Institute and Faculty of Actuaries

Proxy Modelling Using Machine Learning: A Case Study at Royal London

Nick Jackson, Royal London Mutual
Gaurang Mehta, Eva Actuarial and Accounting Consultants Limited

13 November 2018

Institute and Faculty of Actuaries

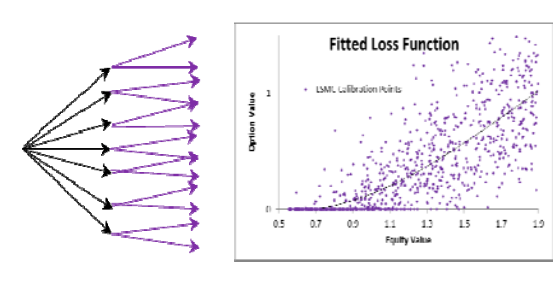
Agenda:

- Introduction and Motivations
- Background to Machine Learning ("ML") Methods
- Model Comparisons
- Lasso Regression – "Optimisation" Grid
- Initial Conclusions
- Q & A

13 November 2018

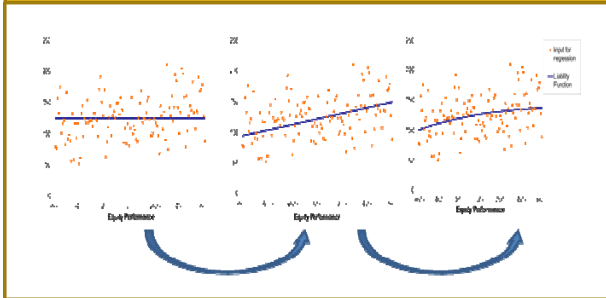
Introduction and Motivations (1)

Royal London ("RL") is developing an all-risk model using Least Square Monte Carlo ("LSMC"):



LSMC uses a very large number of outer scenarios, each with very few inner scenarios.

We currently use a "conventional" forward step-wise algorithm to perform our fit.



R-squared to identify the next most important term; Refit the model; penalty function prevents over-fitting.

Institute and Faculty of Actuaries

Introduction and Motivations (2)

Artificial Intelligence, Machine Learning and "Big Data" are concepts that are becoming increasingly prevalent and accepted throughout a wide spectrum of real-life applications.

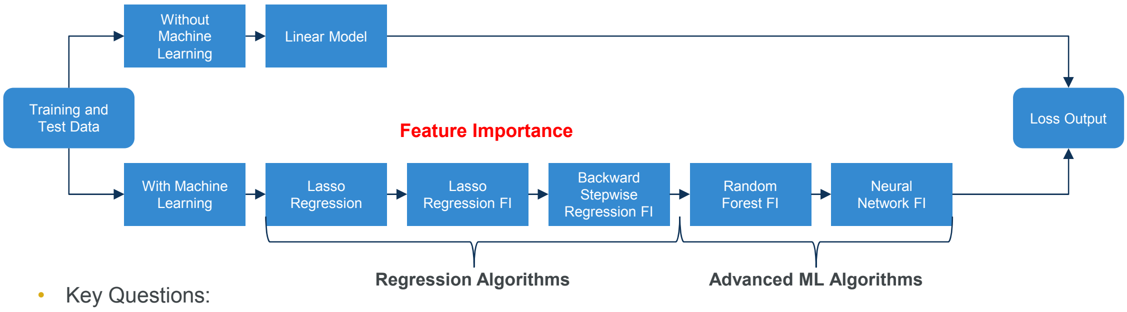
This has become possible in recent times with the significant advances in computer technology, enabling the processing of the huge datasets now available. Examples range from computers beating humans at chess and (the more complex) Go, real-time travel updates ("Google maps") and translation services, Insurance pricing, through to medical diagnoses and driverless cars.

LSMC uses very large datasets and therefore feels like an appropriate problem to which these new cutting edge tools ought to be applied. This could lead to improved fitting, reduced scenario budgets and/or a new way of validating the existing more established fitting processes.

This presentation summarises the results of a Proof-of-Concept ("POC") Machine Learning tool applied to a dataset for one of RL's larger with-profits funds. The objective is to produce an all-risk polynomial to determine the SCR and associated PDF. This initial POC focused on fitting statistics.

Institute and Faculty of Actuaries

Background to Machine Learning Methods
(a) Models Explored



- Key Questions:
  - Model Selection – Which Model to use for Proxy Model calibration?
  - Model Calibration – Under fitting / Over fitting
  - Model Optimisation – Reduction of Cash-flow Bill?
- Approach Used:
  - Max polynomial power = 3
  - Feature engineering – Use of standardized data (Features and Losses)

Institute and Faculty of Actuaries

Background to Machine Learning Methods
(b) Feature Engineering (FE) and Feature Importance (FI)

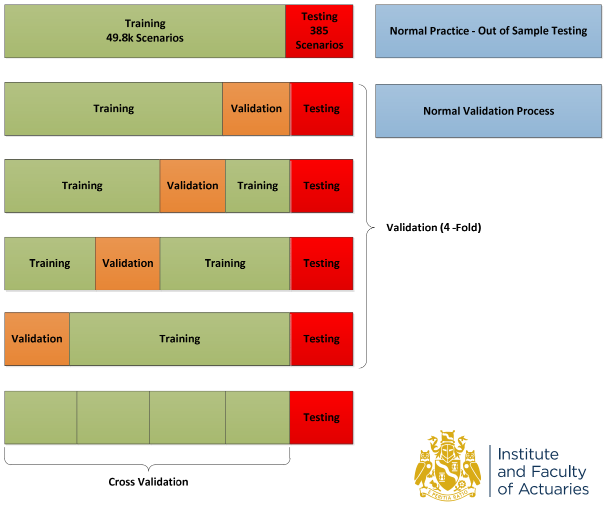
- FE - Creating new features from existing ones:
  - Standardised Data vs. Non-standardised Data
  - Introducing “domain expertise” via deciding interaction features
  - Dummy variables (e.g. Management Actions on or off)
- FI – Exclude unimportant features:
  - Is a filter and helps to mute unnecessary noise
  - Similar to well-known dimension reduction techniques such as PCA, but different
  - Makes models more parsimonious without compromising predictive accuracy
  - Improves performance

Institute and Faculty of Actuaries

Background to Machine Learning Methods

(c) Bias and Variance Trade-offs

- Validation
  - Evaluation of residuals
  - How well model fits to data
  - No indication about model fit to unknown data
- Cross Validation
  - Involves removing part of training data and used for predictions.
  - Process repeated a number of times (4 in this example)
  - Trade-off: Bias vs. Variance
  - Full training dataset used in final fit



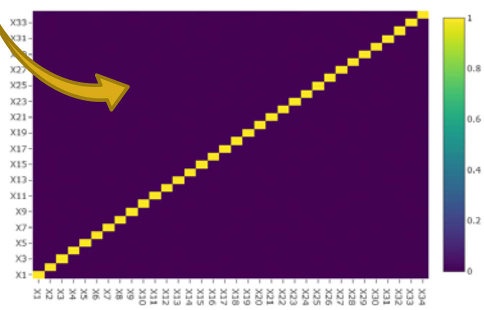
Institute and Faculty of Actuaries

Background to Machine Learning Methods

(d) Understanding Losses Dataset

- Input Features, i.e. Risk Drivers (X1, X2,..): **34**
- Training dataset, i.e. fitting points: **49.8k**
- Validation dataset, i.e. validation scenarios: **385**

- Training Data – No "Structural Multicollinearity":
  - Comfort that model is unlikely to be susceptible to small changes.
  - Increases the precision of the estimate coefficients (i.e. can rely on p-values)



index	L	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
count	49,800	49,800	49,800	49,800	49,800	49,800	49,800	49,800	49,800	49,800	49,800
mean	1.0000	- 0.0001	- 0.0000	- 0.0000	- 0.0000	- 0.0000	- 0.0000	- 0.0000	- 0.0001	0.0000	- 0.0000
std	1.0000	0.6928	0.6928	0.6928	0.6928	0.6928	0.6928	0.6928	0.6928	0.6928	0.6928
min	- 1.0077	- 1.0000	- 1.0001	- 1.0000	- 1.0000	- 1.0000	- 1.0001	- 1.0000	- 1.0000	- 1.0001	- 1.0001
25%	0.1453	- 0.5000	- 0.5000	- 0.5000	- 0.5000	- 0.5000	- 0.5000	- 0.5000	- 0.5000	- 0.5000	- 0.5000
50%	0.1994	- 0.0000	- 0.0000	- 0.0000	- 0.0000	- 0.0000	- 0.0000	- 0.0000	- 0.0000	- 0.0000	- 0.0000
75%	0.3045	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
max	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

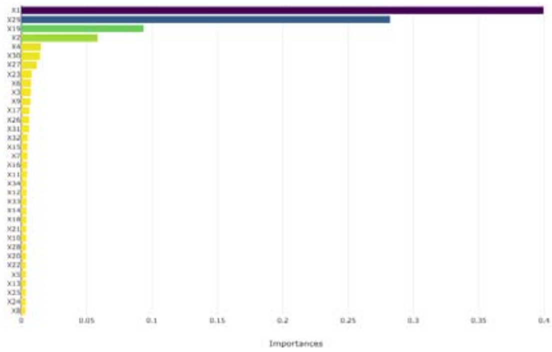
index	L	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
count	385	385	385	385	385	385	385	385	385	385	385
mean	1.0000	0.0070	0.0088	0.0120	1.0000	0.0070	0.0110	0.0158	0.0088	0.0244	0.0427
std	1.0000	0.2463	0.1639	0.1863	0.0000	0.1975	0.1671	0.1994	0.1991	0.1820	0.2397
min	0.0891	- 0.6009	- 2.3878	- 1.4239	1.0000	- 0.6430	- 0.8872	- 0.9820	- 0.9710	- 0.7461	- 0.9319
25%	0.3209	- 0.0898	-	-	1.0000	- 0.0292	-	-	-	-	-
50%	0.5349	-	-	-	1.0000	-	-	-	-	-	-
75%	0.6196	0.0072	0.1558	0.0484	1.0000	-	0.0225	0.0094	0.0330	0.0710	0.0737
max	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Institute and Faculty of Actuaries

Background to Machine Learning Methods

(e) Applying Feature Importance to Training Data

- It's a filtration step used a proposing step
- Set all features → Select the best Subset → Learning algorithm → Performance
- Independent of any ML Algorithm
- Feature importance is one of the most versatile features of ML:
  - simplification of models & shorter training times
  - avoids the "curse" of dimensionality
  - enhances generalisation by reducing overfitting
  - Reduces subjectivity in selecting cross terms

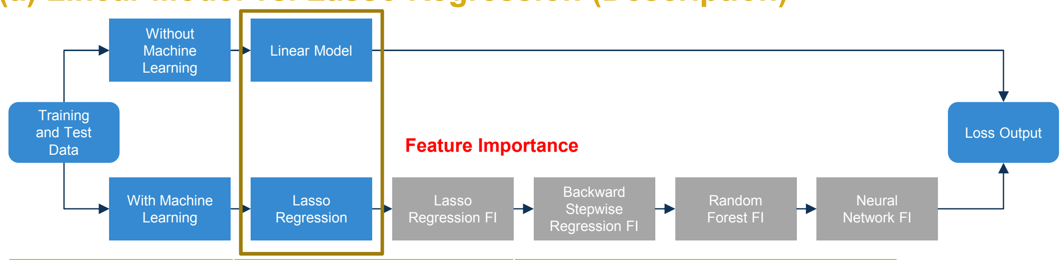


- Top 7 covers 85% → 146 terms (Cross Terms)
- Top 10 covers 90% → 309 terms (Cross Terms)
- Top 20 covers 95% → 1784 terms (Cross Terms)

Institute and Faculty of Actuaries

Model Comparisons

(a) Linear Model vs. Lasso Regression (Description)

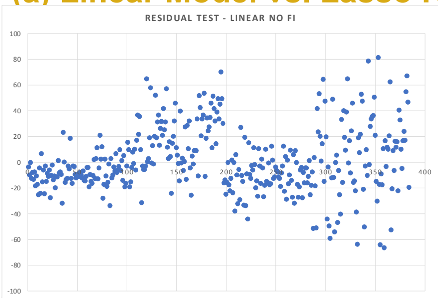


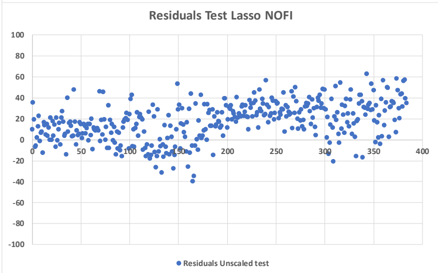
Criteria	Linear Model	Lasso Regression
RSS	$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j * x_{ij})^2$	$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j * x_{ij})^2 + \lambda\sum_{j=1}^{p}\mathrm{I}\beta_j\mathrm{I}$
Variable Selection	Yes	No
Model Interpretation	Easy	Easier
Variance	High	Low
Bias	Low	High

Institute and Faculty of Actuaries

Model Comparisons

(a) Linear Model vs. Lasso Regression (Results)





	Linear Model	Lasso Regression
Features used in fitting	34	34
Combination Terms	7769	7769

	Linear Model	Lasso Regression
R^2	95%	95%
Abs. Max Value (Predicted "True" value)	£81m	£64m
Std Deviation (Predicted "True" value)	25	18

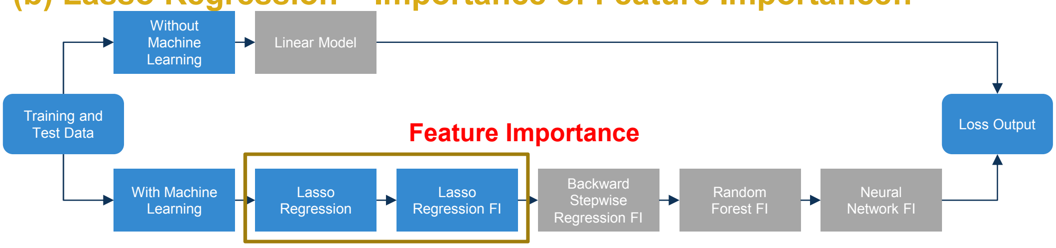
Key points:

- **Lasso performs materially better in comparison to Linear Model**
- **Reduces both Max. Absolute Error and Standard Deviation of residuals**
- **Same R^2 but materially different fitting results**

Institute and Faculty of Actuaries

Model Comparisons

(b) Lasso Regression – Importance of Feature Importance!!



	Lasso Regression	Lasso Regression with FI (10 Features)	Lasso Regression with FI (20 Features)	Lasso Regression with FI (30 Features)
Features used in fitting	34	10	20	30
Terms (excluding intercept)	7769	309	1784	5459
Total Feature Importance	100%	90%	95%	99%

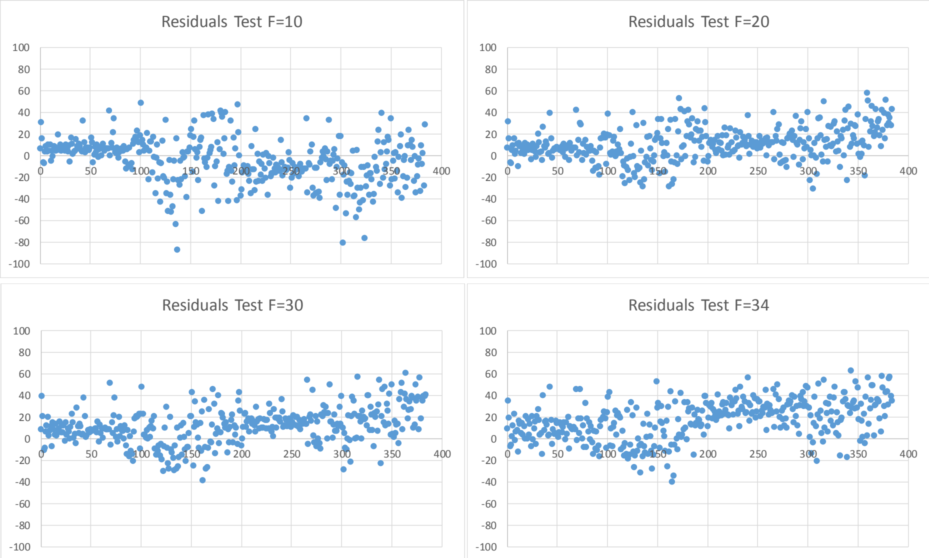
	Lasso Regression	Lasso Regression with FI (10 Features)	Lasso Regression with FI (20 Features)	Lasso Regression with FI (30 Features)
Average R^2	95%	94%	94%	95%
Abs. Max Value (Predicted "True" value)	64	87	57	61
Std Deviation (Predicted "True" value)	18	21	15	17

- **FI leads to:**
  - **more manageable model**
  - **Improvement in fit**
  - **Reduction in run time**

Institute and Faculty of Actuaries

Model Comparisons

(b) Lasso Regression – Importance of Feature Importance (Results)

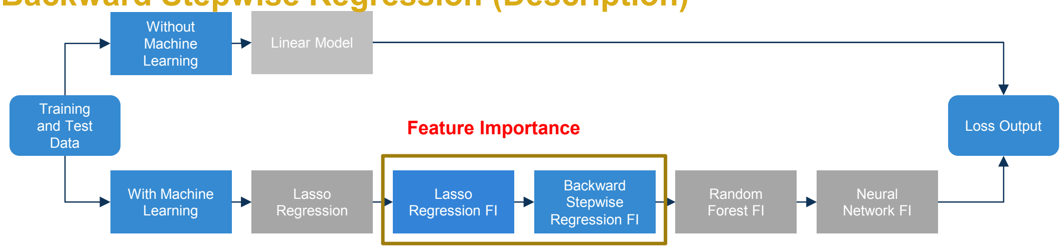


- **10 features covers 85% variation → Not enough;**
- **34 features covers 100% variation and is an improved fit;**
- **20 features covers 95% variation, leading to a further improvement still. This reflects less over-fitting;**
- **Optimum number of features is between 20 and 30.**
- **More to come once we review the remaining models…….**

Institute and Faculty of Actuaries

Model Comparisons

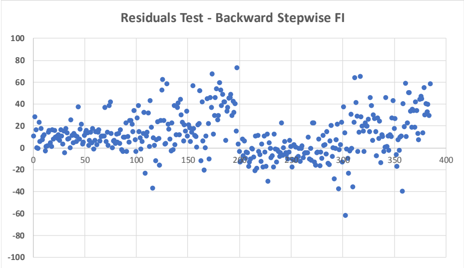
(c) Backward Stepwise Regression (Description)

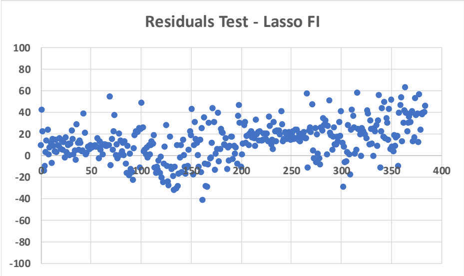


- Approach comes from the same linear model family
- Two widely used approaches – forward and backward stepwise algorithms.
- Feature selection for backward regression by removing statistically unimportant features
- Implementation:
  - Step1: Starts with full polynomial
  - Step2: Removes the statistically insignificant features (AIC, R^2, MSE, etc.)
  - Step3: Repeats step 2 iteratively
  - Step4: Stops when no further features can be removed without any statistical significance

Institute and Faculty of Actuaries

Model Comparisons
(c) Backward Stepwise vs. Lasso Algorithm (Results)





	BSM (with FI)	Lasso (with FI)
Features used in fitting	20	20
Cross Validation	4-Fold	4-Fold
Training Data	35k	35k

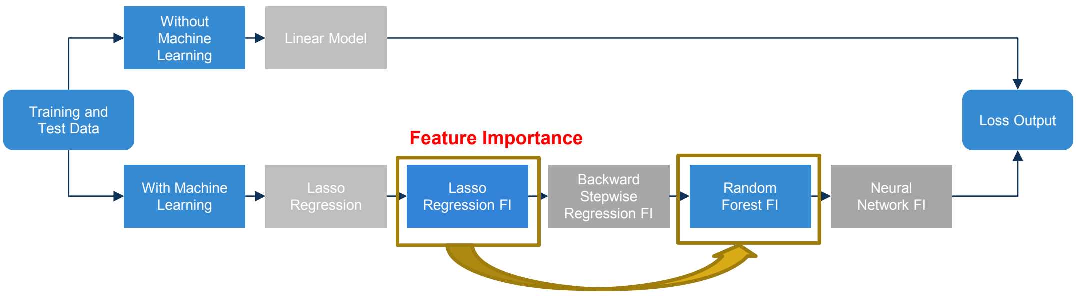
	BSM (with FI)	Lasso (with FI)
AVG. R^2	94.02%	94.26%
Abs. Max Value (Predicted value – True value)	73	58
Std Deviation (Predicted value – True value)	20	16

Key Points:

- Lasso performs better even after applying Feature Importance
- Why?

Institute and Faculty of Actuaries

Model Comparisons
(d) Random Forest Algorithm (Description)

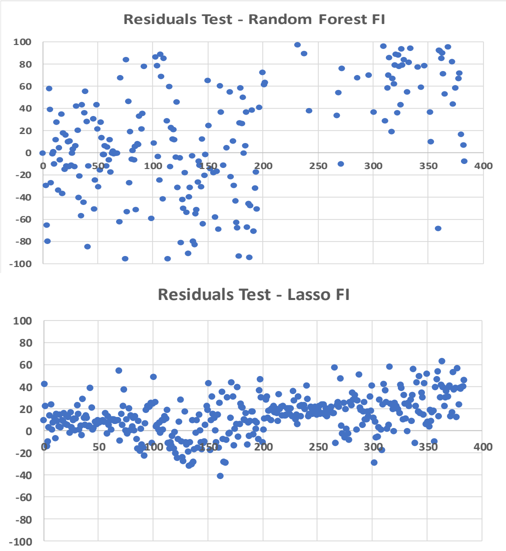


- Widely used as a “classification” algorithm
- Also used for regression purposes
- Works on averaging several noisy but unbiased models & reduces variance

Institute and Faculty of Actuaries

Model Comparisons

(d) Random Forest Algorithm vs. Lasso Algorithm (results)



	Random Forest (with FI)	Lasso Regression (with FI)
Features used in fitting	20	20
Cross Validation	4-Fold	4-Fold
Training Data	35k	35k

	Random Forest (with FI)	Lasso Regression (with FI)
AVG. R^2	85.88%	94.26%
Abs. Max Value (Predicted value – True value)	423	58
Std Deviation (Predicted value – True value)	102	16

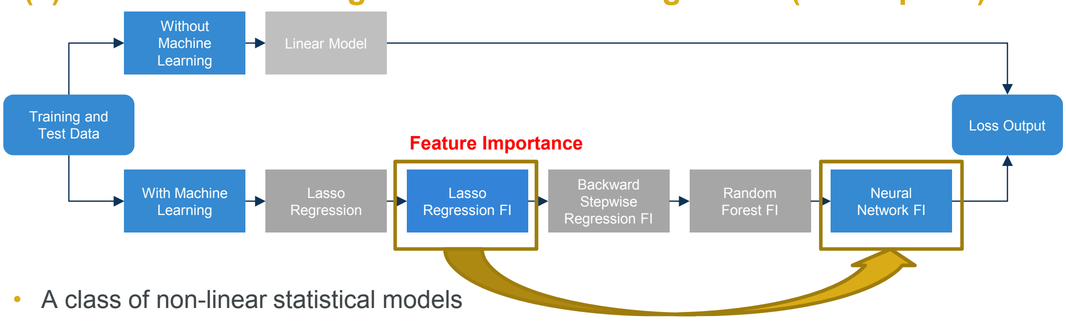
Key Points:

- Random Forest leads to increased standard deviations and Max absolute error
- Random Forest is more appropriate for classification problem then regression problems

Institute and Faculty of Actuaries

Model Comparisons

(e) Neural Network Algorithm vs. Lasso Algorithm (Description)

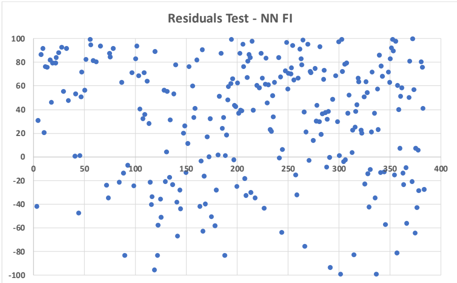


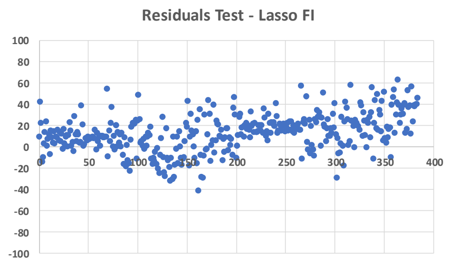
- A class of non-linear statistical models
- Impressive results for some real-life examples
  - Google search
  - Cancer research
  - Driverless cars
- Generally implemented using back propagation, where error term is distributed back up through layers by modifying weights at each node.

Institute and Faculty of Actuaries

Model Comparisons

(e) Neural Network Algorithm vs. Lasso Algorithm (Results)





	Neural Network (with FI)	Lasso Regression (with FI)
Cross Validation	4-Fold	4-Fold
Training Data	35k	35k

	Neural Network (with FI)	Lasso Regression (with FI)
AVG. R^2	94.8%	94.26%
Abs. Max Value (Predicted "True" value)	379	58
Std Deviation (Predicted "True" value)	83	16

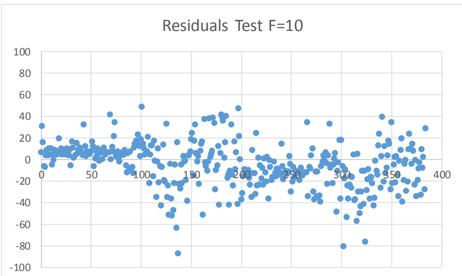
Key Points:

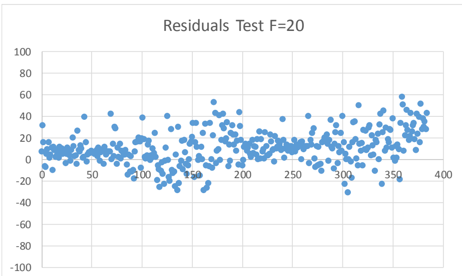
- Neural Network algorithm leads to increased standard deviations and Max absolute error in this application.
- Neural Network algorithm may require further tuning of hyper-parameters for better results.

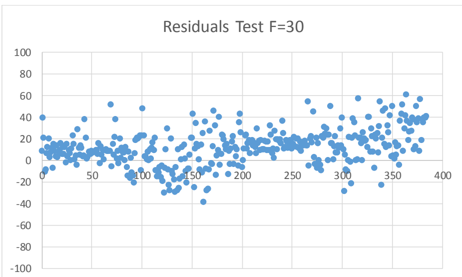
Institute and Faculty of Actuaries

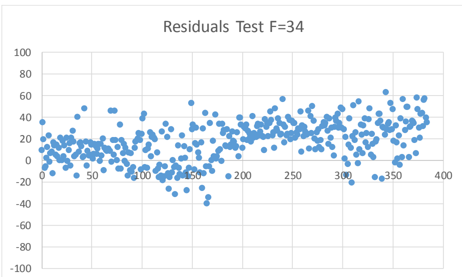
Lasso Regression: "Optimisation" Grid

Refresher: This model gives the best results of the models examined







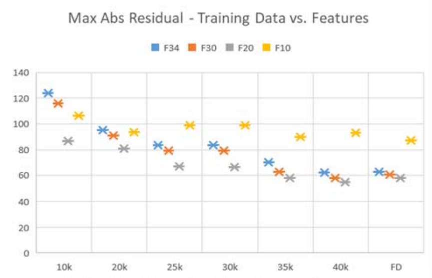


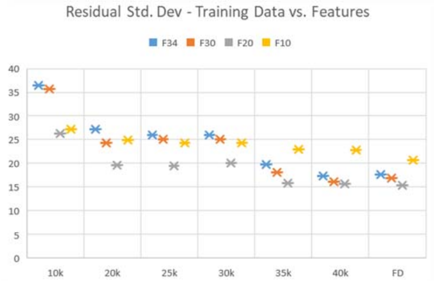
- **10 features covers 85% variation → Not enough;**
- **34 features covers 100% variation and is an improved fit;**
- **20 features covers 95% variation, leading to a further improvement still. This reflects less over-fitting;**
- **Optimum number of features is between 20 and 30.**
- ***What if we also vary the simulation budget?***

Institute and Faculty of Actuaries

Lasso Regression: “Optimisation” Grid

Number of Features vs. Size of Training Dataset





Key Points:

- Increasing number of features improves fit (up to a point)
- Increasing training data set improves fit
- Parameter tuning can reduces/optimises the cash-flow bill
- Sweet-spot here is 35k Sims and 20 Features.

Institute and Faculty of Actuaries

Initial Conclusions

(a) Technical:

- Jury is still out there - there is no single “best” approach (“Horses for Courses!”);
- Analysis of training data is equally important before selecting any approach;
- Use of feature engineering and feature importance are the two key ML techniques which reduce complexity of the existing proxy model and / or improve its accuracy;
- Consider Bias-Variance trade-off, i.e. beware of under/over-fitting; and
- Further technical investigation areas identified, e.g. Auto-encoders for Regression techniques and Stacking/Hyper-parameter optimisation under RF/NN algorithms.

(b) Business:

- Recognising methodology developments in current practice, leading to improved proxy model fits;
- Reduced LSMC simulation budget – cheaper (and quicker) results; and
- Validation of the selected proxy model fit using alternative models.

Institute and Faculty of Actuaries

Q & A

- Questions?

- For further details on ProxyML[1] Software write to gaurang.mehta@evact.co.uk

Institute and Faculty of Actuaries

1. ProxyML is a commercial proprietary software of Eva Actuarial and Accounting Consultants Limited