Analytics & Regulation Beyond GLMs
Modeler's Perspective

31 March 2014

Christopher Cooksey, FCAS, MAAA
EagleEye Analytics

EagleEye Analytics

Analytics & Regulation – Modeler's Perspective

Agenda

1. Prohibited Fields
2. Generalization Power
3. Model Validation

2

1.

Prohibited Fields

Analytics & Regulation – Modeler's Perspective

Prohibited Fields

It is the responsibility of insurance companies and those who work for them to be aware of predictors that are prohibited or limited in their use.

Has advanced analytics changed this?

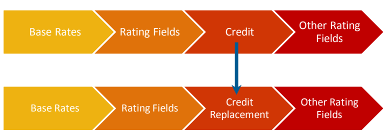
No, but now we have the possibility of "replacing" fields.

4

Analytics & Regulation – Modeler's Perspective

Prohibited Fields

What is meant by "replace"?



Similar amount of signal? Uses same underlying fields (credit report) or different fields? Similar impact on current customers? Similar impact on protected groups?

5

2.

Generalization Power

Analytics & Regulation – Modeler's Perspective

Generalization Power

Fundamental Modeling Question:

How do we know that the model is any good?

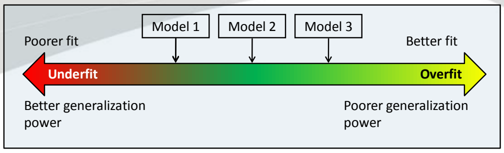
What do we mean by "good"?

How do we know that the model will apply correctly to future data that is collected?

This fundamental question speaks directly to the issue of generalization power, and it exists regardless of what modeling technique is used.

7

Analytics & Regulation – Modeler's Perspective



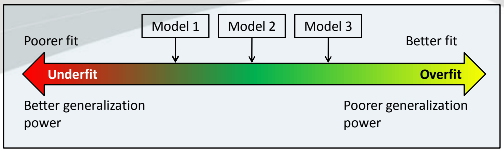
Modeling in general is a constant balancing act between under-fitting and over-fitting the data.

- "Fit" – how well does the model match the given data?
- "Generalization" – how well *will* the model fit new data?

Noise will vary between datasets, but real signal will be constant.

8

Analytics & Regulation – Modeler's Perspective



Fit versus Generalization Power –
these are NOT co-equal goals!

Underfit – Signal left on the table, but generalizes well.
Overfit – All the signal extracted (and more!), but the model under-performs on new data.

9

3.

Model Validation

Analytics & Regulation – Modeler's Perspective

Model validation uses unseen data in order to test the model. It is the most direct way to test a model's generalization power. It is the most direct way to show proof that the "model is any good".

Regulators are really the last person you have to convince…

11

Analytics & Regulation – Modeler's Perspective

Many of the metrics used to evaluate models are not tests against hold-out data.

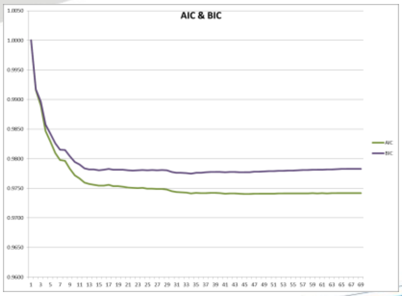
- P-values, confidence intervals, etc. are a direct consequence of an assumed error distribution.
  - *Many times only one predictor is being evaluated.*
  - *Not all modeling techniques make this assumption.*
- Whole mode metrics like AIC and BIC attempt to measure performance on unseen data, but do so indirectly through a complexity penalty.

These approaches only use training data.

12

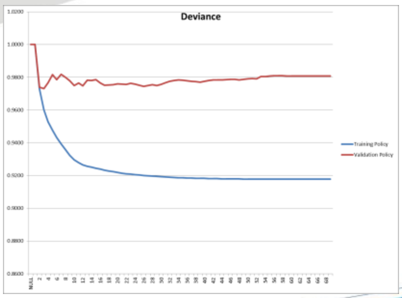
Analytics & Regulation – Modeler's Perspective

Severity – Forward Regression – AIC & BIC Improvement



Analytics & Regulation – Modeler's Perspective

Severity – Forward Regression – Deviance Improvement

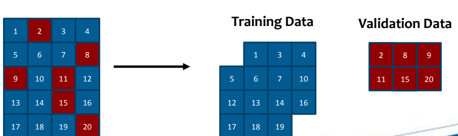


Analytics & Regulation – Modeler's Perspective

Hold-out datasets

Two main methods –

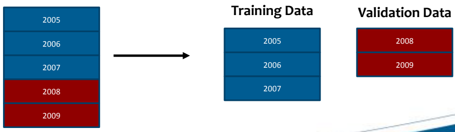
- Out of sample: randomly trained on 70% of data; validated against remaining 30% of data.



Analytics & Regulation – Modeler's Perspective

Hold-out datasets

Two main methods –
- Out of sample: randomly trained on 70% of data; validated against remaining 30% of data.
- Out of time: trained against older years of data; validated against newest years of data.



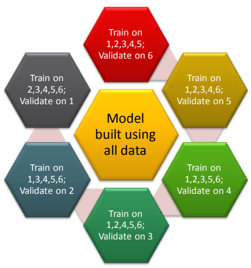
16

Analytics & Regulation – Modeler's Perspective

Cross validation

When there isn't enough volume to set data aside for testing.

Can take various forms, depending on the modeling technique.



17

Analytics & Regulation – Modeler's Perspective

The use of hold-out data suggests two powerful ideas.

First, models can and should be evaluated in their entirety.
How well does the whole model perform on unseen data?

Second, there is a distinction between the inner workings of the model and its output.
- How the model works may be proprietary
- How the model works may be complicated to explain
- Input to the model should be understood and specified to regulators
- Performance of the model on unseen data should be shared and available

18

Questions?

Contact Info
Christopher Cooksey, FCAS, MAAA
EagleEye Analytics

ccooksey@eeanalytics.com

www.eeanalytics.com