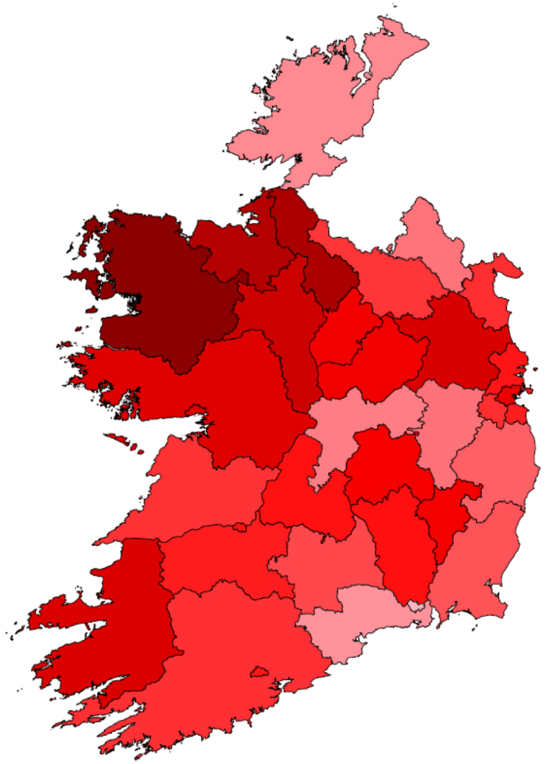
severity

Irish GI insurer data: counties in the frequency model

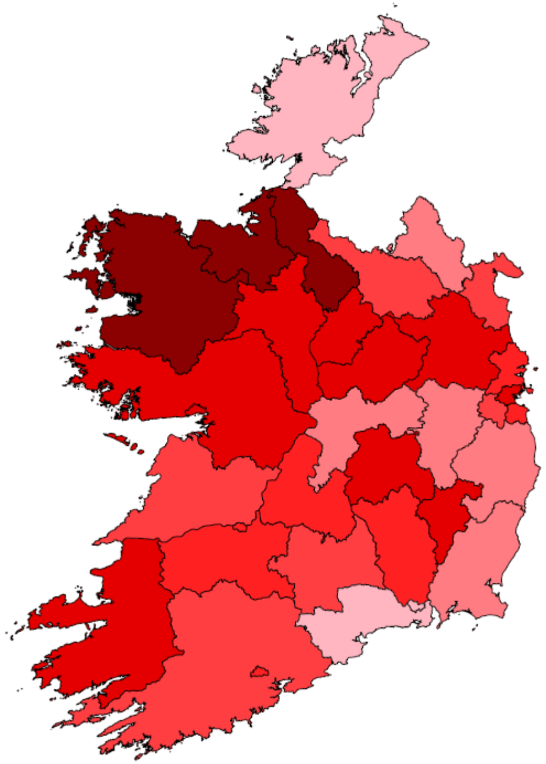
County	model coef.	new coef.
Waterford City	-6.6556	-6.6415
Unknown	-6.6130	-6.6415
Waterford County	-6.6073	-6.6415
Donegal County	-6.5959	-6.6415
Offaly County	-6.5787	-6.5733
Monaghan County	-6.5670	-6.5733
Kildare County	-6.5638	-6.5733
Wicklow County	-6.5397	-6.5733
Wexford County	-6.5217	-6.5733
South Tipperary	-6.5063	-6.5001
Cavan County	-6.4809	-6.5001
Clare County	-6.4764	-6.5001
Cork County	-6.4738	-6.5001
Louth County	-6.4720	-6.5001
South Dublin	-6.4708	-6.5001
Dun Laoghaire-Rathdown	-6.4489	-6.4648
Limerick County	-6.4473	-6.4648
Cork City	-6.4385	-6.4648
Fingal	-6.4379	-6.4648
North Tipperary	-6.4323	-6.4648
Limerick City	-6.4306	-6.4648
Kilkenny County	-6.4299	-6.4648
Laois County	-6.3923	-6.3766
Carlow County	-6.3865	-6.3766
Longford County	-6.3813	-6.3766
Westmeath County	-6.3808	-6.3766
Dublin City	-6.3694	-6.3766
Galway City	-6.3421	-6.3766
Galway County	-6.3415	-6.3766
Kerry County	-6.3323	-6.3766
Meath County	-6.3282	-6.3766
Roscommon County	-6.3031	-6.3766
Sligo County	-6.2503	-6.2106
Leitrim County	-6.2282	-6.2106
Mayo County	-6.1615	-6.2106

Irish counties clustering

Coefficients of Irish counties from the standard GLM to coefficients after clustering:



Frequency: before clustering



Frequency: after clustering

Irish counties clustering

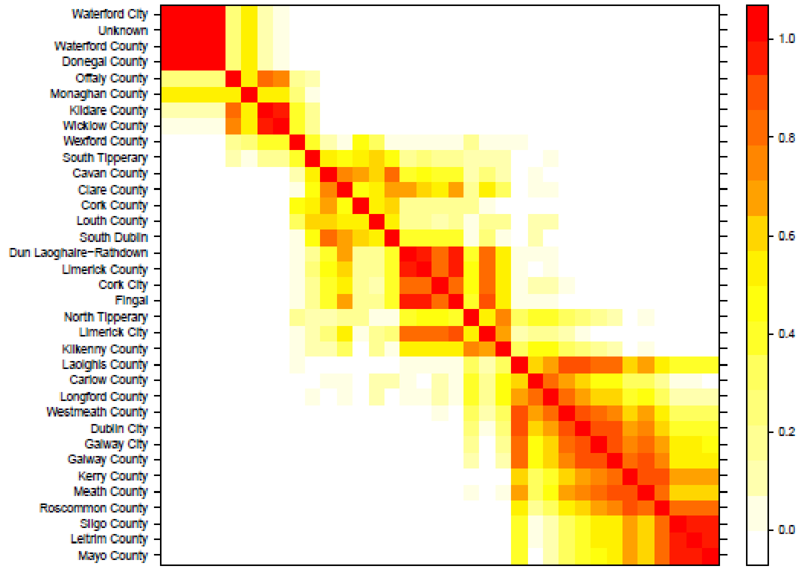
(Subset of) frequency model coefficients for the standard GLM and results of FC-BMA. Categories are of increasing order based on the standard GLM coefficients. Only five models are selected for illustration.

	Std. GLM	BMA	Model 1	Model 2	Model 3	Model 4	Model 5
BIC			62,807.2927	62,807.3039	62,807.3972	62,807.4069	62,807.4294
Model weights			0.0233	0.0232	0.0221	0.0220	0.0218
Waterford City	-6.6556	-6.6359	-6.6414	-6.6399	-6.6326	-6.6341	-6.6311
Unknown	-6.6130	-6.6359	-6.6414	-6.6399	-6.6326	-6.6341	-6.6311
Waterford County	-6.6073	-6.6359	-6.6414	-6.6399	-6.6326	-6.6341	-6.6311
Donegal County	-6.5959	-6.6359	-6.6414	-6.6399	-6.6326	-6.6341	-6.6311
Offaly County	-6.5787	-6.6218	-6.5733	-6.6399	-6.6326	-6.6341	-6.6311
Monaghan County	-6.5670	-6.6080	-6.5733	-6.5732	-6.6326	-6.6341	-6.6311
Kildare County	-6.5638	-6.5695	-6.5733	-6.5732	-6.5689	-6.5674	-6.5645
Wicklow County	-6.5397	-6.5695	-6.5733	-6.5732	-6.5689	-6.5674	-6.5645
Wexford County	-6.5217	-6.5695	-6.5733	-6.5732	-6.5689	-6.5674	-6.5645
South Tipperary	-6.5062	-6.5263	-6.5000	-6.5023	-6.5006	-6.5674	-6.5645
Cavan County	-6.4809	-6.5004	-6.5000	-6.5023	-6.5006	-6.5011	-6.4980
Clare County	-6.4764	-6.5004	-6.5000	-6.5023	-6.5006	-6.5011	-6.4980
Cork County	-6.4738	-6.5004	-6.5000	-6.5023	-6.5006	-6.5011	-6.4980
Louth County	-6.4720	-6.5004	-6.5000	-6.5023	-6.5006	-6.5011	-6.4980
South Dublin	-6.4708	-6.5004	-6.5000	-6.5023	-6.5006	-6.5011	-6.4980

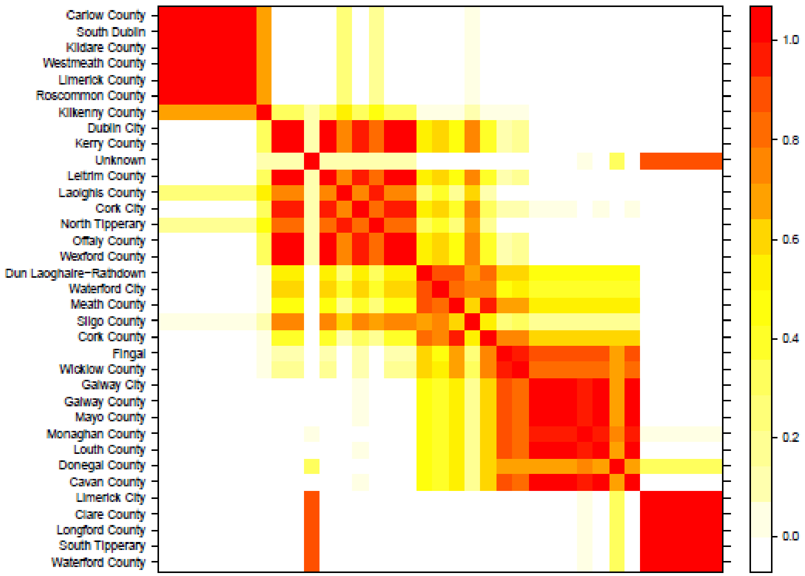
Irish counties clustering uncertainty

Clustering probabilities from the FC-BMA results for the county variable using a binary similarity matrix:

- The frequency part corresponds to the best 66 models selected
- The severity part corresponds to the best 17 models selected.



County clustering probabilities in frequency



County clustering probabilities in severity

Simultaneously collapsing other 19 factors

Frequency

Predictors	Categorical levels	Coefficients
Intercept		-17.2029
Policyholder gender	Female; Male; Neutral	*
Penalty point	0; 7-8	*
	1-6, 9+	-0.1263
Vehicle fuel type	Diesel, Petrol	*
	Unknown	0.7197
Vehicle transmission	Automatic; Manual; Unknown	*
Annual mileage	0-15000; 25001-50000	*
	15001-25000; 50001+	0.2341
Number of registered drivers	1; 2; 3; 4; 5; 6; 7	*
No claim discount (NCD)	0	*
	0.1	-2.2386
	0.2	-1.1407
	0.3; 0.4	12.7243
	0.5; 0.6	10.2648
NCD protection	No; Unknown	*
	Yes	-12.7728
NCD stepback	No	*
	Yes	0.1569
Total excess (€)	0; 300; 600	*
	125	0.3775
Main driver licence category	B; D; I; N	*
	C; F	0.0899

Severity

Predictors	Categorical levels	Coefficients
Intercept		9.6906
Policyholder gender	Female; Male; Neutral	*
Penalty point	0, 1-2, 3-4, 5-6, 7-8, 9+	*
Vehicle fuel type	Diesel; Petrol; Unknown	*
Vehicle transmission	Automatic; Manual; Unknown	*
Annual mileage	0-20000; 25001-40000	*
	20001-25000; 40001+	0.0772
Number of registered drivers	1; 2; 5	*
	3; 4; 6; 7	0.1367
No claim discount (NCD)	0; 0.1; 0.2; 0.3; 0.4; 0.5; 0.6	*
NCD protection	N	*
	Unknown; Y	-0.2960
NCD stepback	N; Y	*
Total excess (€)	0; 125; 300; 600	*
Main driver licence group	B; F; I; N	*
	C; D	0.0952

Metrics used for comparison

- **Gini index:** a measure of the area under the curve where cumulative exposure, ordered by fitted values, is plotted against cumulative response
- **Concordance correlation coefficient:** a measure of the agreement between two variables
- **Wasserstein distance:** a measure of distance between two distributions
- **Kolmogorov-Smirnov test:** a nonparametric test of the equality of two samples
- **KL divergence:** a measure of how one distribution diverges from a second distribution
- **Rooted mean square error:** a measure of the differences between values predicted by a model and the values actually observed.

FC-BMA results of Sweden data set

Prediction comparison (80% training data, 20% testing data) using Gini index, concordance correlation coefficient (CCC), Wasserstein distance (Wass.), Kolmogorov-Smirnov test (KS-test), KL divergence (KL) and mean squared error (MSE).

		Gini index	CCC	Wass.	KS-test	KL	RMSE
Frequency	No-FC	0.8266	0.9968	3.0340	**0.0736 (0.3045)**	0.0122	16.3383
	FC-only	**0.8267**	0.9943	**2.9696**	0.0788 (0.2358)	0.0114	**14.9927**
	FC-BMA	**0.8267**	**0.9973**	4.2012	0.0778 (0.2535)	**0.0113**	21.3629
Severity	No-FC	0.0567	0.0409	1948.3340	0.4489 (0)	0.2191	3840.3720
	FC-only	**0.0576**	**0.0667**	1825.0540	0.4067 (0)	0.2178	**3829.4340**
	FC-BMA	**0.0576**	0.0657	**1822.9450**	**0.4033 (0)**	**0.2178**	**3829.6680**

FC-BMA results of Irish insurer data set

Prediction comparison (80% training data, 20% testing data) using Gini index, concordance correlation coefficient (CCC), Wasserstein distance (Wass.), Kolmogorov-Smirnov test (KS-test), KL divergence (KL) and mean squared error (MSE).

		Gini index	CCC	Wass.	KS-test	KL	RMSE
Frequency	No-FC	0.7000	0.0489	0.0347	**0.9800(0)**	3.6127	0.1428
	FC-only	0.7016	**0.1078**	0.0337	**0.9800(0)**	3.5122	0.1404
	FC-BMA	**0.7019**	0.0977	**0.0335**	0.9806(0)	**3.5013**	**0.1378**
Severity	No-FC	0.5565	0.0559	**575.2057**	**0.2141(0)**	0.4573	4017.6130
	FC-only	0.5745	**0.1602**	855.8074	0.3264(0)	0.3328	2108.5300
	FC-BMA	**0.5747**	0.1602	858.6310	0.3242(0)	**0.3323**	**2106.9610**

“No Free Lunch” theorem!

“if an algorithm performs well on a certain class of problems then it necessarily pays for that with degraded performance on the set of all remaining problems.”

Summary

- FC-BMA deals with model selection and uncertainty and categorical level selection simultaneously
- It helps improve the model parsimony, interpretability and prediction
- Compared with other existing methods in literature, it does not require determination of extra parameters (non-parametric in nature)
- It can be a challenge to obtain the optimum through stochastic optimization and possibly could take a while to reach the optimum.

References

- Faraway, J. (2016). faraway: Function and datasets for books by Julian Faraway. R package version 1.0.7.
- Gertheiss, J. and Tutz, G. (2010). Sparse modeling of categorical explanatory variables. pages 2150-2180.
- Hankin, R. K. S. (2006). Additive integer partitions in R. Journal of Statistical Software, Code Snippets, 16.
- Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. Statistical Science, 14(4):382-417.
- Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. Biometrical Journal, 50(3):346-363.
- Hothorn, T., Bretz, F., Westfall, P., Heiberger, R. M., Schuetzenmeister, A., Scheibe, S., and Hothorn, M. T. (2016). Package `multcomp’. Simultaneous inference in general parametric models. Project for Statistical Computing, Vienna, Austria.
- Malsiner-Walli, G., Pauger, D., and Wagner, H. (2017). Effect fusion using model-based clustering. ArXiv e-prints.
- Ohlsson, E. (2008). Combining generalized linear models and credibility models in practice. Scandinavian Actuarial Journal, 2008(4):301-314.
- Raftery, A., Hoeting, J., Volinsky, C., Painter, I., and Yeung, K. Y. (2015). BMA: Bayesian Model Averaging. R package version 3.18.6.
- She, Y. (2010). Sparse regression with exact clustering. pages 1055-1096.

Questions

Comments

Thank you for listening.