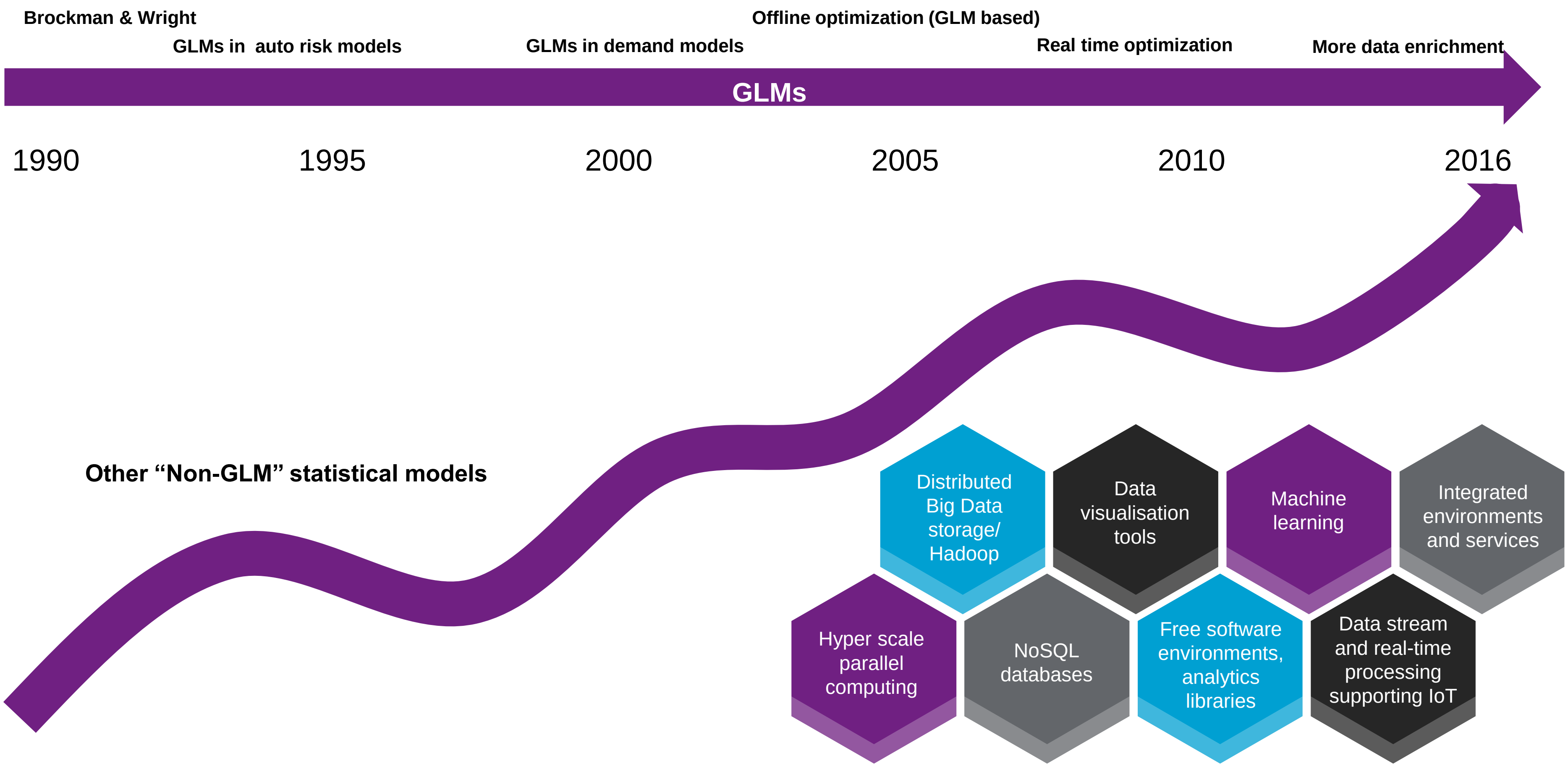
Southwest Actuarial Forum

Exploring Advanced Analytics Solutions in Pricing

December 2016

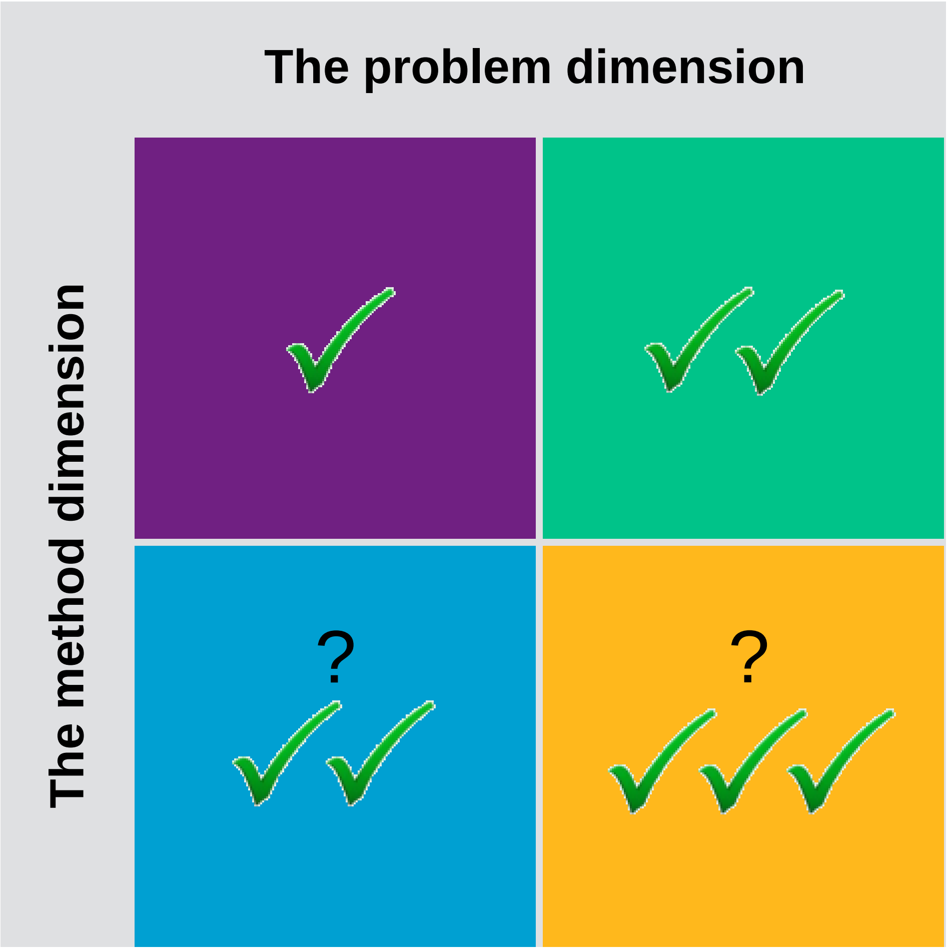

Brockman & Wright
GLMs in auto risk models
GLMs in demand models
Offline optimization (GLM based)
Real time optimization
More data enrichment
GLMs
1990
1995
2000
2005
2010
2016
Other “Non-GLM” statistical models
Distributed Big Data storage/ Hadoop
Data visualisation tools
Machine learning
Integrated environments and services
Hyper scale parallel computing
NoSQL databases
Free software environments, analytics libraries
Data stream and real-time processing supporting IoT

A 2 x 2

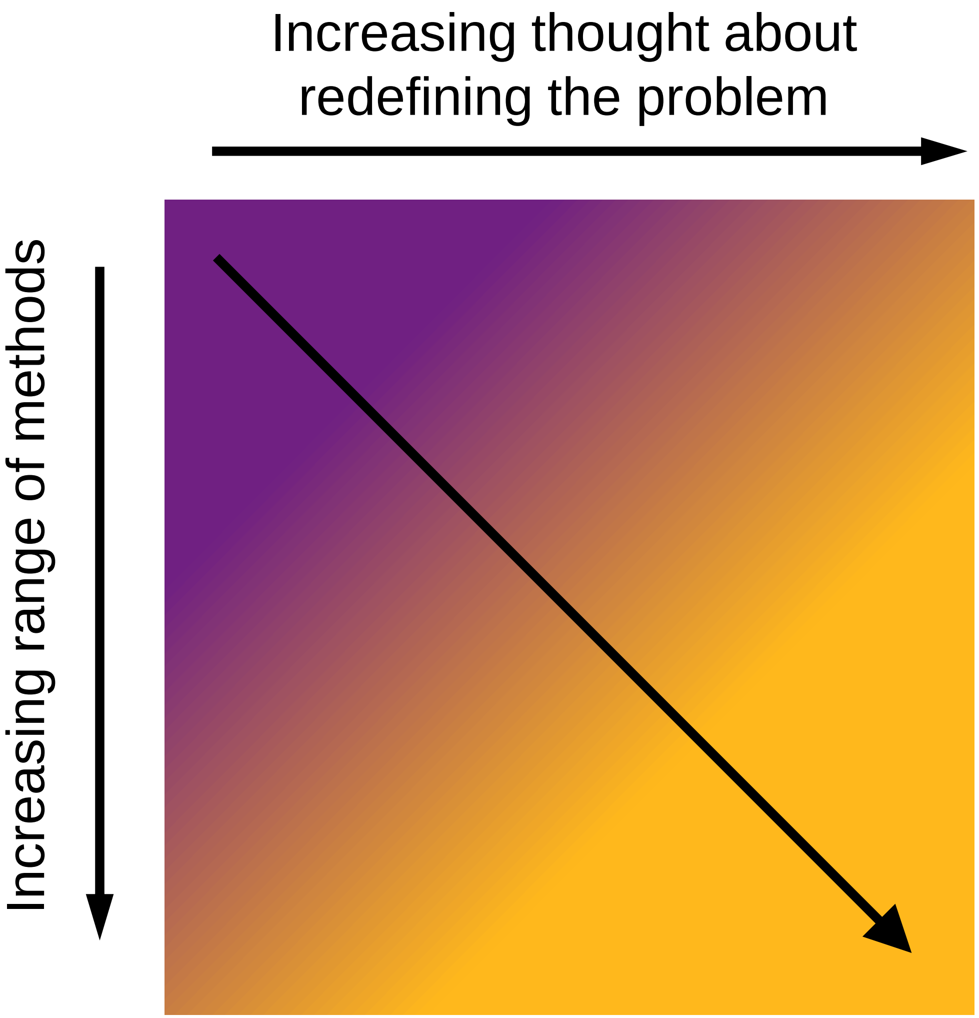
The method dimension	The problem dimension	
	Original problem, solved with original methods	Newly defined problem, solved with original methods
	Original problem, solved with new techniques	Newly defined problem, solved with new techniques

A 2 x 2

Where might the value be?



It's not really a 2 x 2, more a continuous spectrum



What are these other methods?

Ensemble Methods

Classifications Trees

Random Survival Forests

Regression Trees

Gradient Boosting Machines

K-nearest Neighbors

Elastic Net

Neural Networks

Naïve Bayes

Random Forests

K-Means Clustering

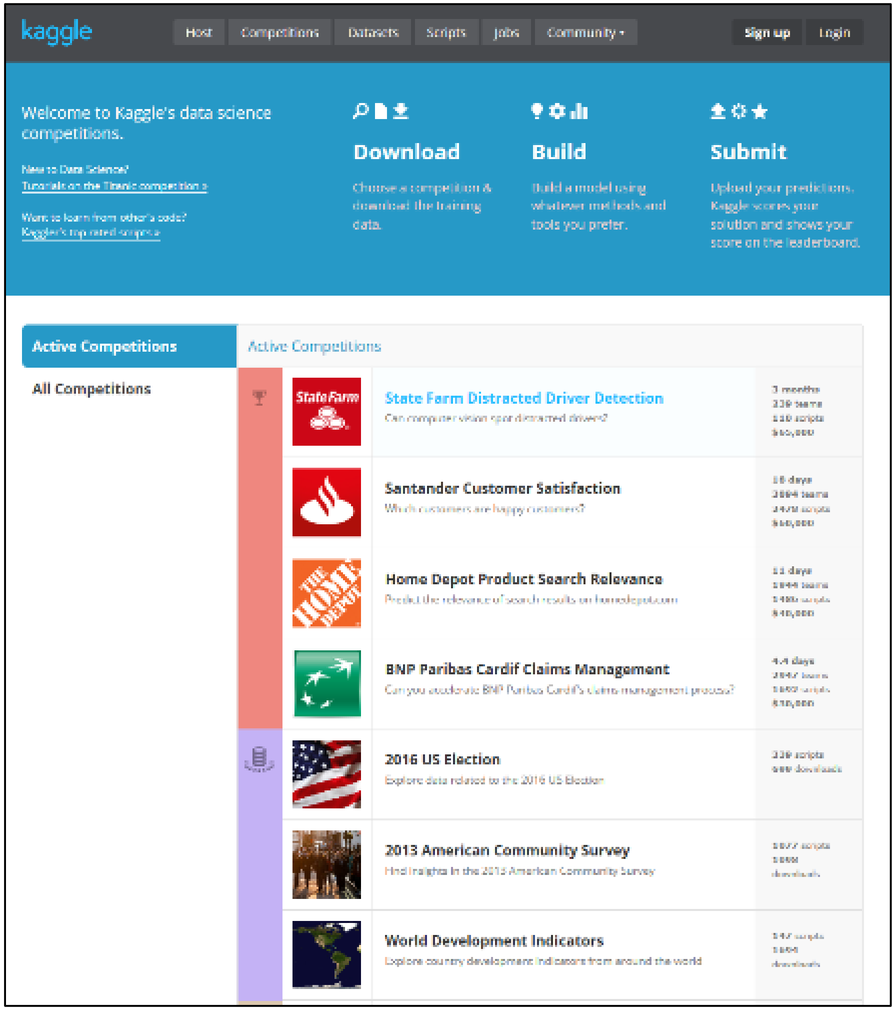
Principal Components Analysis

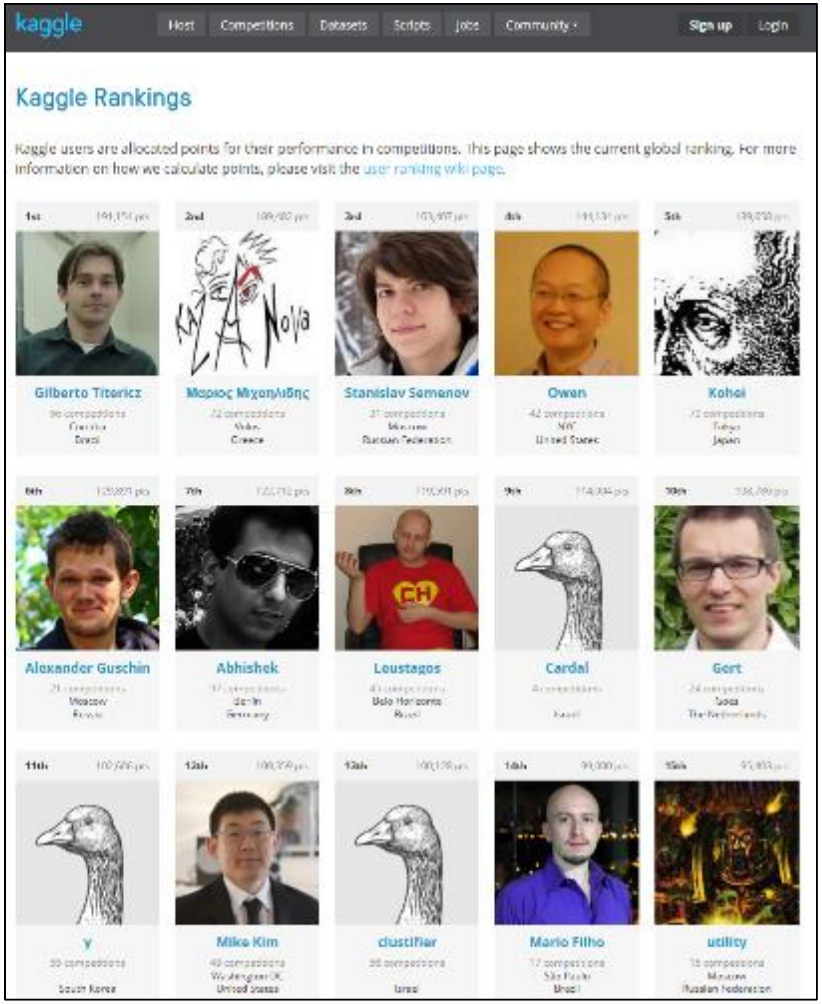
Lasso

Support Vector Machines

Ridge Regression

Kaggle


kaggle
Host
Competitions
Datasets
Scripts
Jobs
Community
Sign up
Login
Welcome to Kaggle's data science competitions.
Download
Choose a competition & download the training data.
Build
Build a model using whatever methods and tools you prefer.
Submit
Upload your predictions. Kaggle scores your solution and shows your score on the leaderboard.
Active Competitions
All Competitions
State Farm Distracted Driver Detection
Santander Customer Satisfaction
Which customers are happy customers?
Home Depot Product Search Relevance
BNP Paribas Cardif Claims Management
2016 US Election
2013 American Community Survey
World Development Indicators


kaggle
Kaggle Rankings
Kaggle users are allocated points for their performance in competitions. This page shows the current global ranking. For more information on how we calculate points, please visit the user ranking wiki page.
Gilberto Titericz
Stanislav Semenov
Owen
Kohei
Alexander Guschin
Abhishek
Leustagos
Cardal
Gert
Mike Kim
clustifier
Mario Filho
utility

Kaggle winning methods (January 2015 to February 2016)

- **Gradient Boosted Machines** was most successful technique across the board
- **Feature Creation/Selection** was noted as biggest contributor to success
  - The nature of Kaggle and the sharing of benchmarks means most competitors use the same algorithms – thus the key differentiator is the improvement gained from good feature creation/selection

Count of method placing "top 3" in competition (for which data was available)

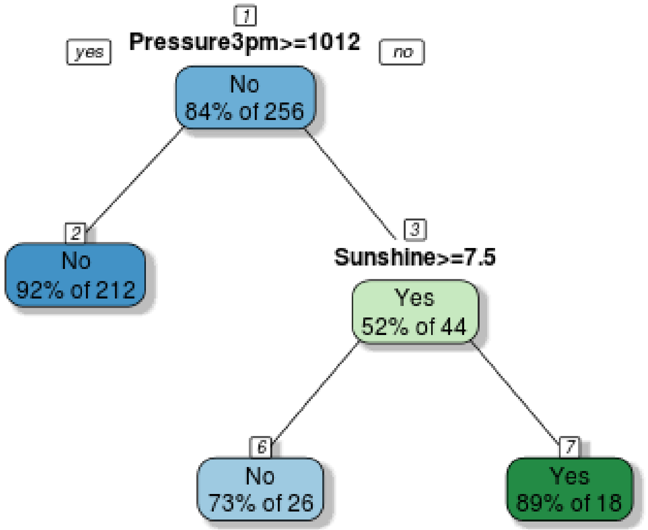
Competition subject	Support Vector Machine	Gradient Boosted Machine	Neural Network	Mixed Method Ensemble	Random Forest	Total
All	1	19	10	10	1	41
Insurance	**-**	**3**	**-**	**4**	**-**	**7**

An overview of Gradient Boosted Machines

- **Boosting** is where models are successively trained on the **residuals** of the previous model
- At each iteration, the model is updated by adding only a fraction (λ) of the new model
- Each iteration performed on a random sample of data points to reduce over-fitting to the training data
- The overall prediction is given by

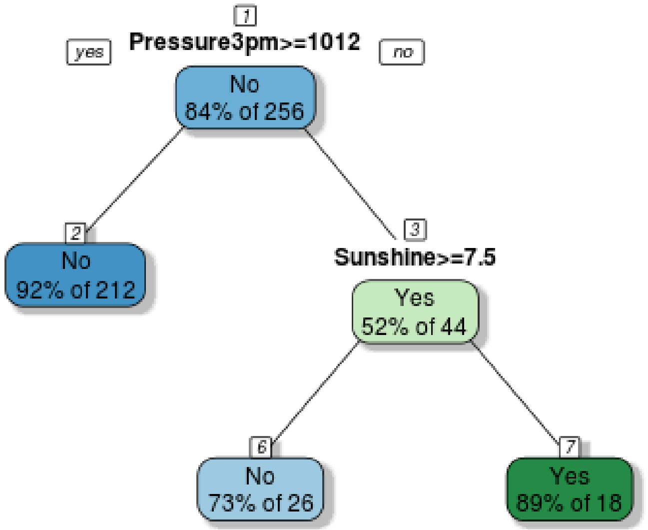
$$f(x) = \lambda \sum_{n=1}^{N} f_n(x)$$

- Base models are usually Decision Trees, but could use other model forms (eg GLMs)
- There are numerous parameters to decide (including λ and the number of trees – this is done via cross validation)



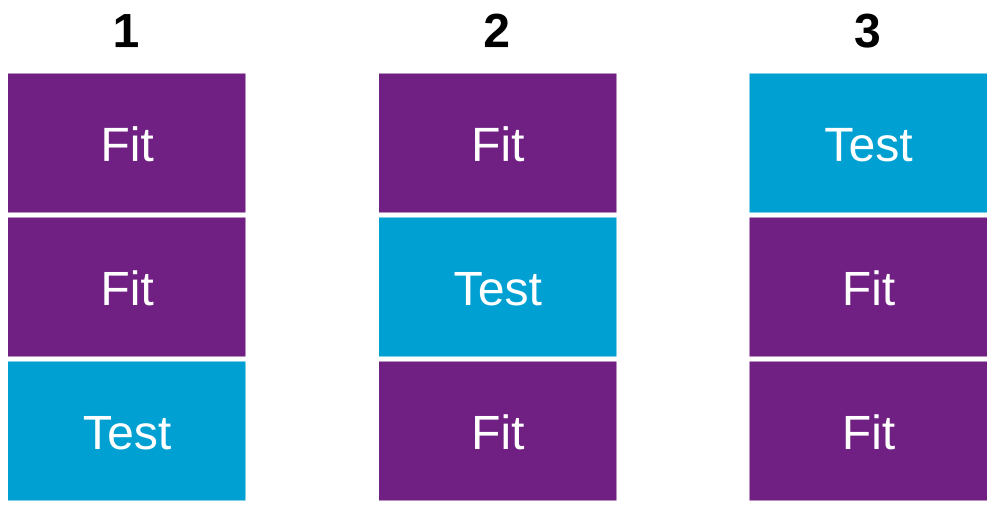
Four main assumptions

- **Learning rate / “shrinkage”**
  - Amount by which the old model predictions are varied for the next model iteration
  - New model =
    Old + (Prediction x Learning rate)
- **Interaction depth**
  - Number of splits allowed on each tree (or the number of terminal nodes – 1)
- **Number of trees** (iterations) allowed
- **Bag fraction**
  - Trees are fitted to a subset of the data (the bag fraction) on a randomized basis
  - Additional noise-reduction can be achieved by using a random subset of the available factors at each iteration



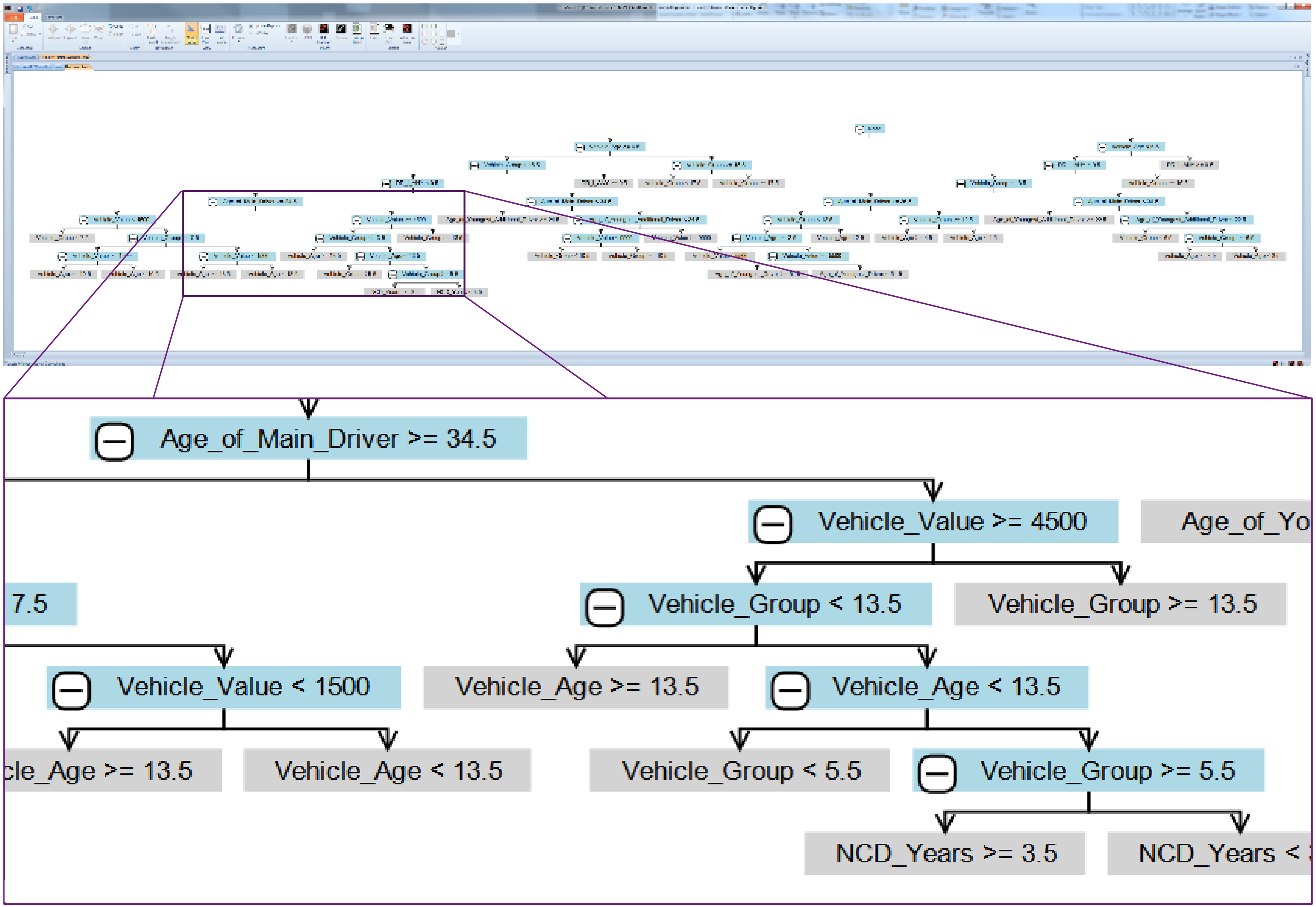
Calibrating the assumptions

- n-fold cross validation used to develop the interaction depth and learning rate assumptions
  - Eg for 3-fold validation, split into 3, fit on purple, test on blue parts, take average

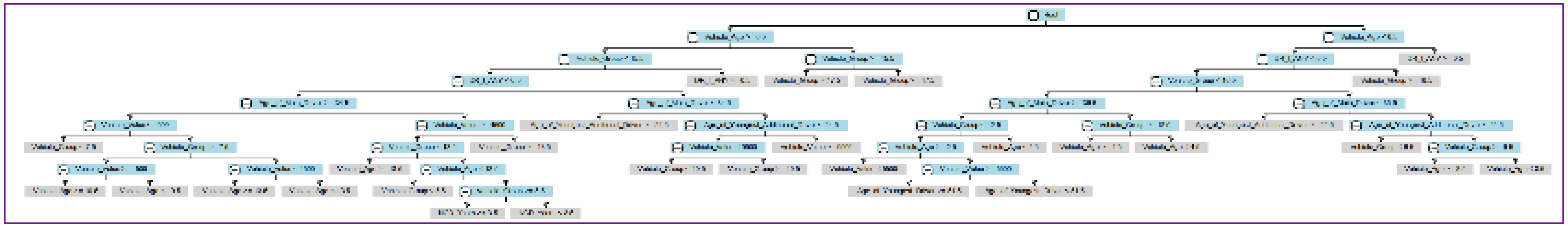


- Resulting plots can be used to determine the optimal assumption choice
  - Including how many trees to run

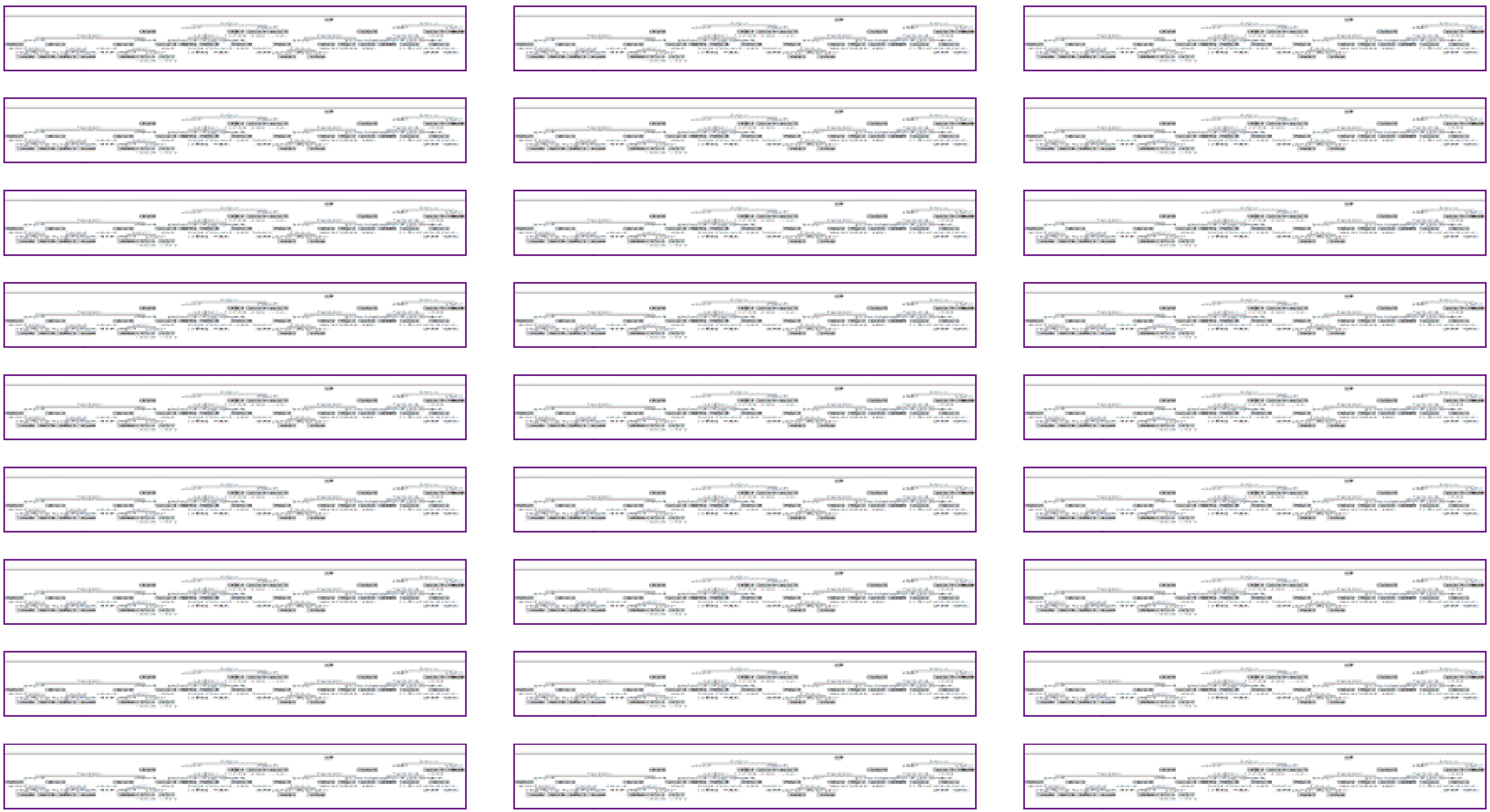
What does the result look like?



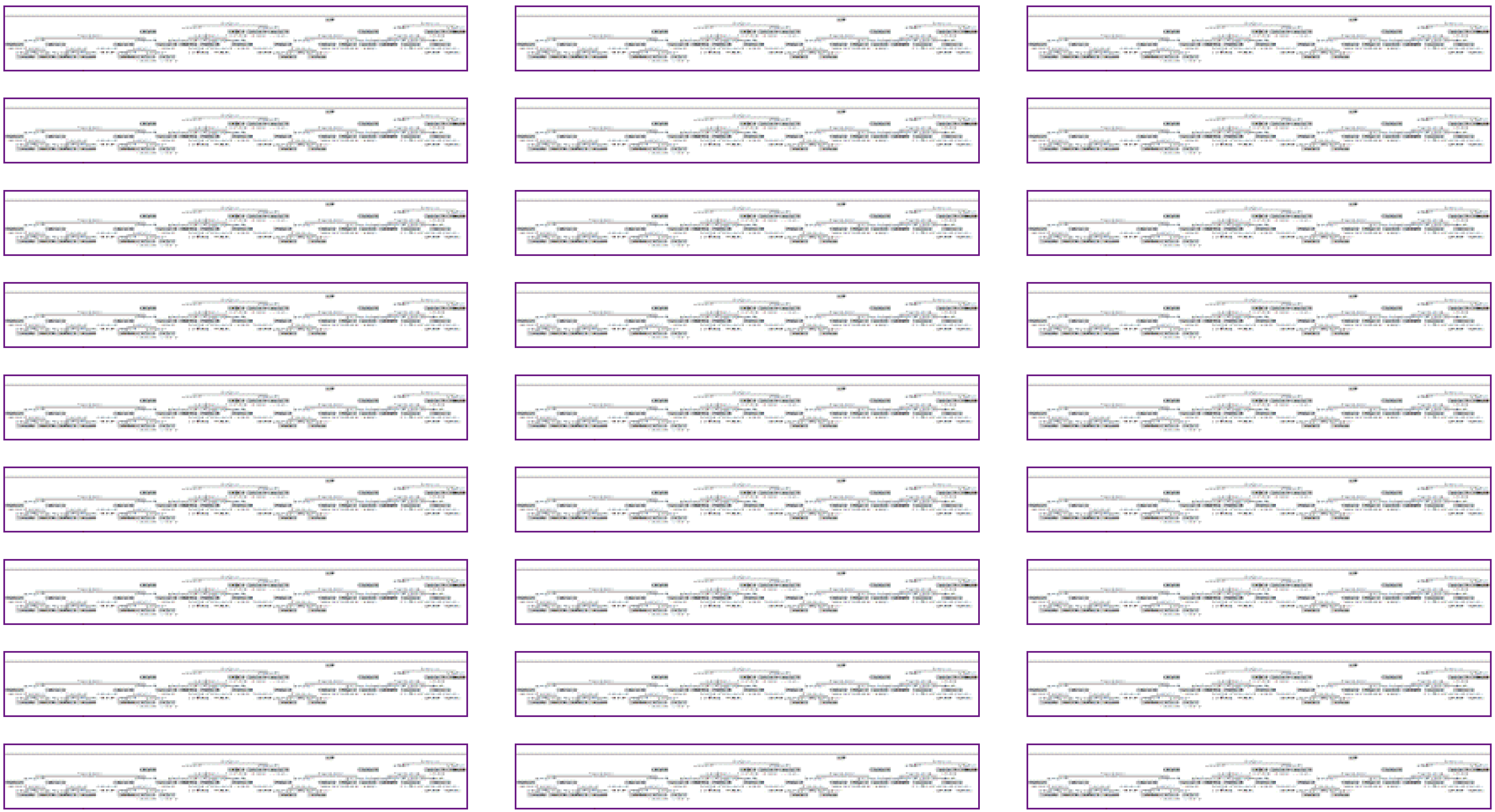
What does the result look like?

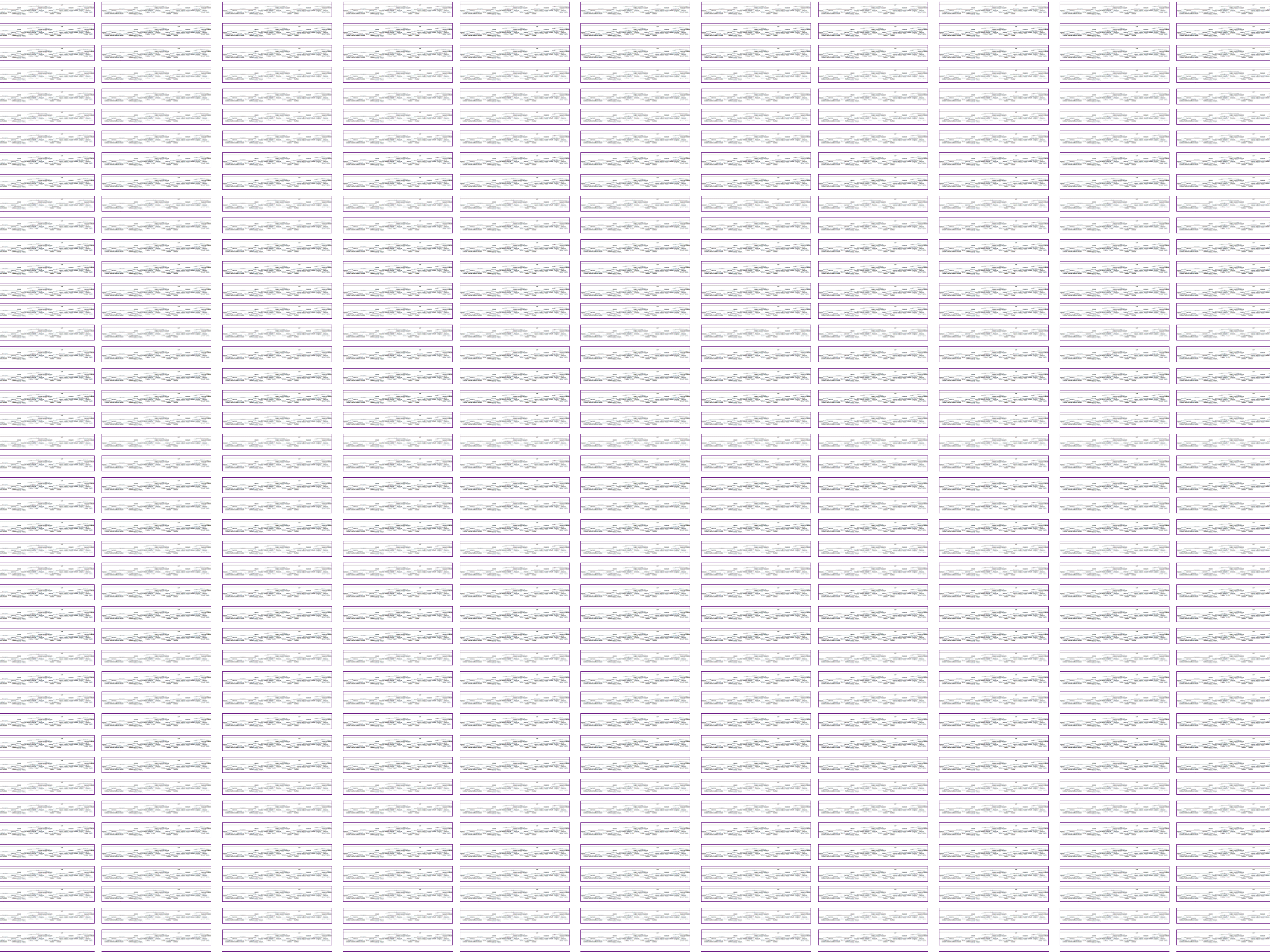


What does the result look like?



What does the result look like?

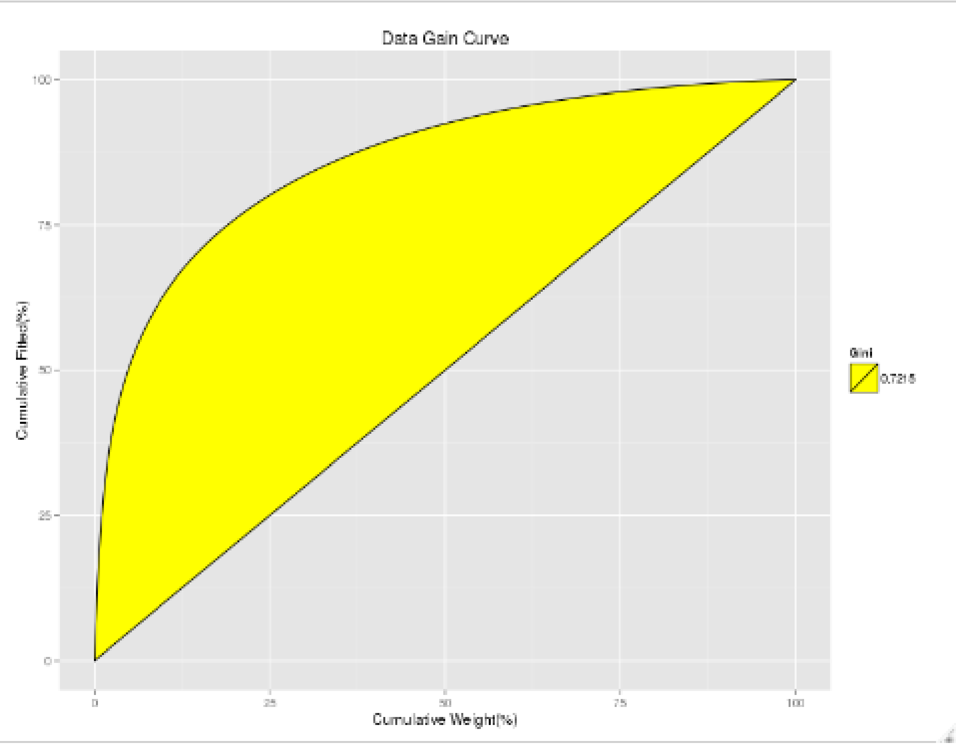




Three (and a half) interesting questions

1. Does the model add value?
2. What does the model mean?
   - Do we even need to know?
3. How can we use the model?

Gini curves



- Rank hold out observations by their **fitted values** (high to low)
- **Plot cumulative response** by cumulative exposure
- A **better model** will explain a **higher proportion of the response** with a **lower proportion of exposure**
- …and will give a **higher Gini coefficient** (yellow area)

Example results

Model	Gini
GLM	0.327

Example results

Model	Gini
GLM	0.327
GBM	0.332

Example results

Model	Gini	Gini improvement
GLM	0.327	0.0%
GBM	0.332	1.7%

Example results

Model	Gini	Gini improvement	Gini rank
GLM	0.327	0.0%	2
GBM	0.332	1.7%	1

Example results

Model	Gini	Gini improvement	Gini rank
GLM (main factor removed)	0.318	-2.6%	4
GLM (minor factor removed)	0.322	-1.3%	3
GLM	0.327	0.0%	2
GBM	0.332	1.7%	1

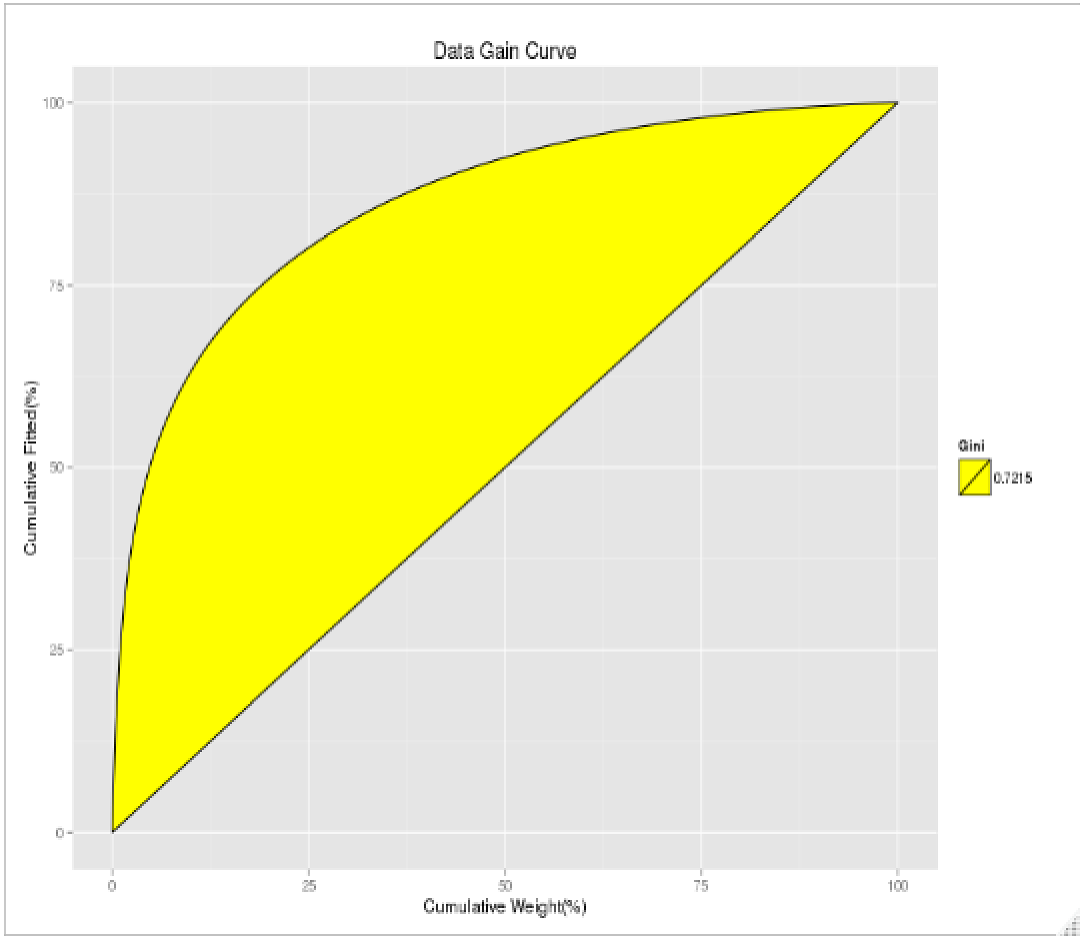
But…

- Think of a model…
- Multiply it by 123
- Square it
- Add 74½ billion

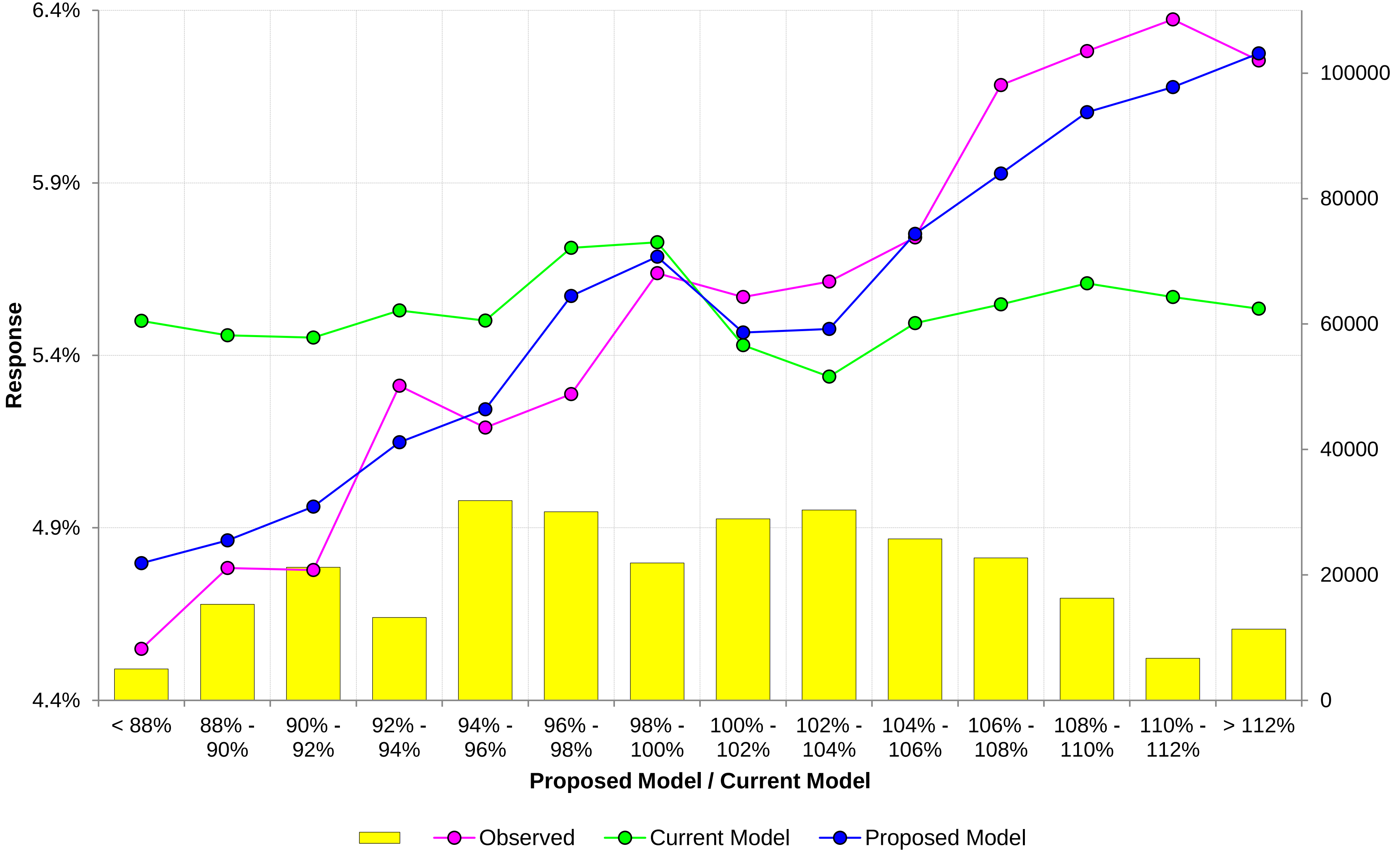
But…

- Think of a model…
- Multiply it by 123
- Square it
- Add 74½ billion

- …and you get the same Gini coefficient!

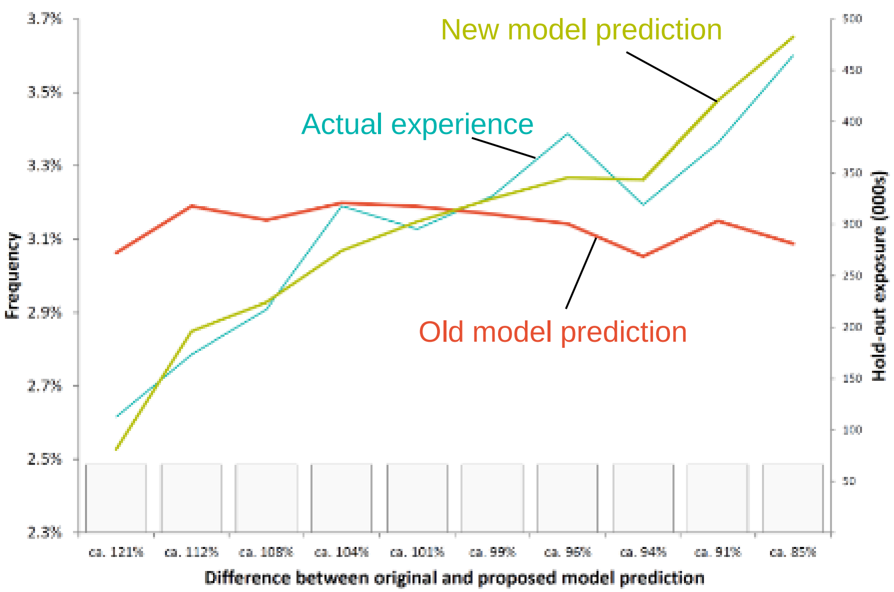


Double lift chart



Financial value estimate

- Errors in insurance pricing are not symmetrical
- Simplified model can approximate this and provide a robust sense-check on value created given
  - an assumed elasticity
  - an assumed cap/floor approach



Simple formula

Old/New	New premium	Expected volume	Actual claims	Increased profit
121%	P_1	V_1	C_1	X_1
…	P_2	V_2	C_2	X_2
…	…	…	…	…
…	P_{99}	V_{99}	C_{99}	X_{99}
85%	P_{100}	V_{100}	C_{100}	X_{100}
			Value created	$ X

Financial value estimate

Model	Gini	Gini improvement	Gini rank	Loss ratio @ elasticity 6	Loss ratio rank	Loss ratio @ elasticity 2	Loss ratio rank
GLM (main factor removed)	0.318	-2.6%	4	-0.9%	4	-0.4%	4
GLM (minor factor removed)	0.322	-1.3%	3	-0.4%	3	-0.2%	3
GLM	0.327	0.0%	2	0.0%	2	0.0%	2
GBM	0.332	1.7%	1	2.8%	1	0.6%	1

Financial value estimate

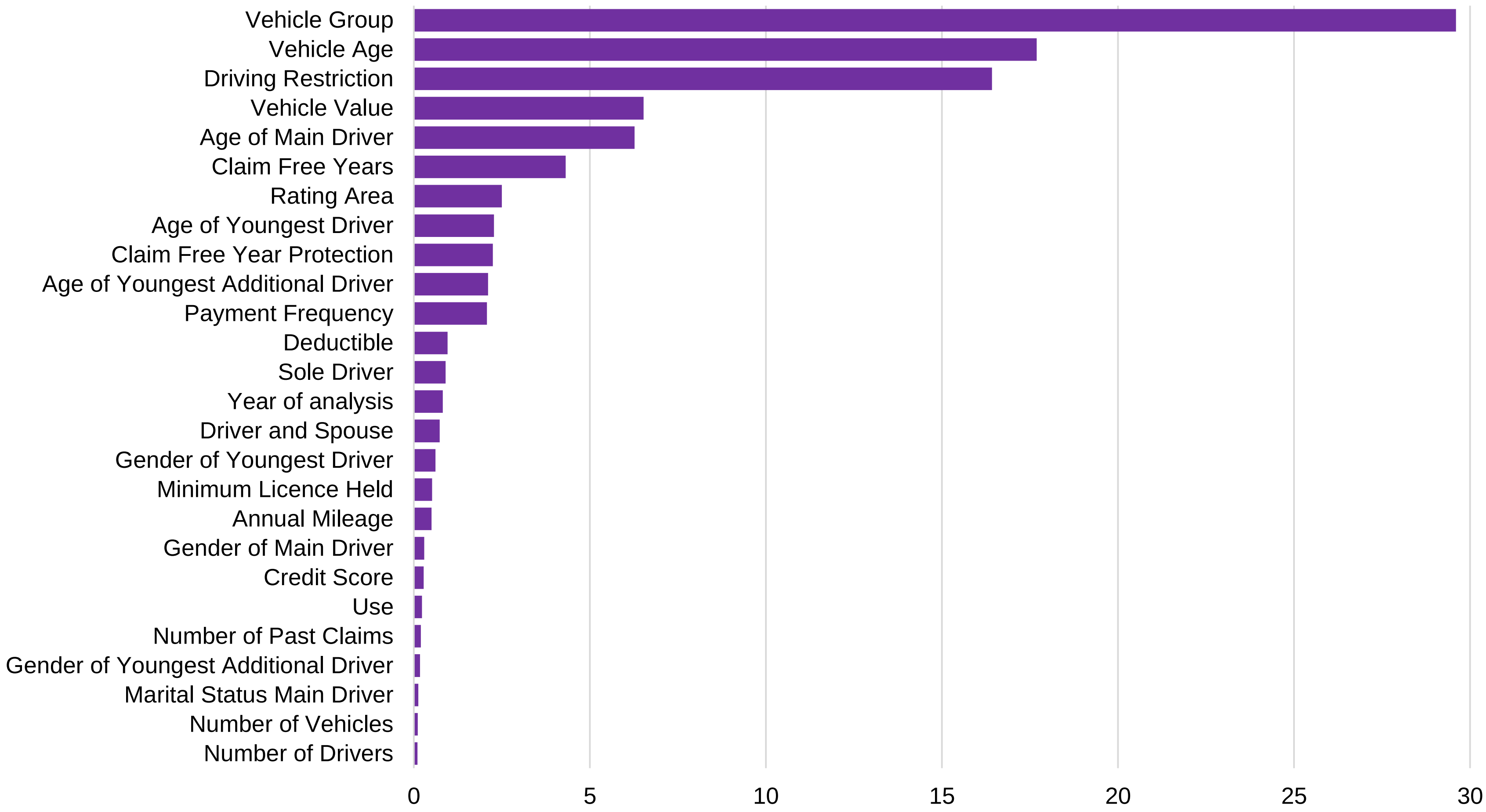
Model	Gini	Gini improvement	Gini rank	Loss ratio @ elasticity 6	Loss ratio rank	Loss ratio @ elasticity 2	Loss ratio rank
GLM (main factor removed)	0.318	-2.6%	5	-0.9%	5	-0.4%	5
GLM (minor factor removed)	0.322	-1.3%	4	-0.4%	4	-0.2%	4
GLM	0.327	0.0%	3	0.0%	3	0.0%	3
GBM	0.332	1.7%	2	2.8%	1	0.6%	2
Ensemble of GBM & GLM	0.338	3.4%	1	2.7%	2	0.7%	1

Three (and a half) interesting questions

1. Does the model add value?
2. What does the model mean?
   - Do we even need to know?
3. How can we use the model?

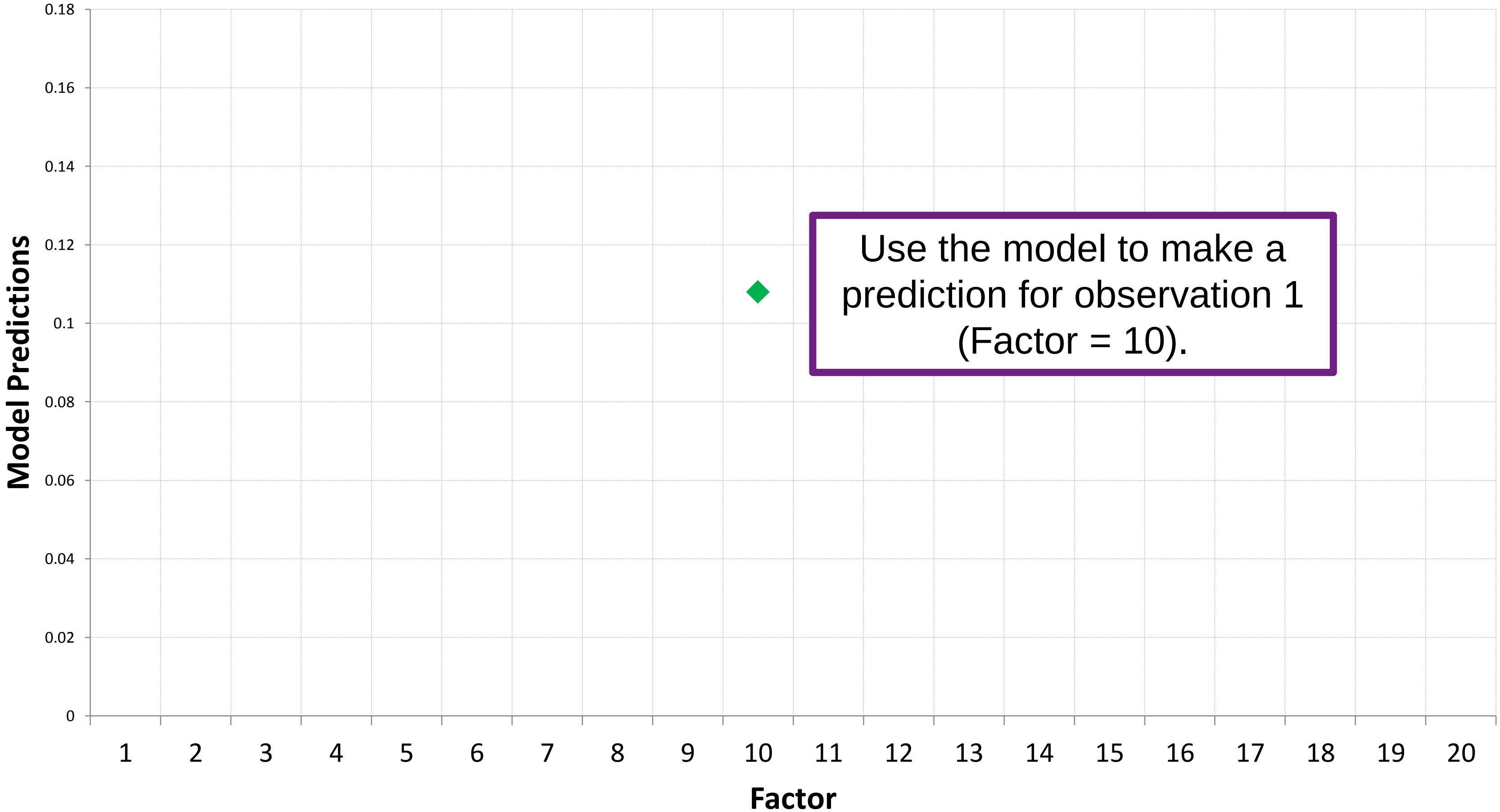
Factor importance – relative influence

The relative influence of a factor can be measured as the total reduction in error attributable to splits by that factor, across all trees in the GBM



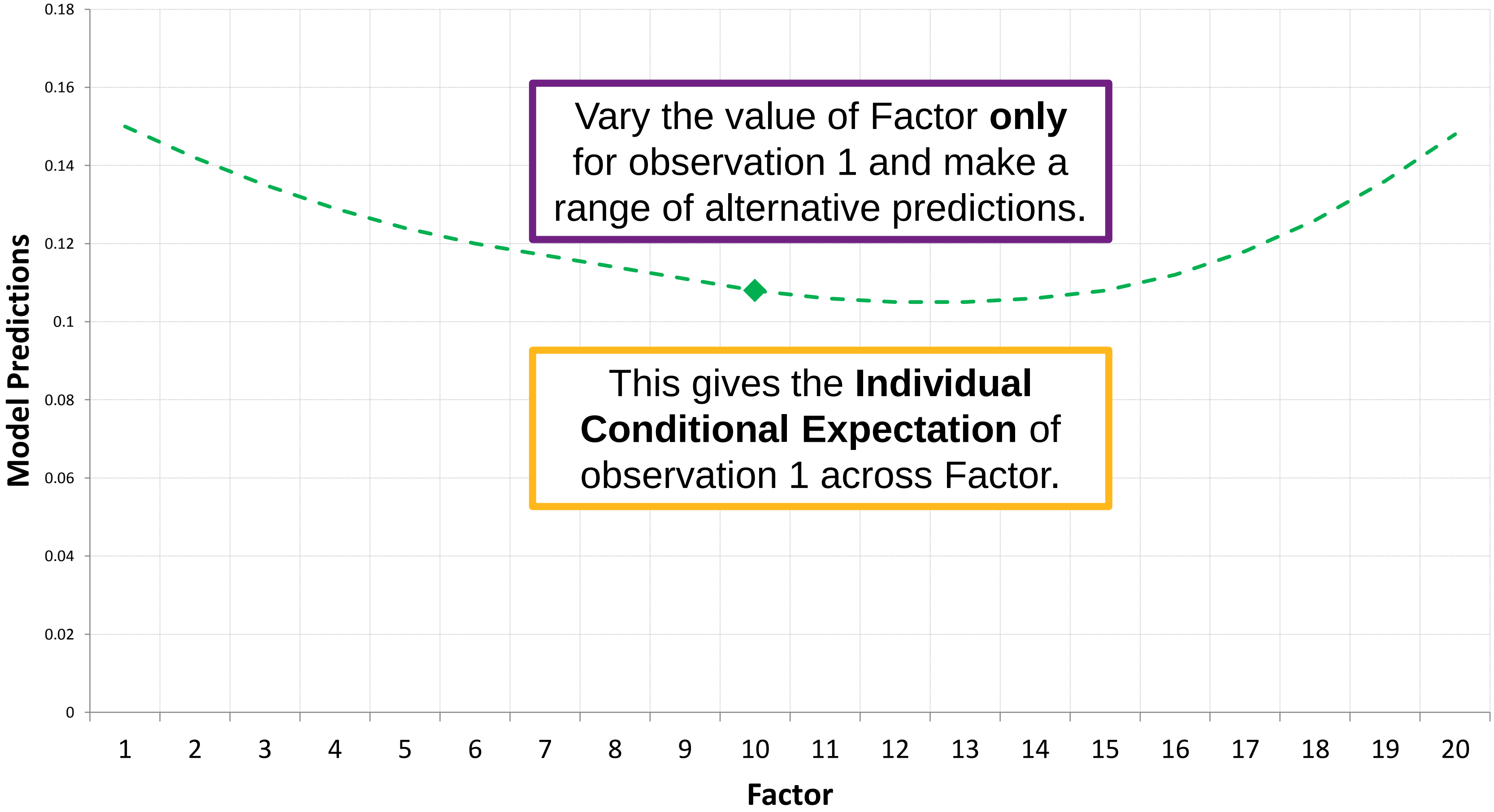
Partial dependency plots

Example



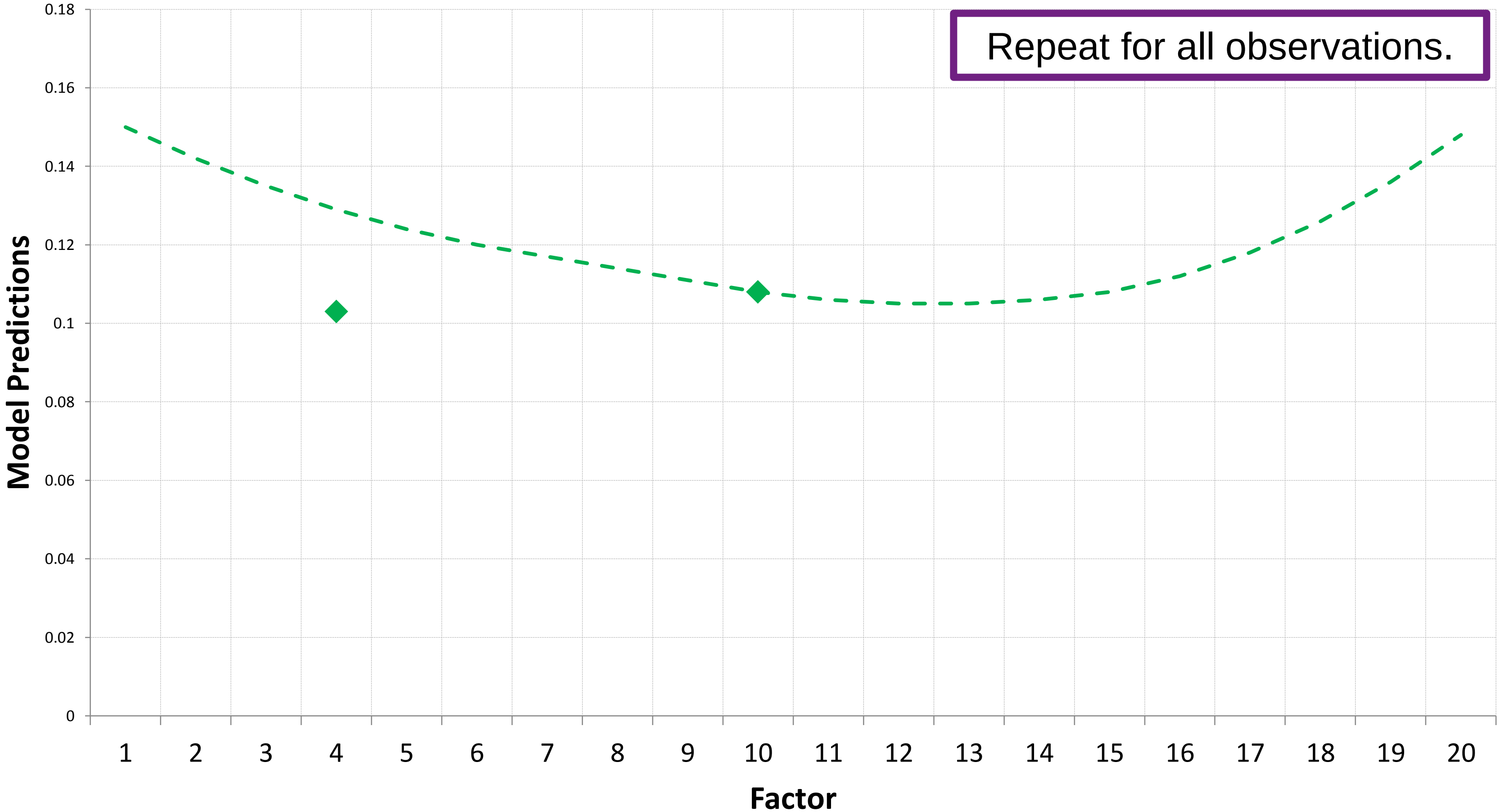
Partial dependency plots

Example



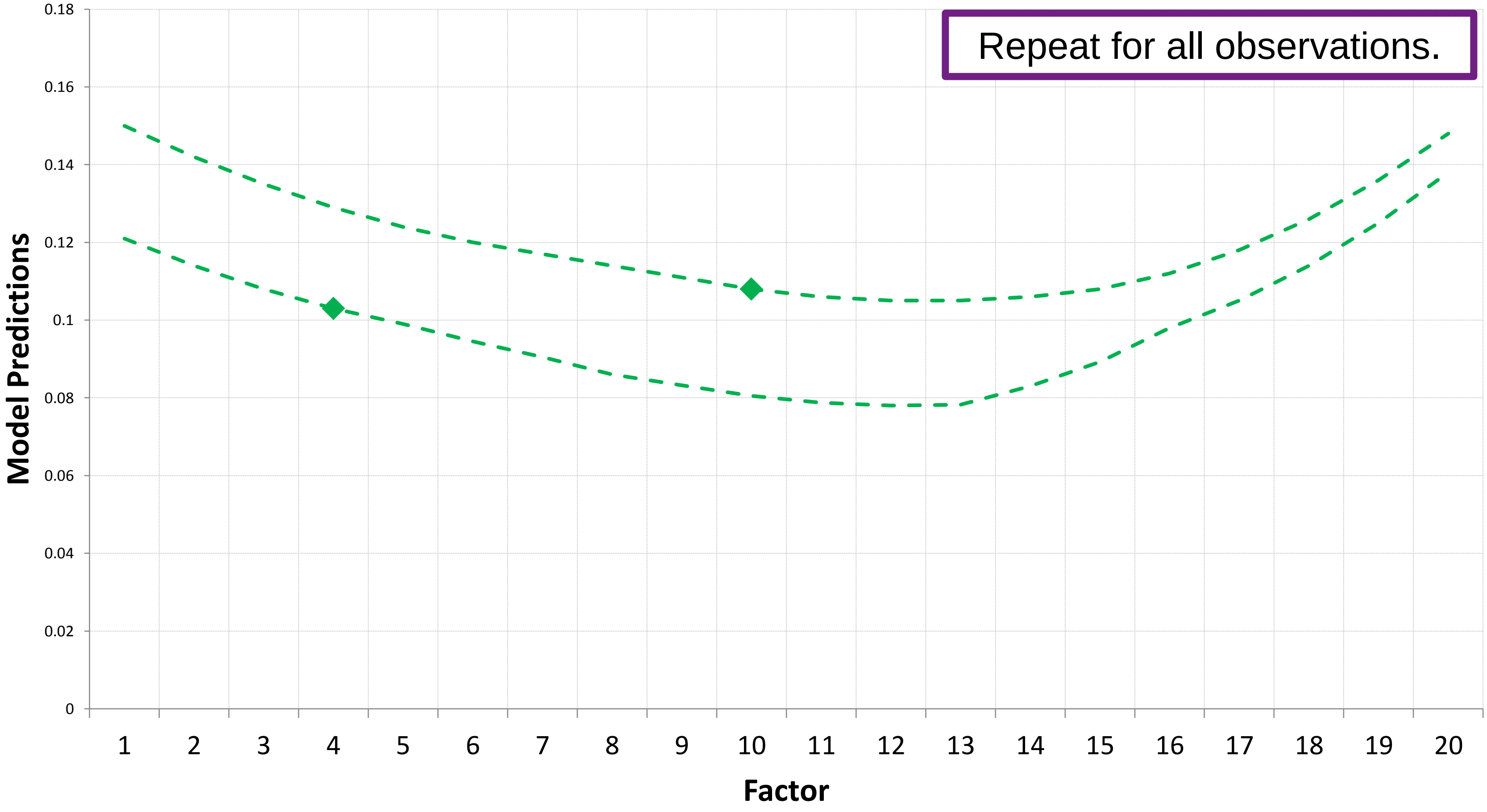
Partial dependency plots

Example



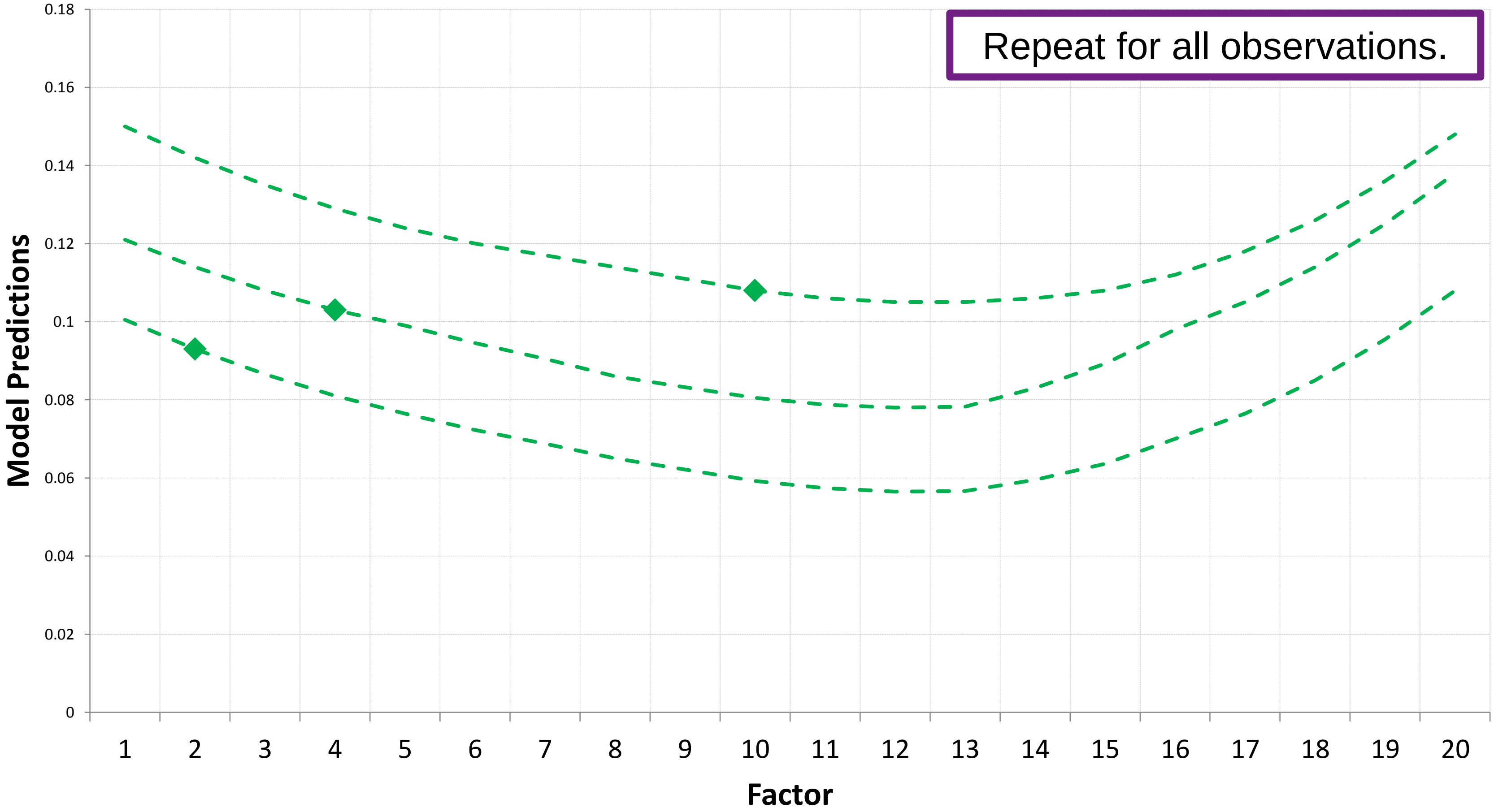
Partial dependency plots

Example



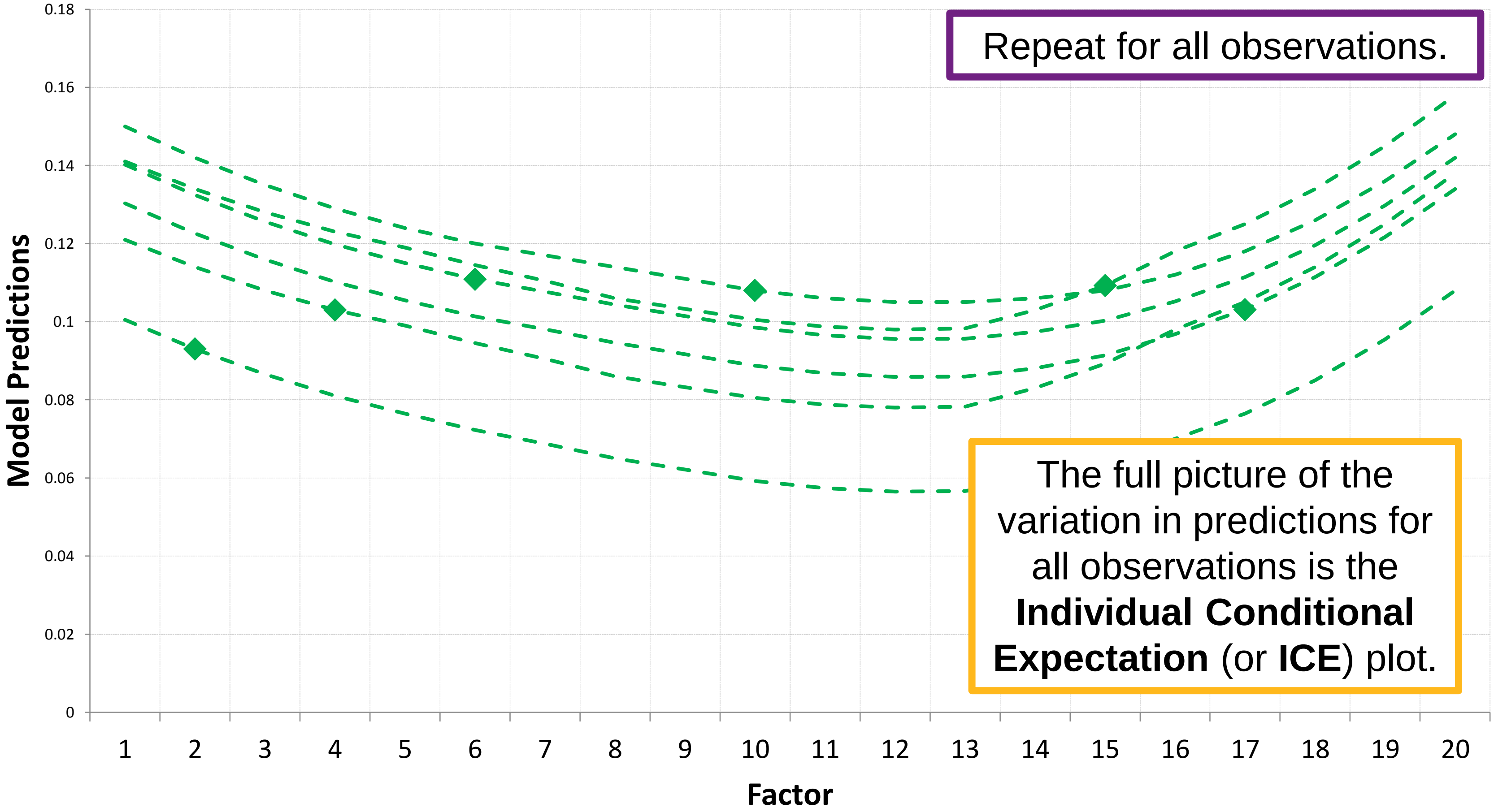
Partial dependency plots

Example



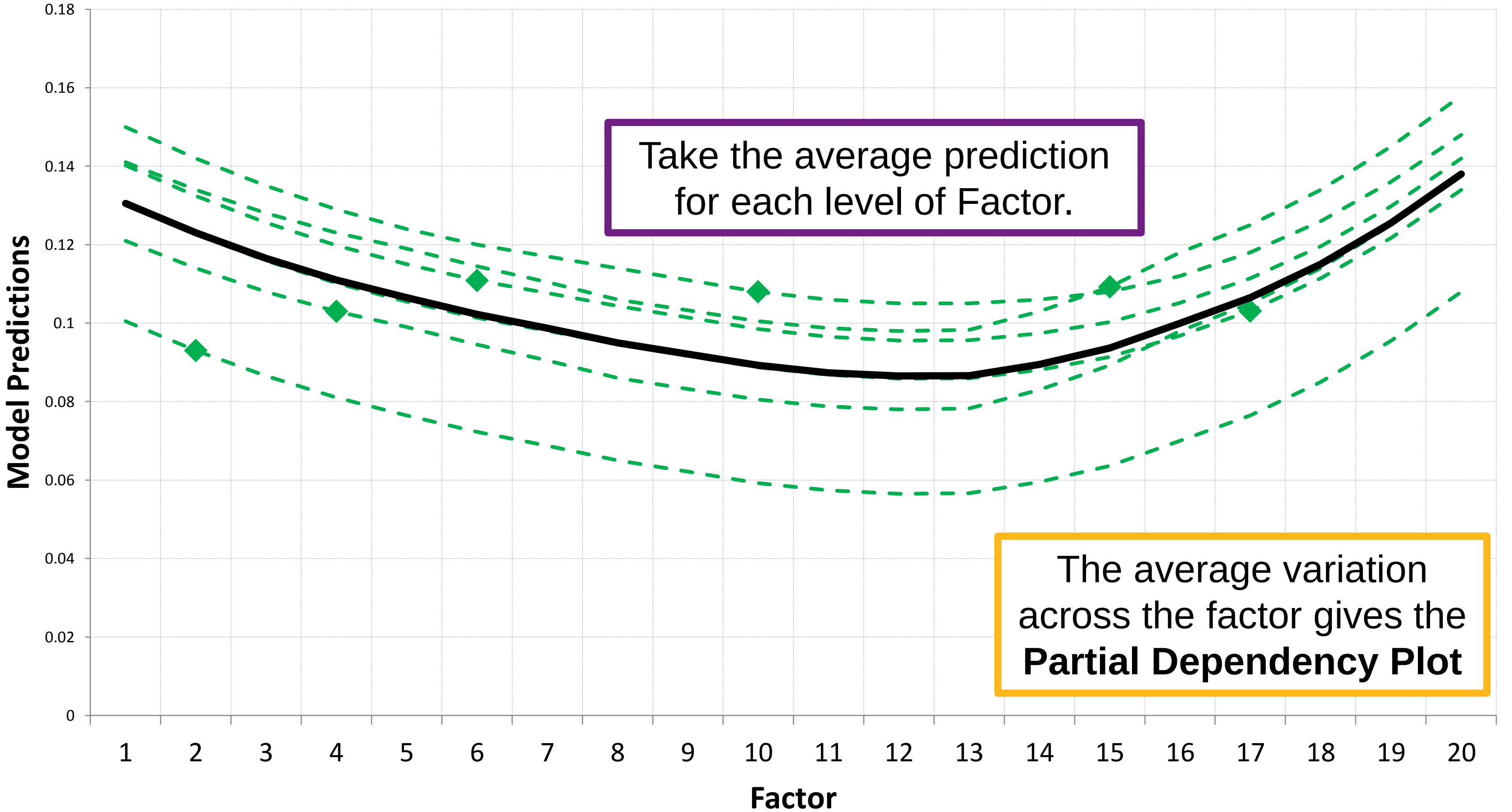
Partial dependency plots

Example



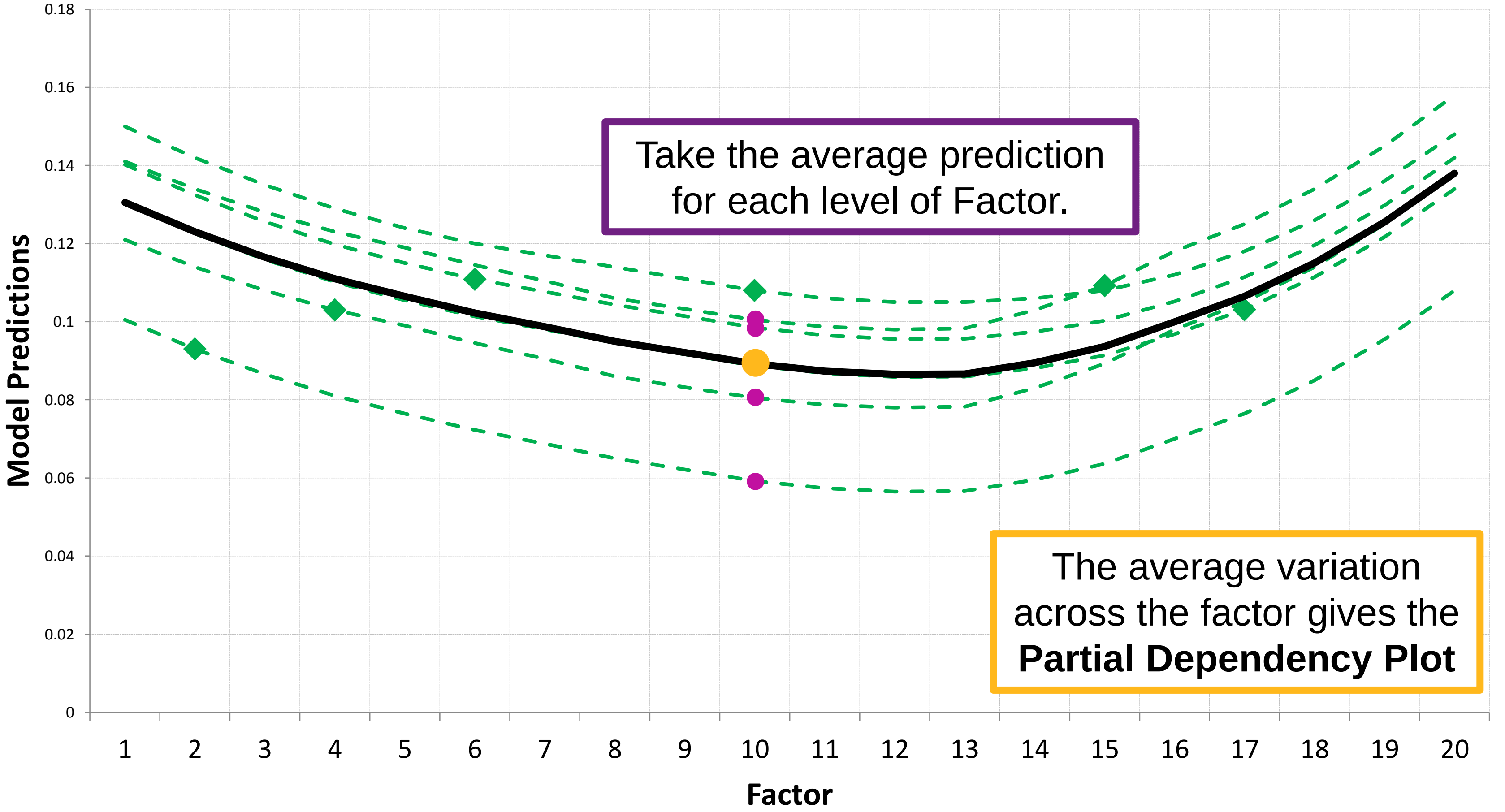
Partial dependency plots

Example

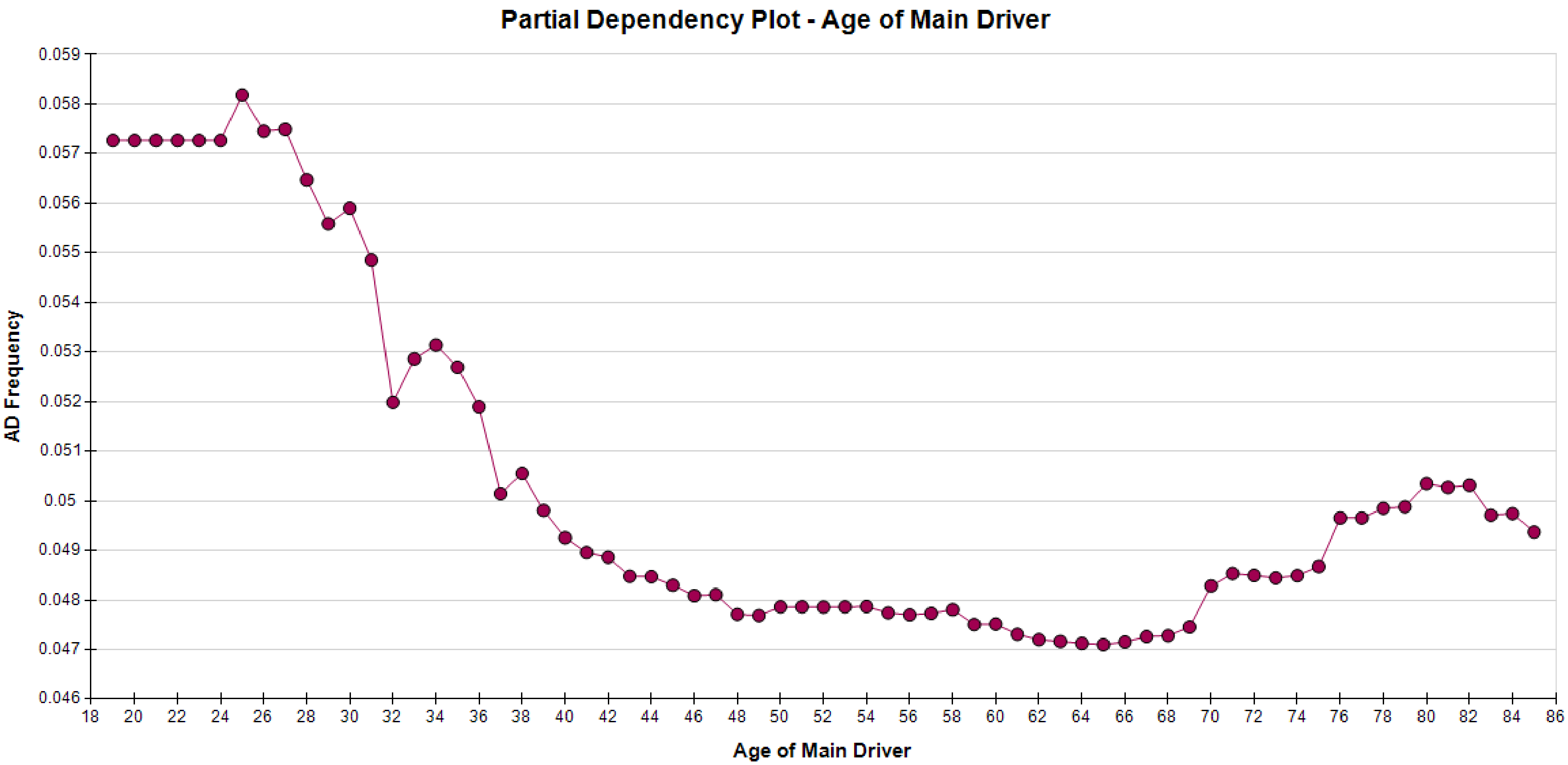


Partial dependency plots

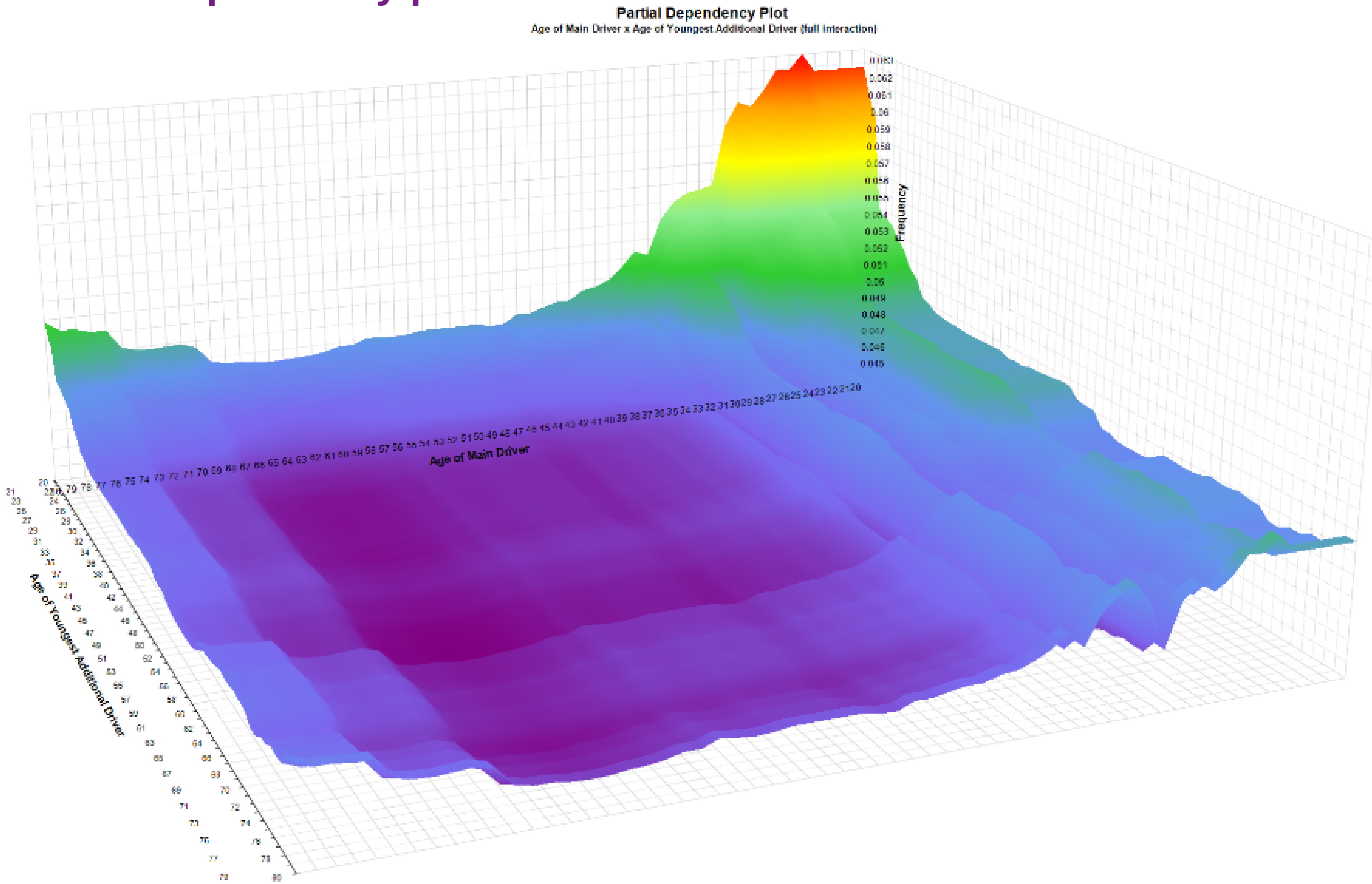
Example



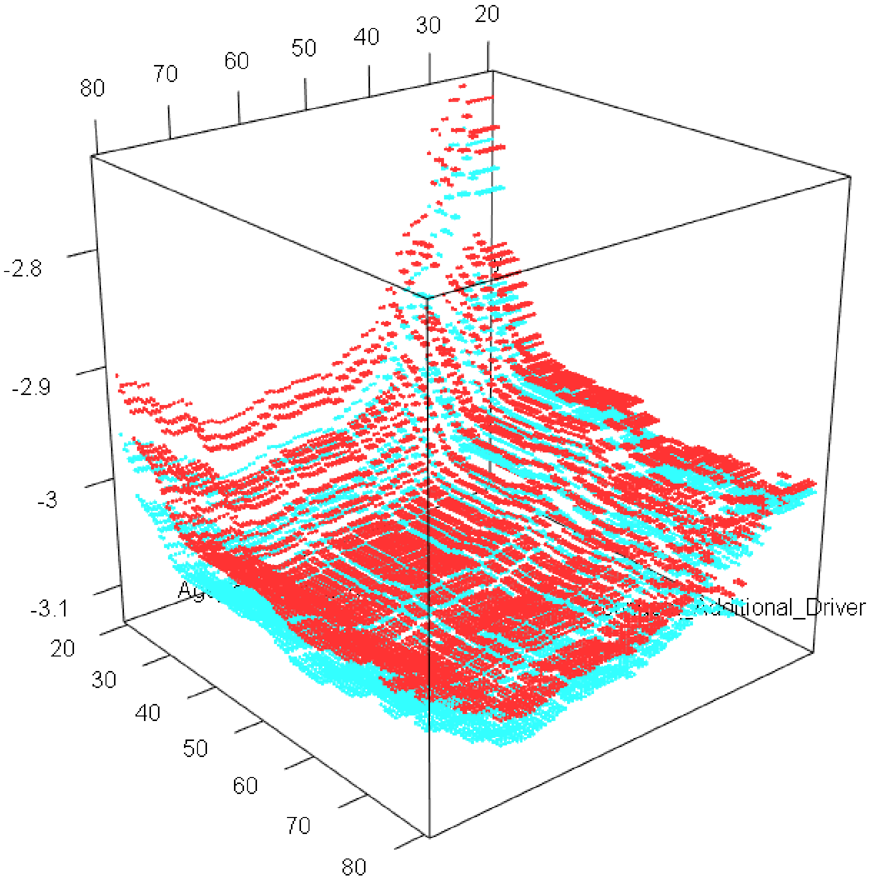
Partial dependency plots



Partial dependency plots



Partial dependency plots



Advantages

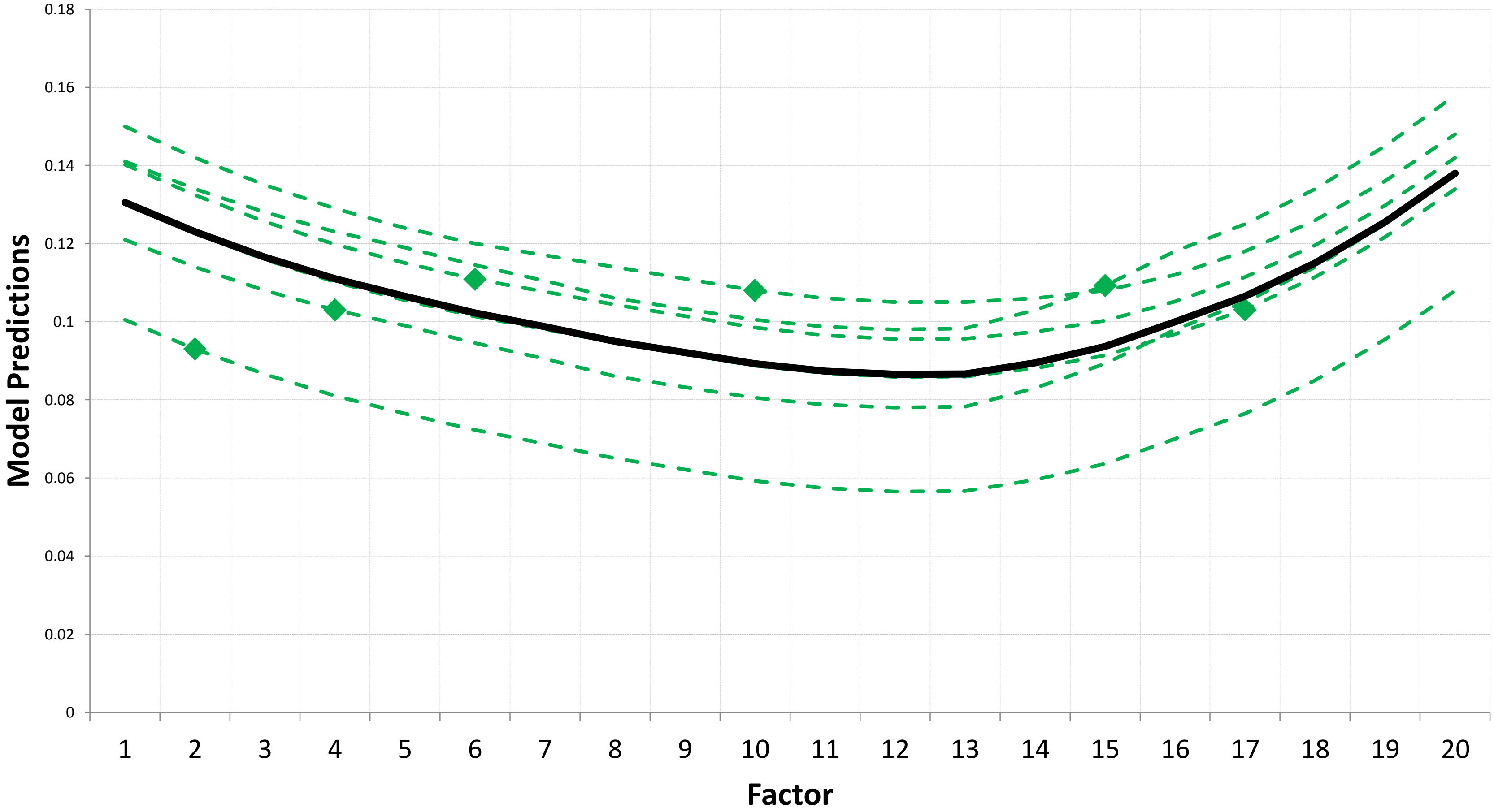
- Qualitative description of properties of relationships
- Most revealing of additive and multiplicative relationships

Disadvantages

- "GLM view of a non-GLM thing"
- Interaction effects outside of the chosen subset may be obfuscated
- eg if X_1X_2 is important and X_2 is averaged out in the partial dependence plot, X_1 may show as being heterogeneous, thus obfuscating the complexity of the modelled relationships

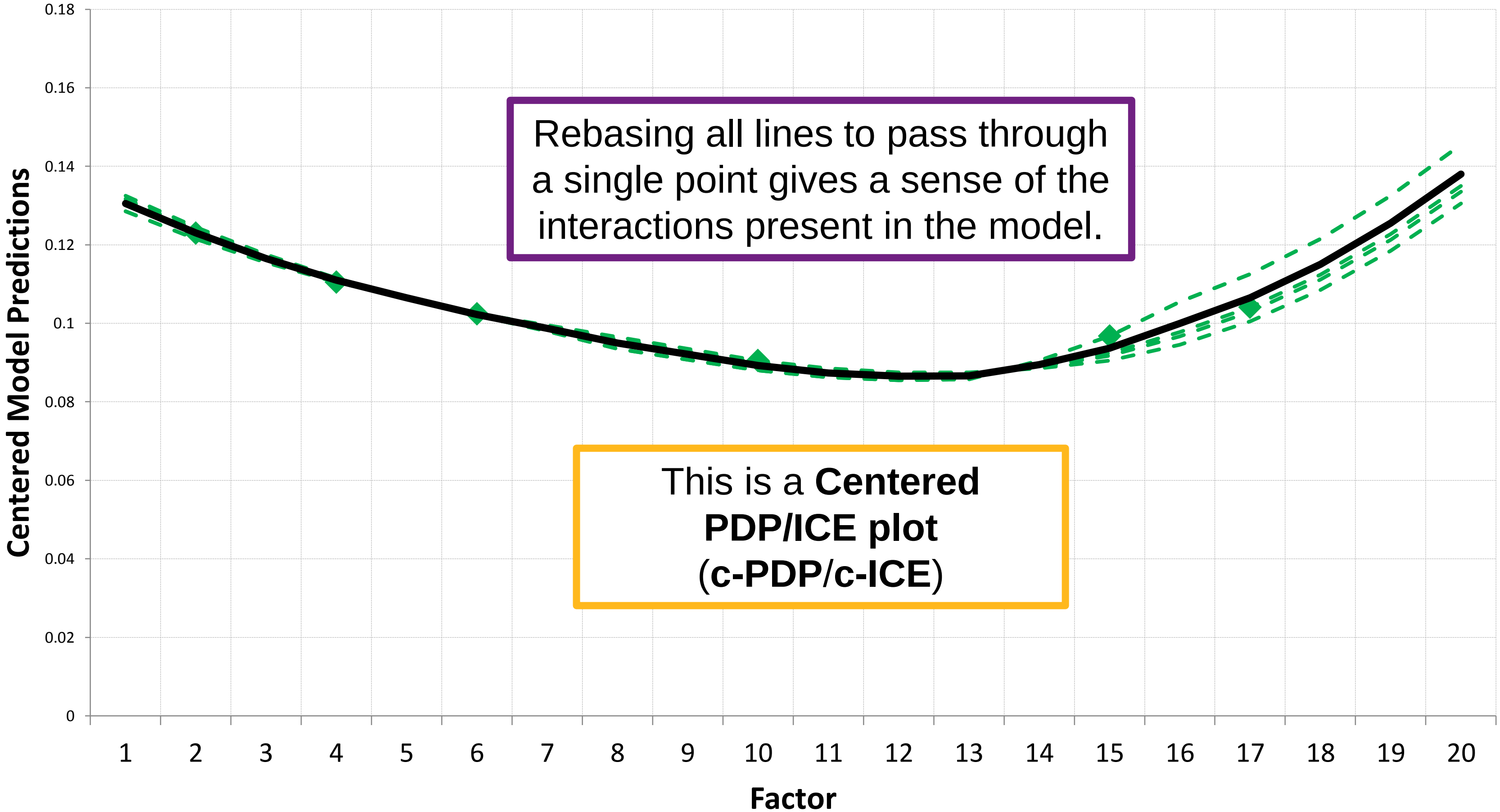
Partial dependency plots

Example



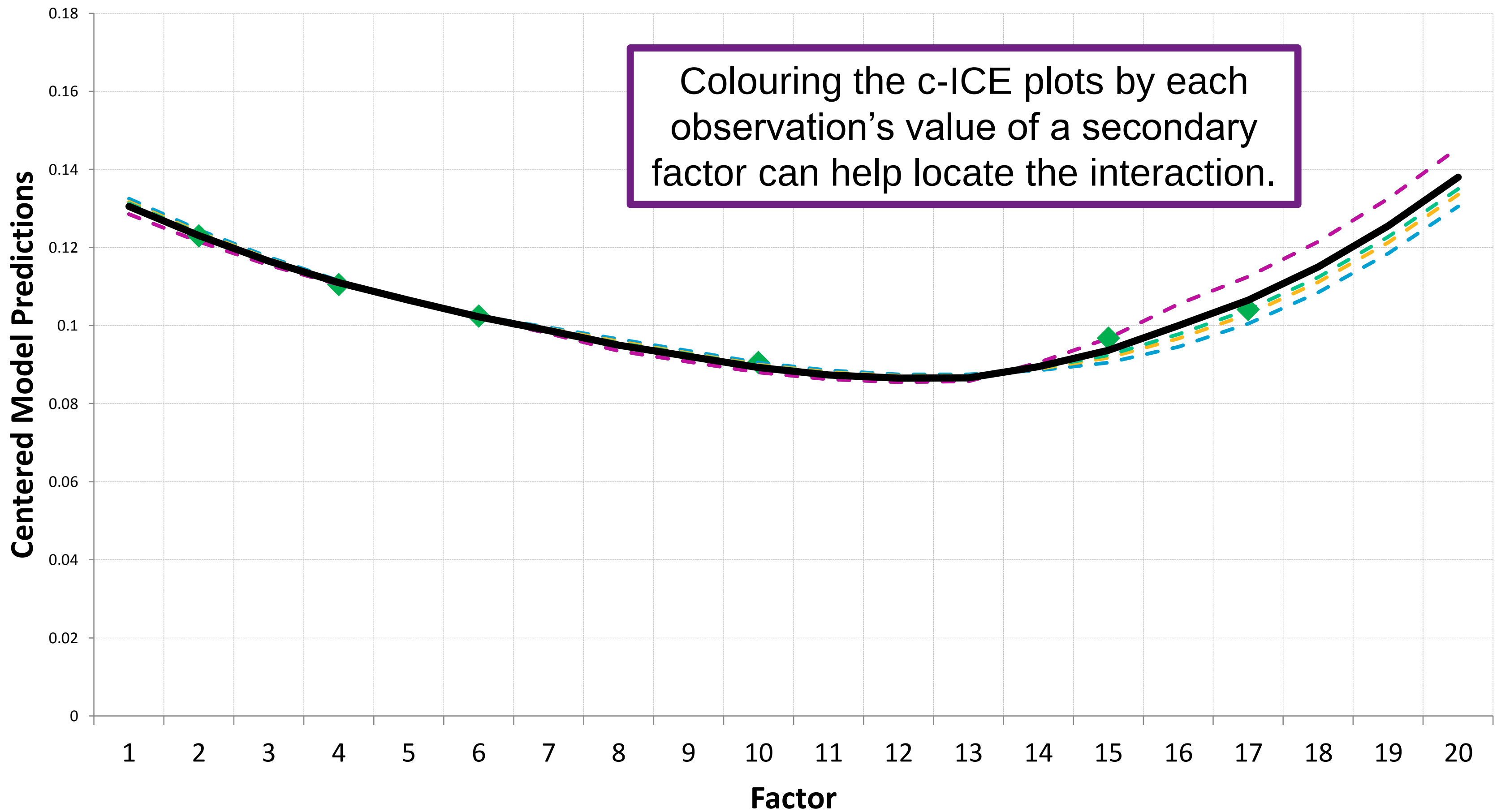
Partial dependency plots

Example

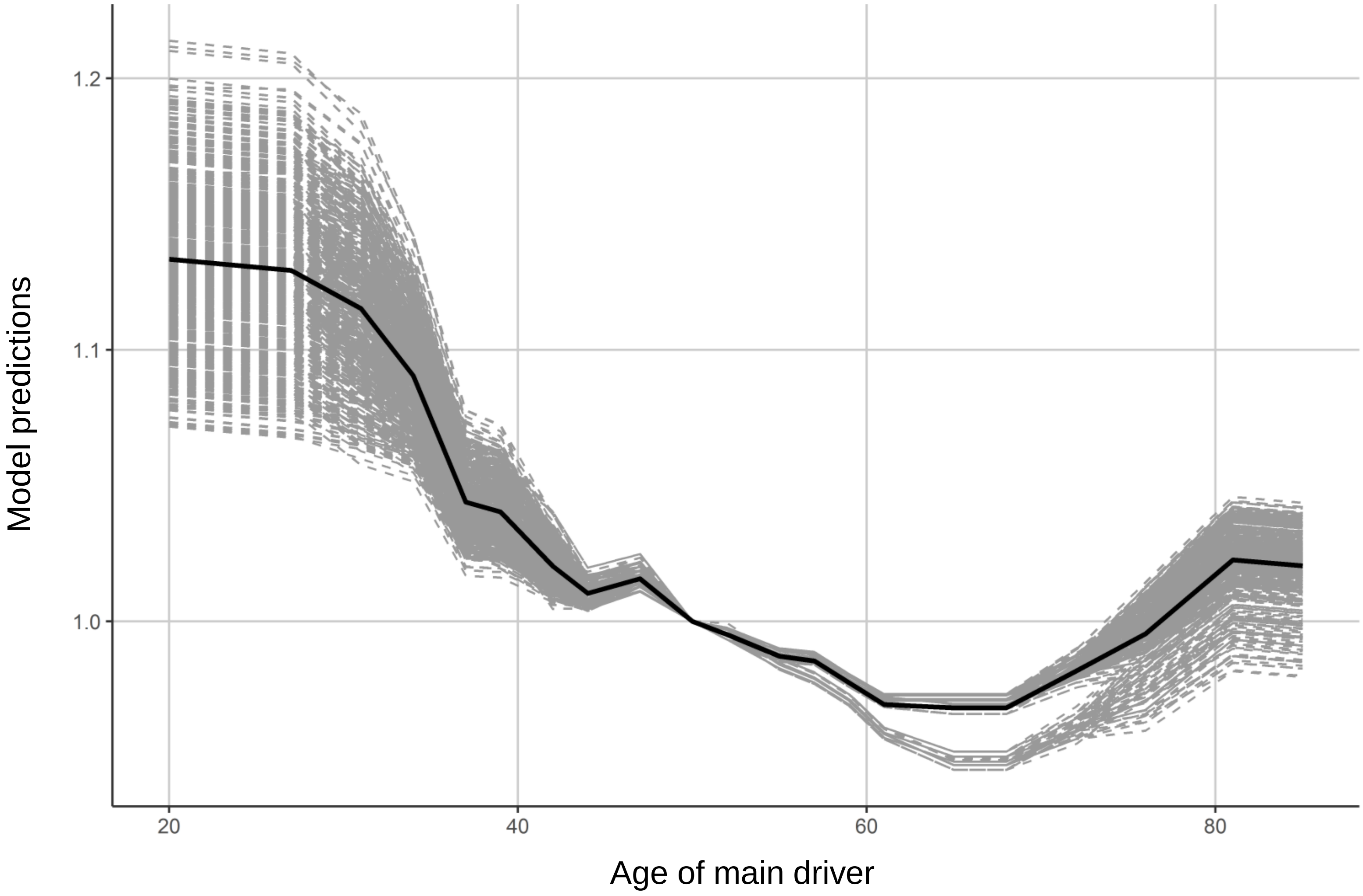


Partial dependency plots

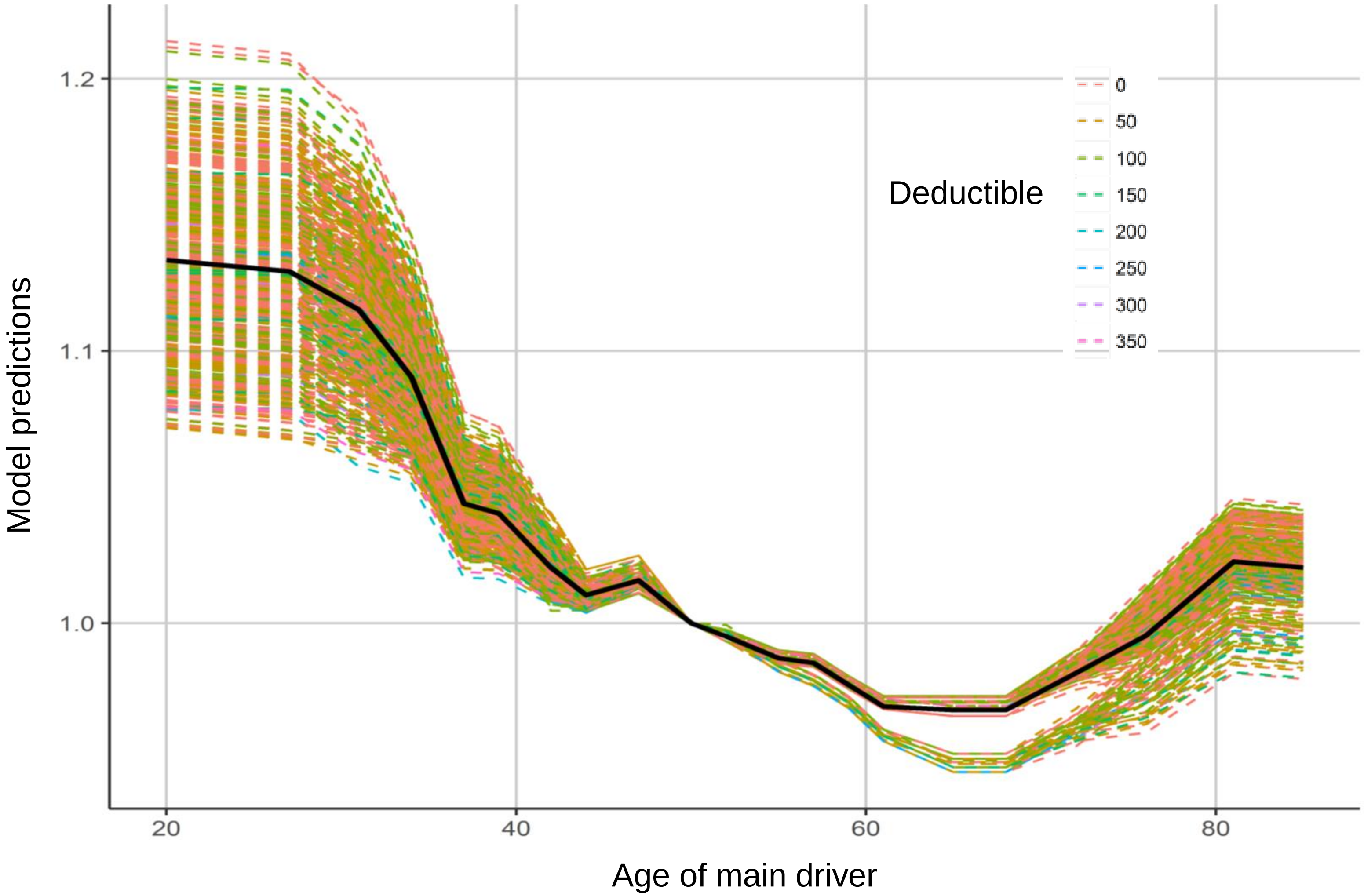
Example



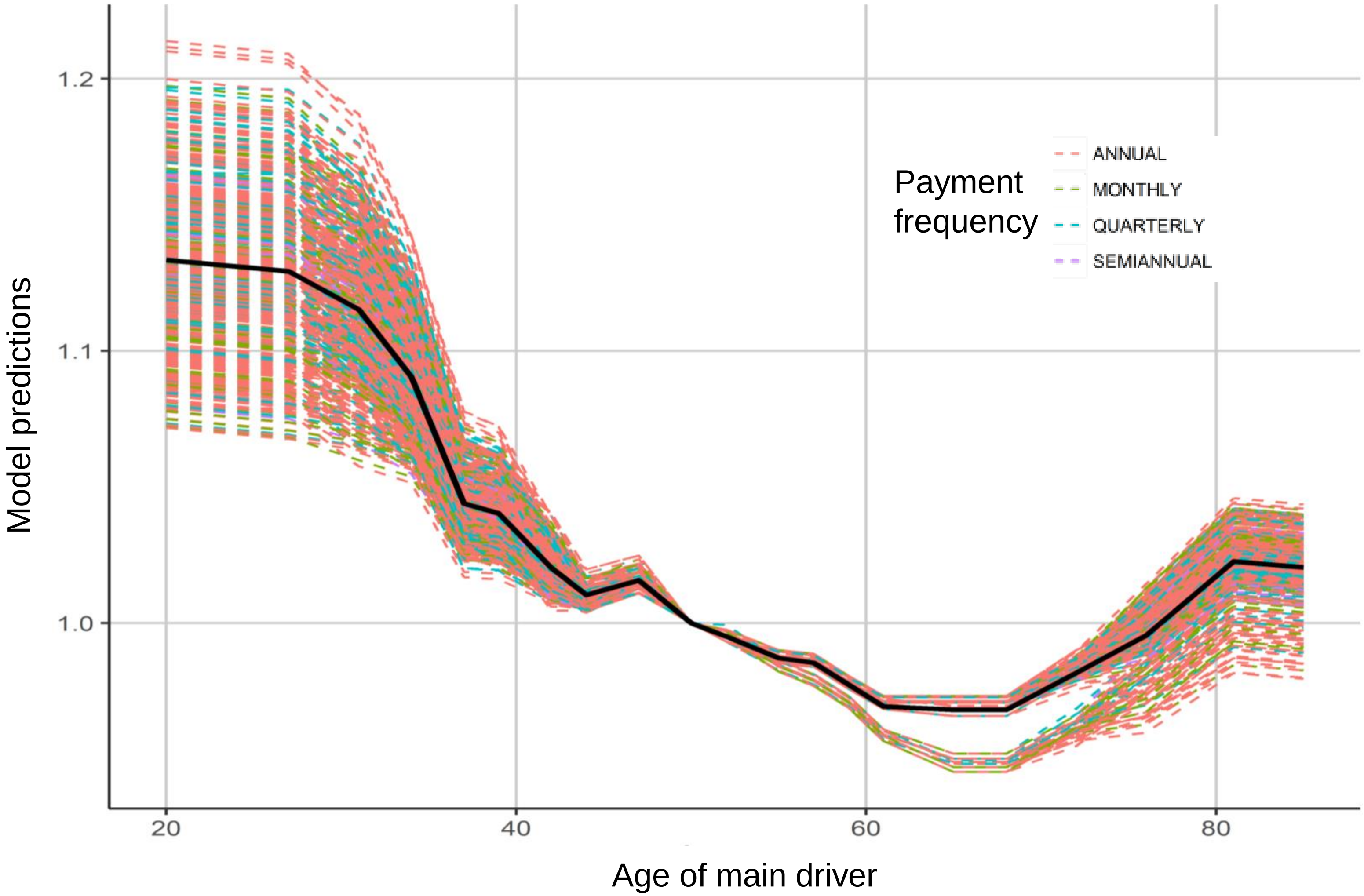
Partial dependency plot with individual conditional expectation plot



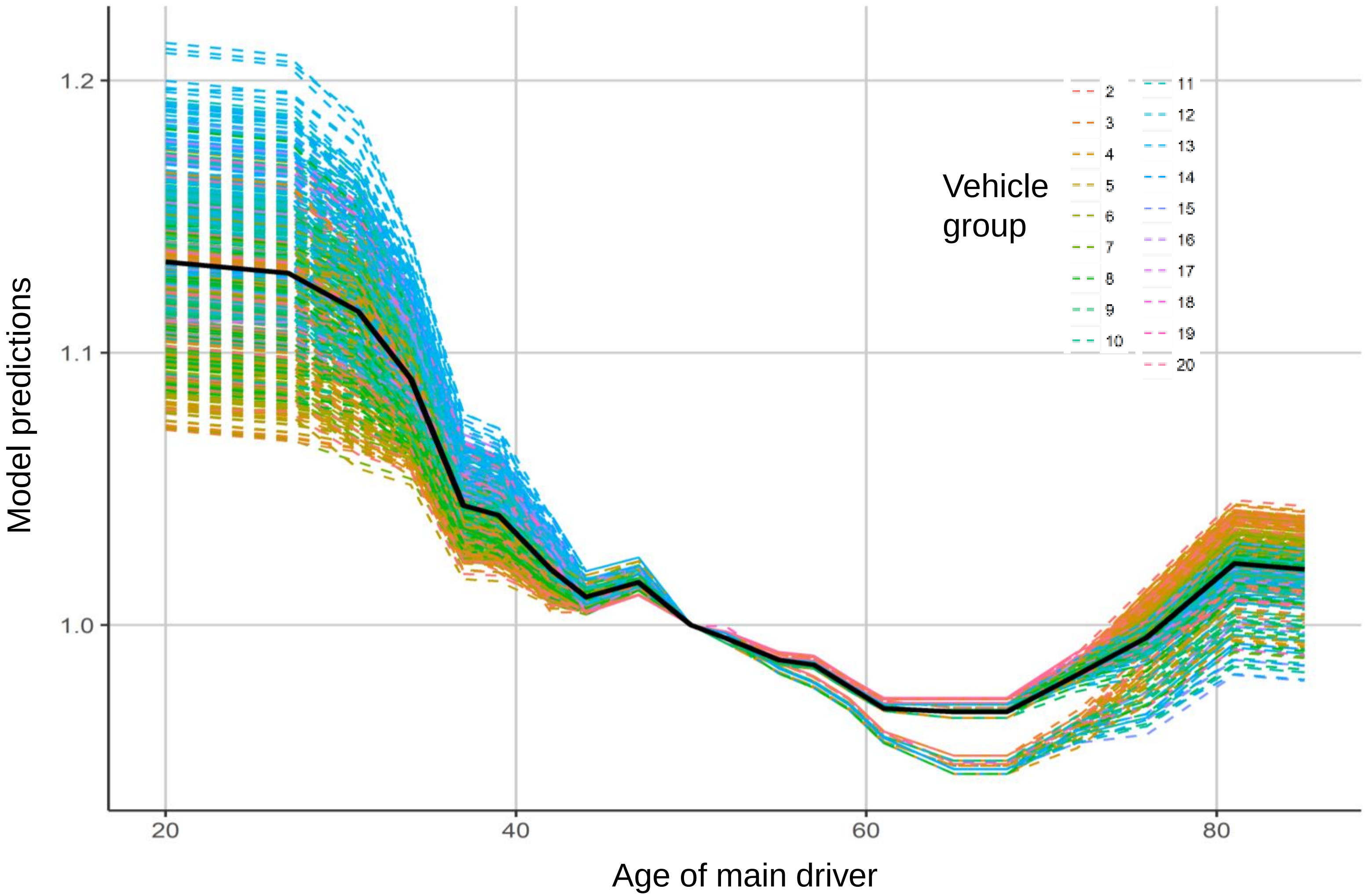
Partial dependency plot with individual conditional expectation plot



Partial dependency plot with individual conditional expectation plot



Partial dependency plot with individual conditional expectation plot



Three (and a half) interesting questions

1. Does the model add value?

2. What does the model mean?

   - Do we even need to know?

3. How can we use the model?

Model down into a GLM form

Use insights to guide GLM

Use non-GLM directly

Three (and a half) interesting questions

1. Does the model add value?

2. What does the model mean?

   - Do we even need to know?

3. How can we use the model?

Model down into a GLM form

Use insights to guide GLM

Use non-GLM directly

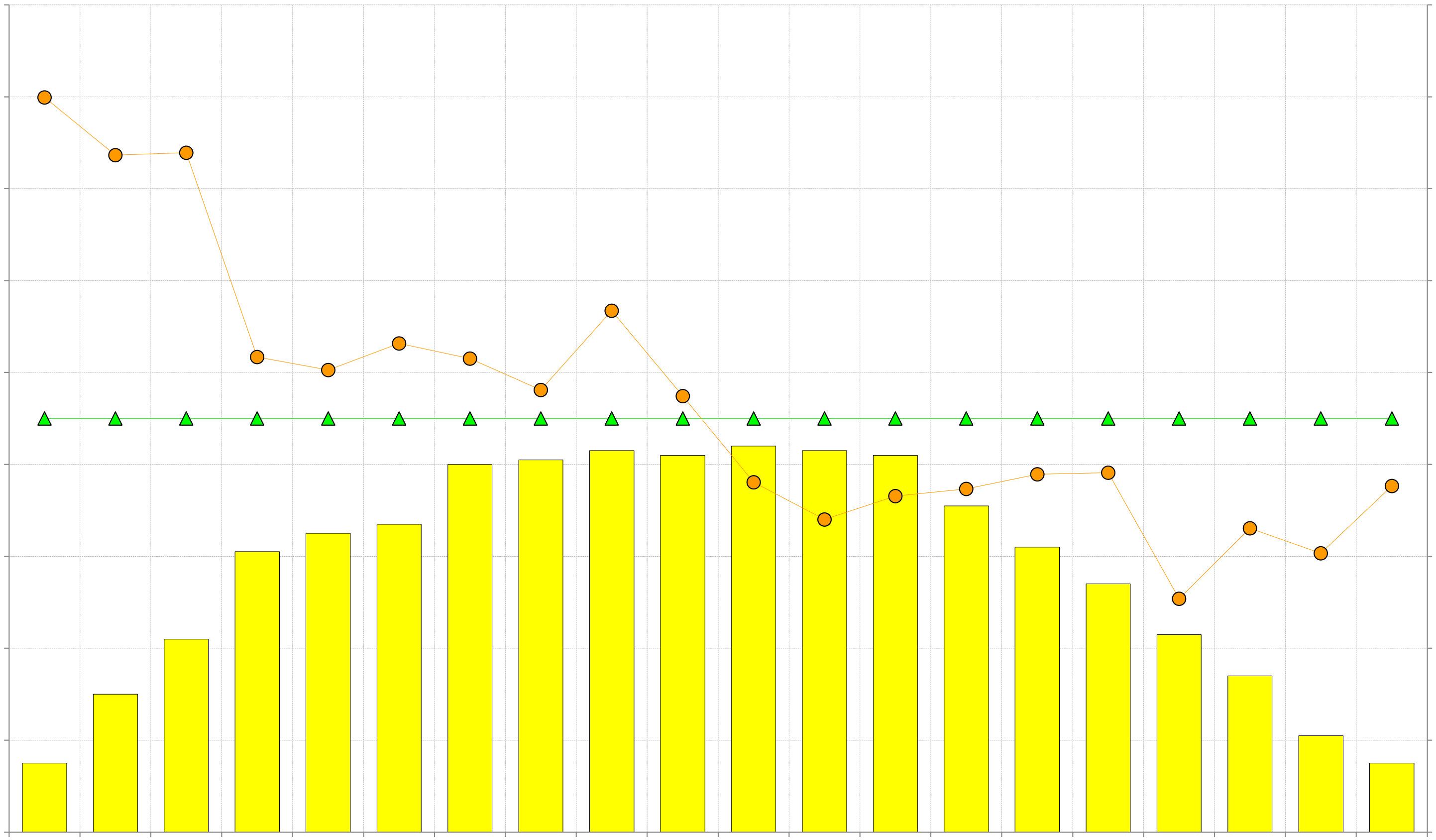
How can we use the model?

Model down into a GLM form | Use insights to guide GLM | Use non-GLM directly

Model	Gini	Gini improvement	Gini rank	Loss ratio @ elasticity 6	Loss ratio rank	Loss ratio @ elasticity 2	Loss ratio rank
GLM (main factor removed)	0.319	-2.4%	6	-0.8%	6	-0.3%	6
GLM (minor factor removed)	0.322	-1.3%	5	-0.3%	5	-0.2%	5
GLM	0.326	0.0%	4	0.0%	4	0.0%	4
GLM fitted to GBM	0.328	0.5%	3	0.9%	3	0.2%	3
GBM	0.332	1.8%	2	2.9%	1	0.6%	2
Ensemble of GBM & GLM	0.338	3.4%	1	2.8%	2	0.7%	1

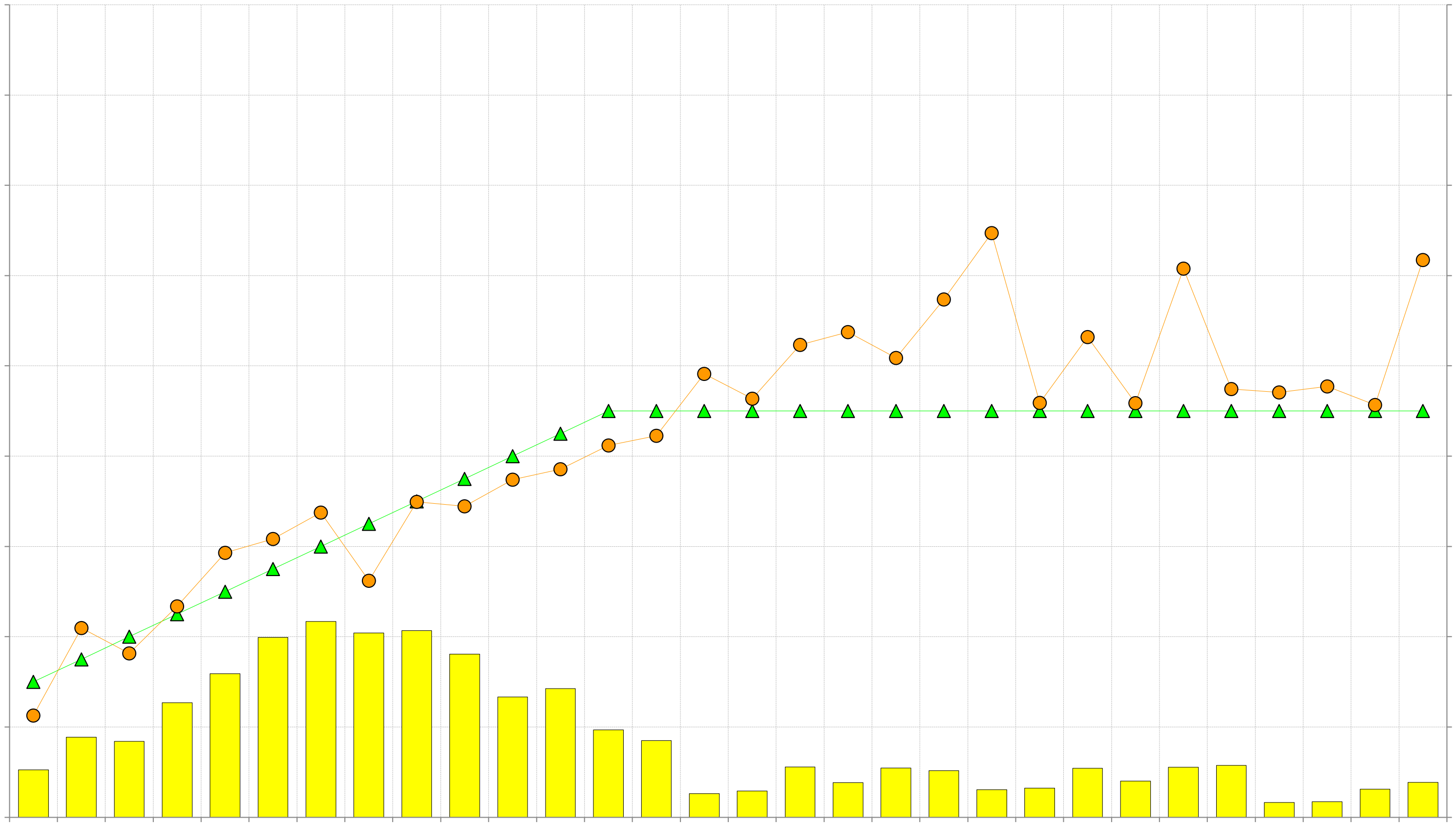
How can we use the model?

Model down into a GLM form	Use insights to guide GLM	Use non-GLM directly



How can we use the model?

Model down into a GLM form	Use insights to guide GLM	Use non-GLM directly

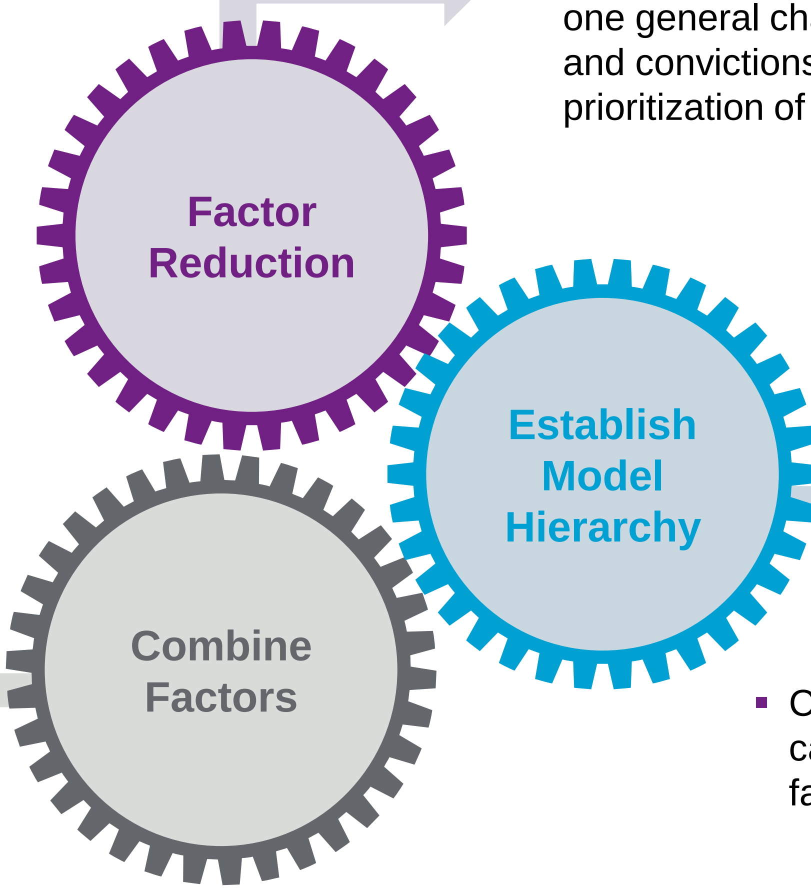


How can we use the model?

Model down into a GLM form	Use insights to guide GLM	Use non-GLM directly

- Given many hundreds of factors describing one general characteristic (eg past claims and convictions) GBM can help with prioritization of candidate factors

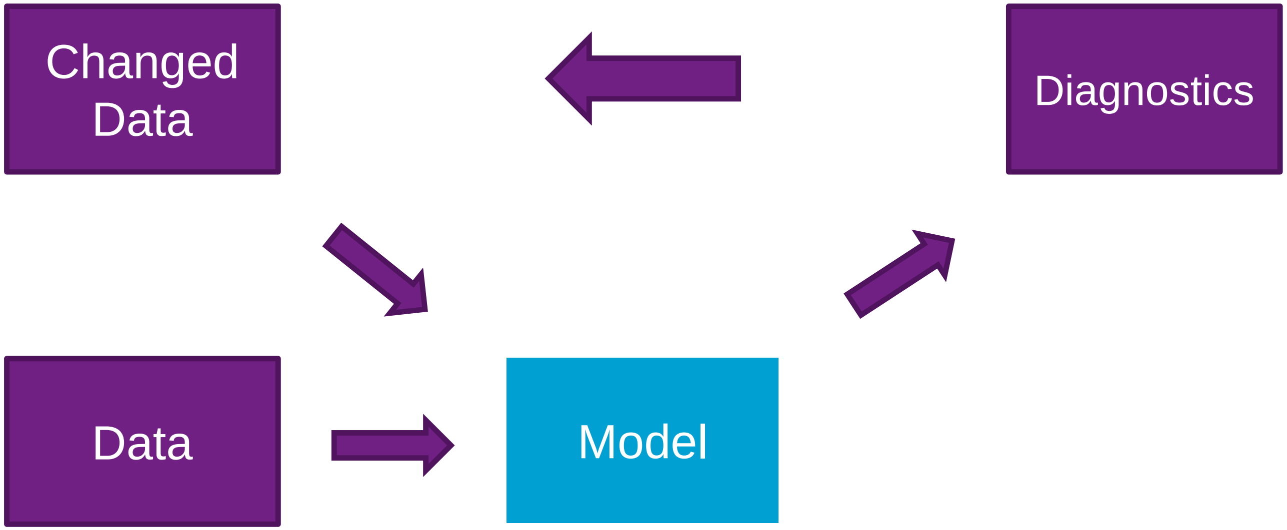
- Findings of GBM can be restated as higher order interactions within a GLM framework
- For example to correct “corners” of a GLM where the model over predicts



- Can support identification of candidate model segmentation factors within a GLM framework

How can we use the model?

Model down into a GLM form	Use insights to guide GLM	Use non-GLM directly



Three (and a half) interesting questions

1. Does the model add value?

2. What does the model mean?

   - Do we even need to know?

3. How can we use the model?

A fair playing field re comprehensibility?

Automated parameterization

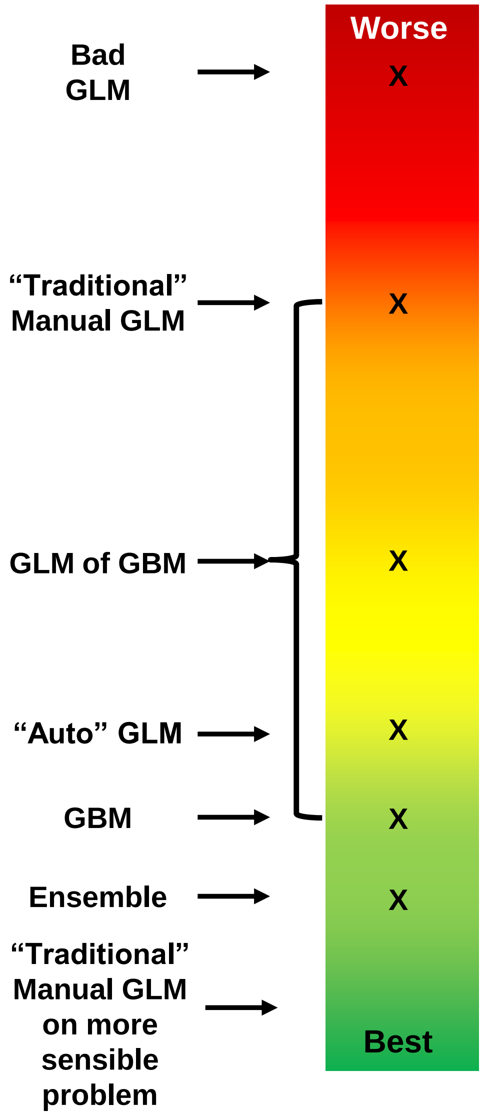
Model	Gini	Gini improvement	Gini rank	Loss ratio @ elasticity 6	Loss ratio rank	Loss ratio @ elasticity 2	Loss ratio rank
GLM (main factor removed)	0.319	-2.4%	7	-0.8%	7	-0.3%	7
GLM (minor factor removed)	0.322	-1.3%	6	-0.3%	6	-0.2%	6
GLM	0.326	0.0%	5	0.0%	5	0.0%	5
GLM fitted to GBM	0.328	0.5%	4	0.9%	4	0.2%	4
GLM with "Auto Saddles"	0.329	0.7%	3	1.0%	3	0.5%	3
GBM	0.332	1.8%	2	2.9%	1	0.6%	2
Ensemble of GBM & GLM	0.338	3.4%	1	2.8%	2	0.7%	1

A simple 2 x 2

The full picture

	The problem dimension	
The method dimension	Original problem, solved with original methods	Newly defined problem, solved with original methods
	Original problem, solved with new techniques	Newly defined problem, solved with new techniques

Conclusions



- If you can..
  - Cope with not seeing the model and instead using broad diagnostics
  - And cope with small segments being wrong
  - And your regulator can as well
  - And you have a rating engine that can implement it
  - And you have the software and hardware to fit to large datasets
- …then there are predictive lift benefits from GBMs et al in pricing
  - In other areas, eg marketing, application is less problematic
- If not, there are ways of finding new insight, implementing within GLMs
- But also if you accept models that are hard to interpret, GLMs can be machine fitted also…
- Perhaps most important don't lose sight of the value of thinking and domain expertise…

Southwest Actuarial Forum

Exploring Advanced Analytics Solutions in Pricing

December 2016

Willis Towers Watson