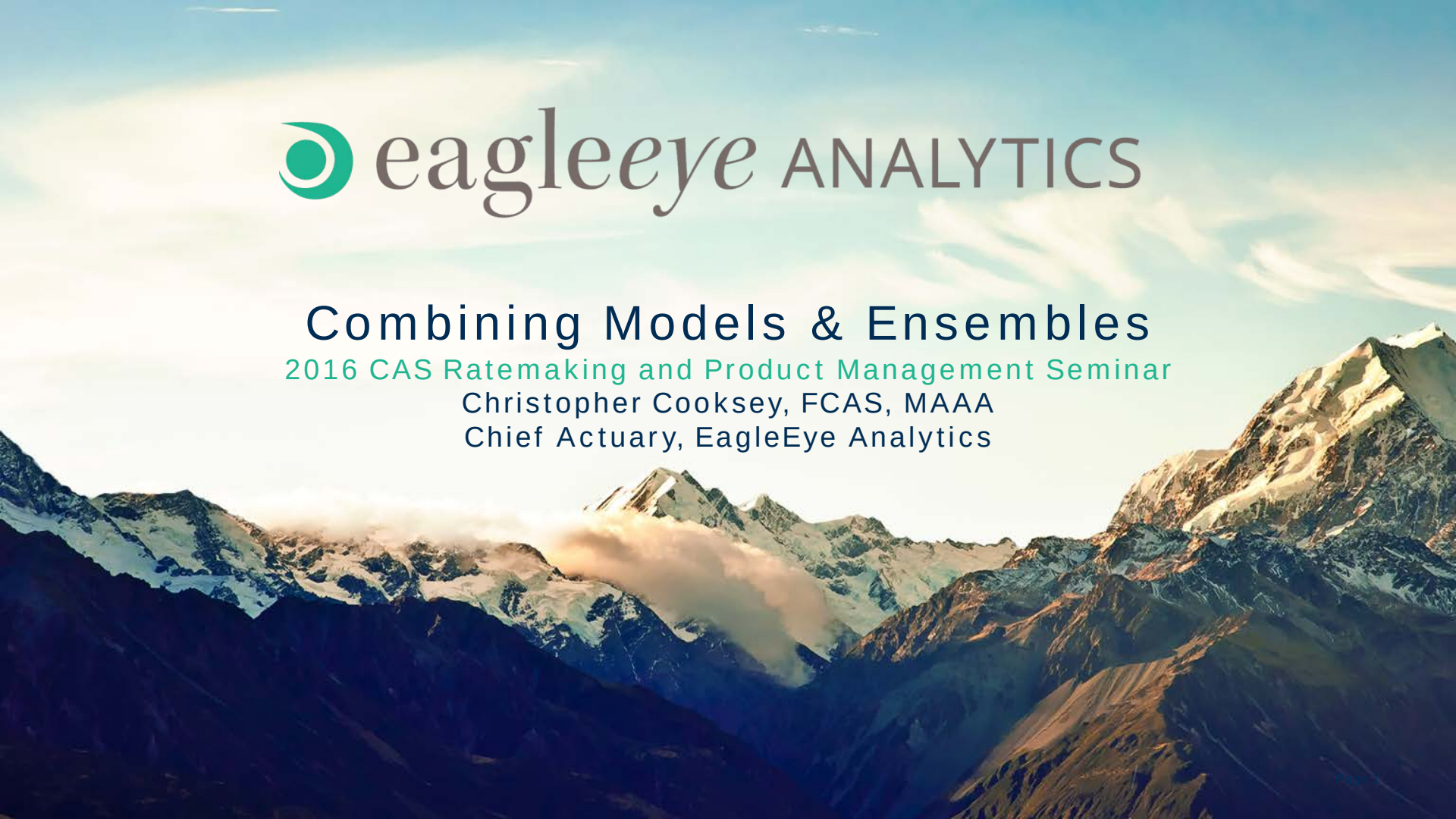


Combining Models & Ensembles

2016 CAS Ratemaking and Product Management Seminar

Christopher Cooksey, FCAS, MAAA
Chief Actuary, EagleEye Analytics

AGENDA

1. **Styles of combining models**
2. **Ensembles**
3. **Objections to ensembles**

What do people mean when they say they are building “a GLM”?

- Solving for the optimal relativities?
- Determining the best structure among the possible predictors?
- Or more generally using a GLM technique to solve a business problem?

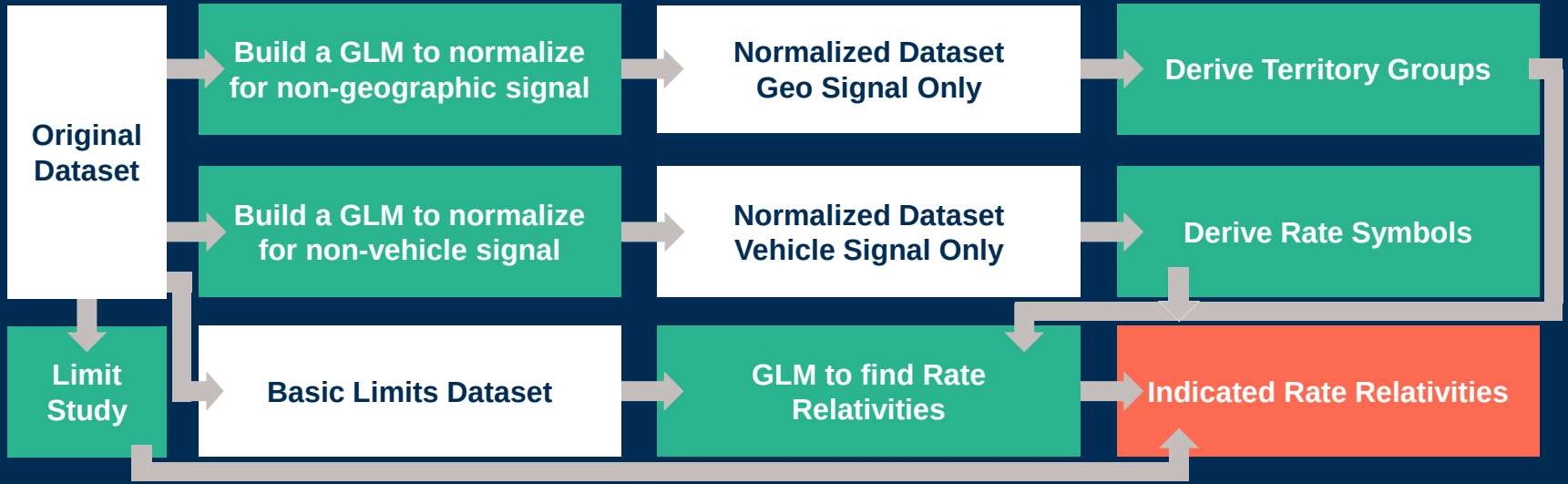
Consider using a GLM technique to build a personal lines auto rating algorithm

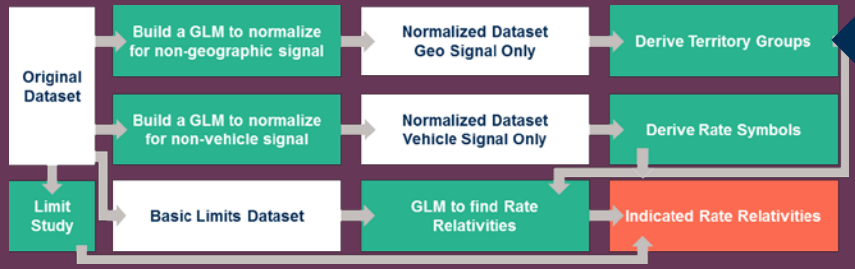
- What do you do with VIN?
- How do you capture geographic signal?
- Do you model limit directly?

How to approach the geographic portion of the signal?

- A separate study on the geographic signal is a fine idea, but doesn't the data include correlations with non-geographic predictors?
- Often we build a GLM for the sole purpose of normalizing the data for everything BUT the geographic signal. This gives us the necessary data to do a geographic study.

Building “a GLM” for an auto rating plan may look like this…





The previous auto example contains a nested model

Sometimes nested models are built to represent specific portions of the signal, but sometimes they are motivated by the need to stabilize the model.

- Loss ratio can be a difficult target; it is inherently a residual metric.
- Building models which target frequency and severity can sometimes help.
- Like territory, the output of the sub-models are used as predictors in the desired model.

Application of predictive analytics in insurance is much wider than rating

Claims Example:

Is attorney involvement predictive of claim severity?

For claims where it is too early for an attorney, would a model predicting attorney involvement improve the model predicting severity?

Here the model is trying to fill in the blank for what we know is an important piece of information.

The signal in the data can be compartmentalized

(A key concept we've been using but not discussing)

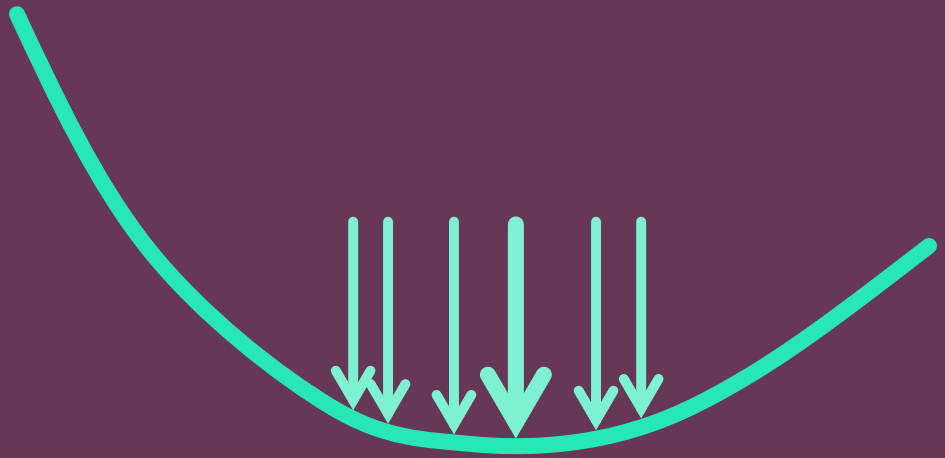
The "geographic portion" of the signal, for example:

- Keeping track of what parts of the signal (information) in the data has been represented is a critical part of modeling.
- A GLM is by its nature a linear model – how about the non-linear portion of the information in the data?

Another Twist

If you have two models, each of which perform similarly from a statistical perspective, which do you choose?

Normally we work with some function to define “best.”



Multiplicity of Models

…there is often a multitude of different descriptions [equations f(x)] in a class of functions giving about the same minimum error rate.

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, Vol. 16, No. 3.

Data will often point with almost equal emphasis on several possible models, and it is important that the statistician recognize and accept this.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*.

AN UNREALISTIC ILLUSTRATION

Ground Rules

1) **We get to know reality & compare our models directly.**
2) **Assume the numbers are frequency relativities.**
3) **Volume is limited; we can only divide the data into three equally-sized groups.**
4) **Model predictions are just the average for each defined group.**

AN UNREALISTIC ILLUSTRATION

2.026	1.948	1.801	1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000
1.948	1.873	1.732	1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000
1.873	1.801	1.665	1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000
1.801	1.732	1.601	1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000
1.732	1.665	1.539	1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000
1.665	1.601	1.480	1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000
1.601	1.539	1.423	1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	0.980
1.539	1.480	1.369	1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980
1.480	1.423	1.316	1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.423	1.369	1.265	1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.369	1.316	1.217	1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960
1.316	1.265	1.170	1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941
1.265	1.217	1.125	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922
1.217	1.170	1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904
1.170	1.125	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886
1.125	1.082	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.082	1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868
1.040	1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851
1.000	1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834
1.000	1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801
1.000	1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785
1.000	1.000	0.980	0.960	0.941	0.922	0.904	0.886	0.868	0.851	0.834	0.817	0.801	0.785	0.769

REALITY

AN UNREALISTIC ILLUSTRATION

MODEL 1

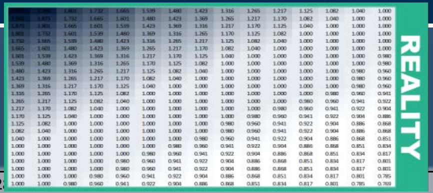
Group relatively homogeneous business together.

Sum of the squared error

= 13.48

1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944
1.260	1.260	1.260	1.260	1.260	1.059	1.059	1.059	1.059	1.059	0.944	0.944	0.944	0.944	0.944

AN UNREALISTIC ILLUSTRATION



MODEL 2

A different way of splitting the data.

Sum of the squared error

= 11.63

1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.2				
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929
0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929	0.929

AN UNREALISTIC ILLUSTRATION

ENSEMBLE
Models 1 & 2

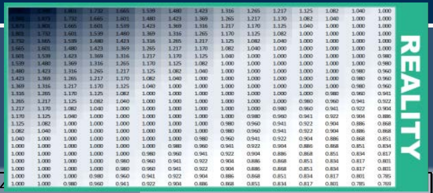
Combining information from models 1 & 2. NOT dividing the data 9 ways.

Sum of the squared error

= 9.02

1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.1[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.274	1.274	1.274	1.274	1.274	1.173	1.173	1.173	1.173	1.173	1.116	1.116	1.116	1.116	1.116
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.154	1.154	1.154	1.154	1.154	1.053	1.053	1.053	1.053	1.053	0.996	0.996	0.996	0.996	0.996
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937
1.094	1.094	1.094	1.094	1.094	0.994	0.994	0.994	0.994	0.994	0.937	0.937	0.937	0.937	0.937

AN UNREALISTIC ILLUSTRATION



ENSEMBLE
Models 1 - 5

Combining information from models 1 - 5.

Sum of the squared error

= 8.47

1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.14				
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.278	1.278	1.278	1.278	1.259	1.218	1.218	1.164	1.164	1.164	1.141	1.141	1.063	1.063	1.063
1.230	1.230	1.230	1.230	1.211	1.170	1.170	1.116	1.116	1.116	1.093	1.093	1.015	1.015	1.015
1.230	1.230	1.230	1.230	1.211	1.170	1.170	1.116	1.116	1.116	1.093	1.093	1.015	1.015	1.015
1.172	1.172	1.172	1.172	1.152	1.058	1.058	1.058	1.058	1.058	1.015	1.015	1.015	1.015	0.975
1.172	1.172	1.172	1.172	1.112	1.018	1.018	1.018	1.018	1.018	0.975	0.975	0.975	0.975	0.975
1.172	1.172	1.172	1.172	1.112	1.018	1.018	1.018	1.018	1.018	0.975	0.975	0.975	0.975	0.975
1.172	1.172	1.172	1.172	1.112	1.018	1.018	1.018	1.018	1.018	0.975	0.975	0.975	0.975	0.975
1.172	1.172	1.172	1.172	1.172	1.054	0.994	0.994	0.974	0.974	0.952	0.952	0.952	0.952	0.952
1.172	1.172	1.172	1.172	1.172	1.054	0.994	0.994	0.974	0.974	0.952	0.952	0.952	0.952	0.952
1.148	1.148	1.148	1.148	1.070	1.030	0.970	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.148	1.148	1.148	1.148	1.070	1.030	0.970	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.148	1.148	1.148	1.148	1.070	1.030	0.970	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.148	1.148	1.148	1.148	1.070	1.030	0.970	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.148	1.148	1.148	1.148	1.070	1.030	1.030	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.070	1.070	1.070	1.070	1.070	1.030	1.030	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.070	1.070	1.070	1.070	1.070	1.030	1.030	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928
1.070	1.070	1.070	1.070	1.070	1.030	1.030	0.970	0.951	0.951	0.928	0.928	0.928	0.928	0.928

AN UNREALISTIC ILLUSTRATION

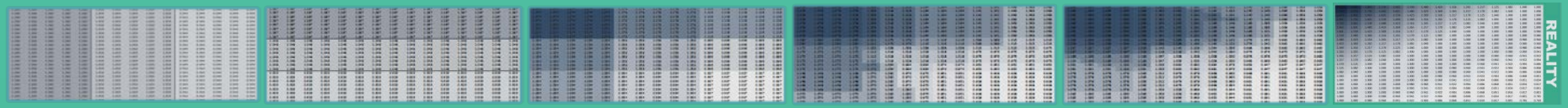
ENSEMBLE Models 1 - 9

Combining information from models 1 - 9.

Sum of the squared error = 7.35

1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.191	1.191	1.152	1.14			
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.191	1.152	1.140	1.140	1.096	1.053	1.053
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.232	1.152	1.140	1.140	1.096	1.053	1.053
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.232	1.193	1.140	1.140	1.096	1.053	1.053
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.196	1.148	1.136	1.095	1.009	1.009	1.009
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.196	1.148	1.136	1.095	1.009	1.009	1.009
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.196	1.148	1.136	1.101	1.009	1.009	1.009
1.296	1.296	1.296	1.296	1.285	1.262	1.262	1.232	1.196	1.148	1.136	1.101	1.009	1.009	1.009
1.269	1.269	1.269	1.269	1.258	1.236	1.236	1.169	1.169	1.122	1.076	1.042	0.992	0.992	0.982
1.269	1.269	1.269	1.269	1.258	1.236	1.236	1.169	1.130	1.122	1.076	1.042	0.992	0.992	0.982
1.237	1.237	1.237	1.237	1.226	1.140	1.140	1.103	1.064	1.056	0.998	0.998	0.992	0.992	0.960
1.237	1.237	1.237	1.237	1.203	1.118	1.118	1.042	1.042	1.034	0.976	0.976	0.970	0.970	0.960
1.166	1.166	1.166	1.166	1.133	1.047	1.008	1.008	1.008	1.000	0.976	0.976	0.970	0.970	0.960
1.124	1.166	1.166	1.166	1.133	1.047	1.005	1.005	1.008	1.000	0.966	0.966	0.960	0.960	0.960
1.091	1.091	1.133	1.133	1.133	1.026	0.992	0.992	0.984	0.976	0.953	0.947	0.947	0.947	0.947
1.091	1.091	1.091	1.133	1.091	1.026	0.992	0.992	0.981	0.976	0.953	0.947	0.947	0.947	0.947
1.078	1.078	1.078	1.078	1.035	1.012	0.979	0.979	0.968	0.960	0.938	0.931	0.931	0.933	0.933
1.078	1.078	1.078	1.078	1.035	1.012	0.979	0.979	0.968	0.960	0.931	0.931	0.931	0.933	0.933
1.078	1.078	1.078	1.078	1.035	1.012	0.979	0.979	0.968	0.960	0.931	0.931	0.931	0.933	0.933
1.078	1.078	1.078	1.078	1.035	1.012	0.979	0.979	0.968	0.954	0.931	0.931	0.931	0.933	0.933
1.078	1.078	1.078	1.078	1.035	1.003	1.003	0.969	0.952	0.944	0.931	0.931	0.931	0.933	0.933
1.035	1.035	1.035	1.035	1.035	1.003	1.003	0.963	0.952	0.944	0.931	0.931	0.931	0.933	0.933
1.035	1.035	1.035	1.035	1.035	1.003	0.996	0.963	0.952	0.944	0.931	0.931	0.931	0.933	0.933
1.035	1.035	1.035	1.035	1.035	0.996	0.996	0.963	0.952	0.944	0.931	0.931	0.931	0.933	0.933

AN UNREALISTIC ILLUSTRATION



"Ensembles remain robust even as they become increasingly complex.

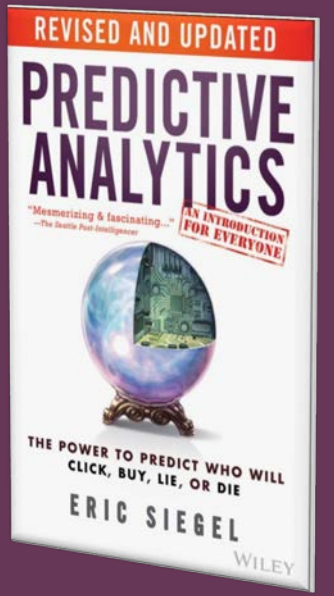
They seem to be immune to this limitation, as if soaked in a magic potion against overlearning."

Siegel, E. (2013). *Predictive Analytics.*

“Ensemble modeling has taken the [Predictive Analytics] industry by storm.

It’s often considered the most important predictive modeling advancement of this century’s first decade.”

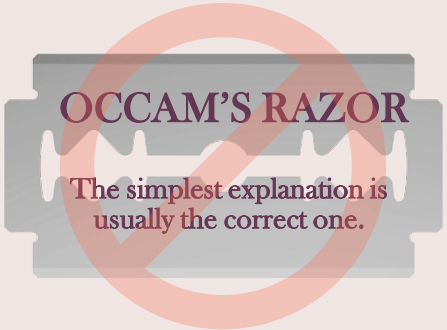
Siegel, E. (2013). *Predictive Analytics.*



OBJECTIONS TO ENSEMBLES

Resistance usually centers around complexity

Simpler is preferred in the absence of certainty, when multiple models perform equally well.



But if an ensemble performs better, then it is simply the better model.

OBJECTIONS TO ENSEMBLES

Framing the question as the choice between accuracy and interpretability is an incorrect interpretation of what the goal of a statistical analysis is.

The point of a model is to get useful information about the relation between the response and predictor variables. Interpretability is a way of getting information.

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, Vol. 16, No. 3.

OBJECTIONS TO ENSEMBLES

(Consider neural nets vs. trees)

All machine learning techniques are equally difficult to explain.

Departments of insurance won't accept them.

Because it can't be explained in simple terms, there is no opportunity for insight.

Anything that is theoretically possible will be achieved in practice, no matter what the technical difficulties are, if it is desired greatly enough.

~ Arthur C Clarke ~

OBJECTIONS TO ENSEMBLES

Don’t think a complex model will be accepted for pricing in your underwriting-driven culture?

Context & Needs for Predictive Analytics in Insurance

- ✓ **Underwriting**
- ✓ **Marketing**
- ✓ **Claims management**
- ✓ **Internal monitoring**

TAKEAWAYS

1) Be aware of your language. Not everyone knows what you mean when you say you are going to “build a GLM”. Don’t short-sell the effort.

2) Combining models is typical. Expect it in any project are doing.

3) Keep track of the signal (information) in your data. Consider what you’ve done to represent relevant portions of it.

4) If your approach is linear, are there ways to capture the non-linear parts of your signal?

5) Results from ensemble approaches are transforming other industries and are worth the effort for insurance predictive modelers to explore.

6) The difficulties around model complexity are real, but can be addressed.

eagleeye ANALYTICS | Talon®

Chris Cooksey

Chief Actuary

ccooksey@EEAnalytics.com
855.757.8500

EEAnalytics.com