A Practical Approach to Variable Selection — A Comparison of Various Techniques

Benjamin Williams
Greg Hansen
Aryeh Baraban
Alessandro Santoni

Abstract: Selecting a useful list of variables for consideration in a predictive model is a critical step in the modeling process and can result in better models. Sifting through and selecting from a long list of candidate variables can be onerous and ineffective, particularly with the increasingly wide variety of external factors now available from third-party providers. This paper explores a variety of variable selection techniques, applied to frequency and severity models of homeowner insurance claims, developed on a dataset with over 350 initial candidate variables. The techniques are evaluated using multiple criteria, including the predictive power of a resulting model (measured using out-of-sample data) and ease of use. A method based on Elastic Net performs well. Random selections perform as well as some more sophisticated methods, for sufficiently long shortlists.

Key Words: variable selection, frequency and severity models, homeowners, Elastic Net regularization

1. INTRODUCTION

Between the data stored by companies and that available from external providers, modelers now have access to hundreds or even thousands of factors. So many factors are available that it is often impractical to consider all of them in a formal predictive modeling context. This situation will only be accentuated in the future, as the number of candidate variables continues to grow. Recognizing which factors to consider in predictive modeling becomes an important problem for which automated approaches are required.

This paper considers this issue in the context of continuous, geo-demographic factors. For each variable selection technique, the long list of factors is reduced to a shortlist, upon which a model is fitted. The techniques are evaluated in terms of various criteria: predictiveness and goodness of fit of the resulting model; ease of set-up and ease of implementation. We do not go into great detail on each of the variable selection methods used, as none of the methods themselves are particularly novel, and such details are available elsewhere.

1.1 Research Context

The area of the science addressed by this paper is Financial and Statistical Methods, Statistical Models and Methods, Predictive Modeling.

1.2 Objective

Variable reduction is an active area of research in insurance modeling, and many papers have been written on the topic. However we consider that the exact question raised here, i.e., how to select, from a (potentially very) long list of factors, a shortlist which will be useful for current predictive modeling techniques, has not received sufficient attention. We make the distinction between this, which we call variable selection, and variable reduction. We consider variable reduction techniques to be those that create "super-factors," small numbers of linear combinations or other functions of the original factors, which may have more predictive power than the original list of factors. An example of what we consider a variable reduction technique is Principal Component Analysis.

The difficulty of fitting models when a large number of variables is available is widely recognized (see for example [1], [2], [4] and [5]). In [1] and [5], the use of a variable clustering technique to reduce a large number of geographic variables into clusters for use in Generalized Linear Models (GLMs) of auto insurance claims experience is discussed. In [2], Principal Component Analysis (PCA) and partial least squares are used to reduce variables on simulated data and the results are compared. [3] and [4] use alternative approaches based on data-mining.

Our analysis seeks to extend these papers by investigating a wide variety of techniques, including some of those discussed in the papers referred to above, and using different methods to compare predictiveness of models.

Our research considers a variety of variable selection techniques applied to a dataset of insurance claims experience, which has previously been randomly divided into training and testing datasets. Each factor selection technique results in a shortlist of factors, which are tested for inclusion in a GLM on the training dataset via an automated approach.

The performance of the resulting models is evaluated in terms of both predictiveness and goodness of fit. The predictiveness is evaluated on the testing data in two ways: 1) ranking Gini coefficients, and 2) comparing selected models via double-lift curves. The techniques are also evaluated in terms of ease of use (including software considerations and processing efficiency). The goodness of fit is ranked by deviance on the testing data.

1.3 Outline

Section 2 discusses in detail the data used, the techniques investigated, and how they were compared. Section 3 provides the results of our analysis. Section 4 gives our conclusions and recommendations for further analysis.

2. A COMPARISON OF VARIABLE SELECTION TECHNIQUES

2.1 Data Used

A dataset of approximately 1.9 million observations corresponding to five years of homeowner claims experience was used in our research. Predictors for each observation included 15 policy-related factors (including typical rating factors such as Age of Dwelling, Policy Tenure, Construction Type, Insurance Score Tier etc.) and over 350 ZIP code or Census Block level geo-demographic variables. This dataset was randomly divided into training and testing datasets on a ratio of 2:1. The geo-demographic variables were anonymized, ordinal, and pre-banded. All variable selection techniques were run and all models fitted on the training dataset, and all comparisons of predictiveness were performed on the testing dataset.

The techniques were analyzed in the context of four different model responses: frequency (number of claims per year of exposure) and severity (average cost per claim) for each of two claim types (water and fire), in order to see if results differed. Due to time constraints, it was not practical to create a shortlist for each method for fire frequency and severity.

2.2 Variable Selection Techniques Considered

The techniques considered belonged to one of three broad types:

1. Modeling methods that, when implemented, create a relevance score which allows the factors to be ranked. A shortlist of length N is created by taking the first N variables.
2. Variations on stepwise modeling techniques.
3. Random selection of variables, used as a baseline against which to gauge the performance of the more sophisticated techniques.

Prior to testing the variable reduction techniques on the 350+ geo-demographic variables, a Base Model was fitted on the 15 policy-related variables using traditional GLM techniques and assumptions. The techniques were then tested on a residual basis, contemplating the signal already explained by the Base Model. Several were also tested directly on the response variable (frequency or severity), without any consideration of the variables in the Base Model.

The following techniques were considered:

- Classification and Regression Trees (CART): the result of a standard implementation of CART by Salford Systems is a usefulness score, which allows the relevance of the variables to be ranked. This method was tested on both a direct (CARTBase) and residual (CARTRes) basis.

- Elastic Net: a penalized regression technique that uses a combination of two different penalty functions; L^2 (i.e., sum of squares) penalty function similar to the penalty used in Ridge regression and L^1 (i.e., sum of absolute values) penalty function similar to the penalty function used in the Least Absolute Shrinkage and Selecting Operator (LASSO) introduced by Tibshirani in 1996. The use of L^1 penalty function allows variables to enter the model one at the time. Variables that are most important in explaining the signal typically enter the model first followed by less important variables. The ranking of variables was based on the order in which they entered the model. This method was tested on both a direct (ENetBase) and residual (ENetRes) basis.
- AIC Improvement Rank: each factor under consideration is added to the model as a first, second and third degree polynomial. Candidates are ranked according to the AIC improvement of the best-performing polynomial. This method was tested on both a direct (AICRankBase) and residual (AICRankRes) basis.
- Stepwise GLM based on AIC Improvement with Correlated Variables Removed (GLMCorr): similar to AIC Improvement Rank. At each step, every factor under consideration is individually added to the model as a first degree polynomial. The best candidate (as determined by AIC improvement) is added to the model, and all strongly correlated variables are removed from further consideration. For this test, "strongly correlated" was subjectively defined as a correlation coefficient of greater than 0.35. This method was only tested on a residual basis.
- Stepwise Least Squares Regression with Correlated Variables Removed (LSRCorr): similar to the GLMCorr but using Least Squares Regression instead of GLMs to significantly improve processing speed. This method was only tested on a residual basis.
- Variable Clustering (Varclus): a standard implementation of the Varclus procedure in SAS, to create N clusters. The variable from each cluster with the lowest $1 - R^2$ ratio was selected.
- Random List (Rand): a random selection of factors from the list of available factors

2.3 Testing Predictiveness and Goodness of Fit

The factors in each shortlist were tested for inclusion in a GLM on the training dataset. The basics of GLMs are beyond the scope of this paper, see [6] for reference. In order to prevent the modeler's judgment from biasing the results, an automated modeling technique, forward regression using AIC improvement, was used. Starting from the same Base Model discussed above, which contained only policy factors, each factor in the shortlist was considered for inclusion as a first degree polynomial with unknown values grouped with the base level. The regression halted when the addition of no remaining unused factors resulted in an improved model.

Predictiveness of the shortlist provided by each technique was analyzed by testing how well the GLM fitted using the shortlist predicted out-of-sample data. Many methods exist to compare predictiveness of models (see, for example [7] for further discussion). As mentioned previously, we have limited ourselves to two methods: Gini coefficient[1] and Double Lift charts[2].

Goodness of fit of the shortlist provided by each technique was analyzed by ranking the deviance of the GLM fitted using the shortlist when applied to out-of-sample data.

2.4 Other Criteria

While we consider that the most fundamental property of a good variable selection technique is that it provides a shortlist of factors that result in a highly predictive model, other desirable properties are:

- Ease of set-up (in terms of software considerations and setting up the analysis)
- Processing efficiency (i.e., speed)

It should be noted that we did not attempt to undertake an exhaustive survey of all software currently available to carry out a variable selection technique.

---

[1] A Gini Coefficient is calculated from a Gains Curve, which is a plot of cumulative exposure, ordered by fitted values, against cumulative response. The Gini Coefficient is twice the area between the Gains Curve and the 45 degree line. The higher the Gini Coefficient, the more predictive a model.

[2] A Double Lift Chart compares two fitted model results on a given dataset. On the horizontal axis is the percentage difference between the fitted values, divided into bands. On the vertical axes are the average observed and fitted values in each percentage difference segment, and the exposure in each segment. The more predictive of the two models is that which more closely follows the observed values.
See [8] for a more detailed discussion on Gini coefficients and double lift charts.

3. RESULTS AND DISCUSSION

3.1 Testing Predictiveness and Goodness of Fit

In the following tables and graphs, N is the length of the shortlist on which each model was fitted. Most of the shortlists in the main table are of length 50 or 5. These lengths were decided upon by the authors. The few methods that have different length shortlists do so because they reached their stopping conditions prior to identifying 50 candidate variables.

Table 1. Comparison of Gini coefficients

		Water		Fire	
Method	N=	Frequency	Severity	Frequency	Severity
AICRankBase	50	0.2914	0.1413	0.2764	0.1018
AICRankRes	50	0.2869	0.1422	0.2766	0.1273
CARTBase	5	0.2982	0.1409		
CARTBase	50	0.3010	0.1473		
CARTRes	50	0.3043	0.1459		
ENetBase	5	0.2939	0.1394		
ENetBase	50	0.3060	0.1425	0.2806	0.1257
ENetRes	5	0.2999	0.1376		
ENetRes	50	0.3086	0.1475	0.2820	0.1168
GLMCorr	25	0.2997	0.1450		
GLMCorr	5	0.2997	0.1392		
Rand	5	0.2924	0.1390		
Rand	50	0.3079	0.1443	0.2711	0.1585
VarClus	50	0.3020	0.1462	0.2692	0.1521
LSRCorr	50	0.3073	0.1455		
LSRCorr	48				0.1166
LSRCorr	45			0.2781	

Chart 1. Ranking of Methods by Gini Coefficient for Water Frequency Models

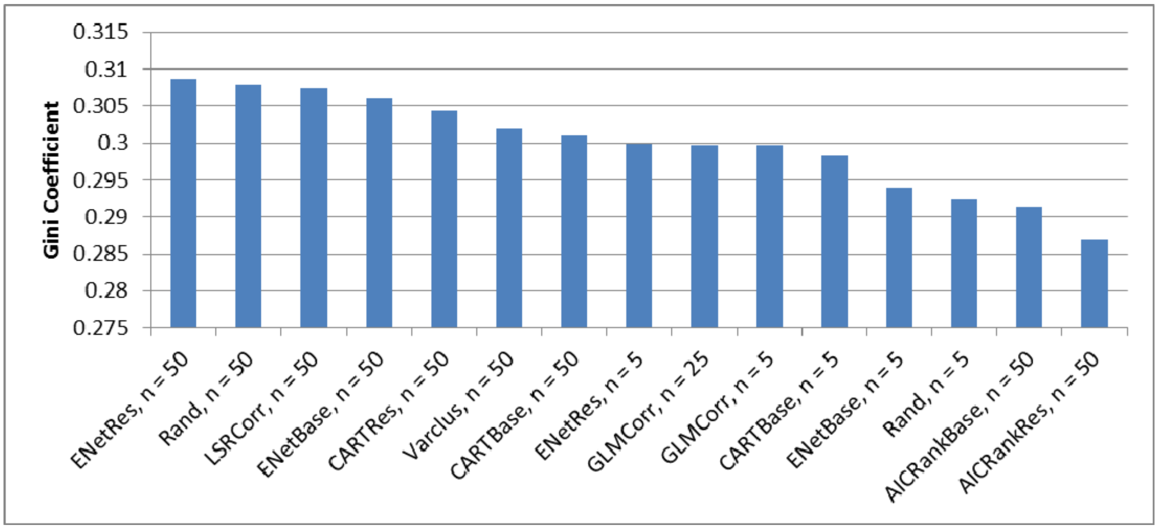


Chart 2. Ranking of Methods by Gini Coefficient for Water Severity Models

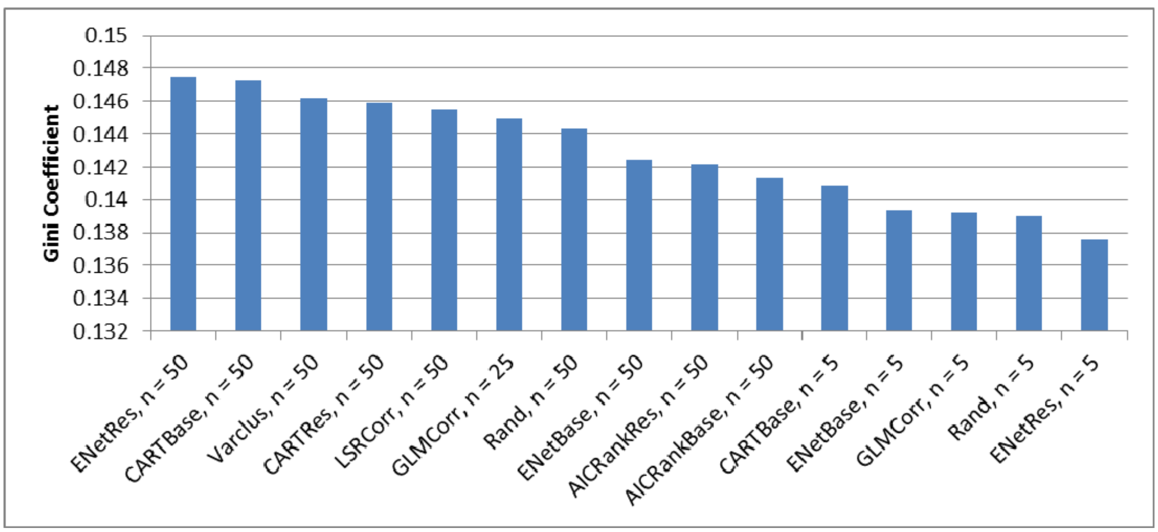


Chart 3. Ranking of Methods by Gini Coefficient for Fire Frequency Models

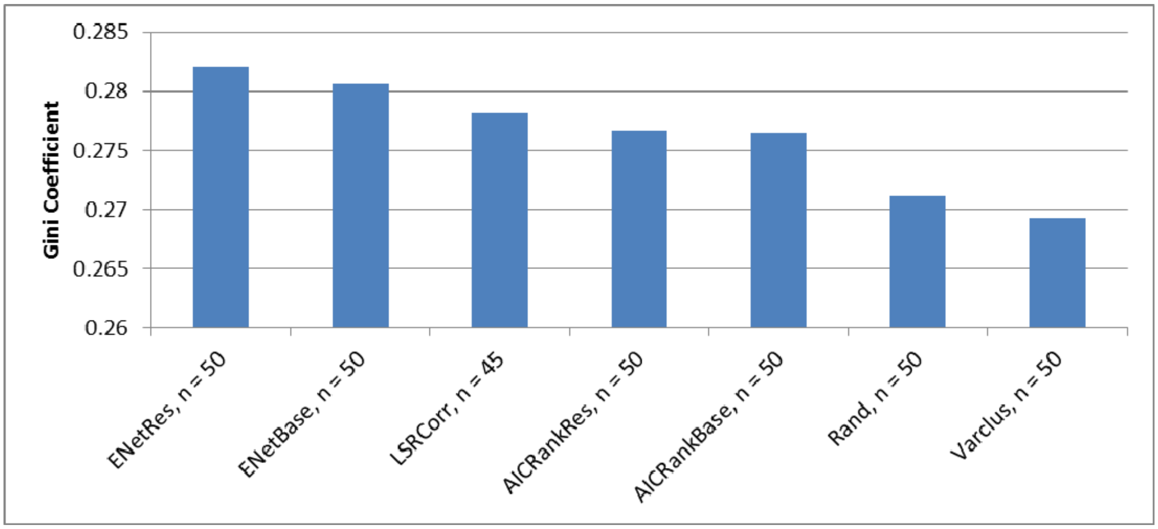
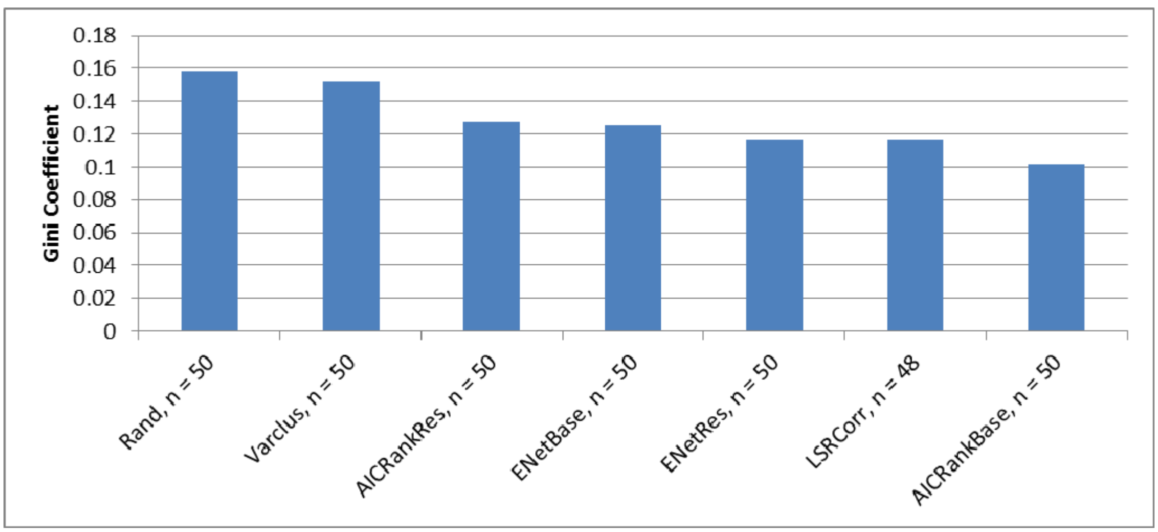


Chart 4. Ranking of Methods by Gini Coefficient for Fire Severity Models

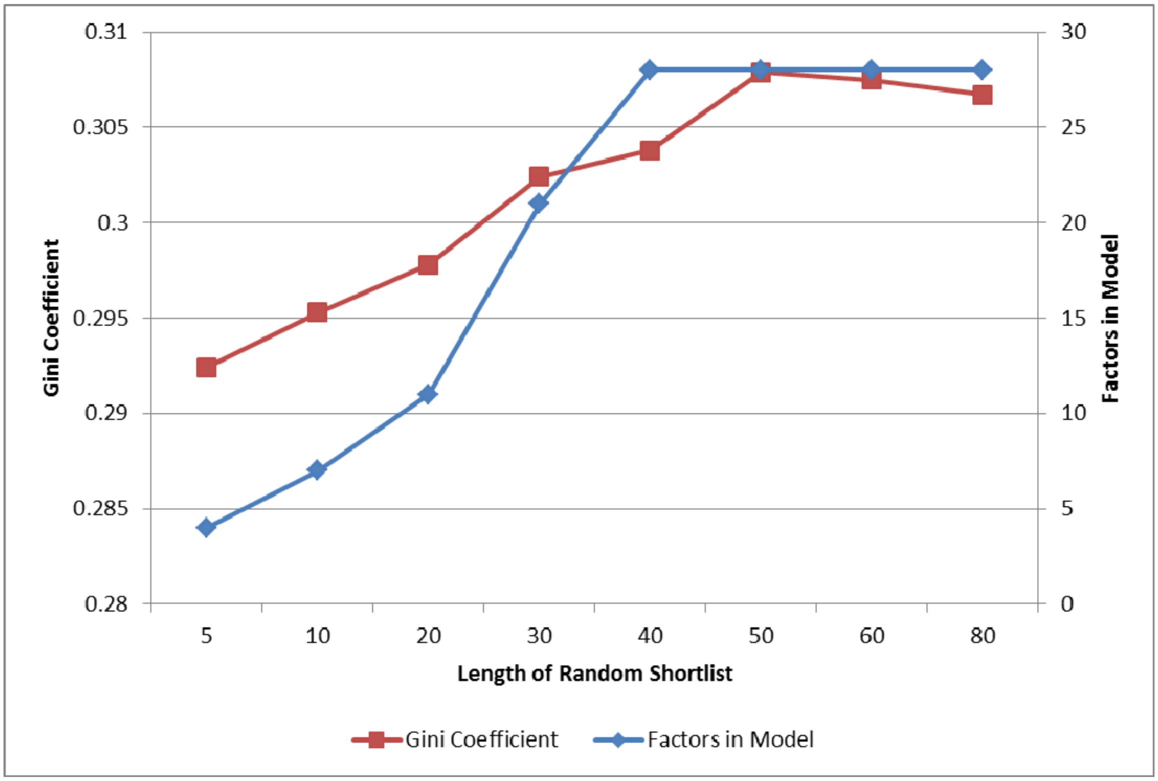


In the absence of an easy-to-apply theory of the distribution of Gini coefficients, it is not clear if any of the methods have performed significantly better than the others. However, we do make the following observations:

1. ENetRes N = 50 achieved the best results on both Water and Fire Frequency and Water Severity
2. ENetBase N = 50, VarClus and the different LSRCorr shortlists performed well.
3. Where the same method was used to generate two shortlists, the model fitted on the longer shortlist was, in most cases, more predictive.
4. A Random shortlist performed as well as many more sophisticated methods on the same length shortlist, for sufficiently long shortlists.
5. The results for Fire Severity are very different from the other models, in terms of ranking of methods and range of Gini coefficients. We believe that this is because of the comparatively small number of Fire claims in the modeling data.

Observations 3 and 4 inspired us to repeat our tests on random shortlists of different lengths. It should be noted here that random shortlists are nested, in the sense that our random shortlist of 10 included our random shortlist of 5, as well as 5 additional randomly-selected factors. The results led to an interesting conclusion, as displayed in Chart 5 below. This chart compares, for various sets of random shortlists (on the horizontal axis), the number of factors retained in the GLM (the right vertical axis) and the Gini coefficient (the left vertical axis). As the shortlists get longer, the number of factors retained in the model by the automated modeling technique plateaus (at about N = 40), while the Gini coefficient continues to improve.

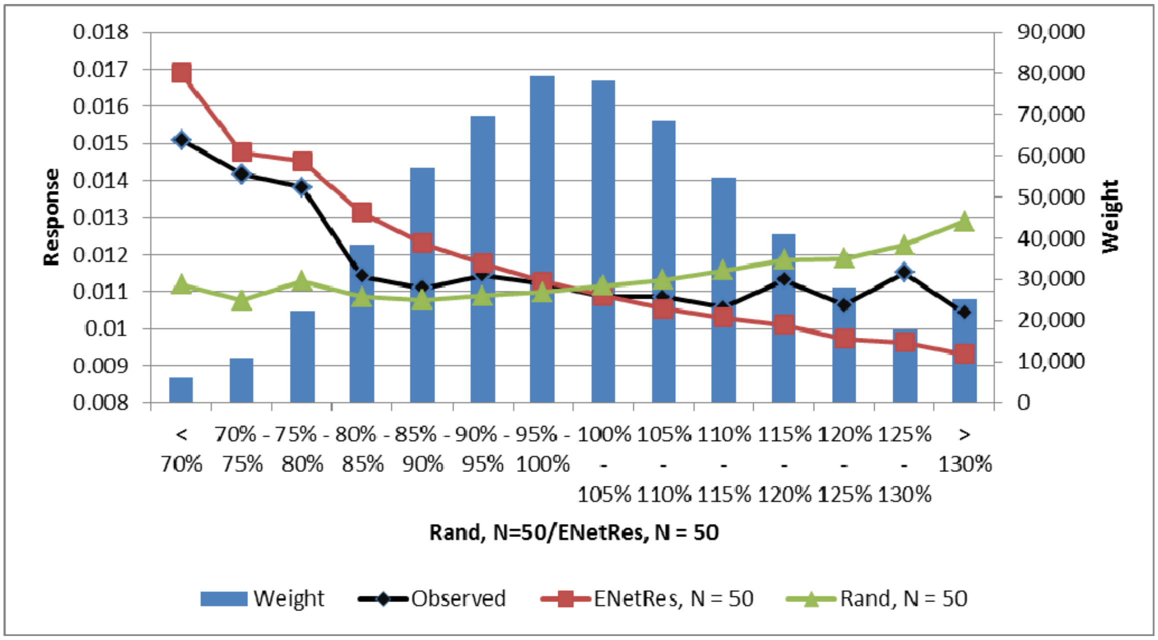
Chart 5. Number of factors in Water Frequency model and Gini coefficient for random shortlists of different lengths



We interpret this as follows. The longer the list of variables the automated modeling technique can choose from, the greater the chances that it will settle on a "best", or at least a "reasonable" predictive set. In the limiting case, the shortlist would include all available factors, and the automated modeling technique would produce the best possible model. This is also related to the structure of correlations among the different factors. We return to this point in Section 4.

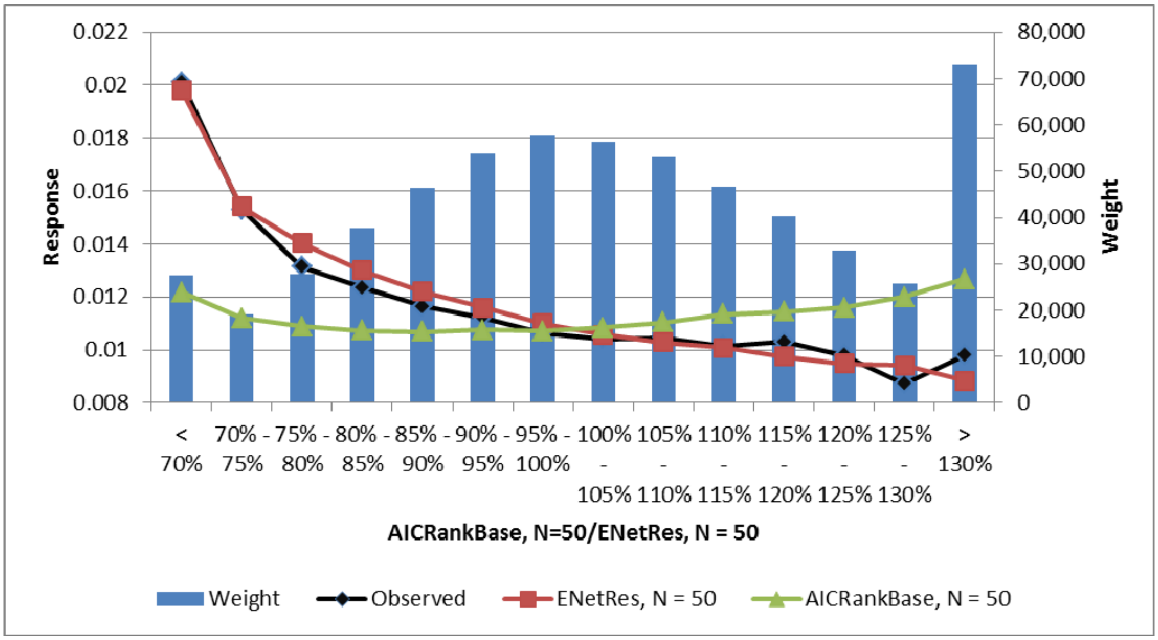
As noted above, a Gini coefficient is only one criterion for judging predictiveness of models. We now consider the double lift chart, to see if it allows us any additional insights.

Chart 6. Double Lift Chart ENetRes, N = 50 vs. Rand, N = 50, Water Frequency



In Chart 6, the two best performing methods in terms of Gini coefficient ranking for Water Frequency are compared. For ratio bands containing the majority of exposure, including 90%-115%, the observed response is closer to ENetRes, N = 50 than to Rand, N = 50, and we conclude that ENetRes, N = 50 is more predictive than Rand, N = 50. This agrees with the Gini coefficient ranking. However, we do note that in some well-populated ratio bands, for example 80%-90% and 115%-120%, the observed response is closer to Rand, N = 50 than to ENetRes, N = 50. It could be argued that ENetRes, N = 50 performs better where the models are not very different, but there is not much difference between the models where the differences between them are more pronounced (i.e., outside the ratio bands 90%-110%). On this basis, we do not consider this a strong victory for ENetRes, N = 50. This may reflect the small difference in Gini coefficient for these two methods.

Chart 7. Double Lift Chart ENetRes, N = 50 vs. AICRankBase, N = 50, Water Frequency



In Chart 7, the best and one of the poorest performing methods in terms of Gini coefficient ranking for Water Frequency are compared. In no ratio band does AICRankBase, N = 50 perform better than ENetRes, N = 50, and in most ratio bands it performs much worse. This appears to confirm the Gini coefficient ranking.

The analysis of other double lift charts tended to confirm the Gini coefficient ranking.

We now evaluate the methods in terms of goodness of fit of the resulting models.

Table 2. Model deviance

		Water		Fire	
Method	N=	Frequency	Severity	Frequency	Severity
AICRankBase	50	59,587	7,559	14,251	3,738
AICRankRes	50	59,634	7,562	14,251	3,712
CARTBase	5	59,495	7,571		
CARTBase	50	59,279	7,627		
CARTRes	50	59,234	7,641		
ENetBase	5	59,541	7,581		
ENetBase	50	59,362	7,655	14,237	3,715
ENetRes	5	59,466	7,589		
ENetRes	50	59,145	7,632	14,230	3,749
GLMCorr	25	59,374	7,544		
GLMCorr	5	59,457	7,577		
Rand	5	59,564	7,577		
Rand	50	59,345	7,648	14,259	3,735
VarClus	50	59,350	7,648	14,260	3,733
LSRCorr	50	59,233	7,656		
LSRCorr	48				3,703
LSRCorr	45			14,238	

Chart 8. Ranking of Methods by Deviance for Water Frequency Models

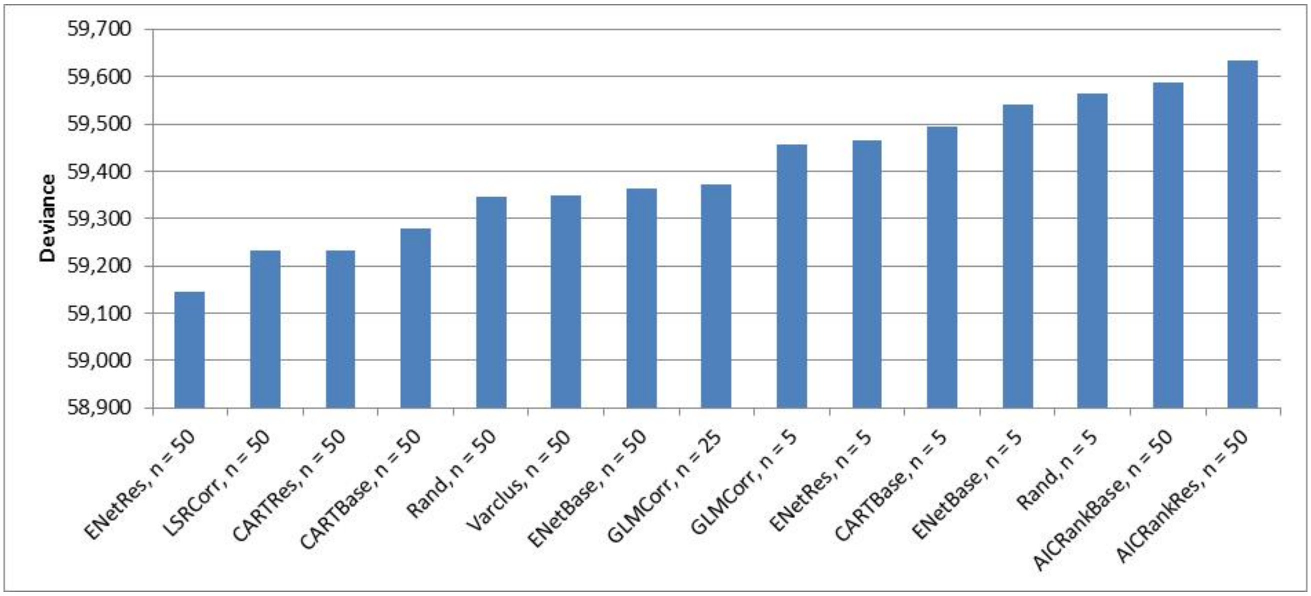


Chart 9. Ranking of Methods by Deviance for Water Severity Models

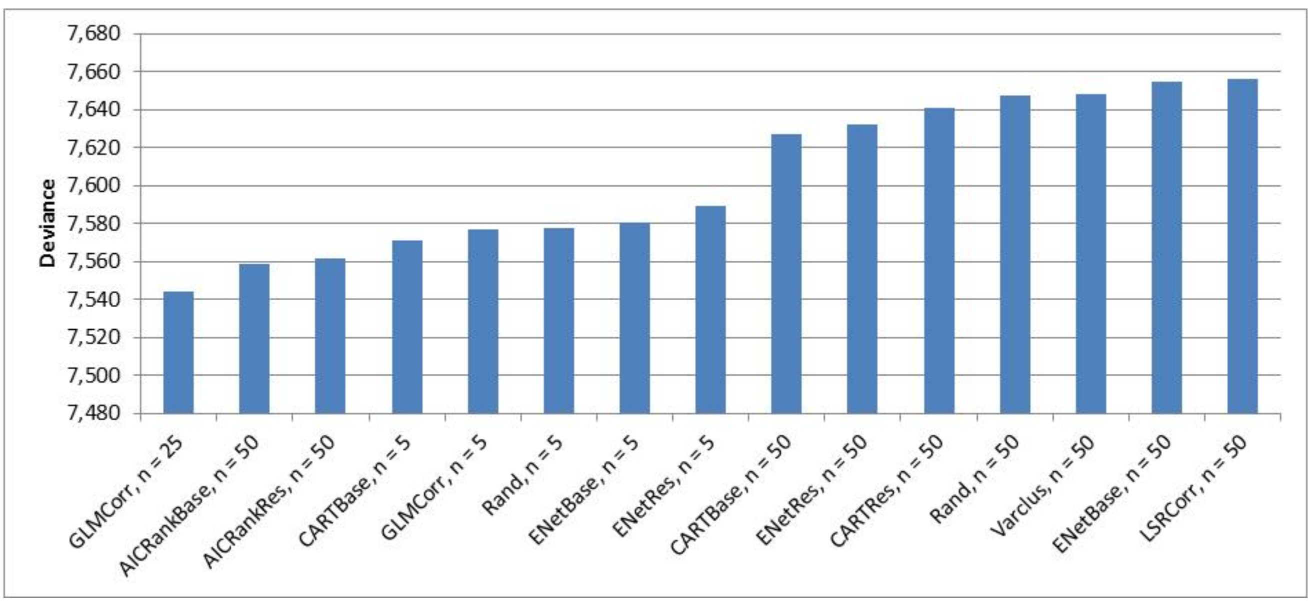


Chart 10. Ranking of Methods by Deviance for Fire Frequency Models

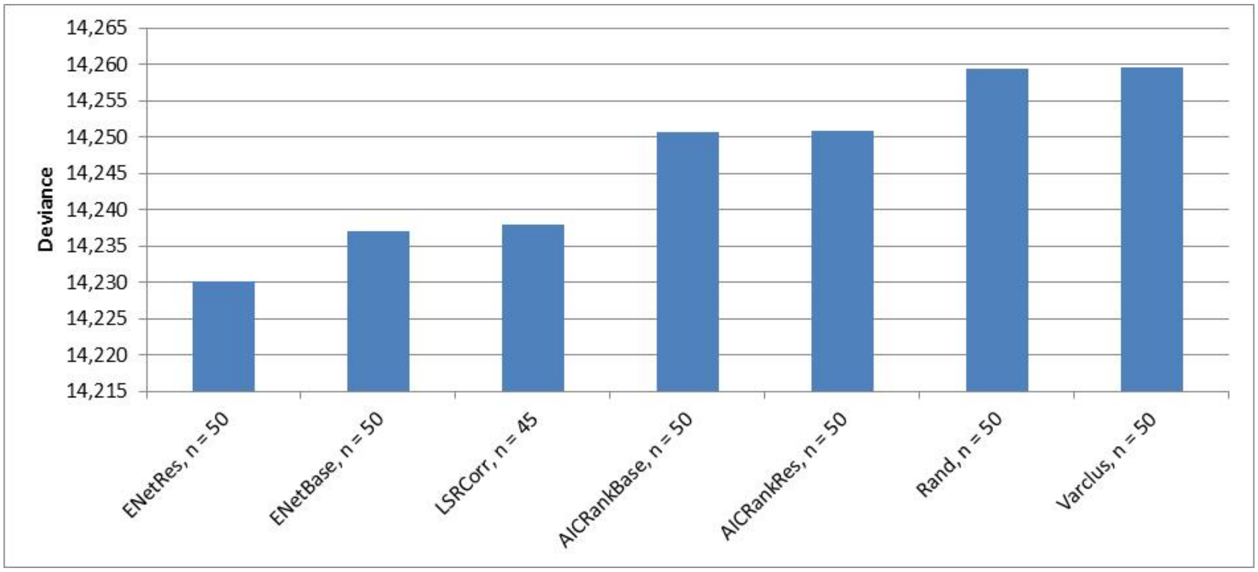
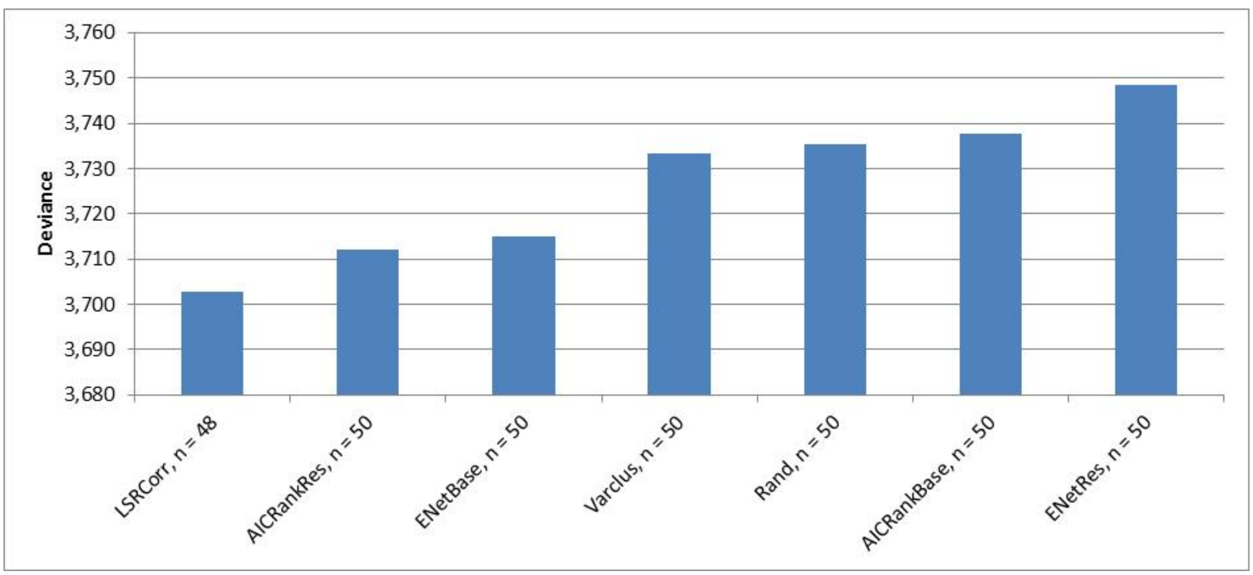


Chart 11. Ranking of Methods by Deviance for Fire Severity Models



While the range of deviances produced by the best and worst-performing methods in each case is not large, we make the following observations based on the above results:

1. When generated by the same method, longer shortlists produced lower deviances. We are not surprised by this result.
2. ENetRes N = 50 achieved the best results on both Water and Fire Frequency.
3. LSRCorr and GLMCorr also performed well.

In general, evaluating models based on either goodness of fit or predictiveness produced similar conclusions.

3.2 Other Criteria

Judging methods on ease of setting up and processing speed can be difficult, as users are typically restricted by the software and hardware available to them, and a given piece of software or better hardware could greatly simplify the set-up, or improve the speed. However we do think that it is important to give some indication of the amount of time required to set up and process each type of method. In each case, we assume the availability of a user with a level of expertise sufficient to implement the process - in practice, an important consideration. We have not conducted a survey of limitations here (e.g., that R requires all calculations to be done in memory, therefore limiting the size of the dataset studied).

Table 3 summarizes the software we used to perform each method, along with our assessment of its ease of implementation and processing speed.

Table 3. Comparison of methods by implementation and speed

Method	Software Used	Complexity to Set-Up	Processing Speed
AICRank	Emblem	Easy	Average
CART	CART	Easy	Fast
ENetBase	R	Average	Fast
GLMCorr	SAS	Easy	Slow
LSRCorr	SAS	Easy	Fast
Rand	-	Trivial	None required
VarClus	SAS	Easy	Fast

In the above table, Slow means over 8 hours, Average means 2-8 hours, and Fast means less than 2 hours. We leave these ranges deliberately wide as they are strongly dependent on the hardware used, as well as the software.

4. CONCLUSIONS AND FURTHER RESEARCH OPPORTUNITIES

Given that Elastic Net on Residuals performed consistently well across most of our tests, we consider it to be a strong candidate for situations similar to those studied here (i.e., personal lines frequency and severity, with a large number of ordinal rating factors). Other strong candidates in these situations are ENetBase, GLMCorr and LSRCorr. We also note that GLMCorr and LSRCorr may be preferred to methods involving Elastic Net, because of its comparative ease of implementation.

Given that this paper is experimental in nature and our experiments were limited to frequency and severity models of two homeowner perils, we do not consider that we are able to draw strong conclusions about what methods would be most suitable in any given situation. Drawing such conclusions would require the analysis of similar tests to those carried out here on many different datasets. We believe that it would be illustrative to observe how results vary for datasets of different sizes (we recall the observation made earlier that results for Fire Severity were very different than those for other models, and our belief that this is related to the limited number of data points). Analyzing a broader range of factors (for example categorical factors, and not just geo-demographic factors) could also provide interesting results. We consider these to be useful further lines of research.

We were surprised by the performance of random shortlists, which are the simplest to implement. We do note that conclusions reached depend on the length of the shortlist. On a very long shortlist, (N=50), a random list performed, in most cases, almost as well as any of the other methods tested. On a very short shortlist (N=5), a random shortlist did not perform as well as other methods with the same length shortlist. We ran tests for various lengths of shortlists, and saw that the number of factors included in the model by the automated modeling technique tended to plateau (see chart 5.) These two observations on random shortlists lead us to conclude that, assuming the availability of computing power able to fit models on shortlists in a reasonable amount of time, another viable approach is to randomly introduce candidate variables to a model until the number of significant factors plateaus.

Of course, the analysis of shortlists of varying lengths need not be restricted to the "random" method. The random method required a shortlist of about 50 variables for Water Frequency before the Gini Coefficient plateaued (see chart 5.) A more efficient method might plateau at the same (or higher) level of predictiveness with a much shorter shortlist. Future research could evaluate variable reduction methods based on their efficiency in addition to predictiveness and goodness of fit.

The performance of random shortlists also led us to consider the conditions under which a random shortlist will perform as well as other methods. We believe that this depends on the underlying data, in particular the lift provided by each factor, and the correlations between all factors. This can be illustrated with the following mental experiment: In an extreme case in which all factors are perfectly correlated, a random list of any length will work as well as any other technique. Consider, on the other hand, the opposite extreme of a dataset with hundreds of uncorrelated factors, only one of which provides any lift at all. In this case, most of the variable selection techniques discussed within this paper would successfully find the "needle in the haystack", whereas a random shortlist would only find it by chance. We consider that investigating the relationship between predictiveness of factors, correlations, and the usefulness of random shortlists is a worthwhile line of future research.

Because most of these techniques are relatively easy to develop and quick to execute, we see no reason why they could not be used in conjunction with each other. For example, consider the following hybrid process:

1. Fit a base model, using traditional techniques on a subset of factors believed to be relevant.
2. Employ ENetRes to narrow the complete list of candidate factors down to a more manageable shortlist.
3. Incorporate the shortlist of factors into the model, exploring traditional techniques such as splines, interactions, and spatial smoothing.
4. Employ LSRCorr, residual to the model developed in step 3, to seek out any additional factors that may have been missed in Step 2.

The extra step of re-scanning the remaining factors with a new selection method should further reduce the risk inherent in any automated selection process. We propose a refinement of this approach as another potential area for future research.

Finally, we feel it is important to distinguish between automated *variable selection* and automated *modeling*. While we used an automated modeling process for the purposes of this research (see section 2.3), we propose that, in practice, such methods should only be used to complement to more traditional modeling techniques. For example, in our research, we were able to significantly improve the predictiveness of models fitted using an automated process through the addition of interaction terms and spatial smoothing of residuals. The use of automated variable selection should allow the modeler more time to refine and improve the models while reducing the risk of altogether missing an important factor.

Acknowledgments

The authors would like to thank Claudine Modlin and Christopher Allard for their involvement in the research, Will Frierson and Mohamad Hindawi for implementing CART and Elastic Net respectively, and Westfield Insurance Company and Explore Information Services for providing the data upon which the research in this paper is based.

5. REFERENCES

[1.] Maitra, Saikat, and Jun Yan, "Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression," Casualty Actuarial Society *2008 Discussion Paper Program,* http://www.casact.org/pubs/dpp/dpp08/08dpp76.pdf

[2.] Sanche, Robert, and Kevin Lonergan, "Variable Reduction for Predictive Modeling with Clustering" Casualty Actuarial Society *Forum*, Winter 2006, http://casualtyactuarialsociety.net/pubs/forum/06wforum/06w93.pdf

[3.] Kolyshkina, Dr. Inna, Sylvia Wong, and Steven Lim, "Enhancing Generalised Linear Models with Data Mining," Casualty Actuarial Society *2004 Discussion Paper Program*, http://www.casact.org/pubs/dpp/dpp04/04dpp279.pdf.

[4.] Guo, Lijia, Ph.D., "Applying Data Mining Techniques in Property/Casualty Insurance", Casualty Actuarial Society *Forum,* Winter 2003, http://www.casact.org/pubs/forum/03wforum/03wf001.pdf.

[5.] Nelson, Brian D., "Variable Reduction for Modeling using PROC VARCLUS, in Statistics, Data Analysis, and Data Mining," *Proceedings of the Twenty-Sixth Annual SAS® Users Group International Conference,* 2001*,* http://www2.sas.com/proceedings/sugi26/p261-26.pdf

[6.] Anderson, Duncan, Sholom Feldblum, Claudine Modlin, Doris Schirmacher, Ernesto Schirmacher, Neeza Thandi "A Practitioner's Guide to Generalized Linear Models, Third Edition, February 2007, http://www.casact.org/pubs/dpp/dpp04/04dpp1.pdf.

[7.] Frees, Edward W., Glenn Meyers, A. David Cummings, "Predictive Modeling of Multi-Peril Homeowners Insurance," *Variance* 6:1, 2012, http://www.variancejournal.org/issues/06-01/11.pdf.

[8.] Tevet, David, "Exploring Model Lift: Is Your Model Worth Implementing?" *Actuarial Review*, Volume 40, Number 2, May 2013, http://www.casact.org/newsletter/index.cfm?fa=viewart&id=6540.

Abbreviations and Notations

The following abbreviations are used in referring to the different variable selection techniques considered:

AICRankBase: AIC Improvement Rank on Response

AICRankRes: AIC Improvement Rank on Residual

CARTBase: CART on Response

CARTRes: CART on Residuals

ENetBase: Elastic Net on Response

ENetRes: Elastic Net on Residuals

GLMCorr: Stepwise GLM based on AICC Improvement with Correlated Variables Removed

LSRCorr: Stepwise Least Squares Regression with Correlated Variables Removed

Rand: Random List

Varclus: Variable Clustering

BIOGRAPHIES OF THE AUTHORS

Benjamin Williams is a consultant in the New York office of Towers Watson. He is a specialist in predictive modeling and applications to Property and Casualty pricing. He has a master´s degree in mathematics from the Instituto Politécnico Nacional in Mexico and a master´s degree in physics from the University of Canterbury in New Zealand.

Greg Hansen leads the Actuarial Research team at Westfield Insurance Company. He is primarily responsible for developing and enhancing pricing and underwriting models for both commercial and personal lines. He has a degree in Mathematics from Oberlin College. He is a Fellow of the CAS and Member of the American Academy of Actuaries. Mr. Hansen has presented at several industry events on the topic of predictive modeling.

Aryeh Baraban works in the New York office of Towers Watson. He is involved with pricing and reserving work for both personal and commercial lines of business. He holds a bachelor's degree in Mathematics from Yale University and a master's degree from The Johns Hopkins University. He volunteers with The Actuarial Foundation.

Alessandro Santoni is the leader of the Mergers and Acquisitions for Property and Casualty in the Americas and is the Director of Towers Watson's Property and Casualty practice in New York. His main areas of expertise include the review of claims reserves for insurance and reinsurance companies, setting of rates for personal lines and capital modeling. He has worked with clients in Europe, U.S. and the Middle East.