ACTUARIAL SOCIETY OF SOUTH AFRICA

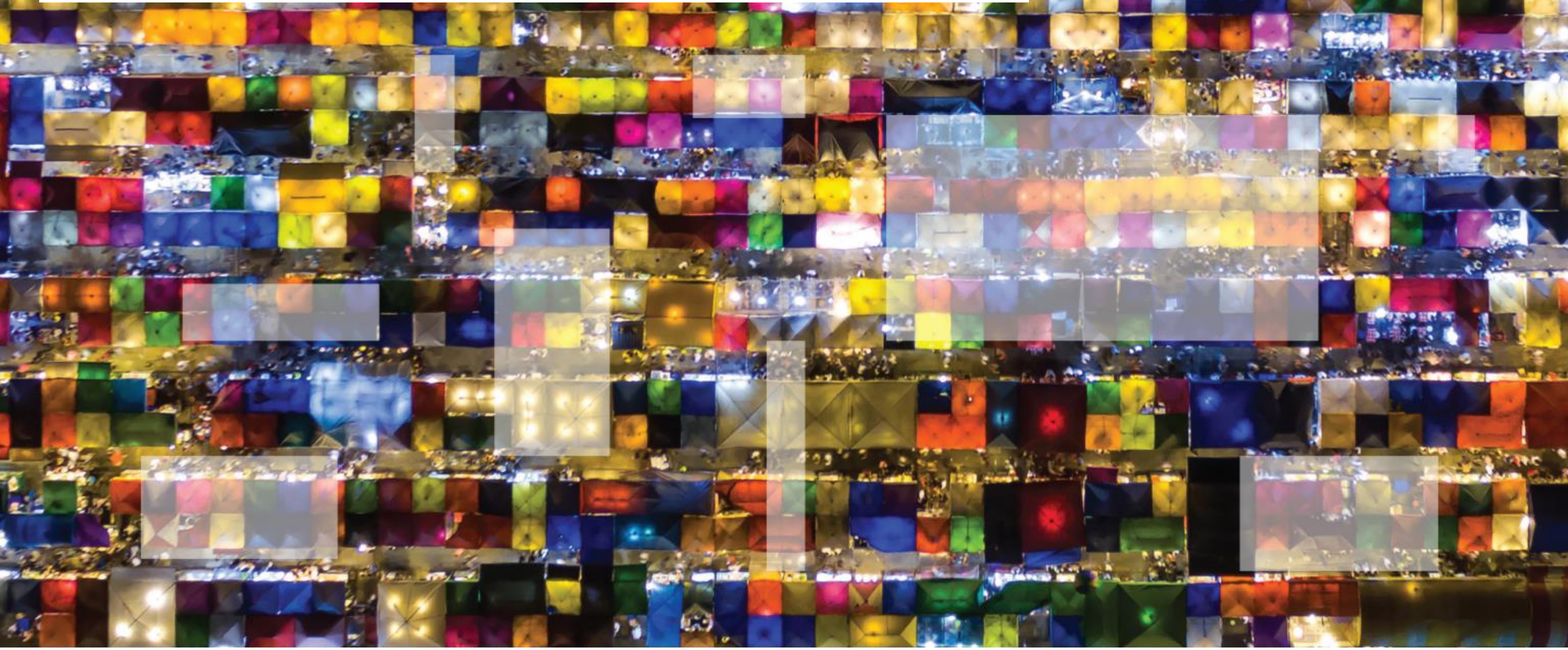
QUANTIFYING RISK, ENABLING OPPORTUNITY

Machine Learning in Pricing Models

Christian Clemmensen
Willis Towers Watson
7 September 2018

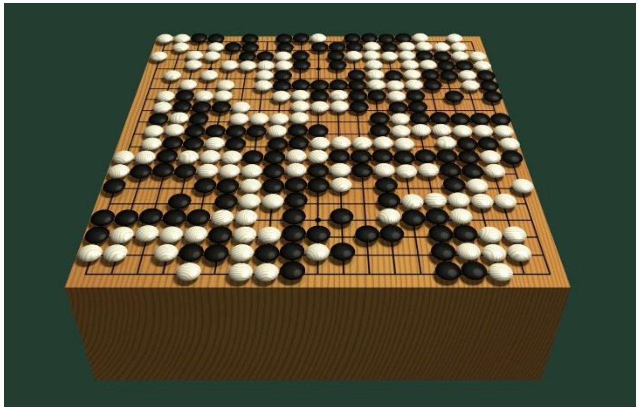
Machine Learning in Pricing Models

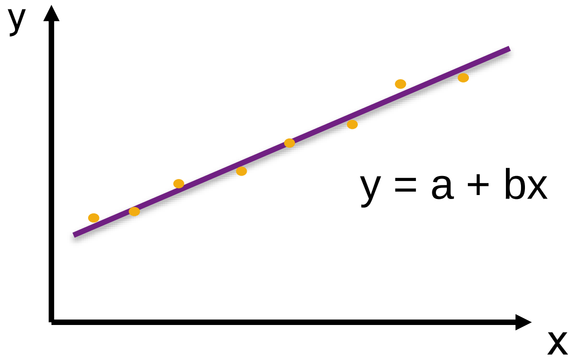
7 September 2018



Willis Towers Watson

Who's interested in what?



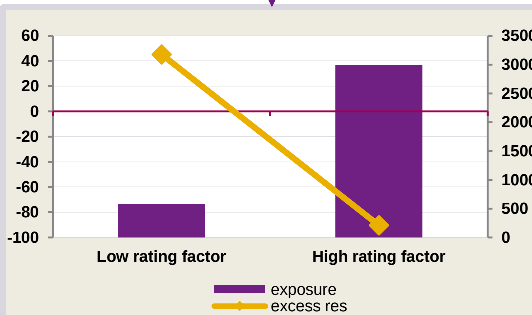


Applications of machine learning in the insurance sector

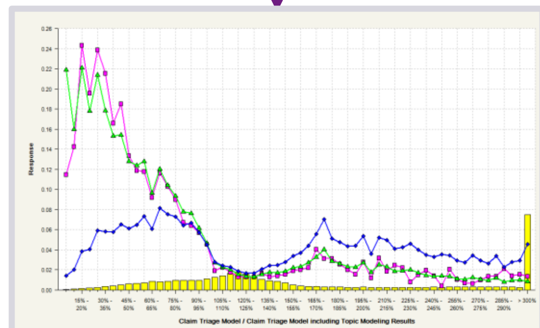
Marketing and Distribution → **Customer management** → **Underwriting and risk management** → **Pricing** → **Claims management** → **Asset management**

- Product development (UBI)
- Retention segmentation and program design
- Agent/Broker performance evaluation
- Fraud detection

P&C to Life cross-selling

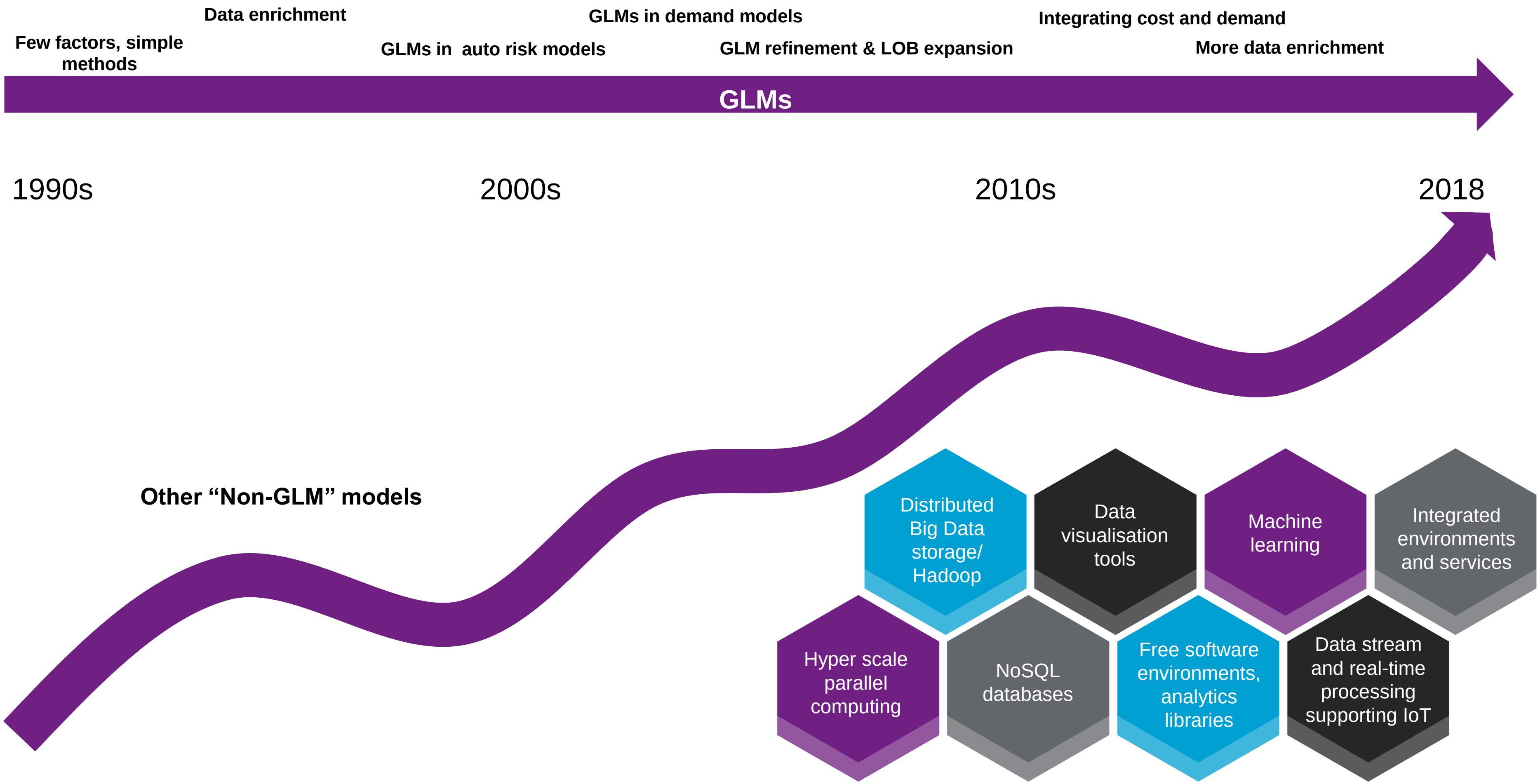


Web-scraping and feature design in commercial lines



Topic modeling and large loss modeling

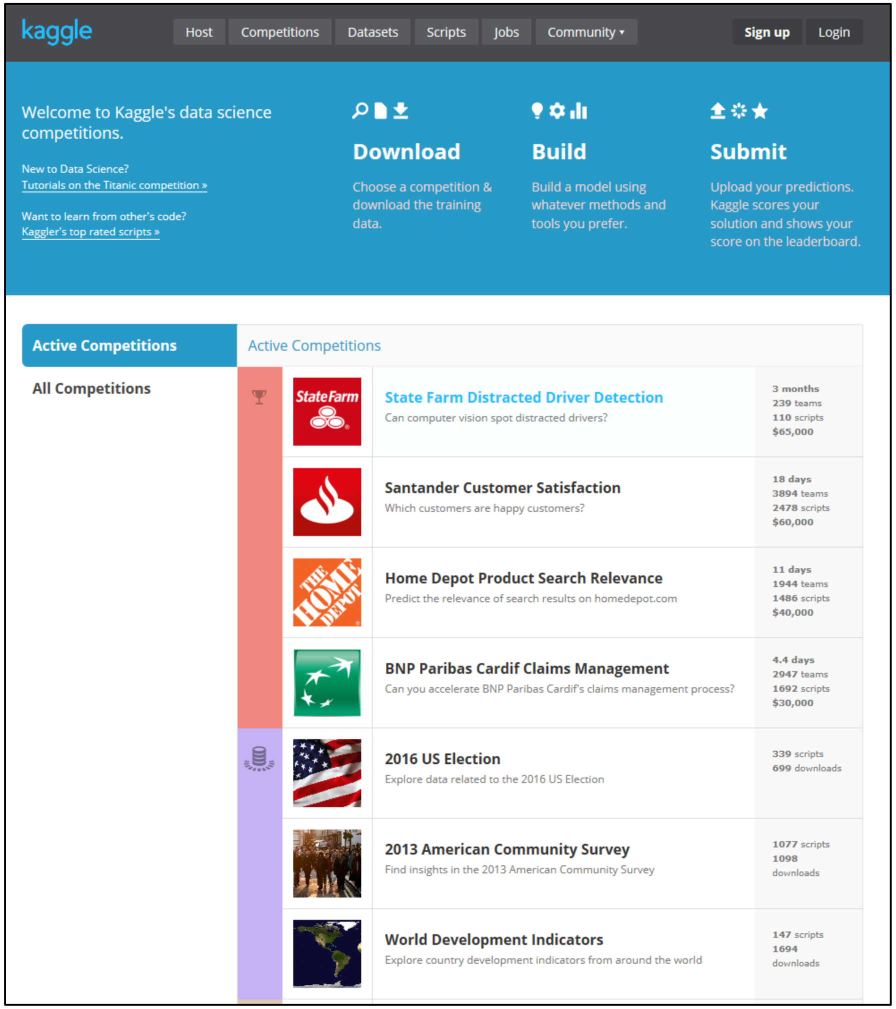
This is not new….



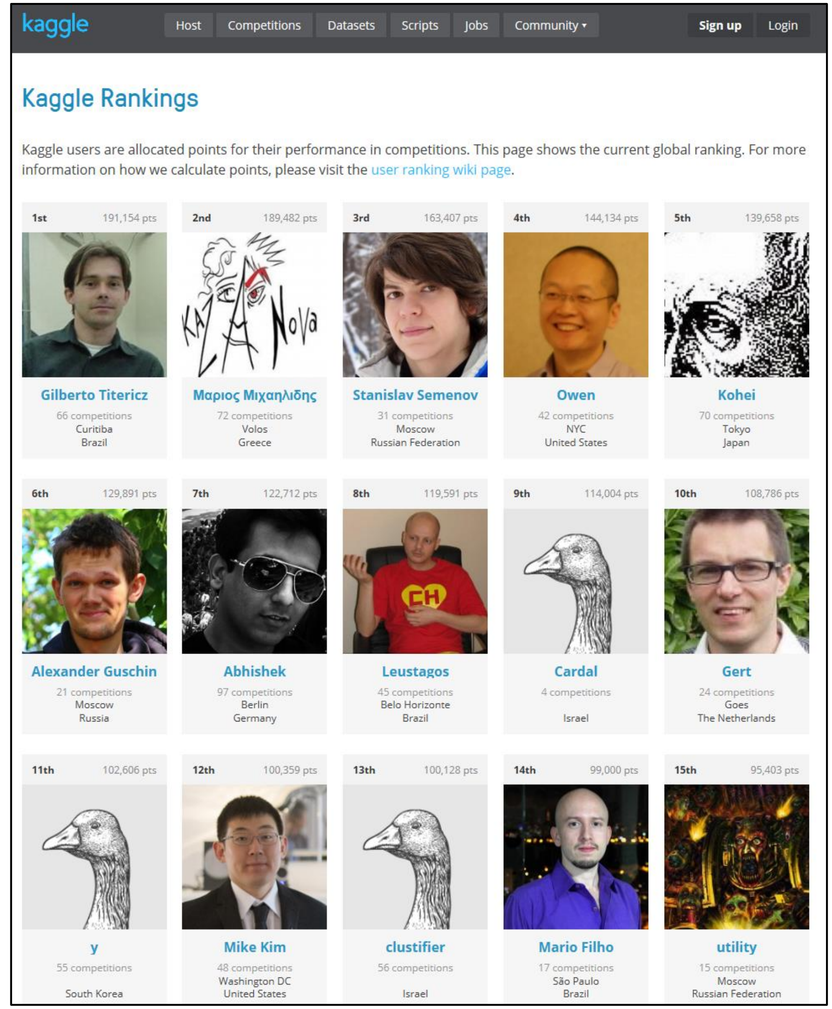
What are these machine learning methods?

- Ensembles
- Classifications Trees
- "Earth"
- 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

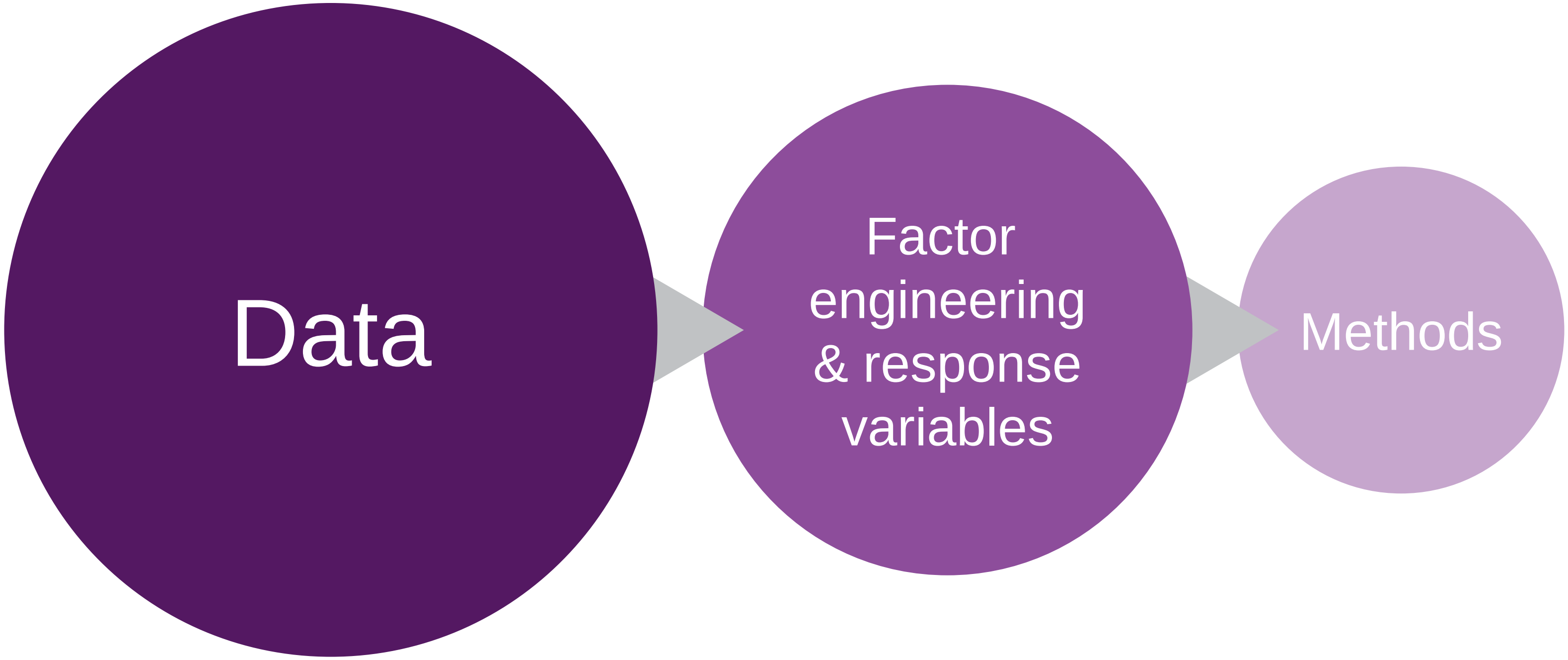


kaggle
Host
Competitions
Datasets
Scripts
Jobs
Community
Sign up
Login
Welcome to Kaggle's data science competitions.
New to Data Science?
Tutorials on the Titanic competition »
Want to learn from other's code?
Kaggler's top rated scripts »
Download
Choose a competition & download the training data.
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
Active Competitions
State Farm Distracted Driver Detection
Can computer vision spot distracted drivers?
3 months
239 teams
110 scripts
$65,000
Santander Customer Satisfaction
Which customers are happy customers?
18 days
3894 teams
2478 scripts
$60,000
Home Depot Product Search Relevance
Predict the relevance of search results on homedepot.com
11 days
1944 teams
1486 scripts
$40,000
BNP Paribas Cardif Claims Management
Can you accelerate BNP Paribas Cardif's claims management process?
4.4 days
2947 teams
1692 scripts
$30,000
2016 US Election
Explore data related to the 2016 US Election
339 scripts
699 downloads
2013 American Community Survey
Find insights in the 2013 American Community Survey
1077 scripts
1098 downloads
World Development Indicators
Explore country development indicators from around the world
147 scripts
1694 downloads

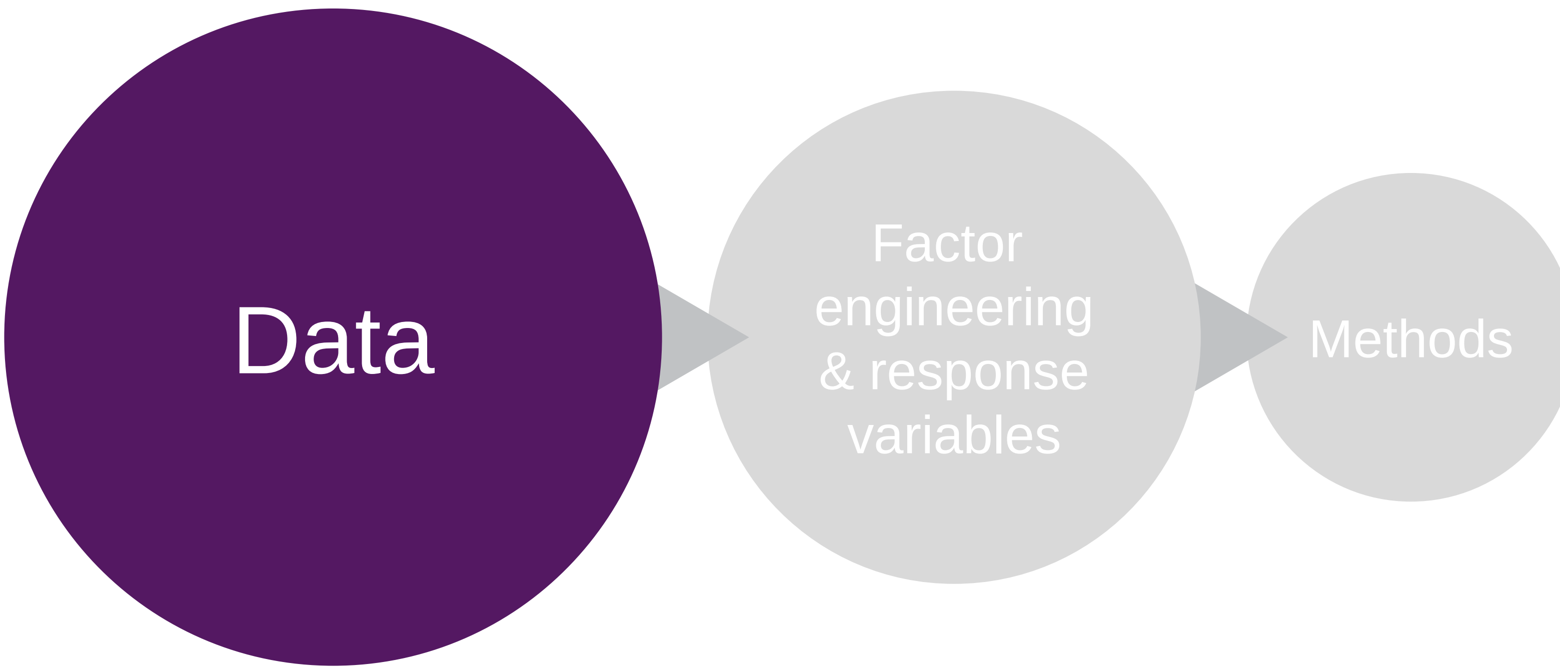


kaggle
Host
Competitions
Datasets
Scripts
Jobs
Community
Sign up
Login
Kaggle Rankings
Kaggle users are allocated points for their performance in competitions. This page shows the current global ranking. For more information on how we calculate points, please visit the user ranking wiki page.
1st 191,154 pts Gilberto Titericz 66 competitions Curitiba Brazil
2nd 189,482 pts Μαριος Μιχαηλιδης 72 competitions Volos Greece
3rd 163,407 pts Stanislav Semenov 31 competitions Moscow Russian Federation
4th 144,134 pts Owen 42 competitions NYC United States
5th 139,658 pts Kohei 70 competitions Tokyo Japan
6th 129,891 pts Alexander Guschin 21 competitions Moscow Russia
7th 122,712 pts Abhishek 97 competitions Berlin Germany
8th 119,591 pts Leustagos 45 competitions Belo Horizonte Brazil
9th 114,004 pts Cardal 4 competitions Israel
10th 108,786 pts Gert 24 competitions Goes The Netherlands
11th 102,606 pts y 55 competitions South Korea
12th 100,359 pts Mike Kim 48 competitions Washington DC United States
13th 100,128 pts clustifier 56 competitions Israel
14th 99,000 pts Mario Filho 17 competitions São Paulo Brazil
15th 95,403 pts utility 15 competitions Moscow Russian Federation

Is it really all about the method?



Is it really all about the method?



Is it really all about the method?

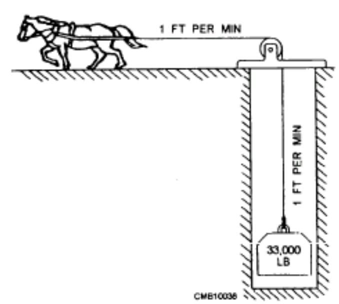
Data

Physical facticity

E.g., height, length, weight

Mechanical nature

E.g., engine size, fuel type



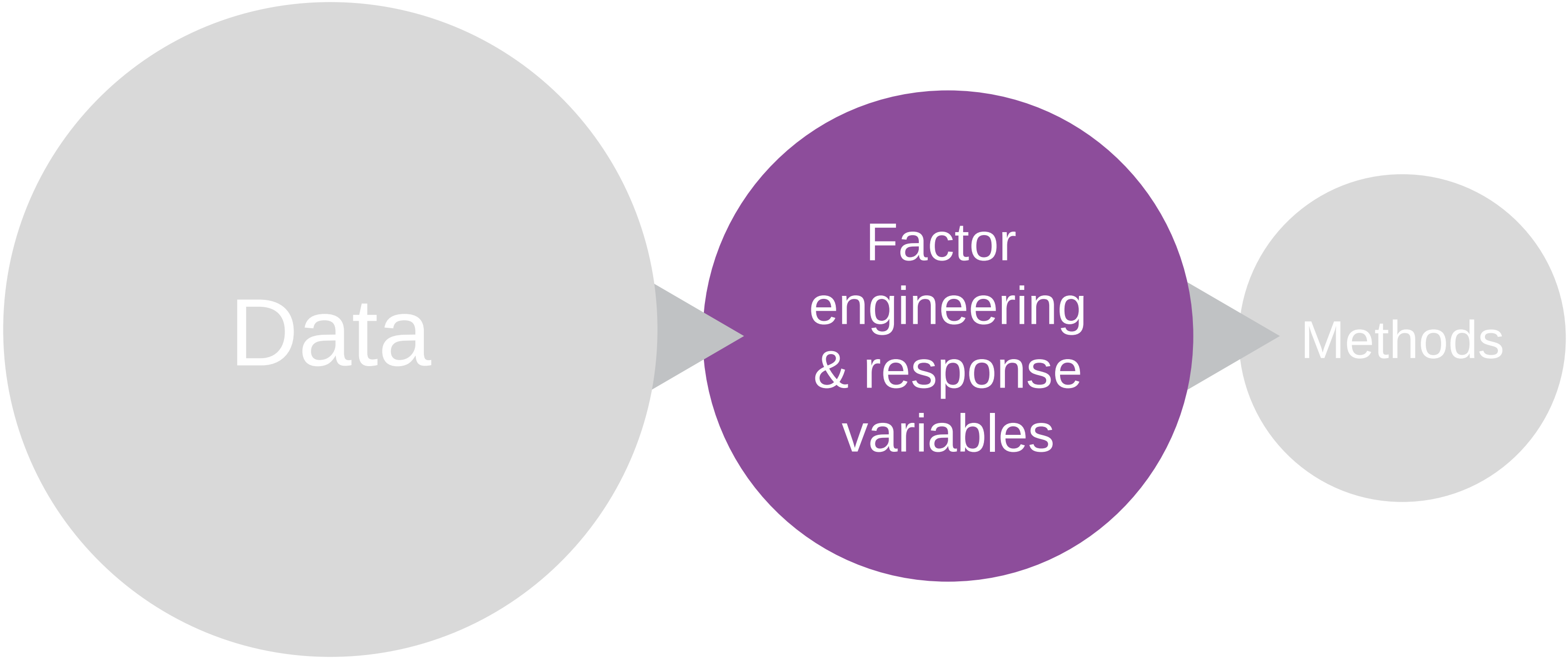
Qualitative descriptors

E.g., body type, model range

Performance

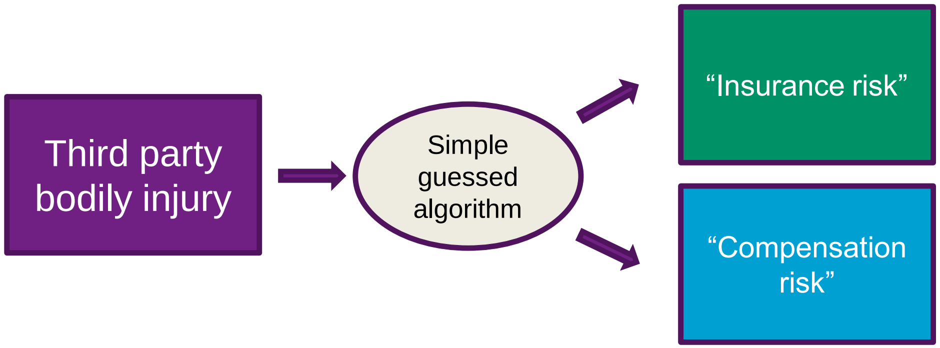
E.g., maximum speed, torque, BHP

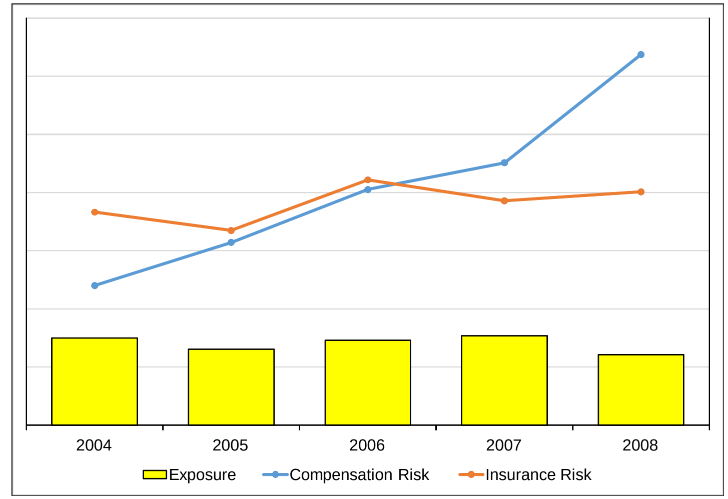
Is it really all about the method?

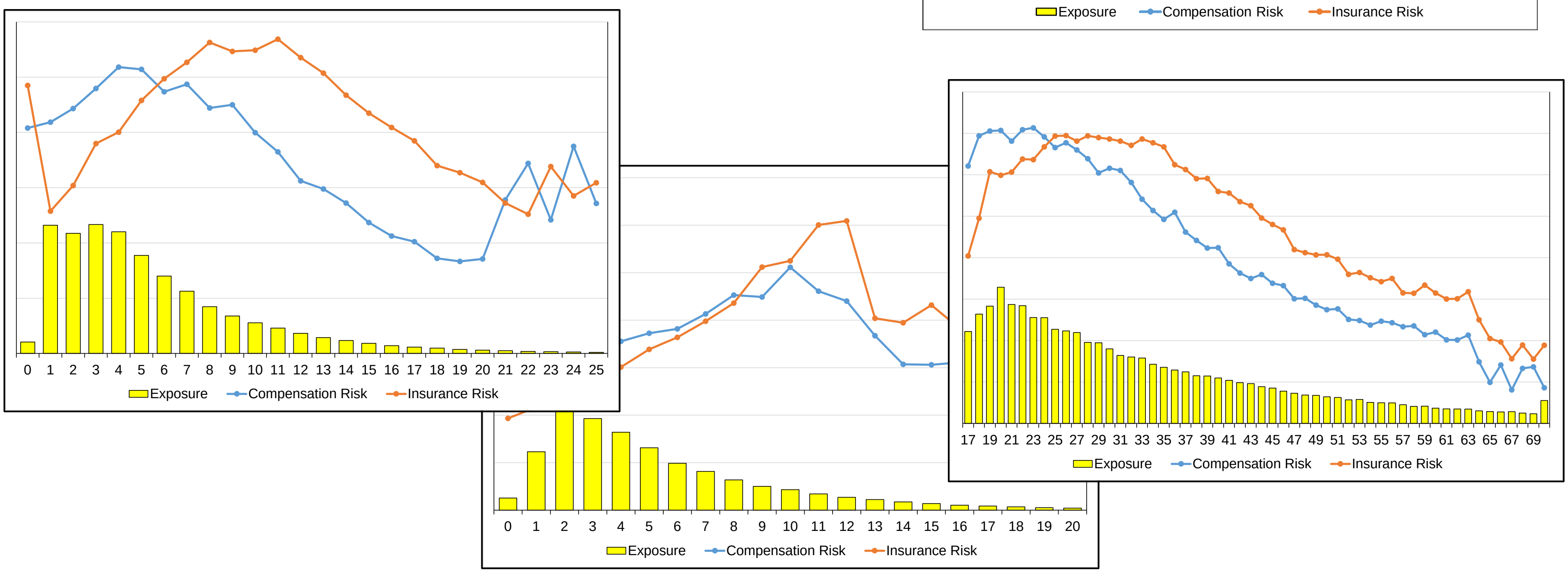


Is it really all about the method?

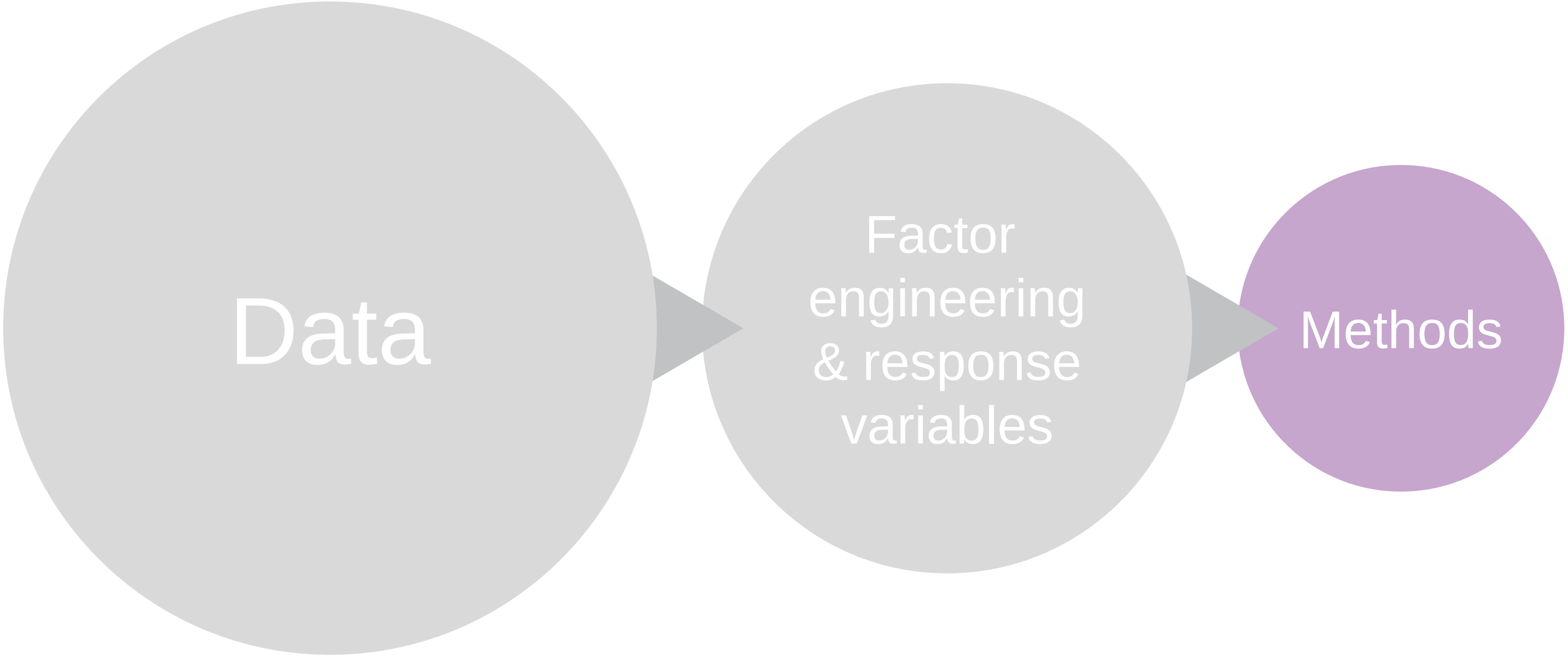
Response selection







Is it really all about the method?

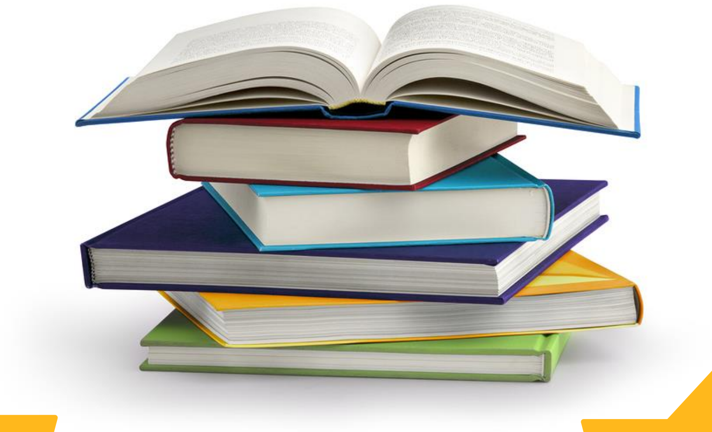


How do you know if a method works?

AIC

Gini

RMSE

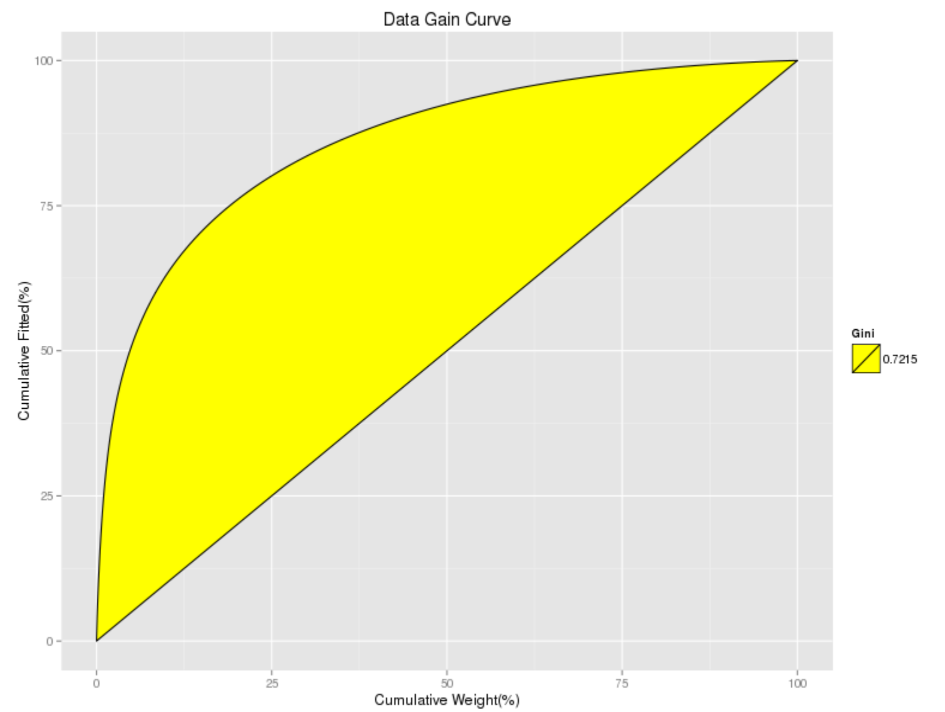


MAE

Log loss

How do you measure value?

Gini



- 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
New Model	0.330

Example results

Model	Gini	Gini improvement
GLM	0.327	0.0%
New Model	0.330	1.0%

Example results

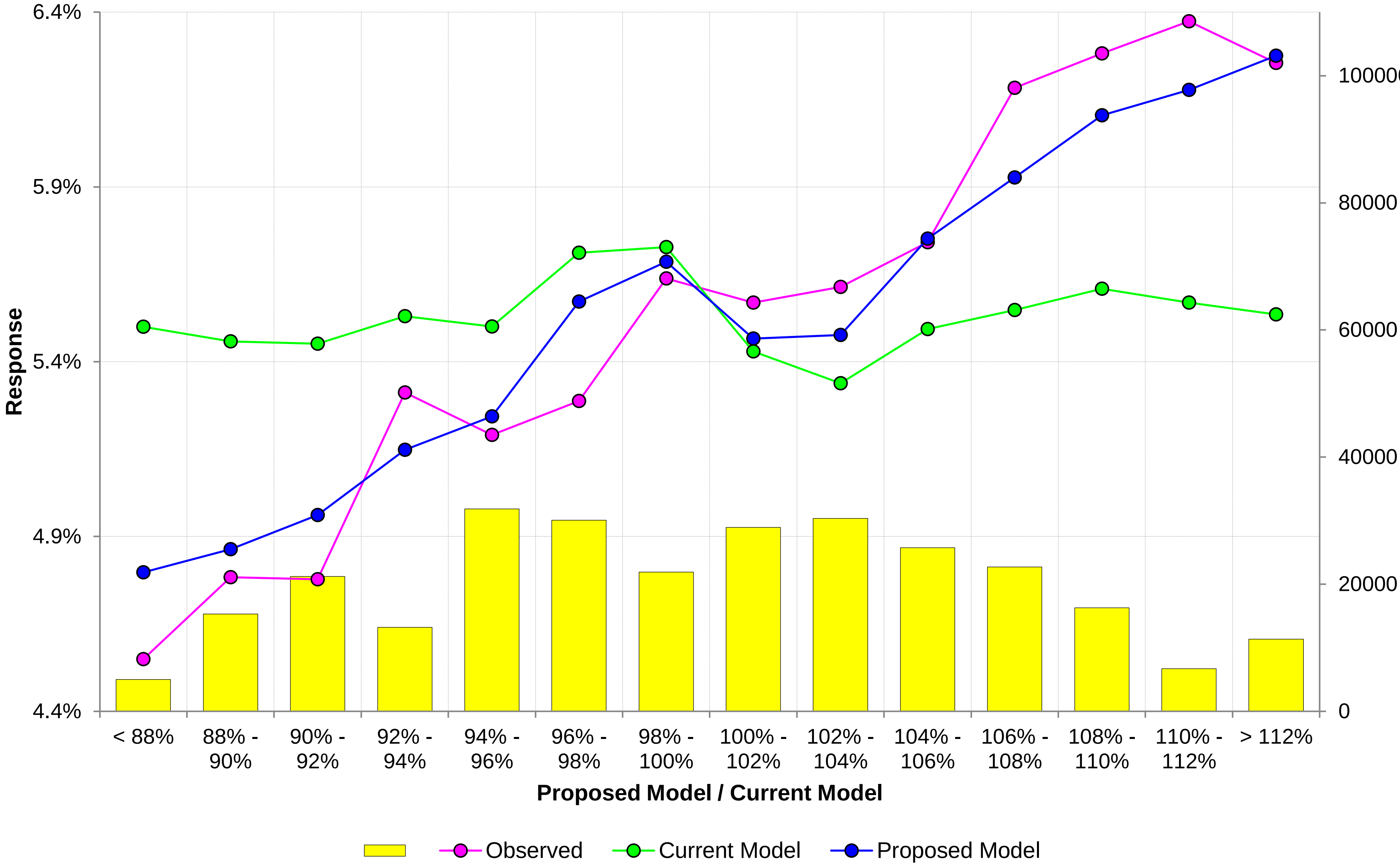
Model	Gini	Gini improvement	Gini rank
GLM (main factor removed)	0.318	-2.6%	4
GLM (minor factor removed)	0.322	-1.3%	3
GLM	0.327	0.0%	2
New Model	0.330	1.0%	1

But…

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

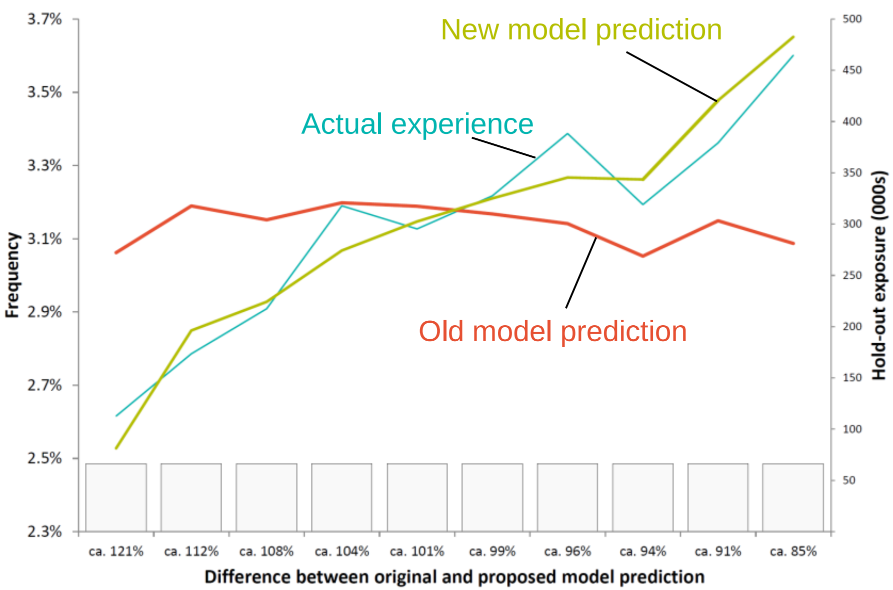
- …and you get the same Gini coefficient!

Double lift chart



Financial value estimate

- Errors in insurance pricing are not symmetrical
- Financial benefit can be estimated
- Consider ***actual experience*** in out of sample data for each percentile of old vs new model fitted values
- Estimate financial benefit that would have been attained
  - given an assumed elasticity
  - given business rules such as an assumed cap/floor approach



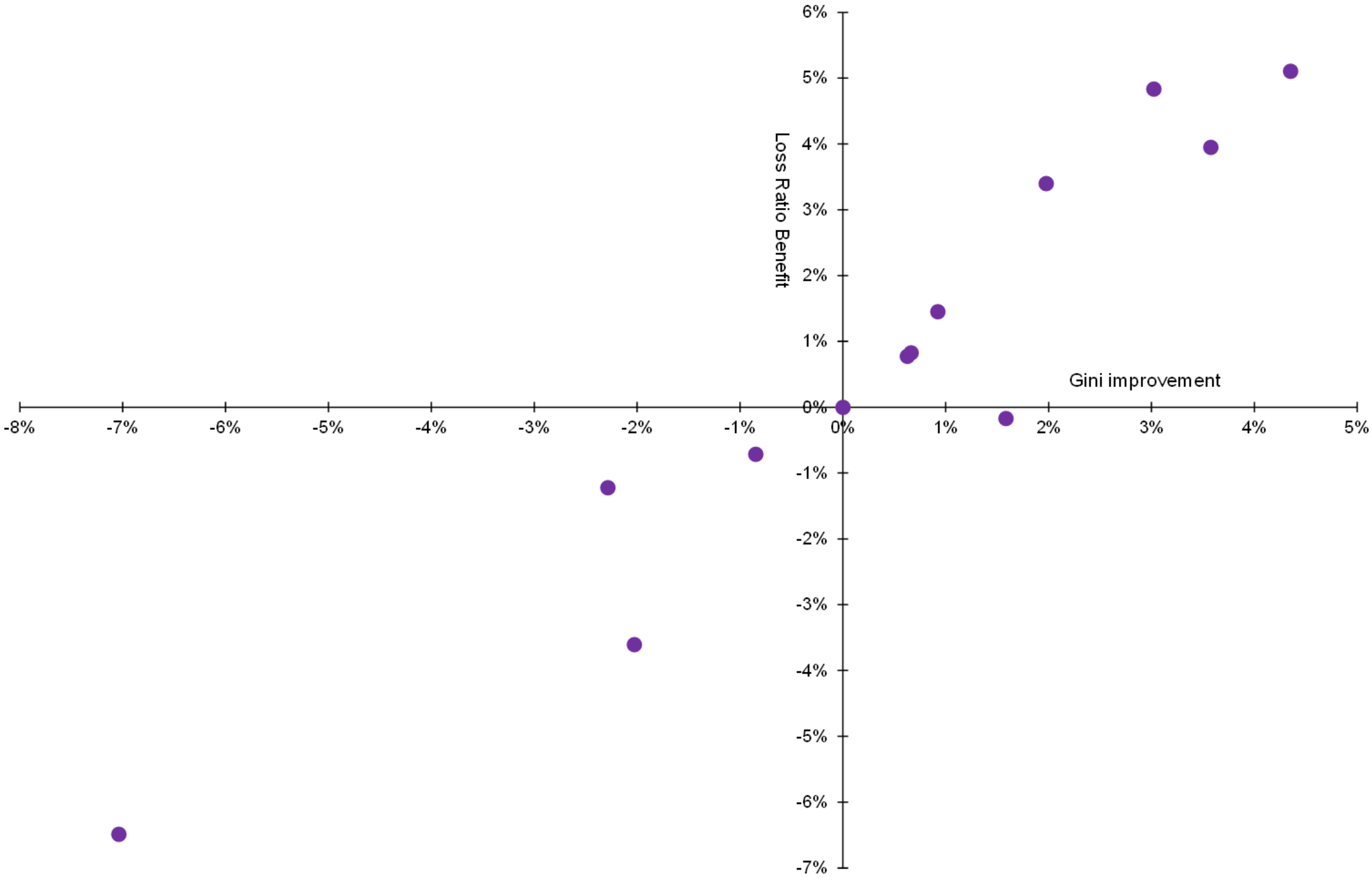
Simple formula

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

Example results

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
New Model	0.330	1.0%	1	2.2%	1	0.5%	1

Financial value vs Gini

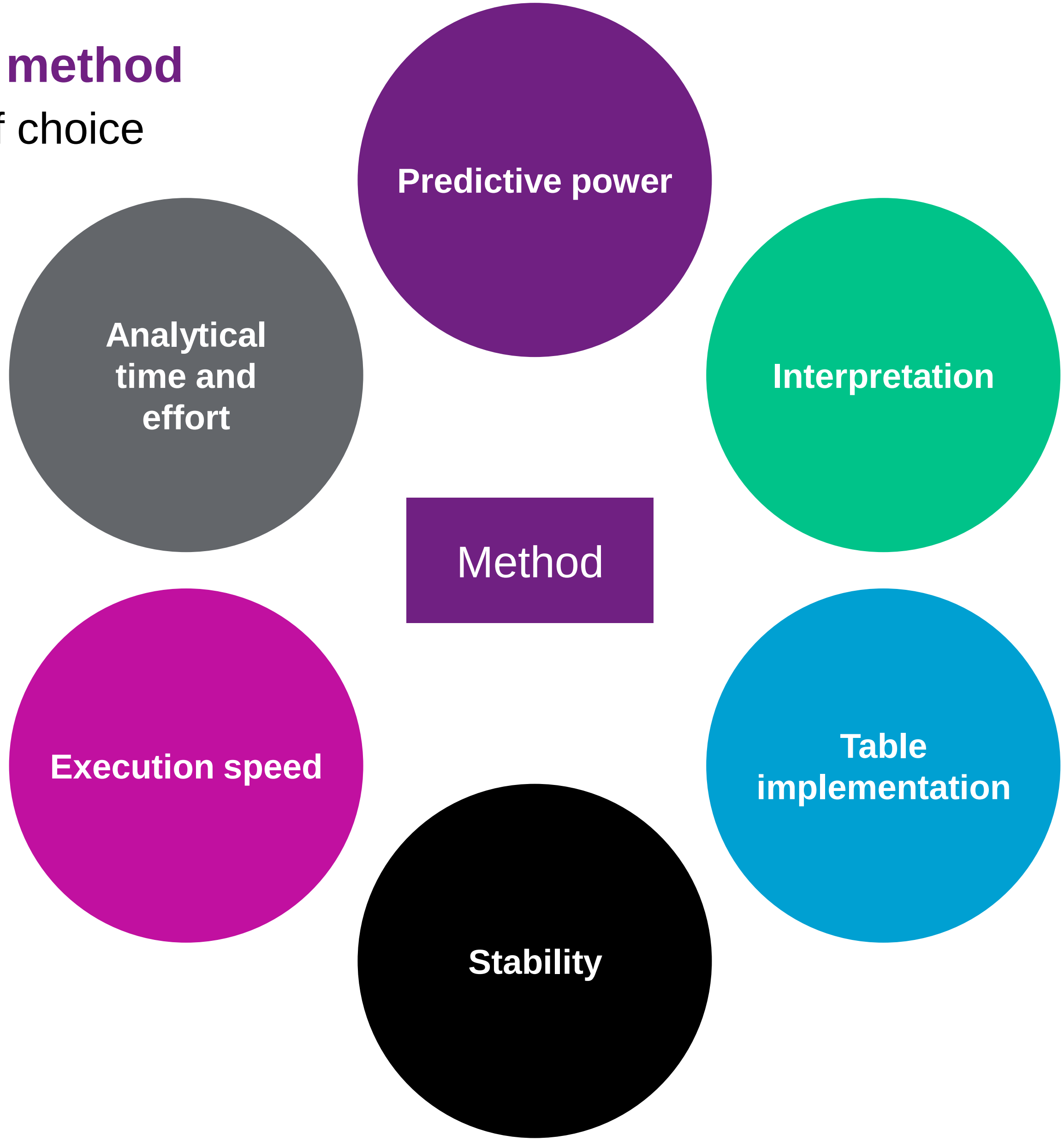


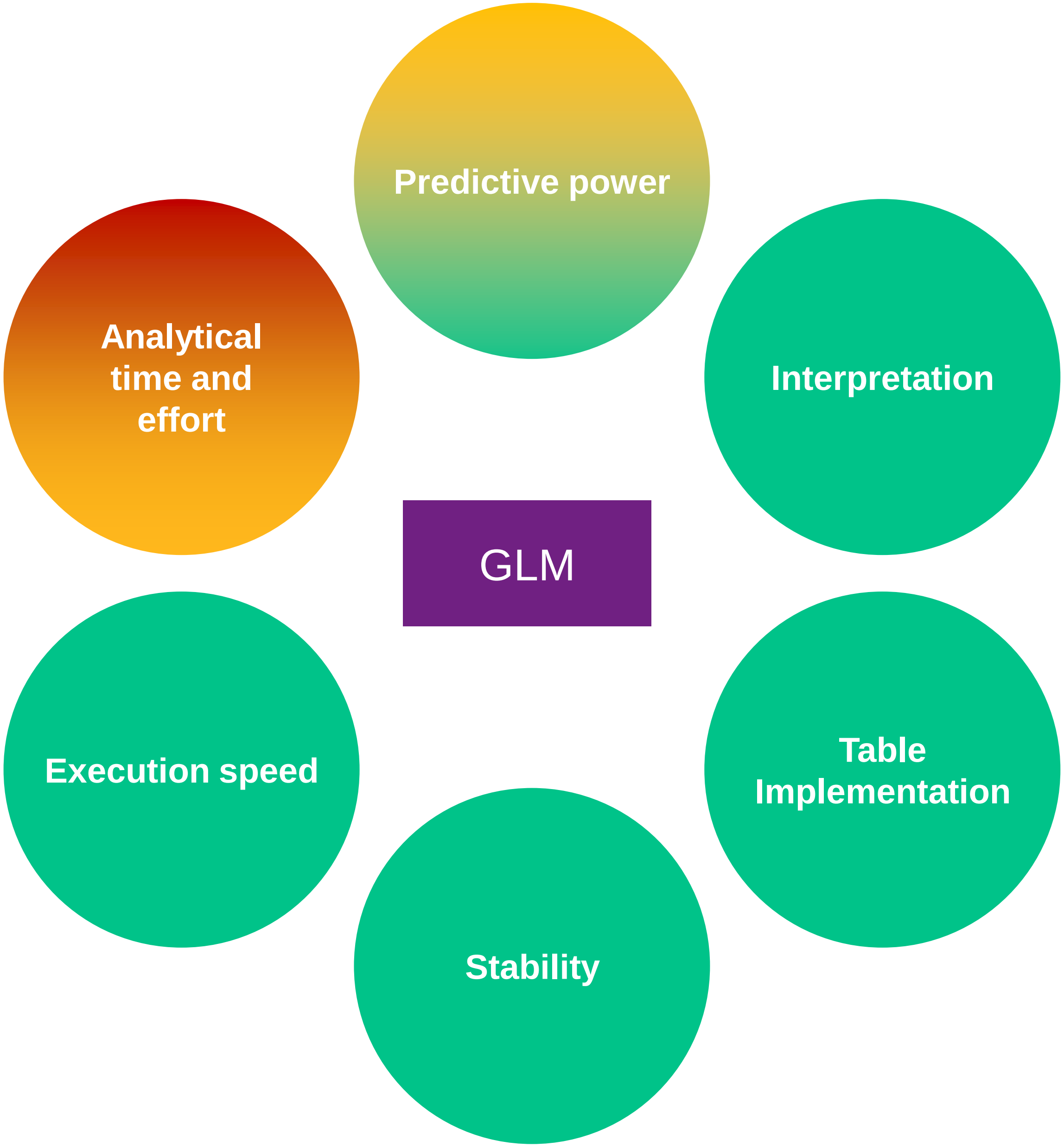
Is there more to it…?

Predictive power

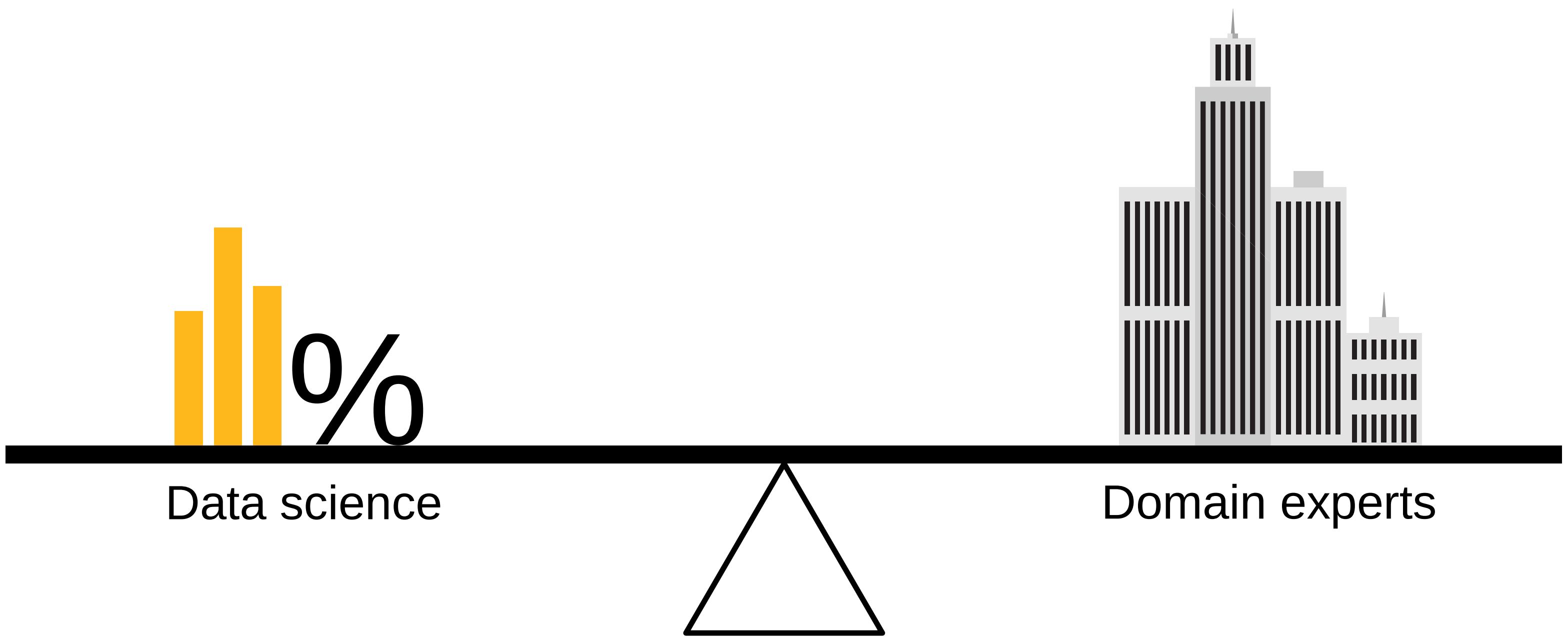
Choosing a method

Dimensions of choice

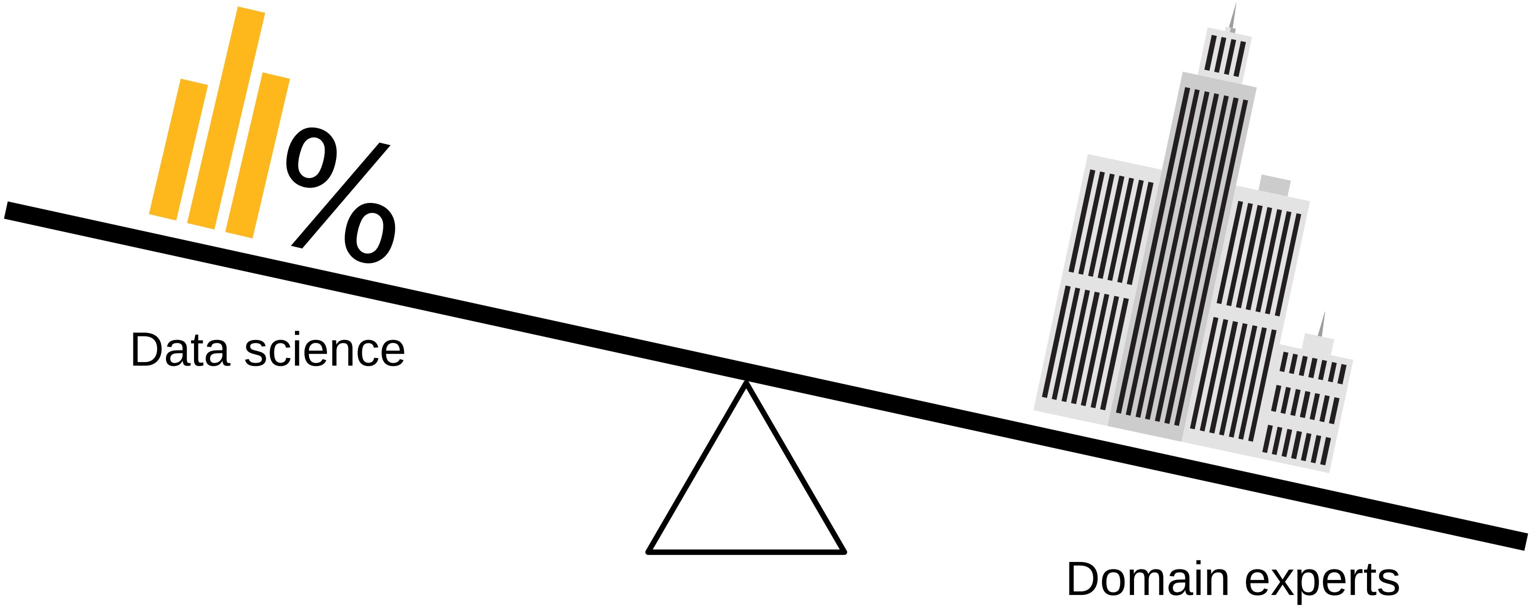



Predictive power
Analytical time and effort
Interpretation
GLM
Execution speed
Table Implementation
Stability

What do you use where?



It’s domain expertise that helps decide



Some machine learning methods

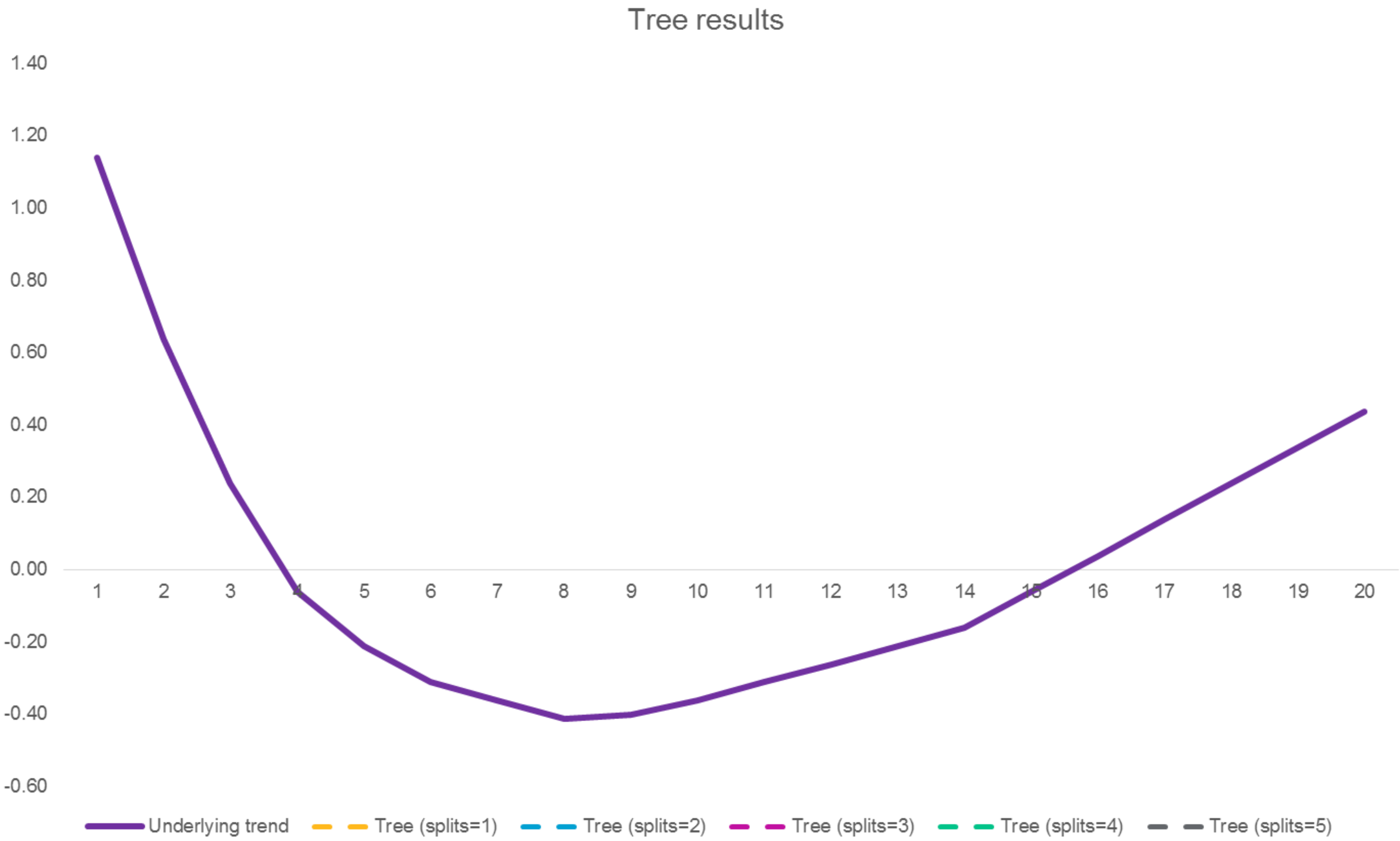
- Ensembles
- Classifications Trees
- "Earth"
- Regression Trees
- Gradient Boosting Machines
- K-nearest Neighbors
- Elastic Net
- Neural Networks
- Naïve Bayes
- Random Forests
- K-Means Clustering
- Principal Components Analysis
- Lasso
- Support Vector Machines
- Ridge Regression

Focus on Trees

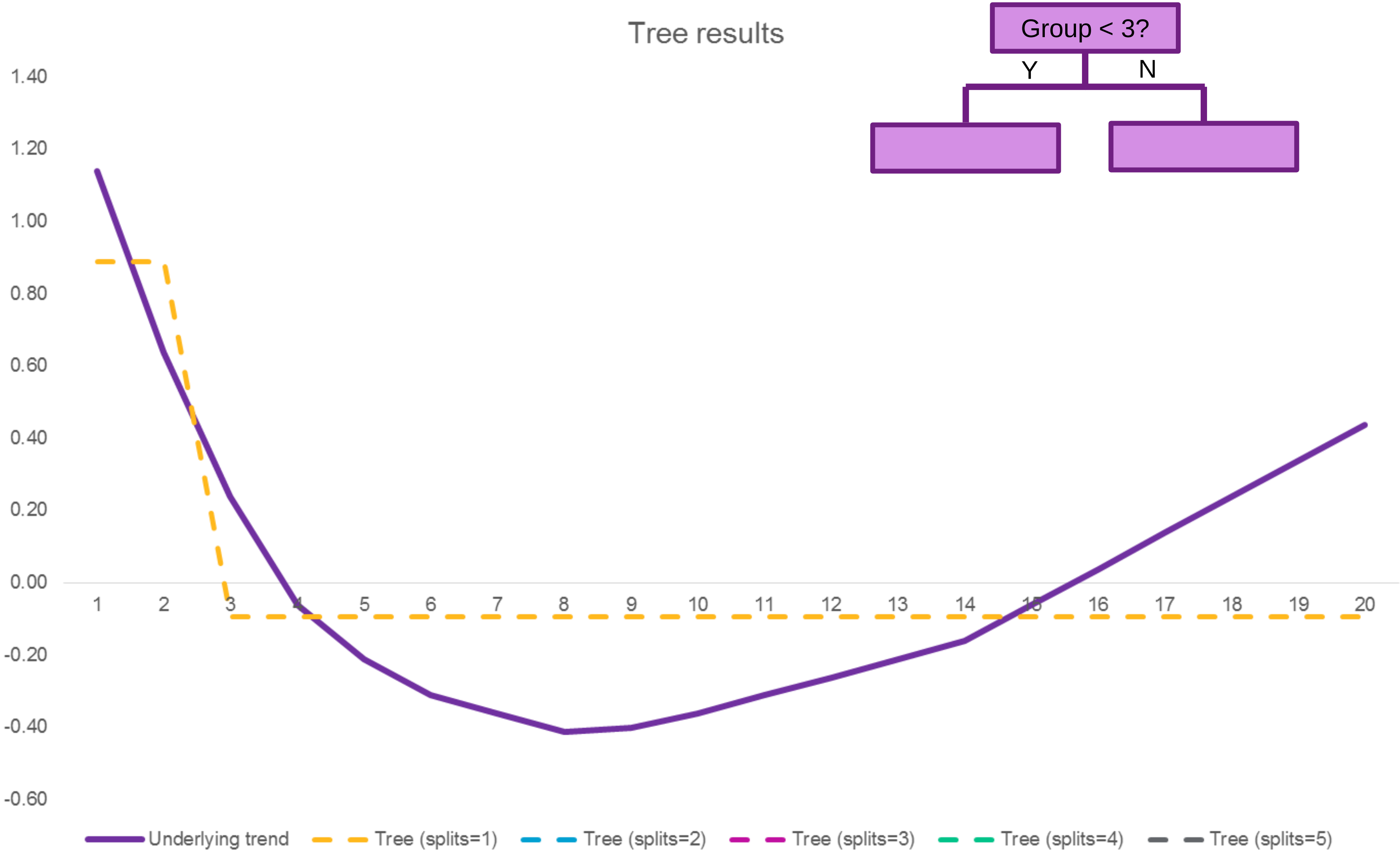
Classifications Trees

Regression Trees

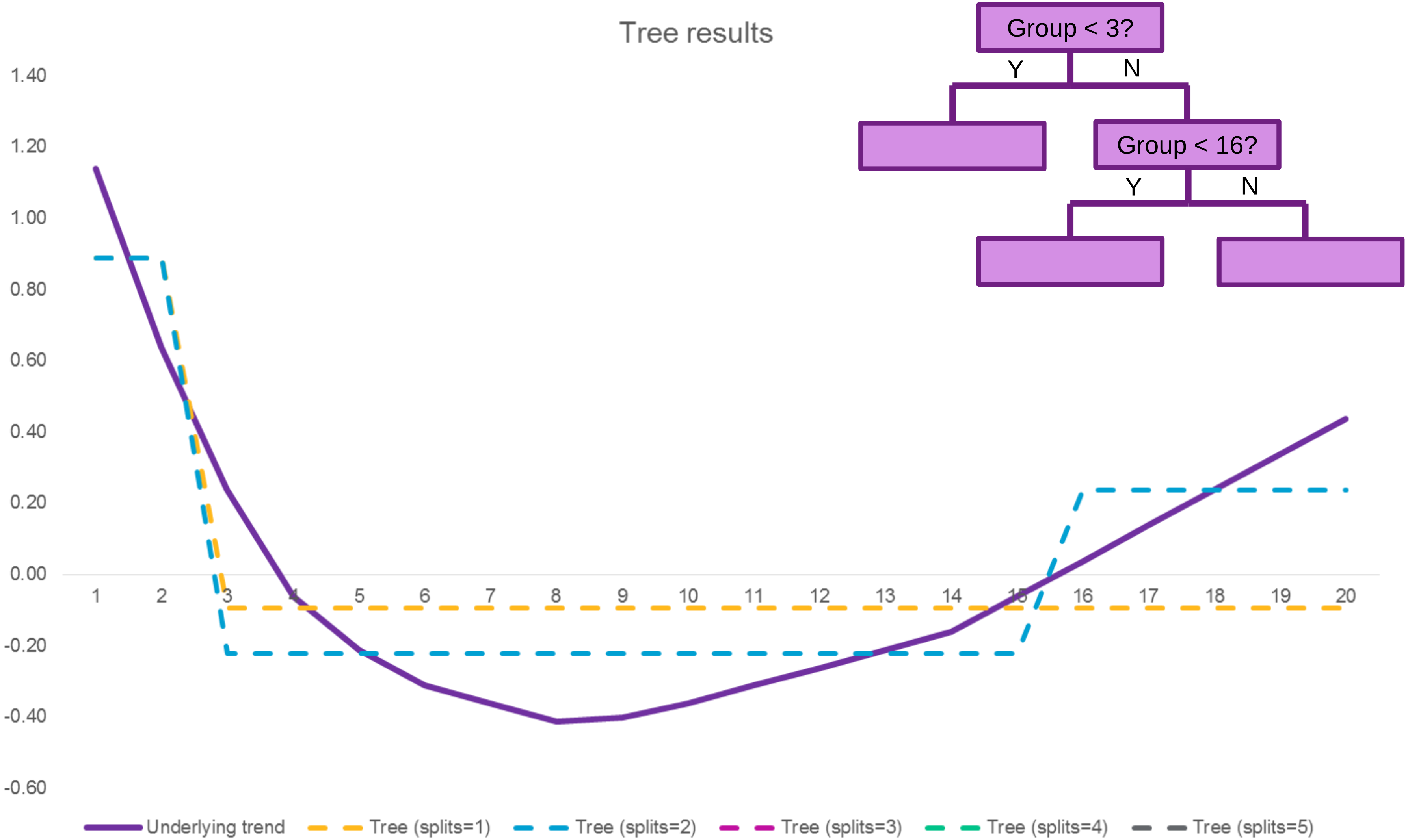
A simple Tree example



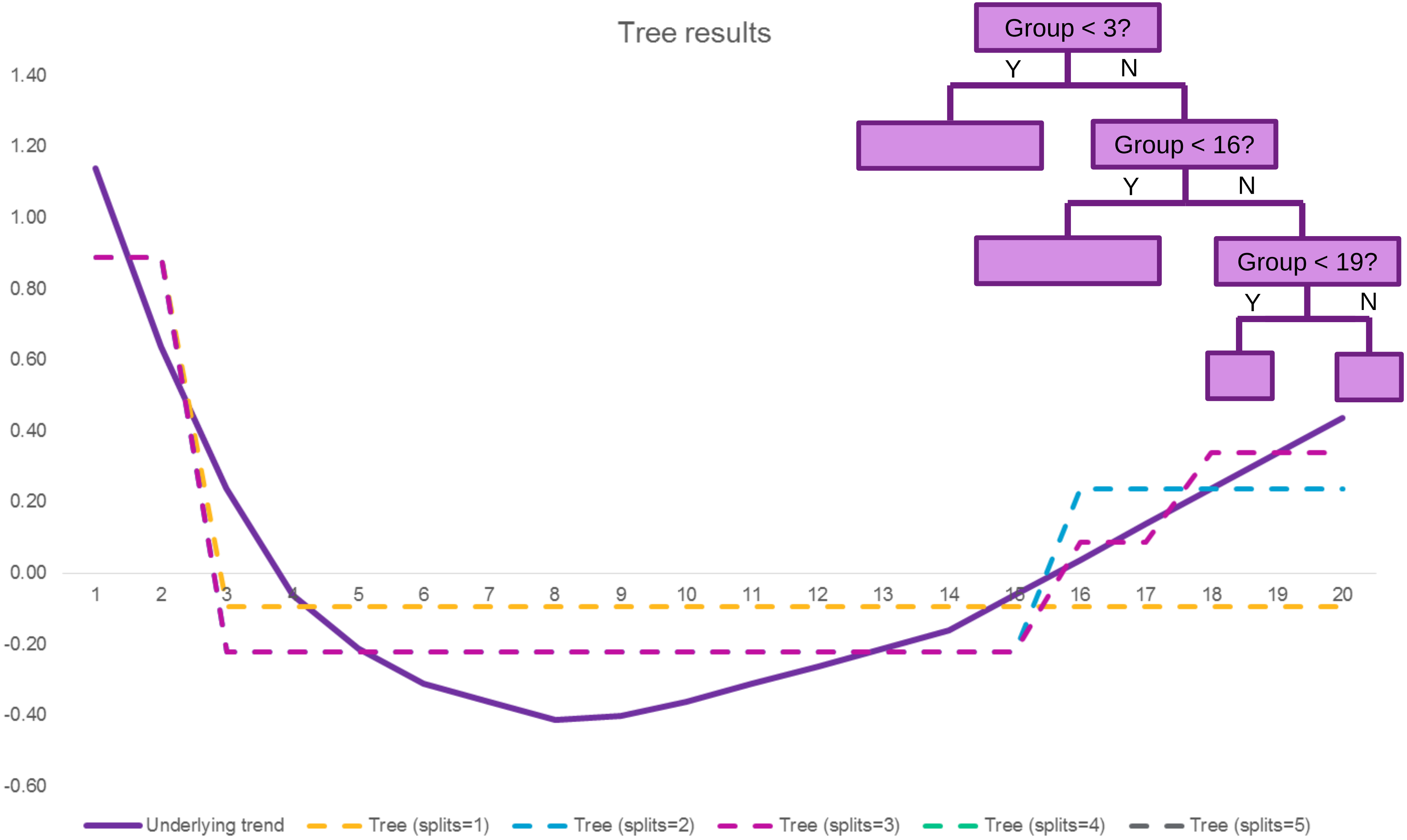
A simple Tree example



A simple Tree example



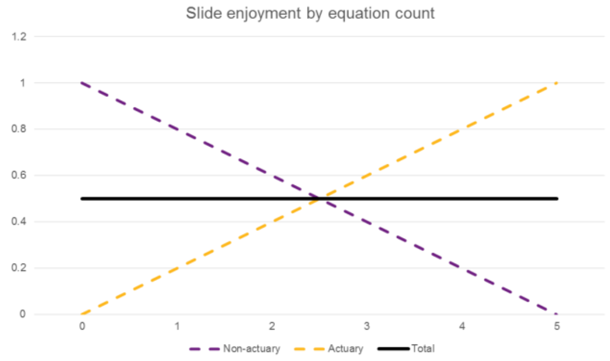
A simple Tree example



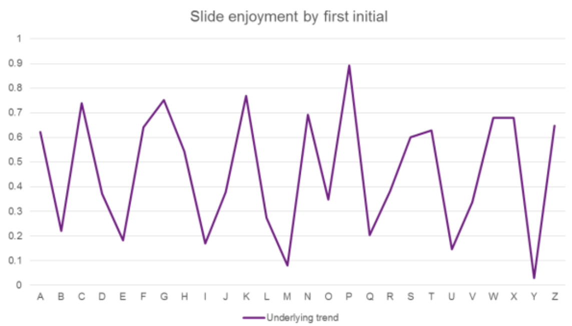
Shortcomings of using trees

Summary

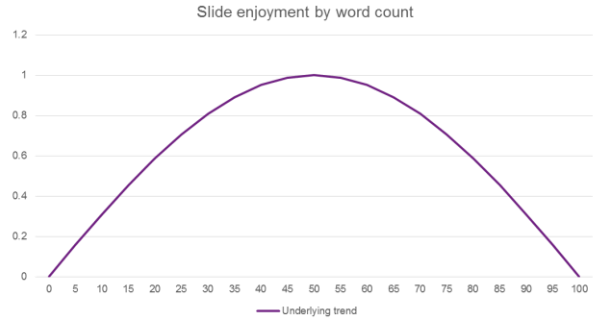
They may miss interactions…

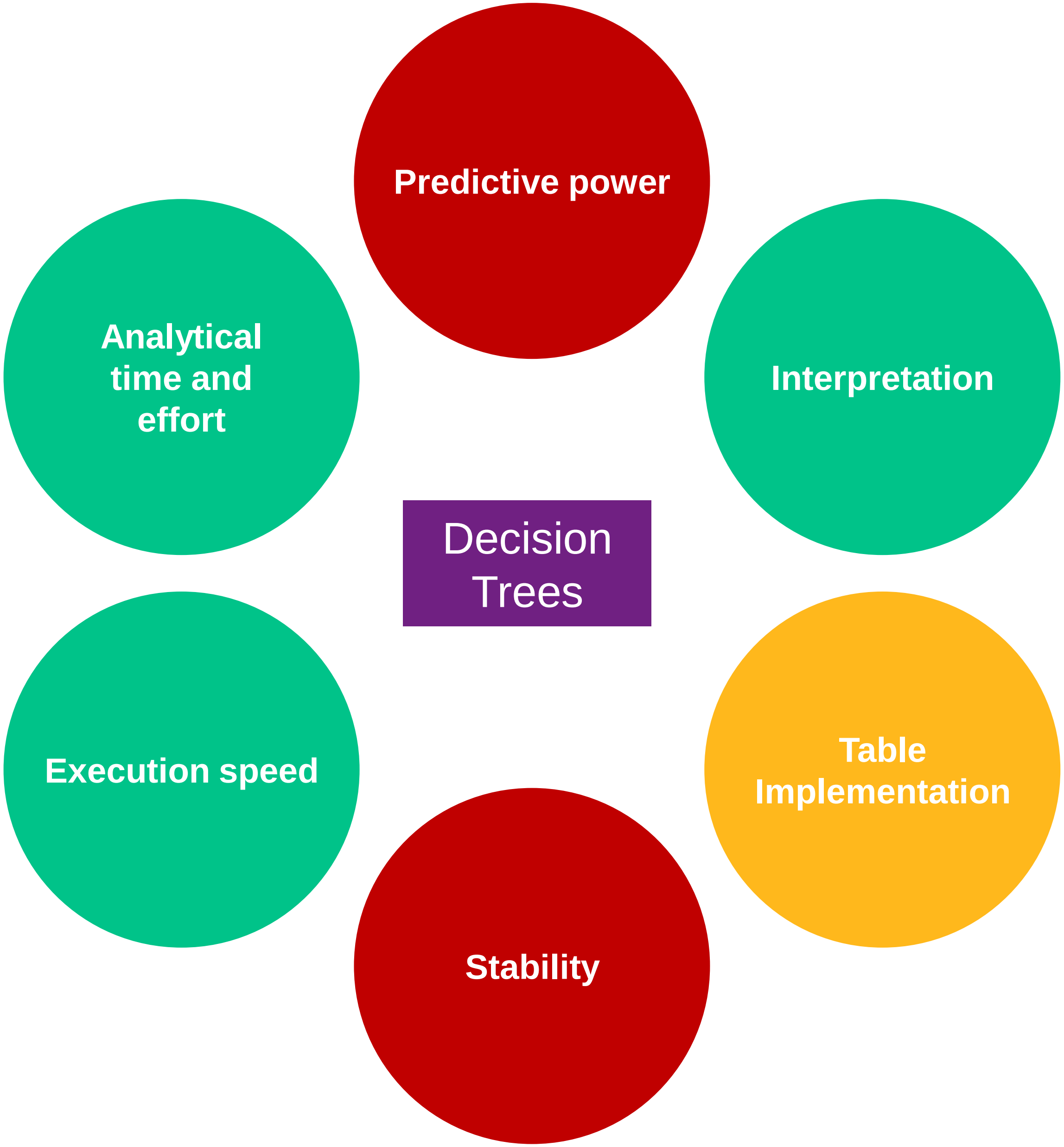


… they may struggles with categorical variables….



…and they can be bad at turning points




Predictive power
Analytical time and effort
Interpretation
Decision Trees
Execution speed
Table Implementation
Stability

Some machine learning methods

- Ensembles
- Classifications Trees
- "Earth"
- 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

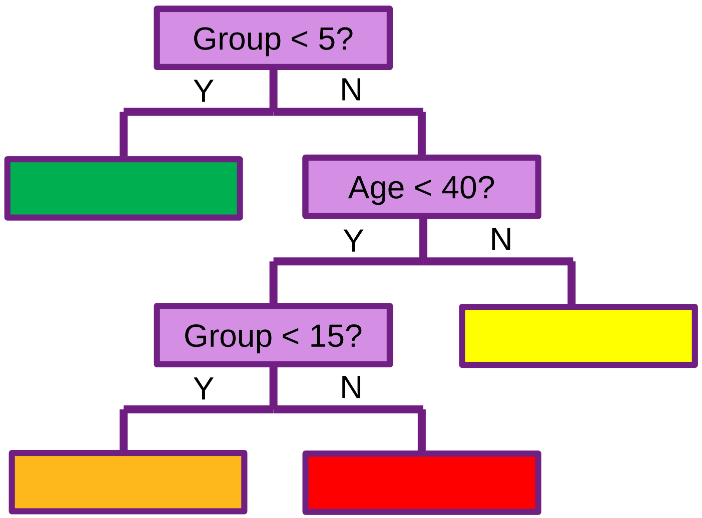
Focus on Gradient Boosting Machines

Gradient
Boosting
Machines

Gradient Boosted Machine or “GBM”

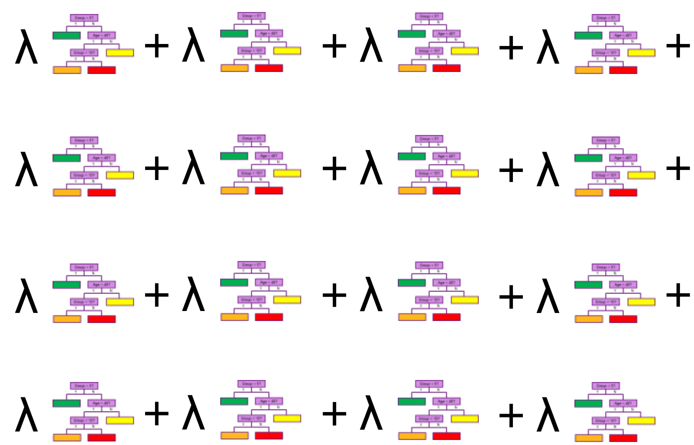
A tree

$$f_i(x)$$



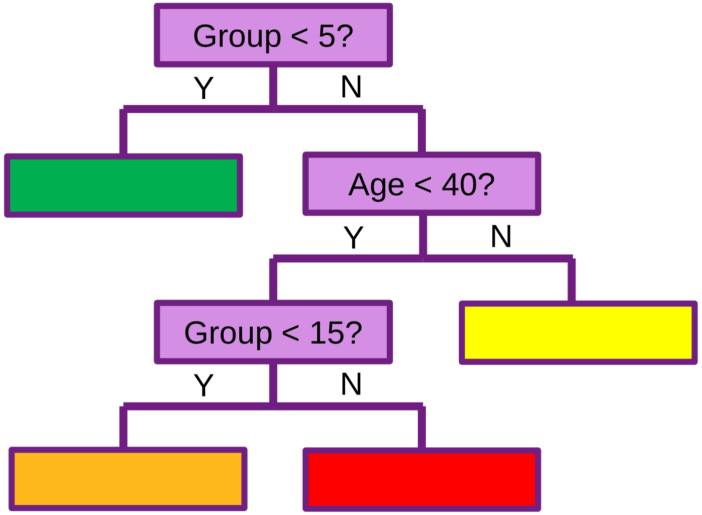
A GBM

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

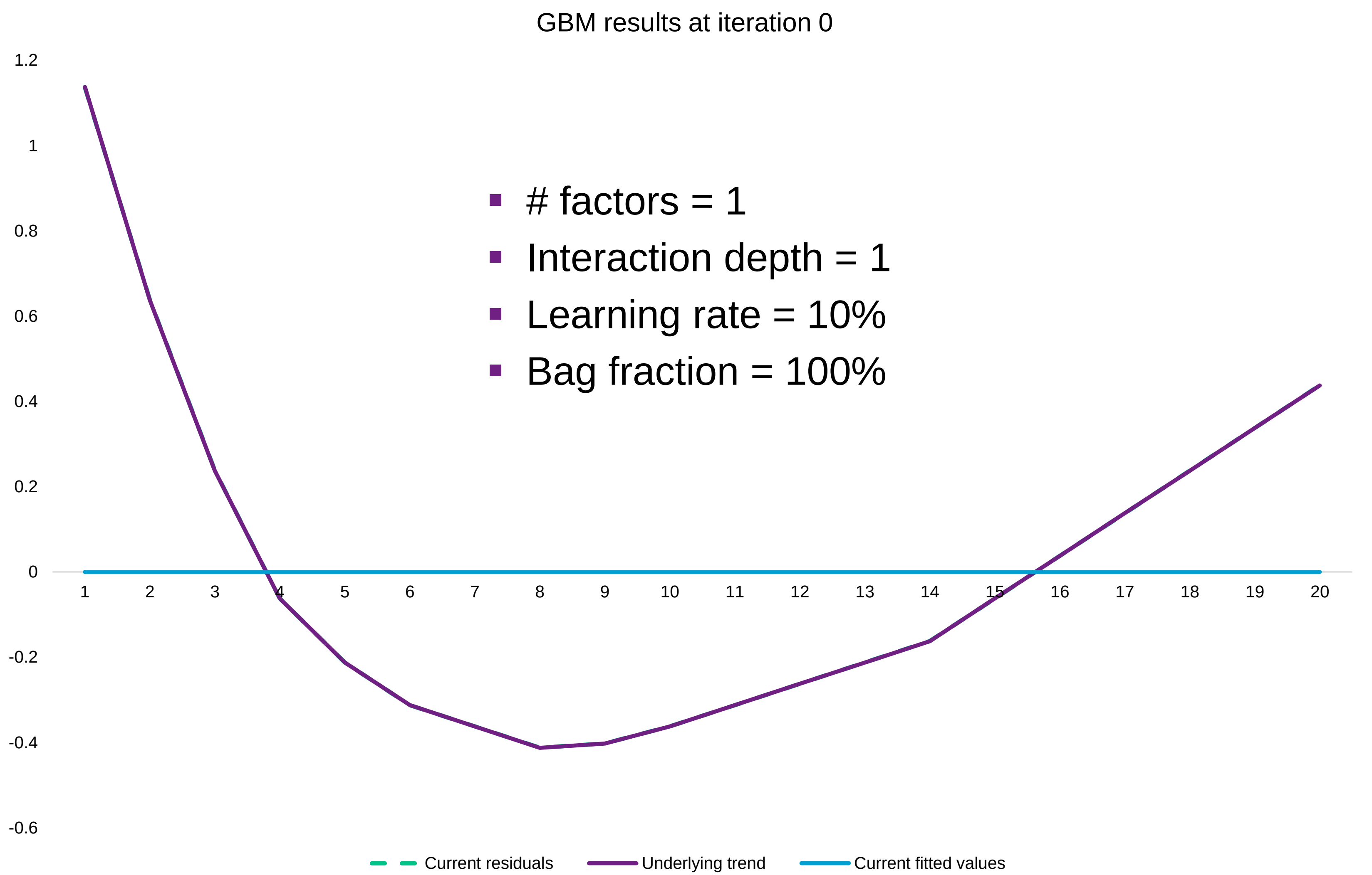


Four main assumptions

- λ **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)
- **N 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

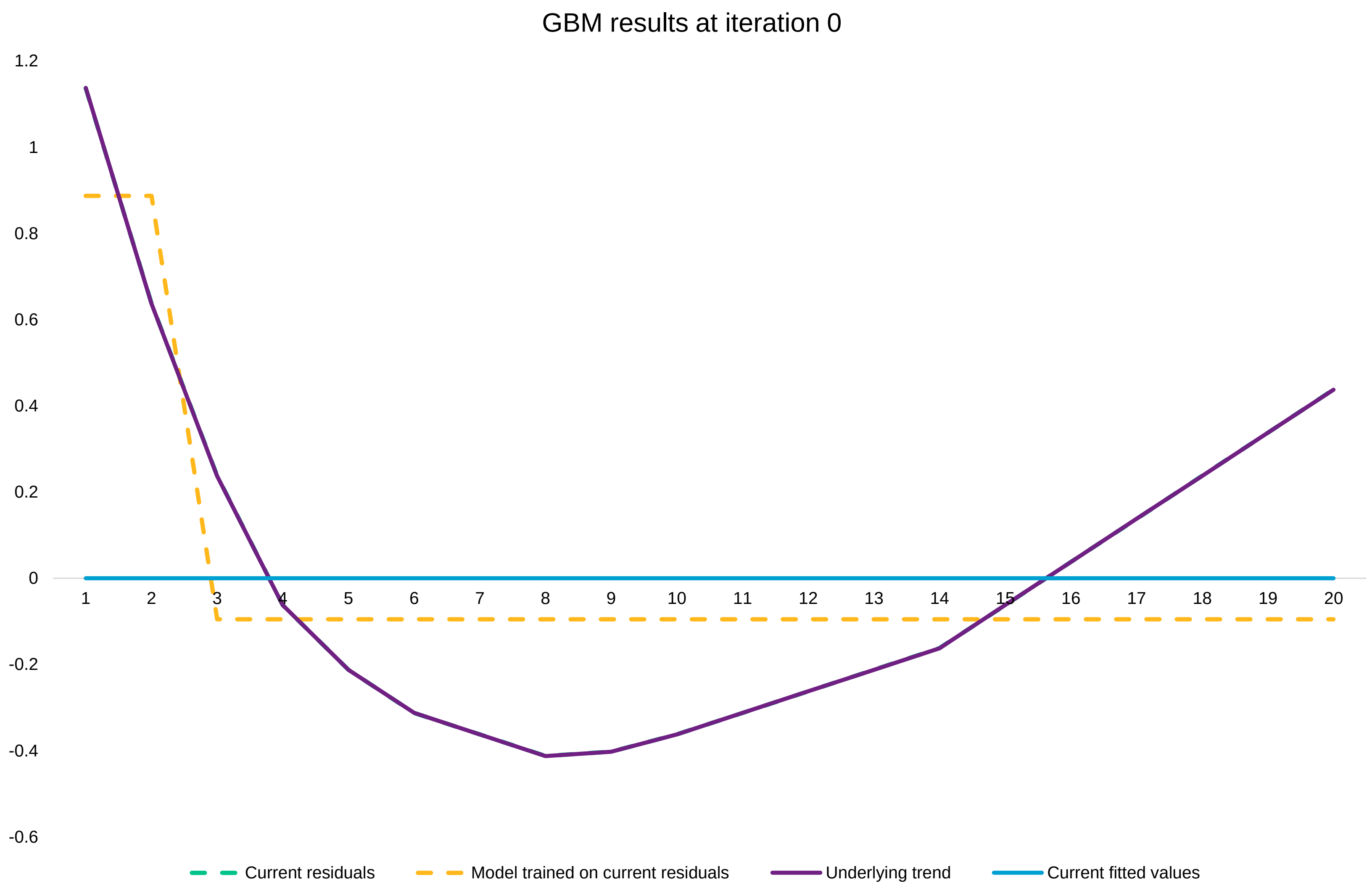


A simple GBM example

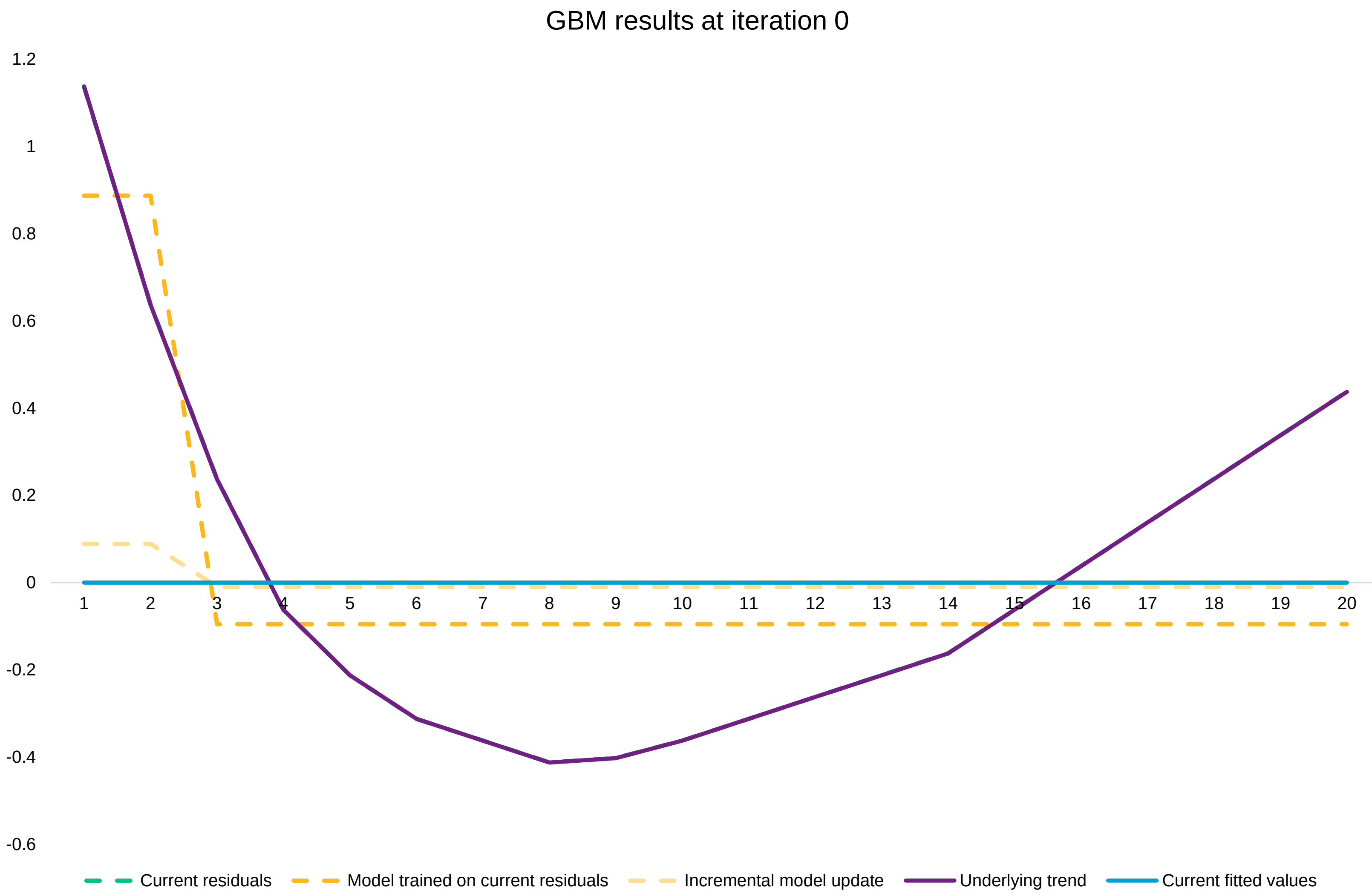


- # factors = 1
- Interaction depth = 1
- Learning rate = 10%
- Bag fraction = 100%

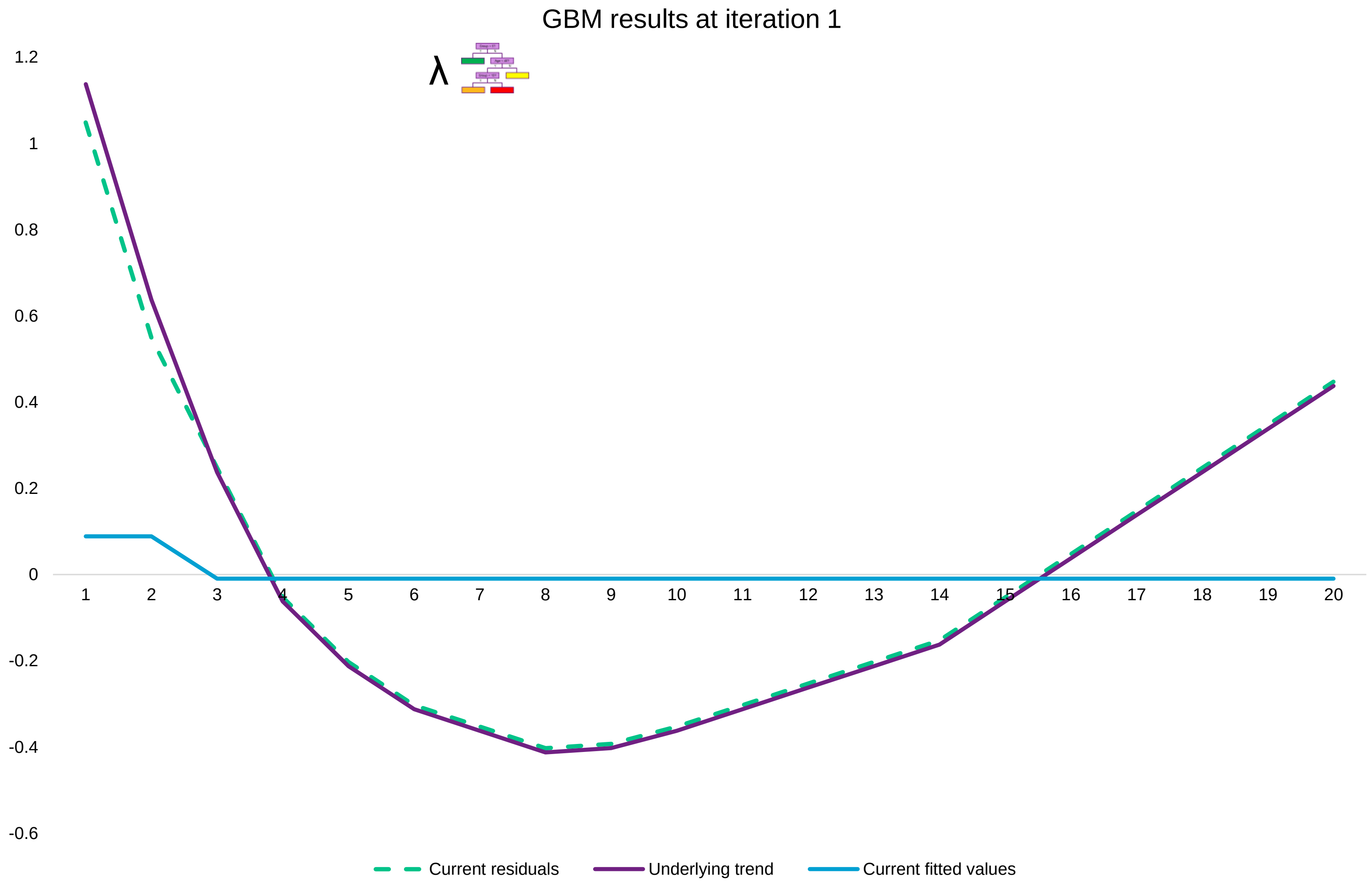
A simple GBM example



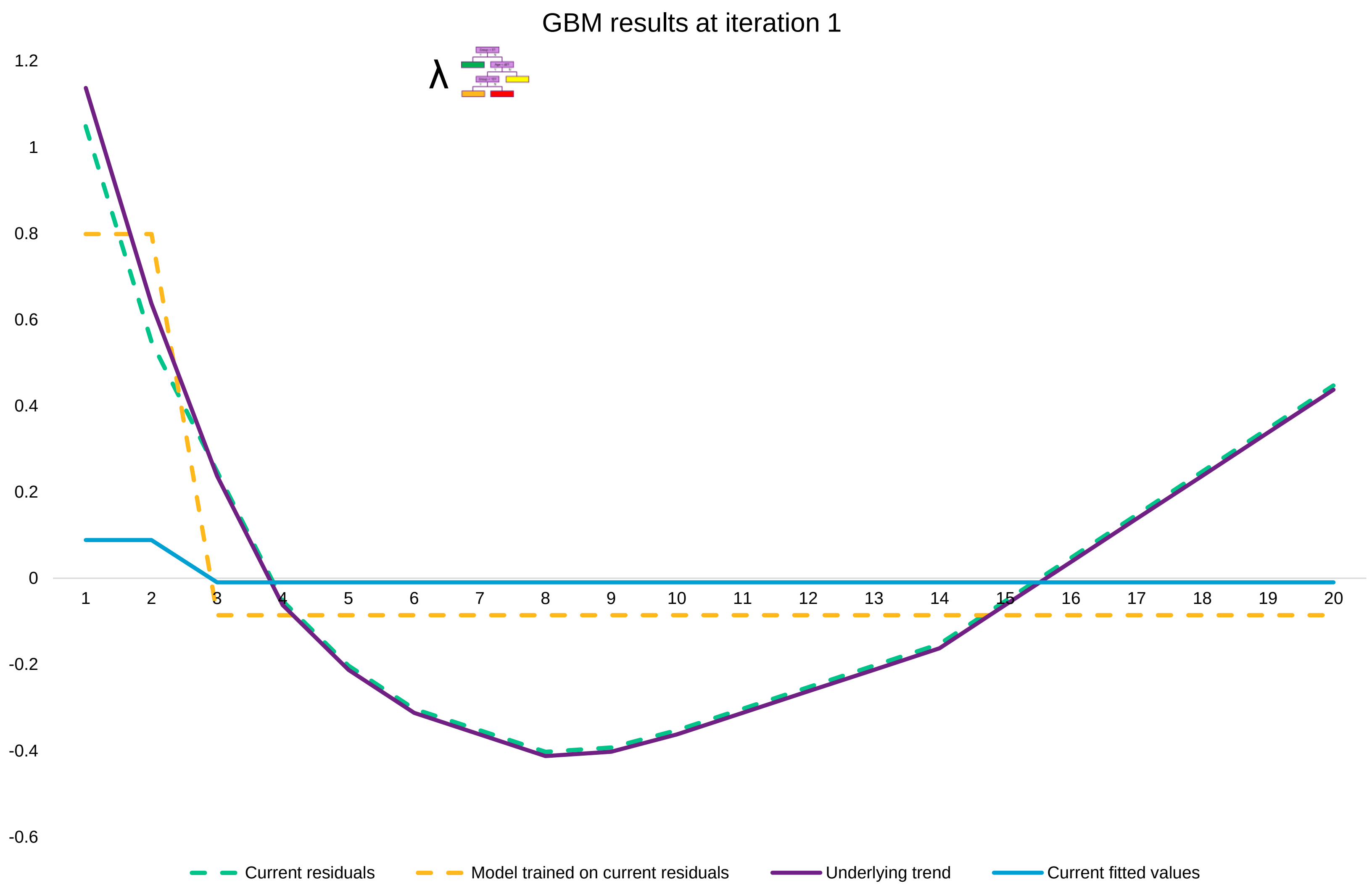
A simple GBM example



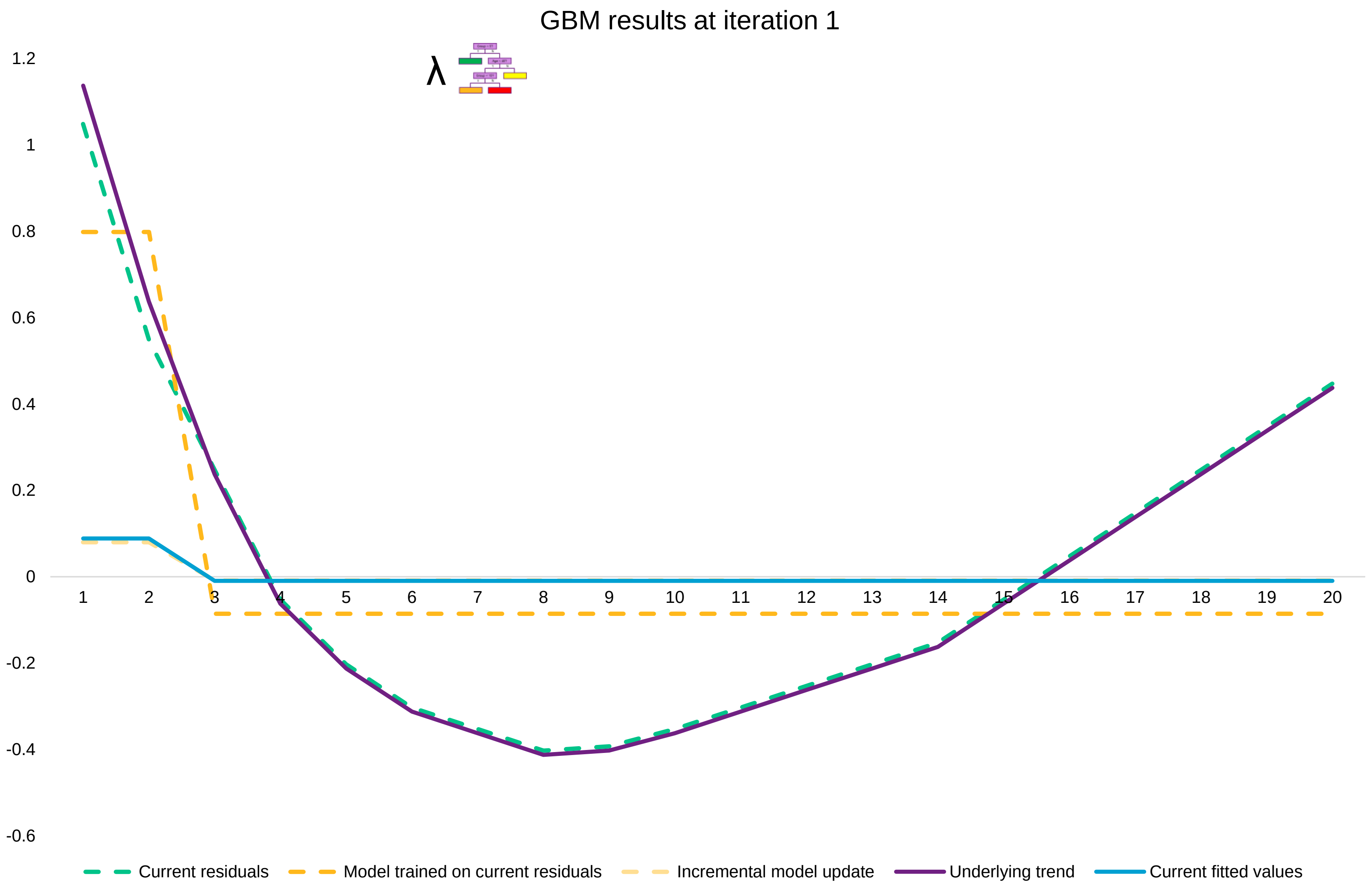
A simple GBM example



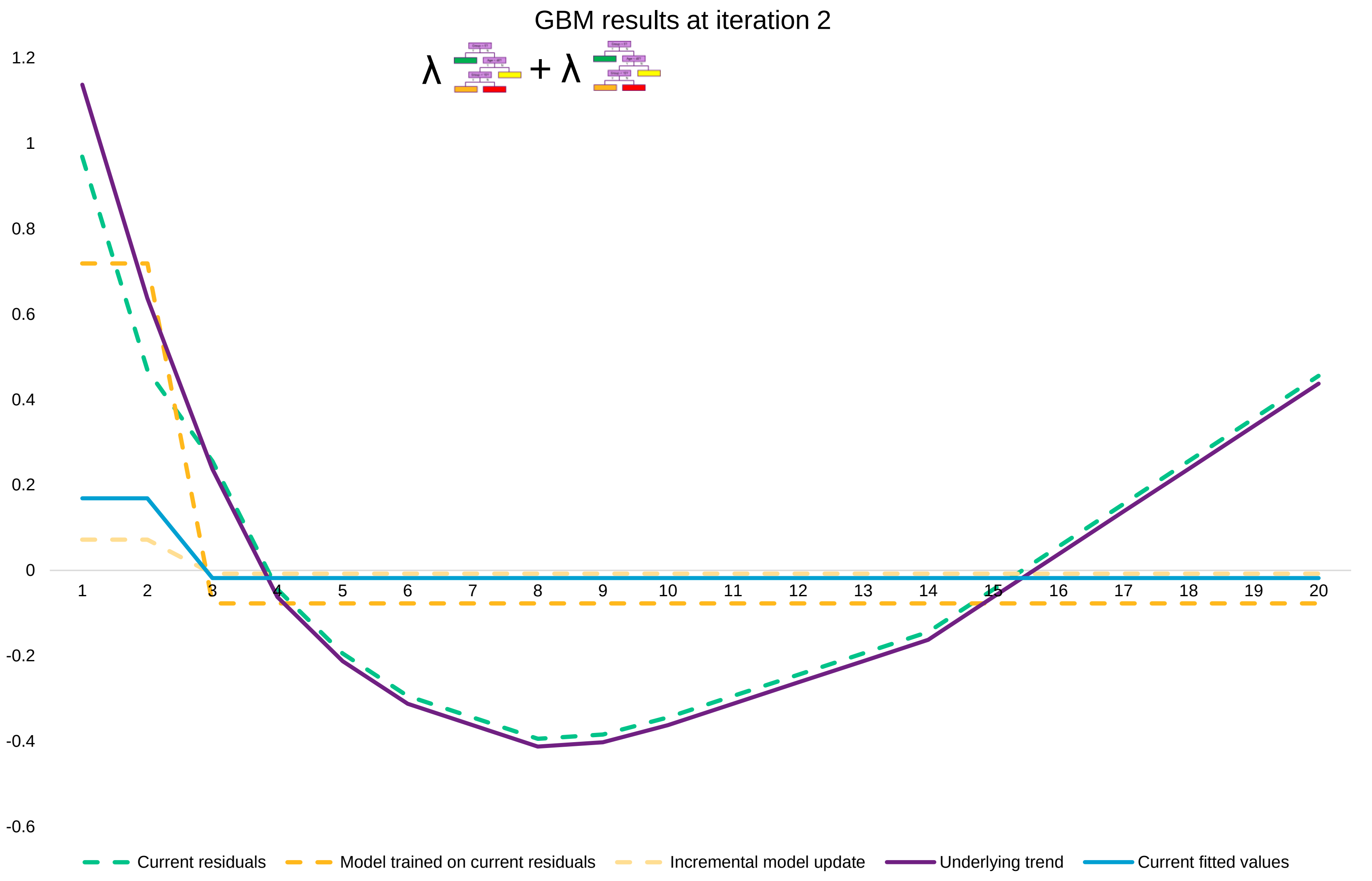
A simple GBM example



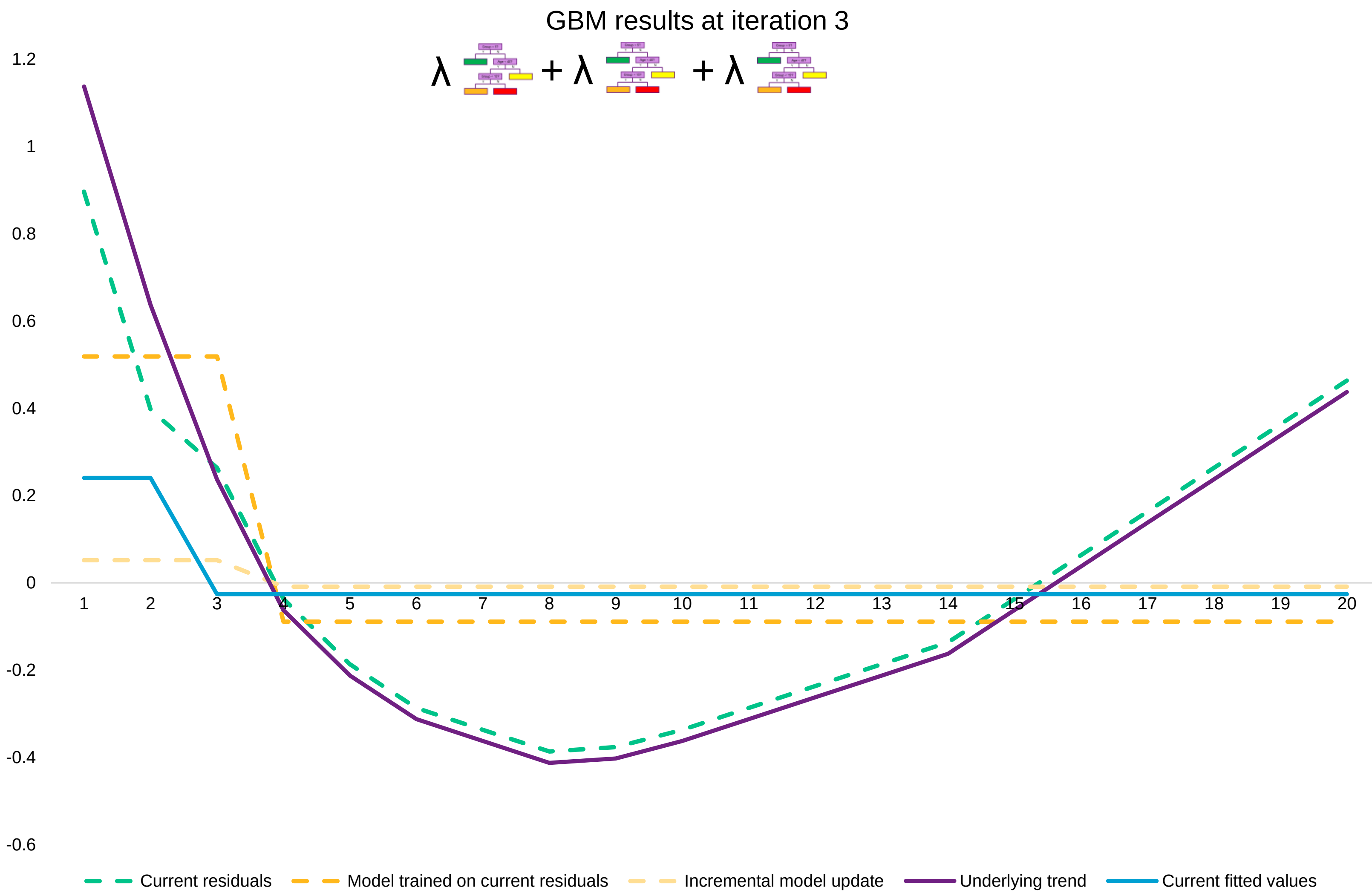
A simple GBM example



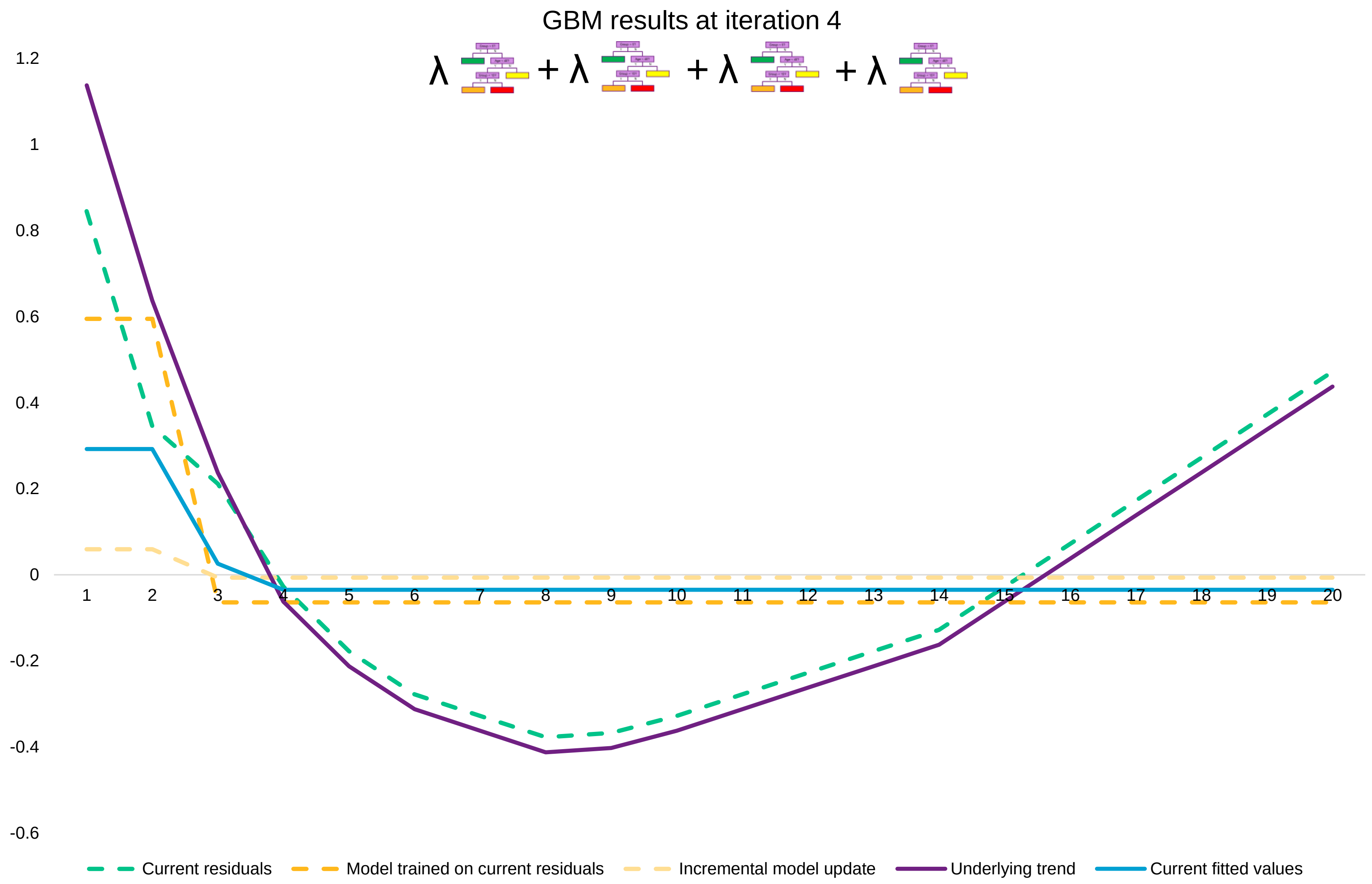
A simple GBM example



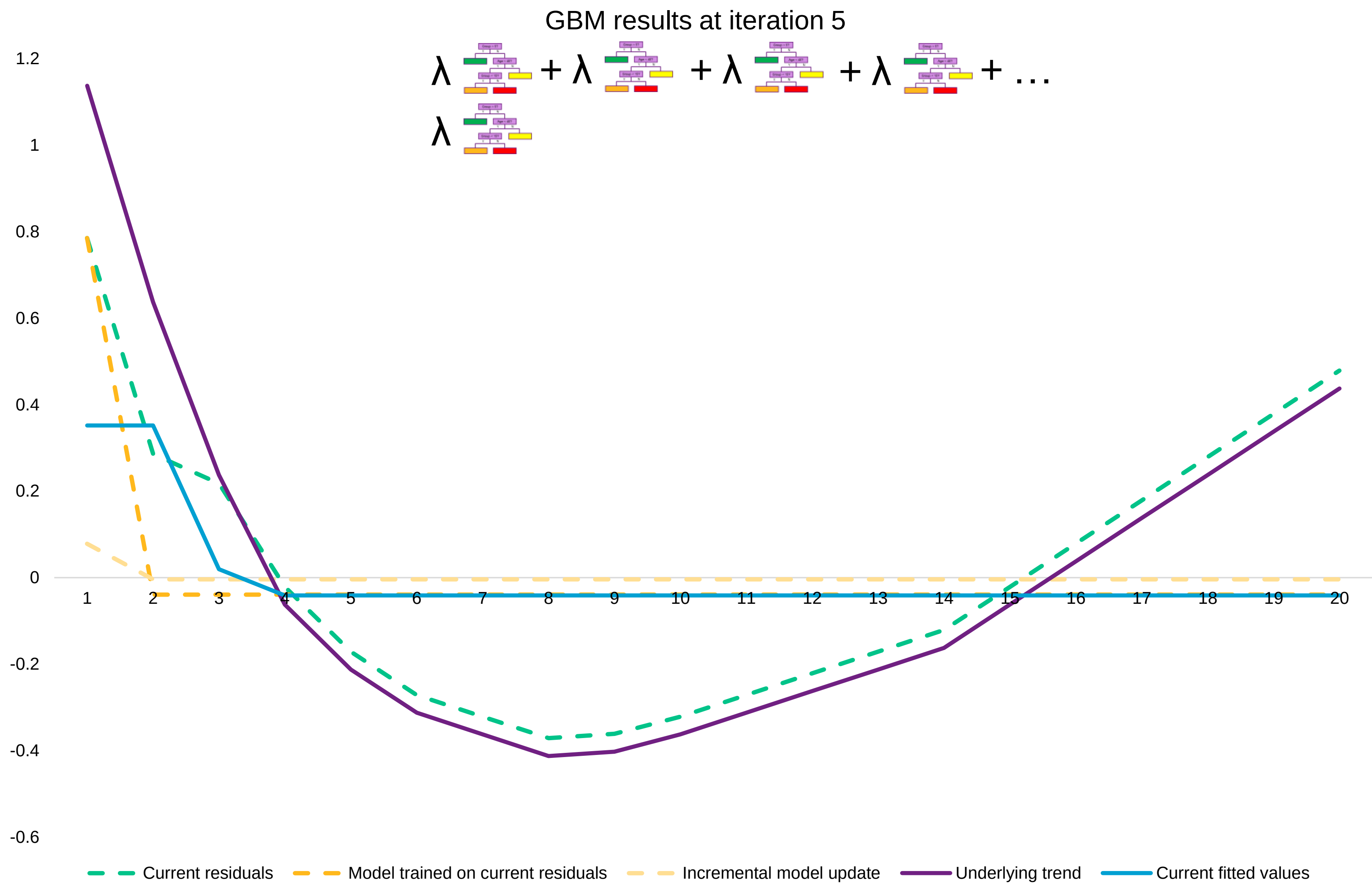
A simple GBM example



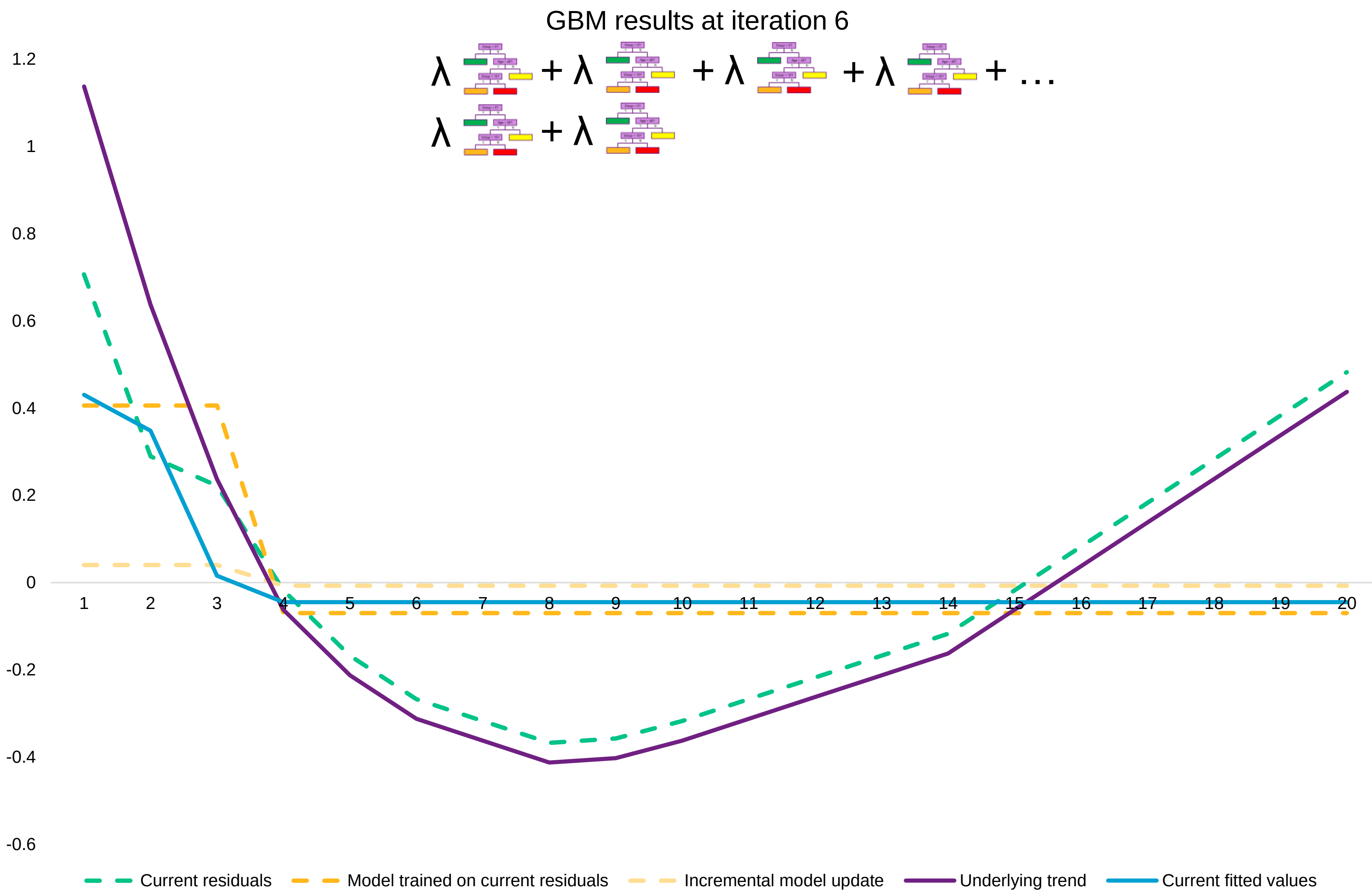
A simple GBM example



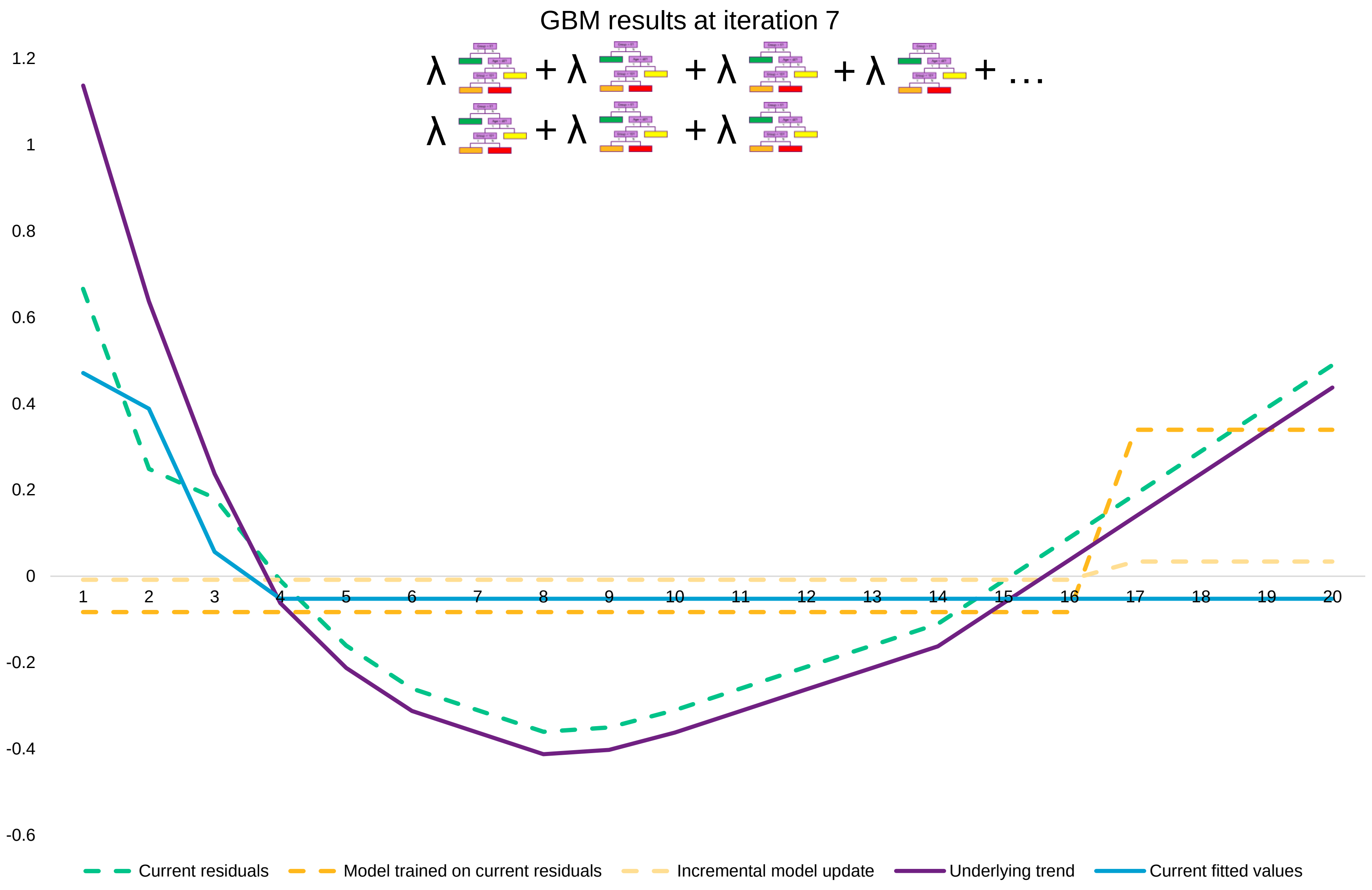
A simple GBM example



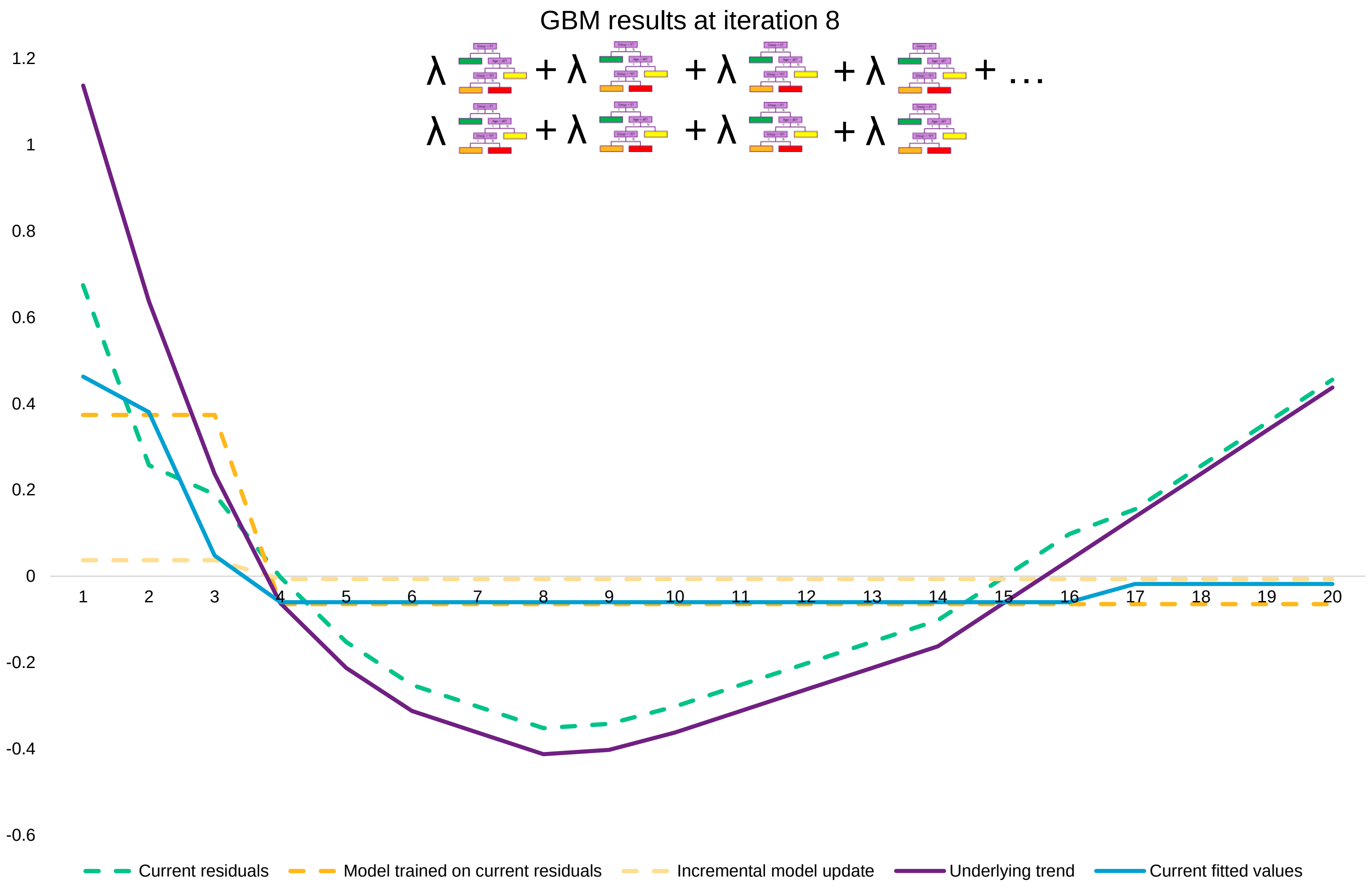
A simple GBM example



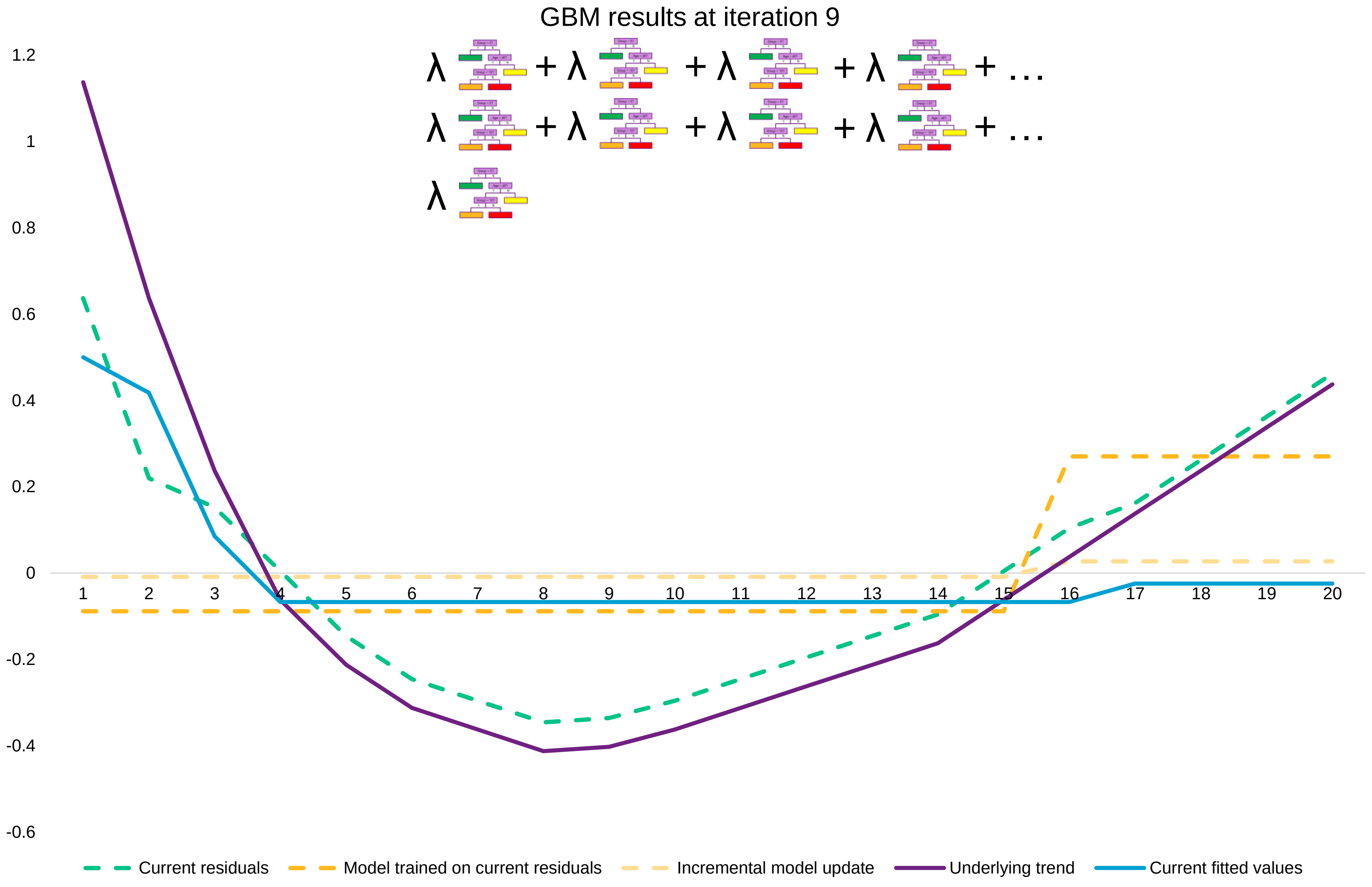
A simple GBM example



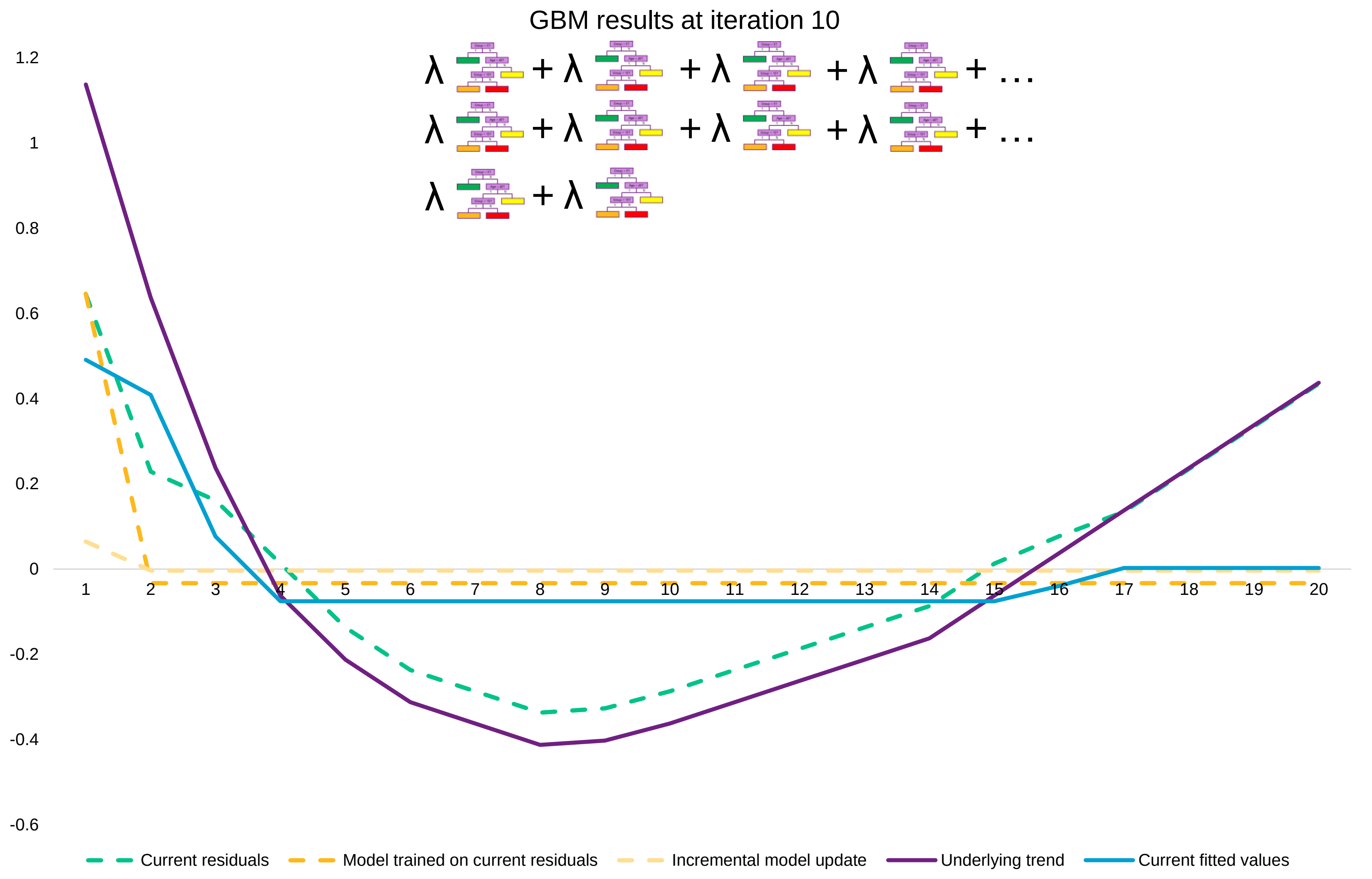
A simple GBM example



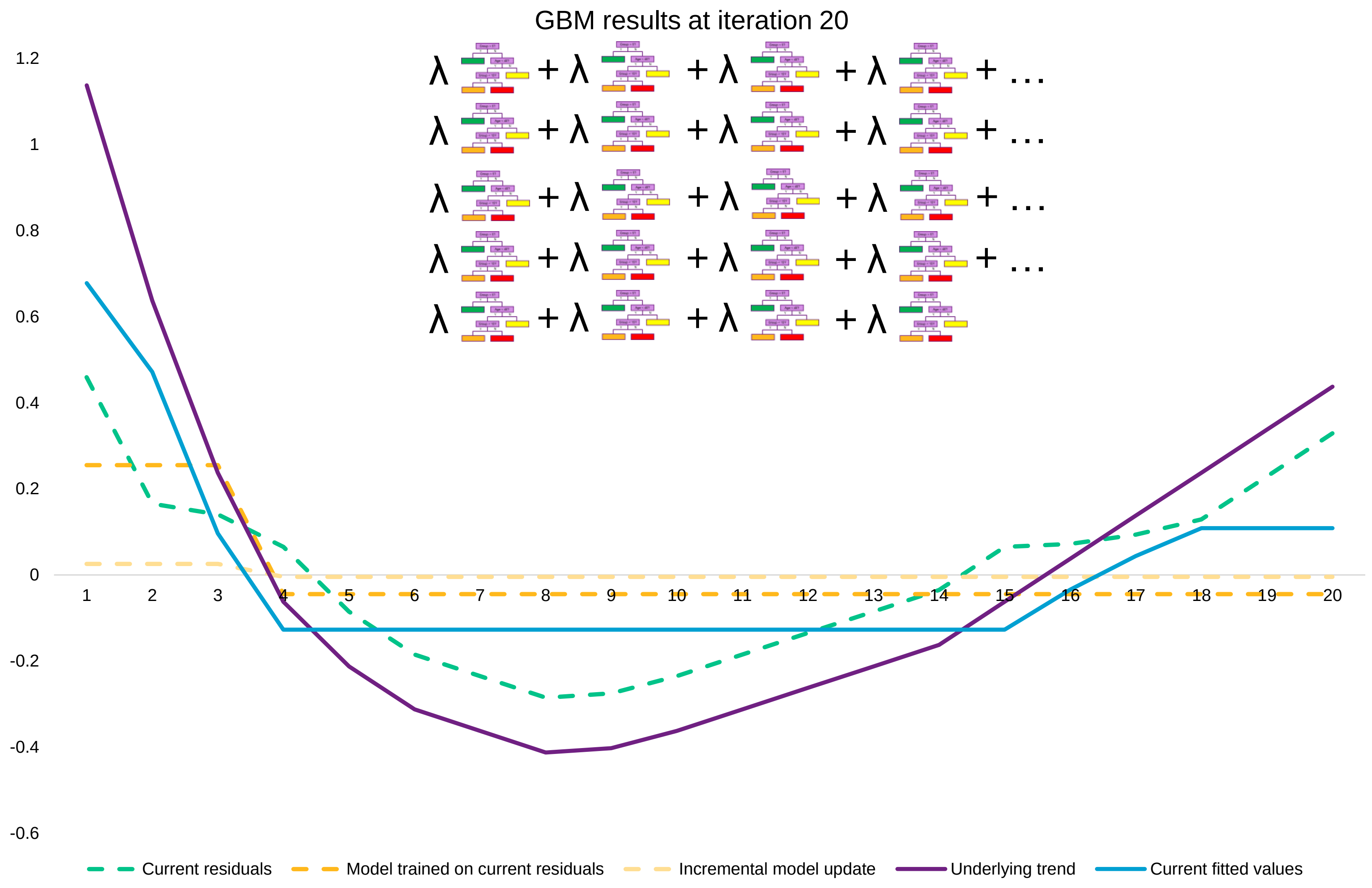
A simple GBM example



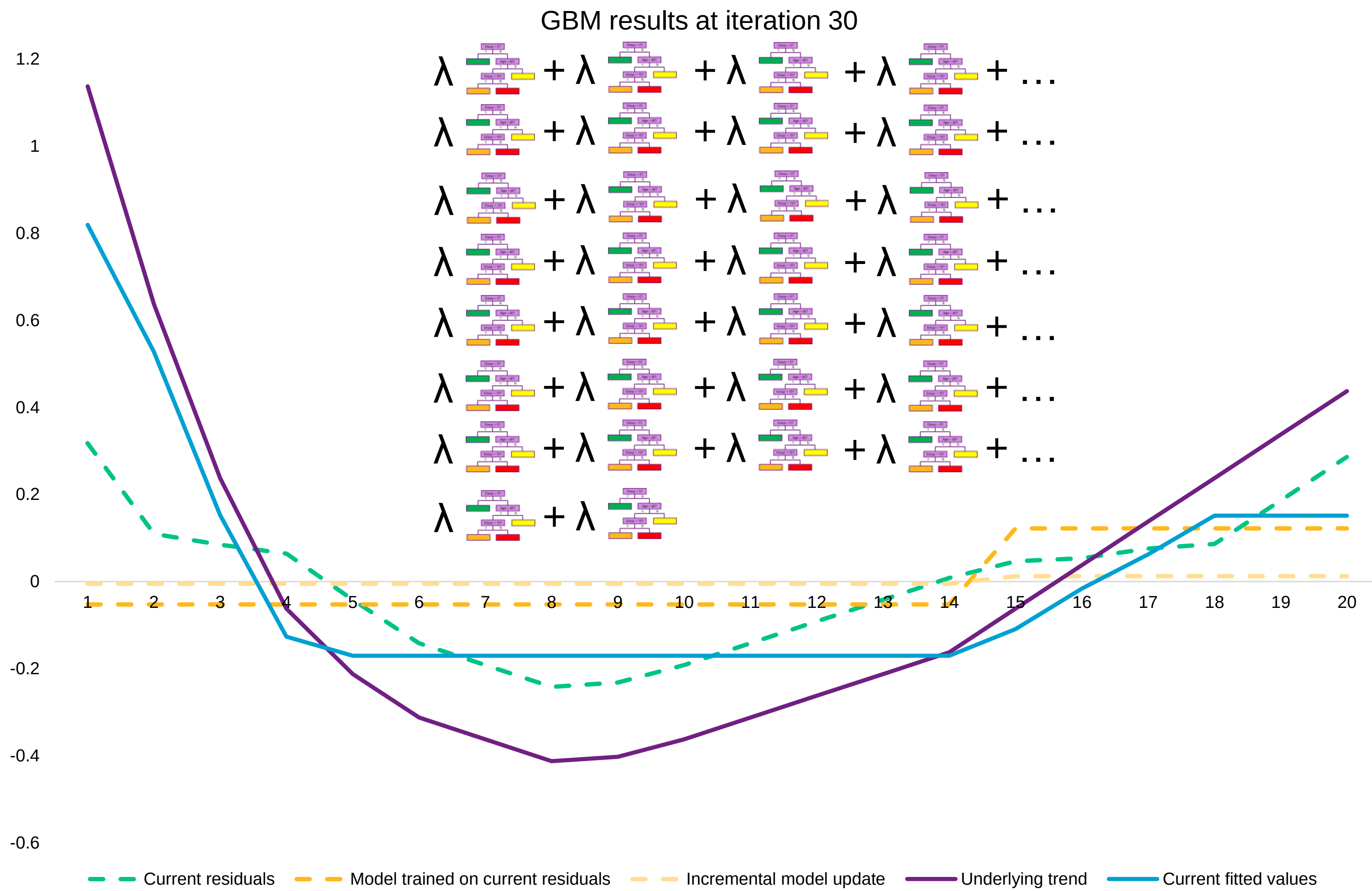
A simple GBM example



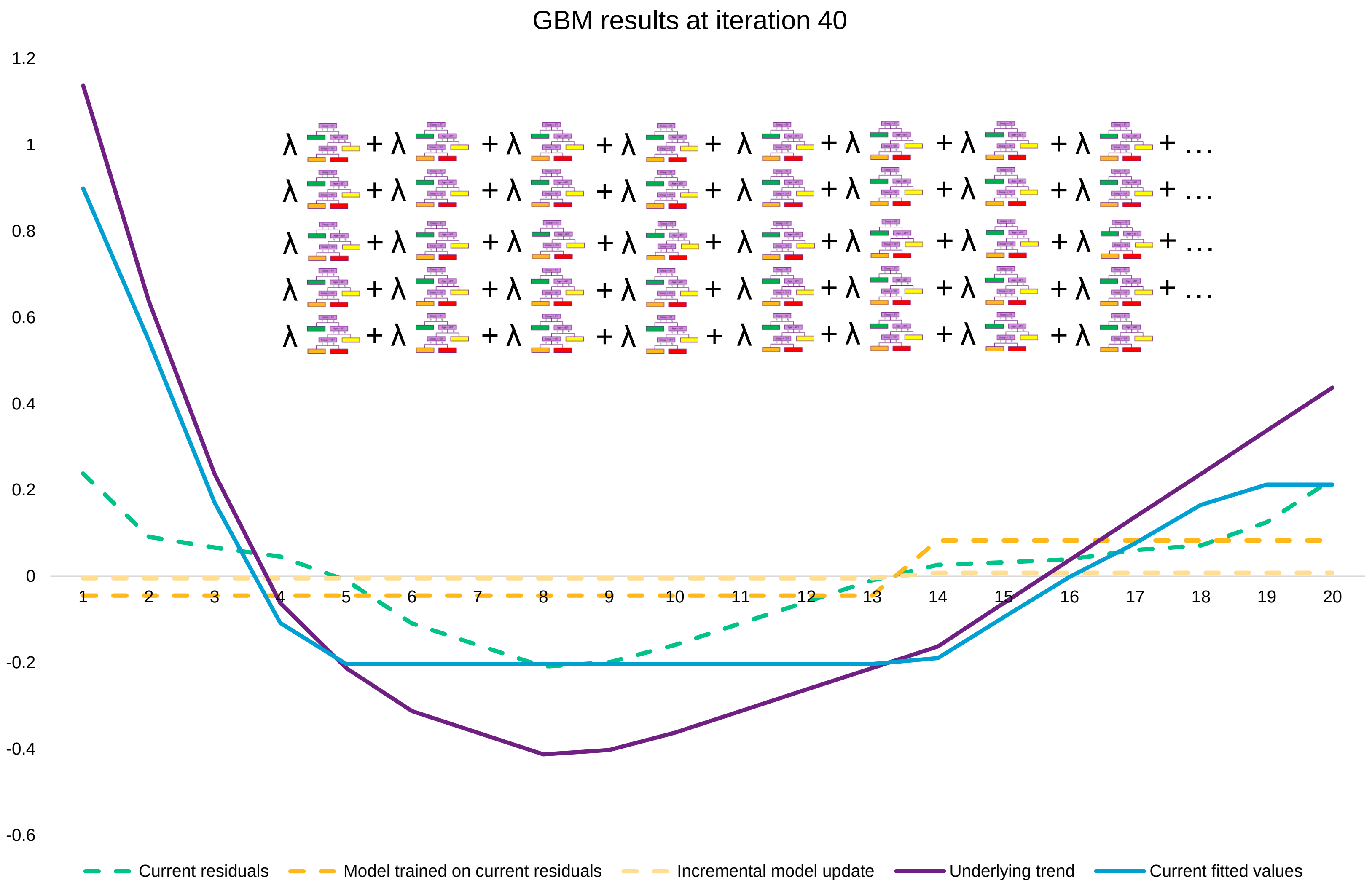
A simple GBM example



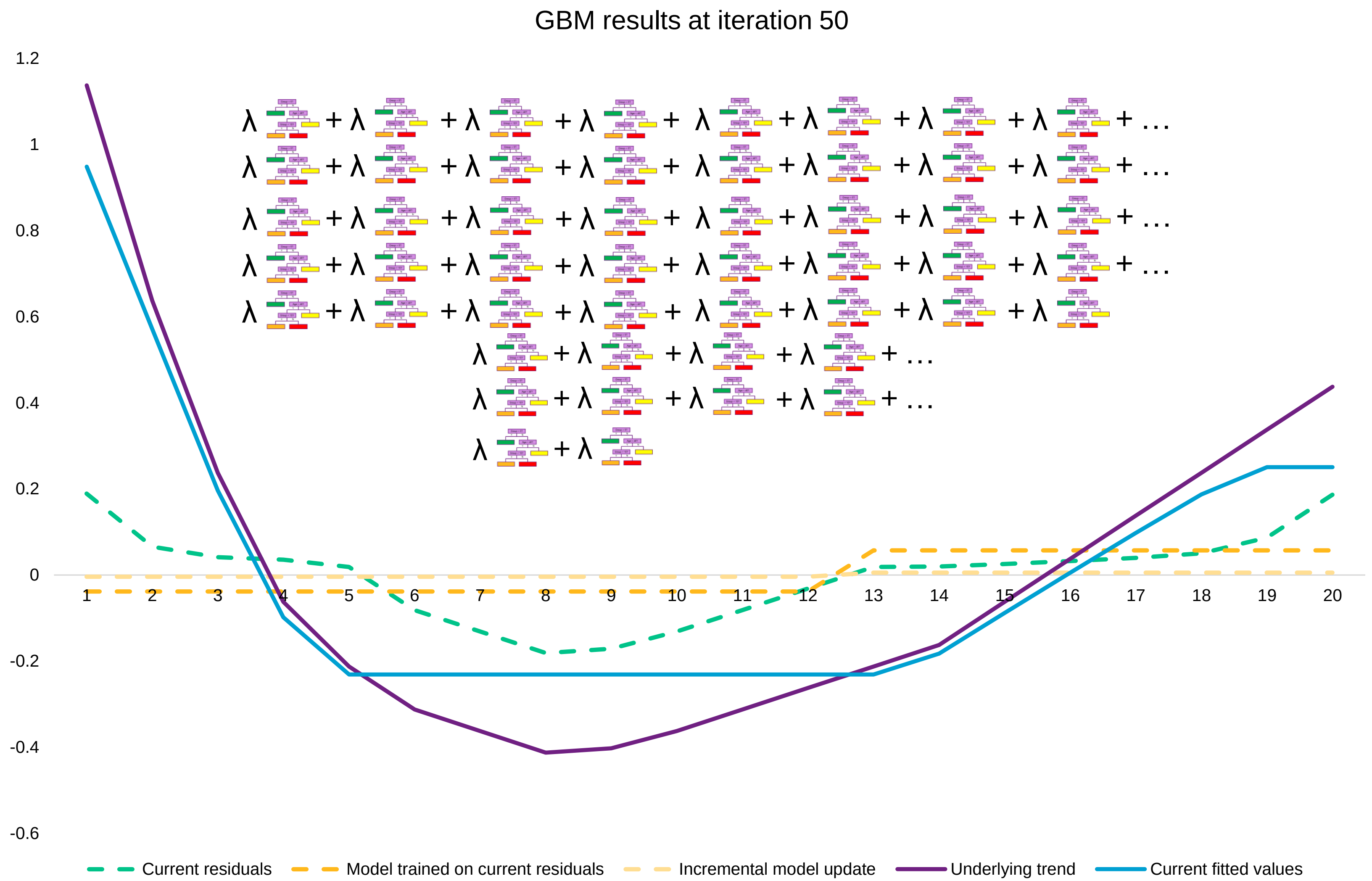
A simple GBM example



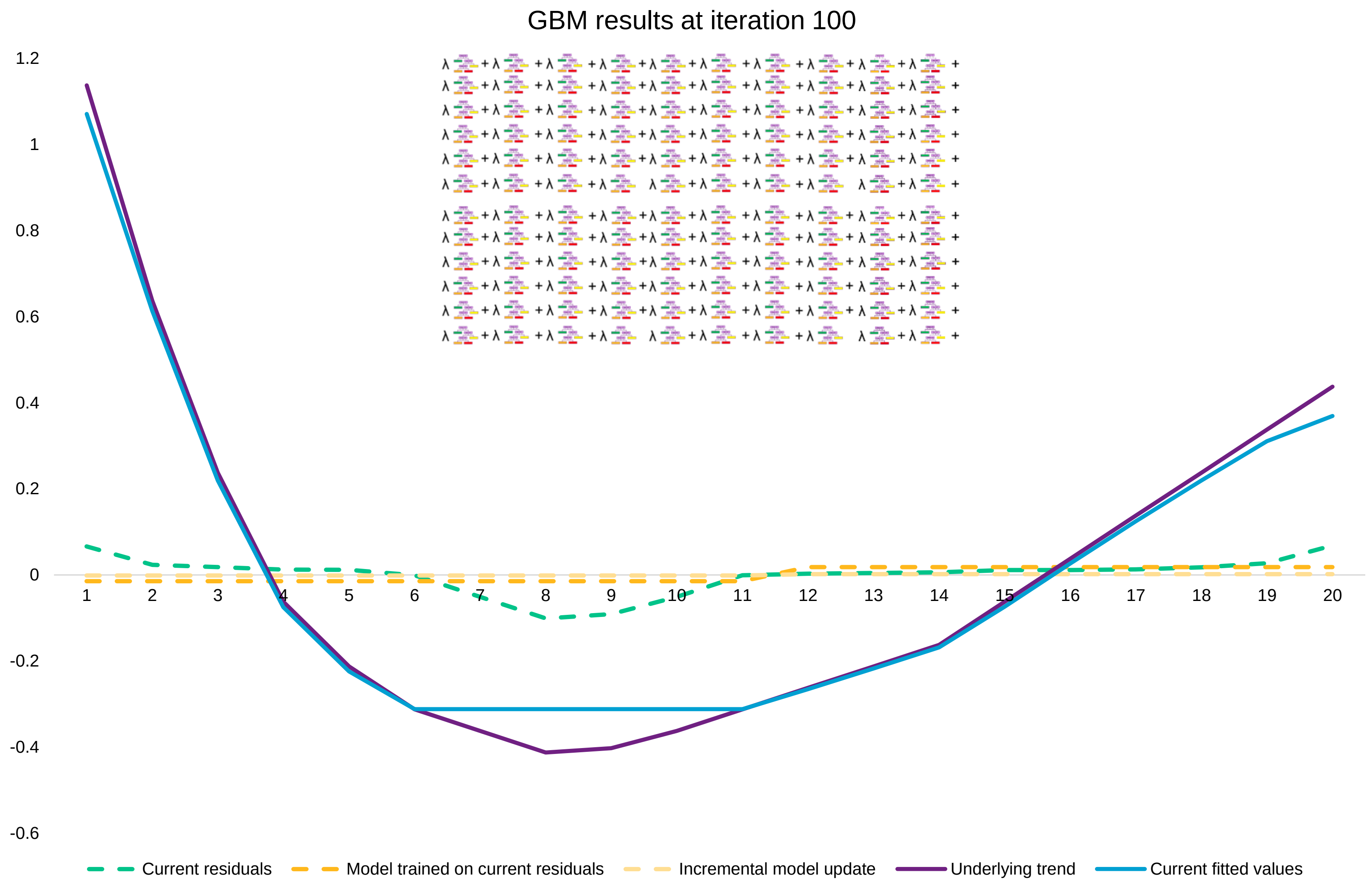
A simple GBM example



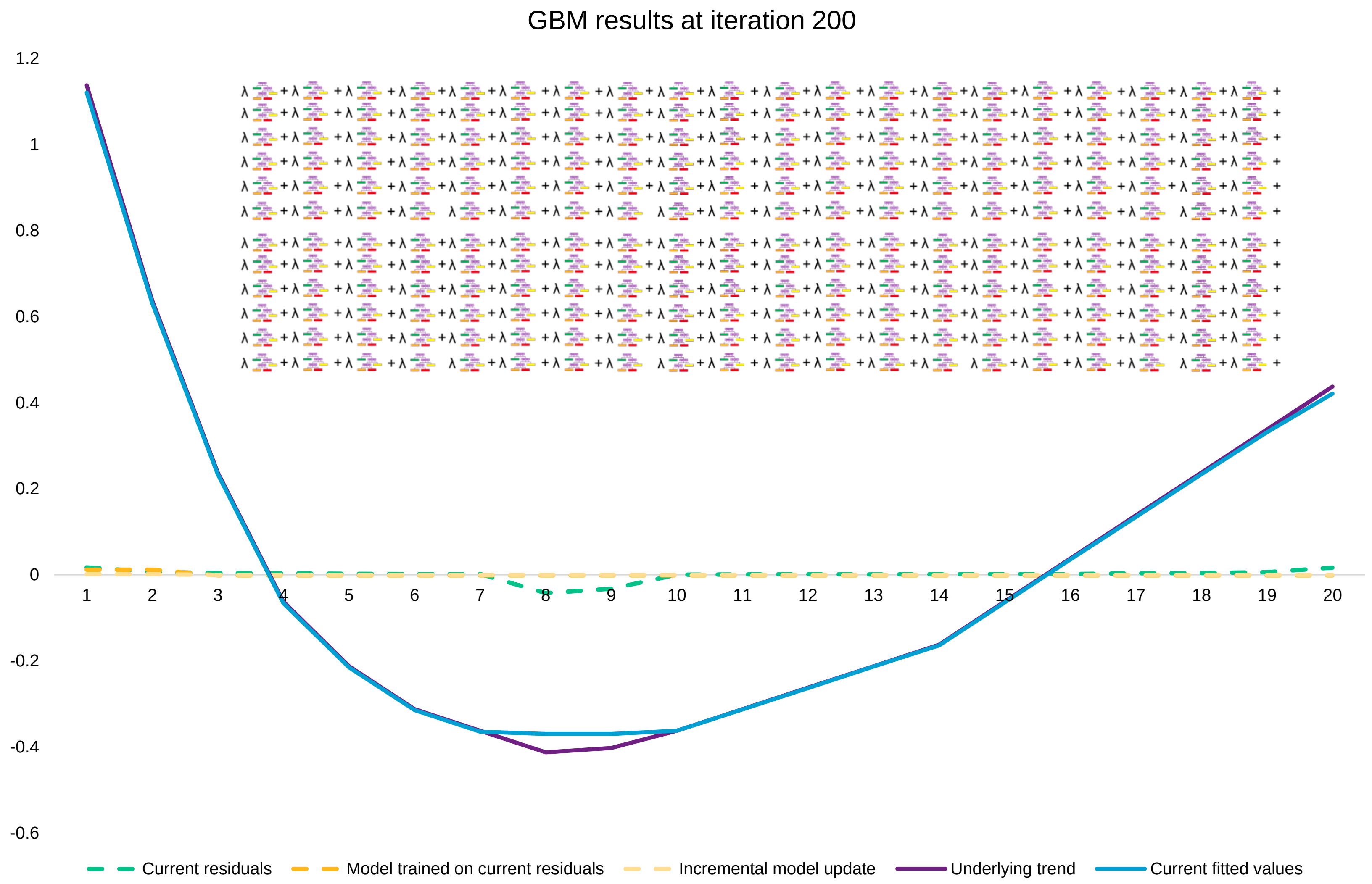
A simple GBM example



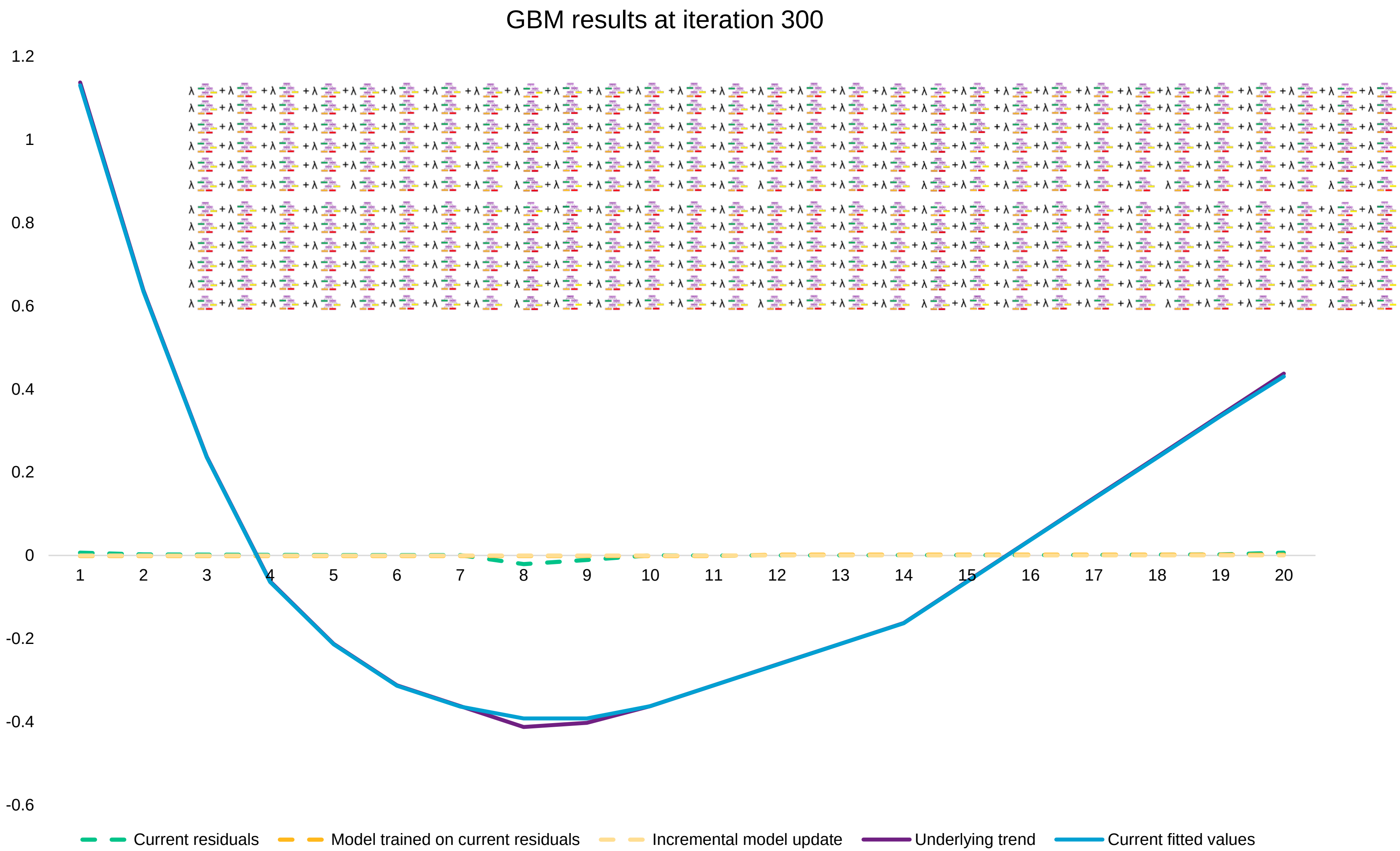
A simple GBM example



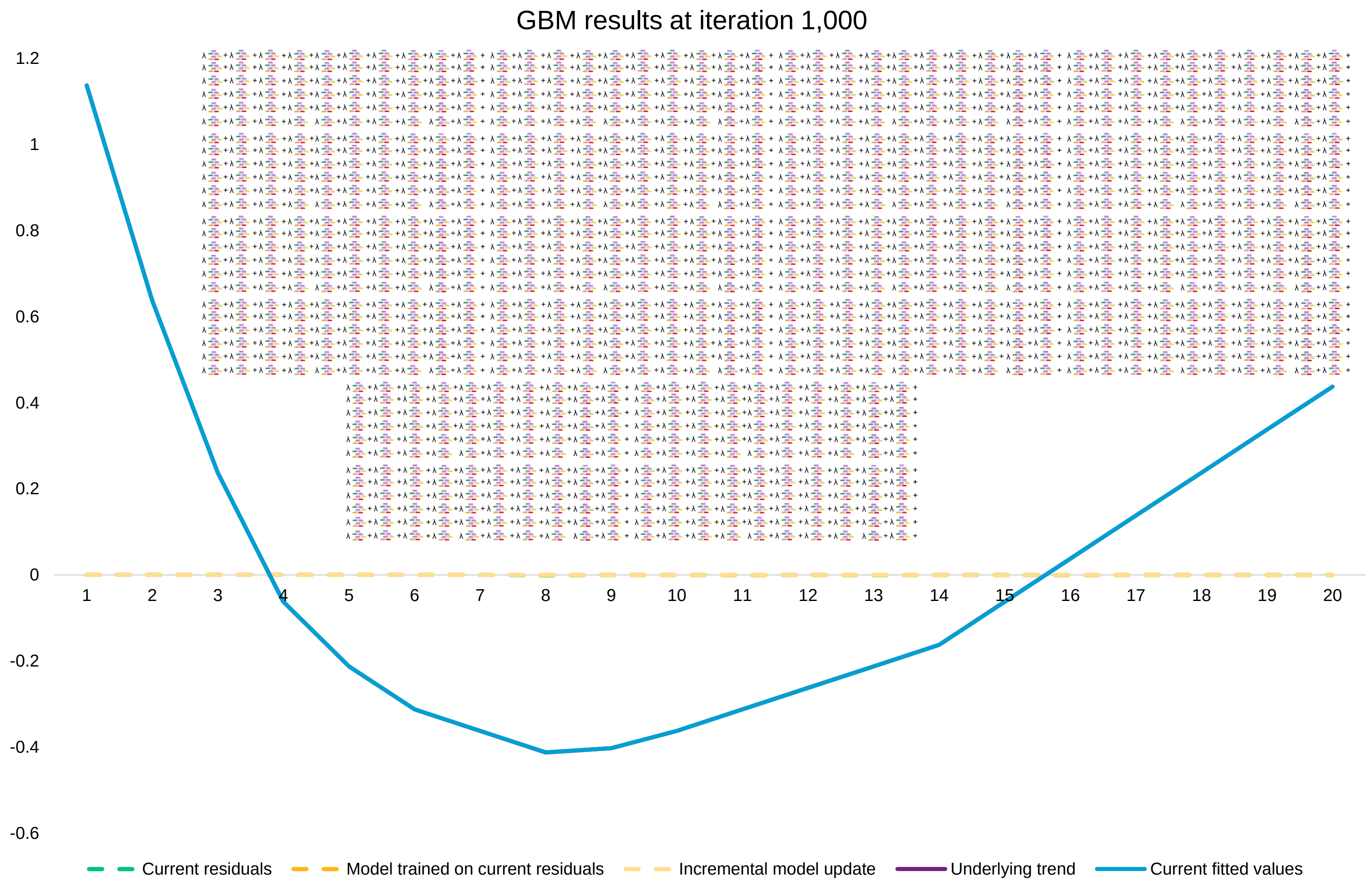
A simple GBM example



A simple GBM example

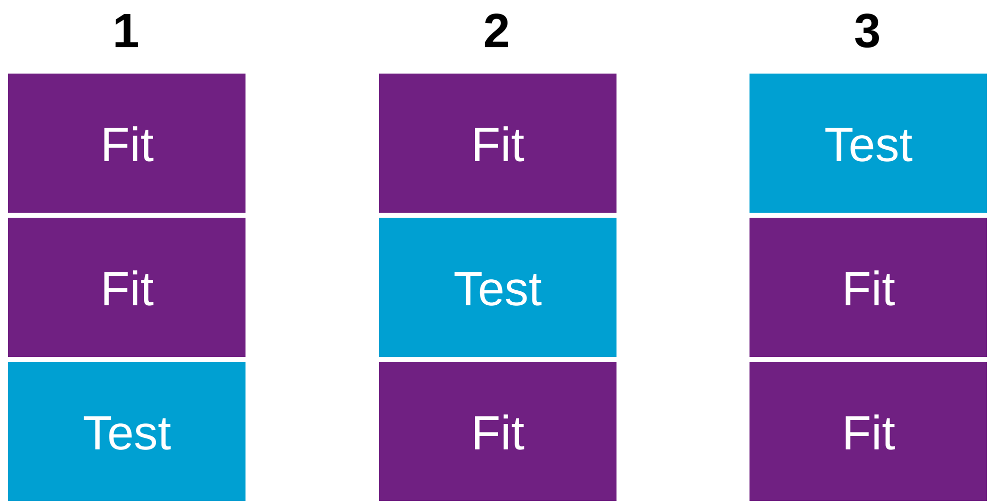


A simple GBM example



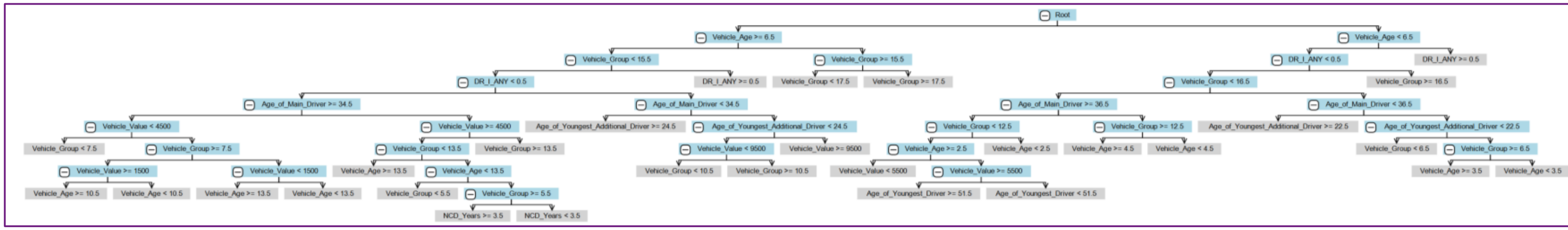
Calibrating the assumptions

- *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

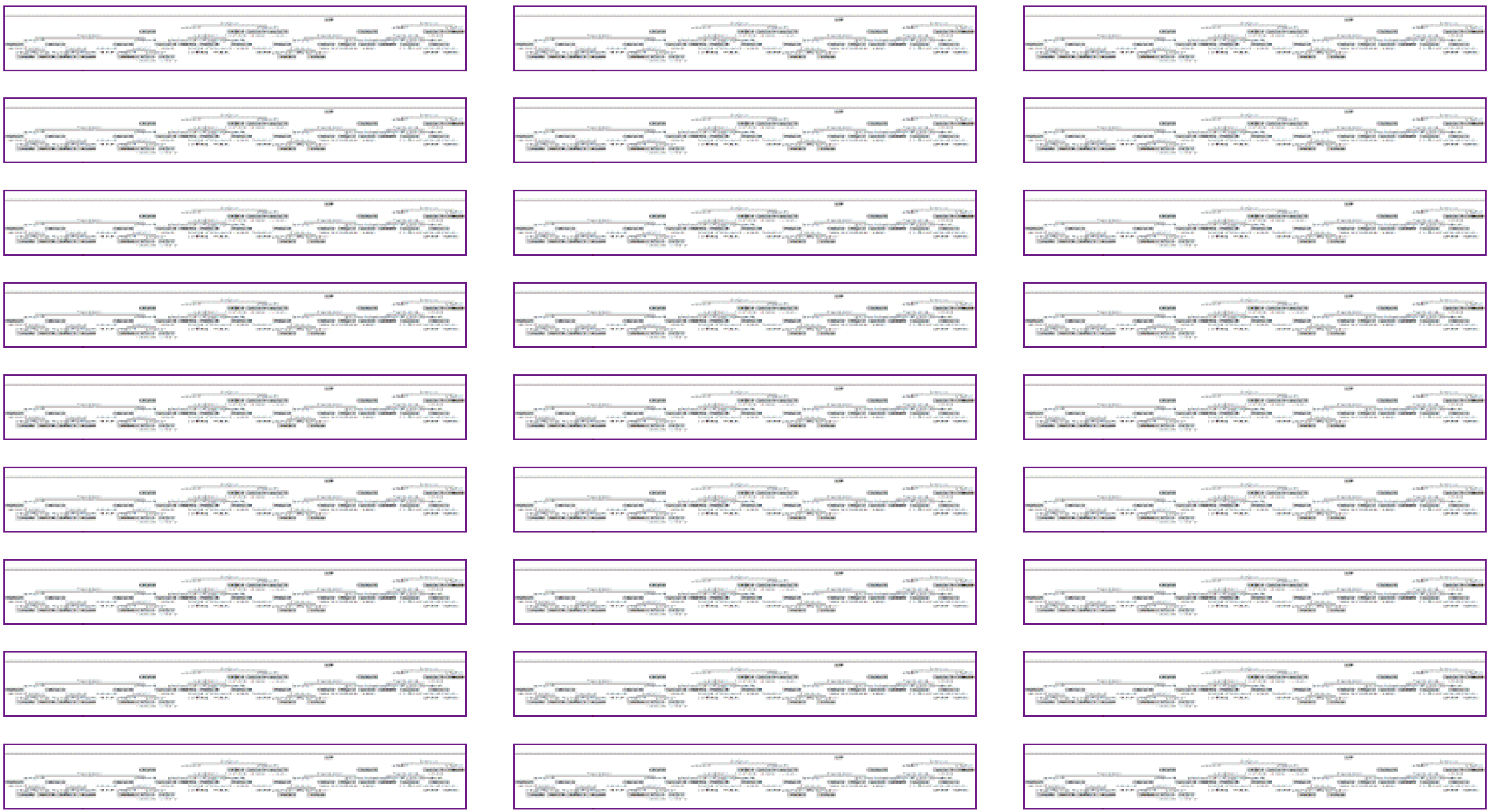


- Resulting plots can be used to determine the optimal assumption choice
  - Including how many trees to run

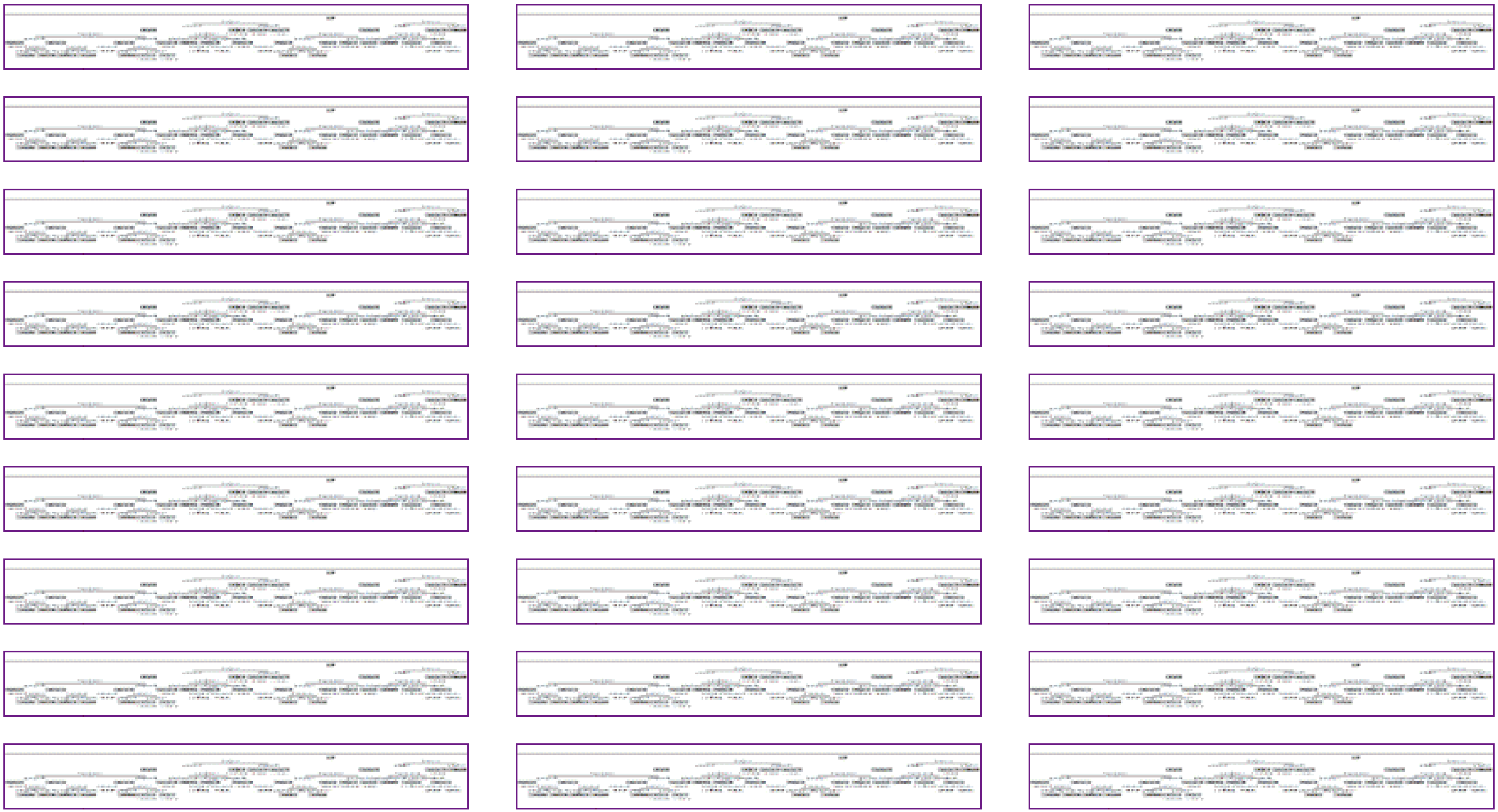
What does a GBM look like?

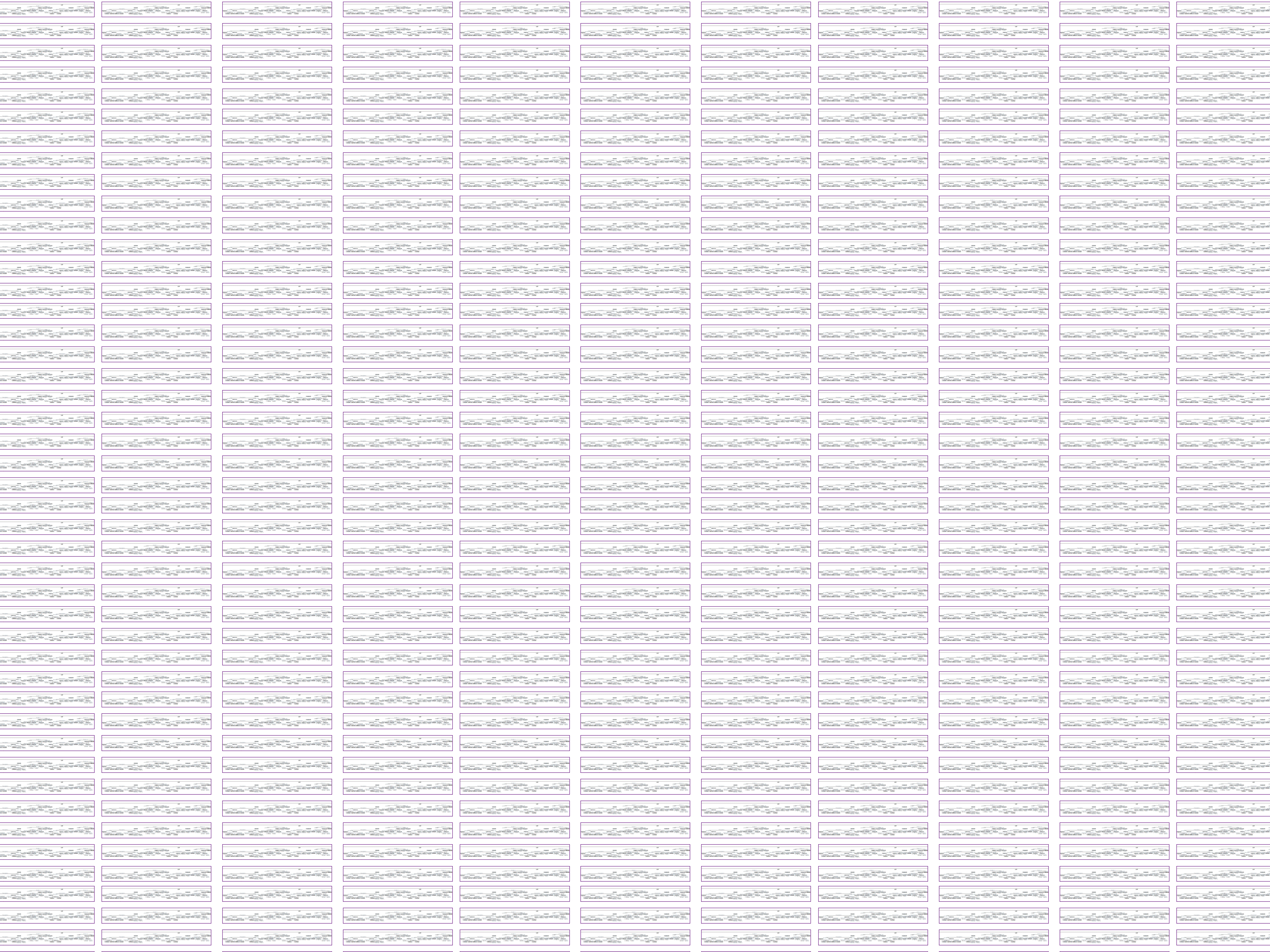


What does a GBM look like?



What does a GBM look like?





- Does it work?
- How does it work?

GBM – value add

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

Ensembles

Ensembles

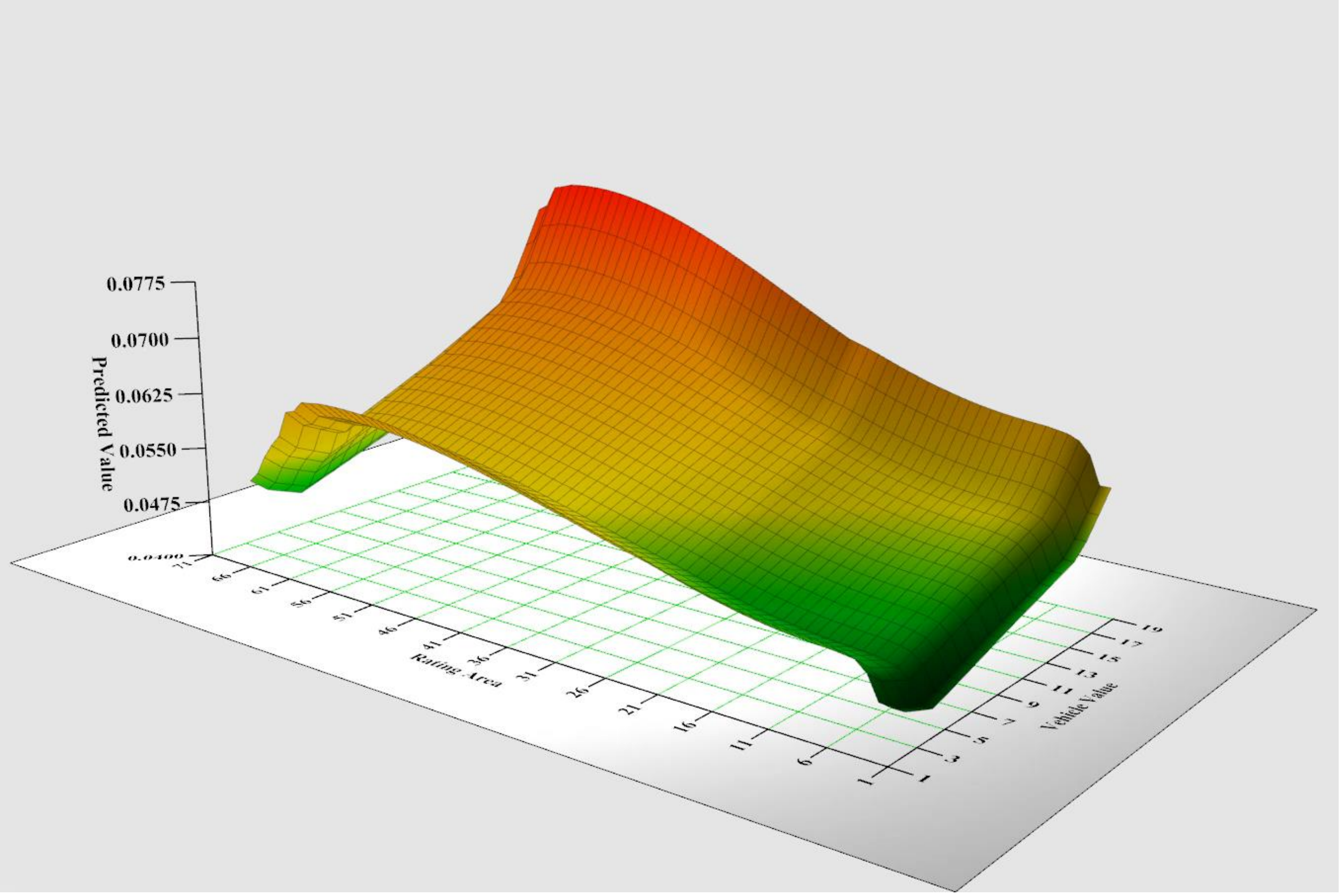
Adding an ensemble

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

What about a modelled down GBM?

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

What about an automated GLM?

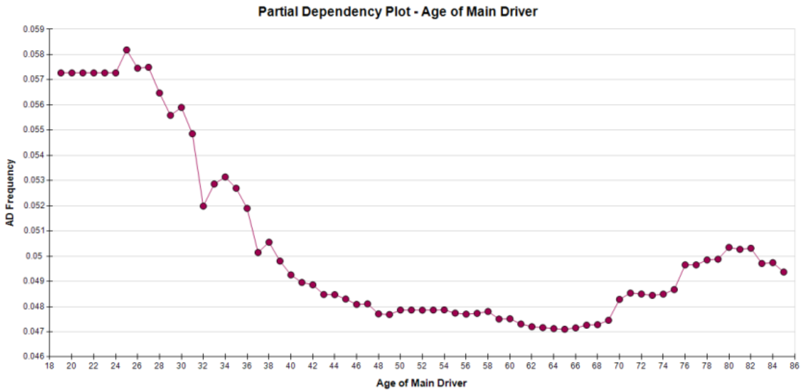


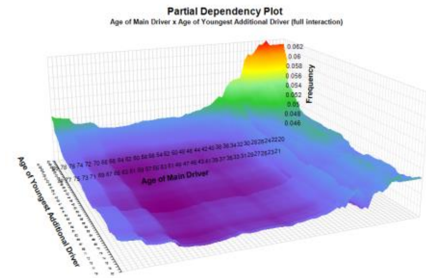
What about an automated GLM?

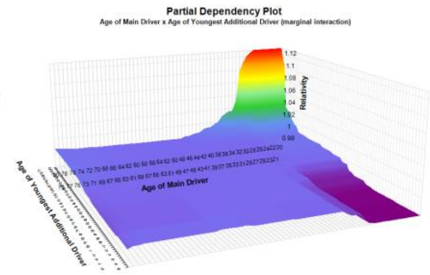
Model	Gini	Gini improvement	Gini rank	Loss ratio @ elasticity 6	Loss ratio rank	Loss ratio @ elasticity 2	Loss ratio rank
GLM (main factor removed)	0.319	-2.4%	7	-0.8%	7	-0.3%	7
GLM (minor factor removed)	0.322	-1.3%	6	-0.3%	6	-0.2%	6
GLM	0.326	0.0%	5	0.0%	5	0.0%	5
GLM fitted to GBM	0.328	0.5%	4	0.9%	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

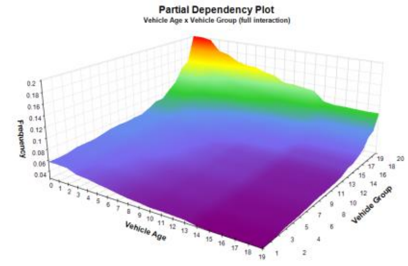
- Does it work?
- How does it work?

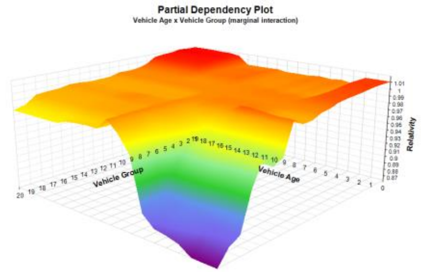
Partial dependency plots etc

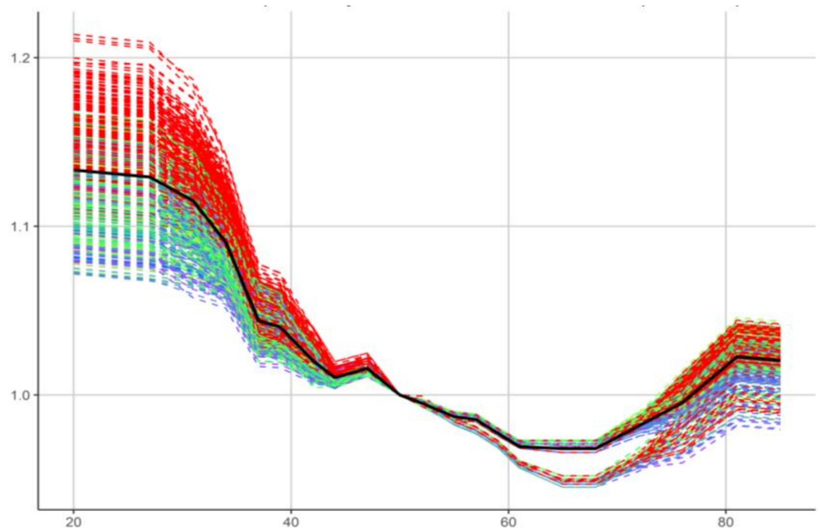


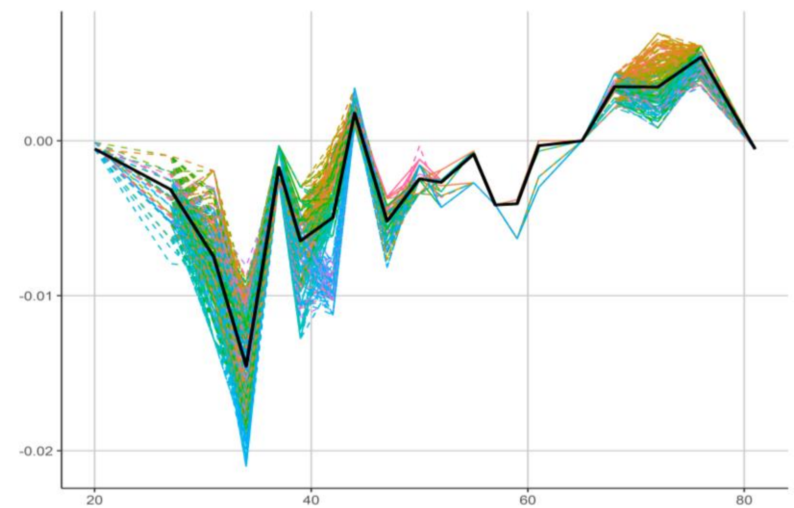


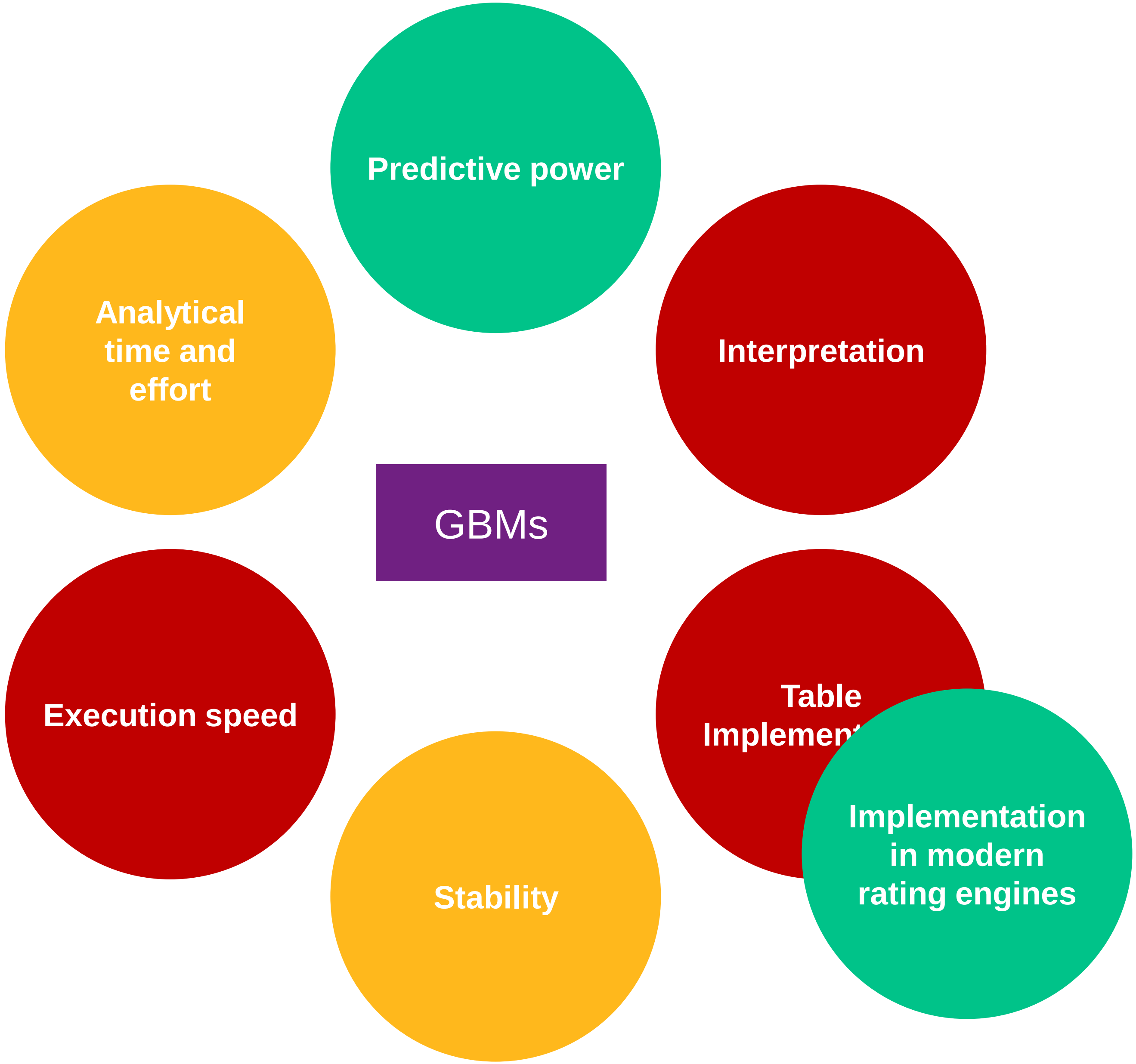


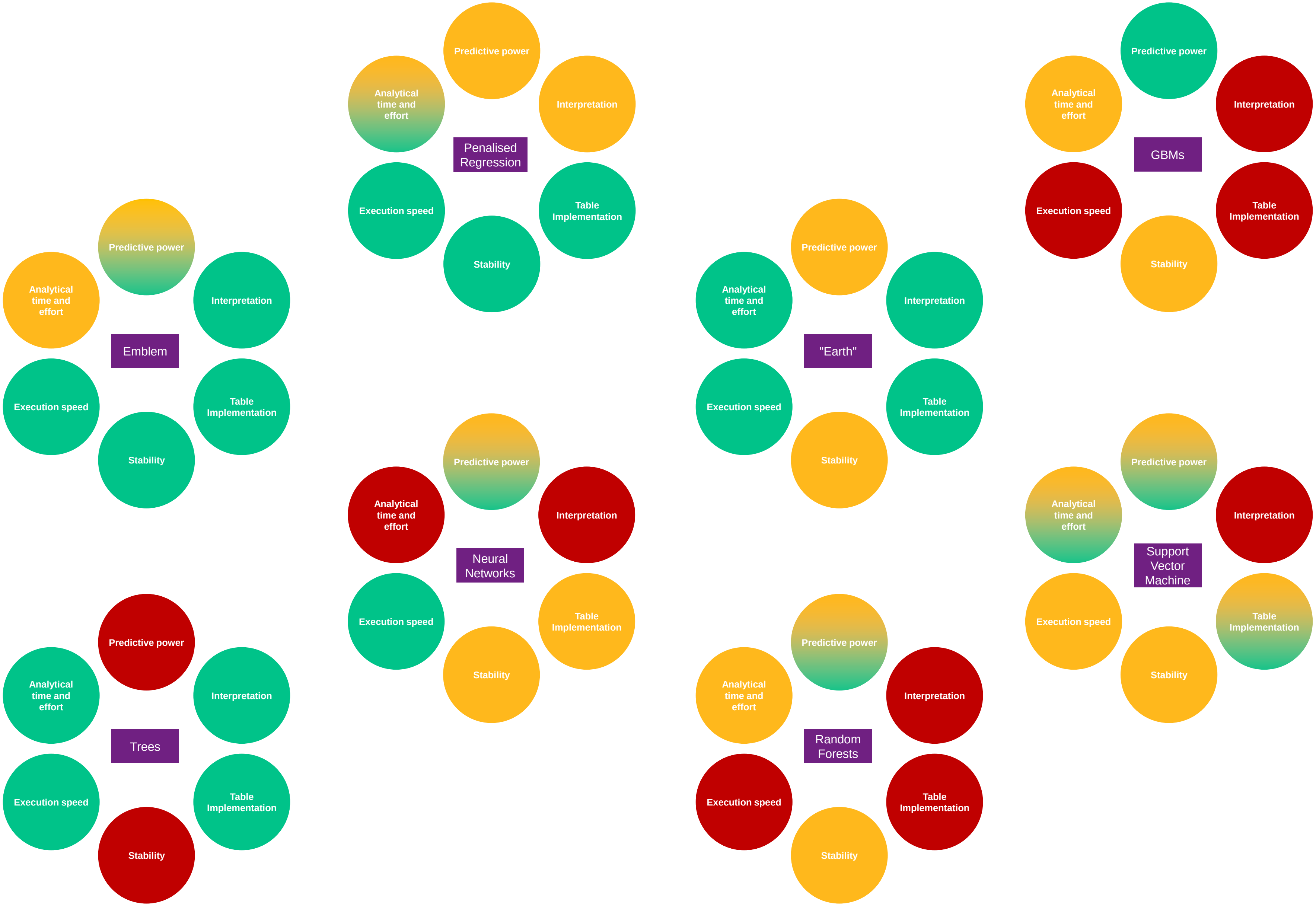









Predictive power
Analytical time and effort
Interpretation
GBMs
Execution speed
Table
Implemen
Implementation in modern rating engines
Stability


Penalised Regression
Predictive power
Interpretation
Table Implementation
Stability
Execution speed
Analytical time and effort
GBMs
Predictive power
Interpretation
Table Implementation
Stability
Execution speed
Analytical time and effort
Emblem
Predictive power
Interpretation
Table Implementation
Stability
Execution speed
Analytical time and effort
"Earth"
Predictive power
Interpretation
Table Implementation
Stability
Execution speed
Analytical time and effort
Neural Networks
Predictive power
Interpretation
Table Implementation
Stability
Execution speed
Analytical time and effort
Support Vector Machine
Predictive power
Interpretation
Table Implementation
Stability
Execution speed
Analytical time and effort
Trees
Predictive power
Interpretation
Table Implementation
Stability
Execution speed
Analytical time and effort
Random Forests
Predictive power
Interpretation
Table Implementation
Stability
Execution speed
Analytical time and effort

Machine Learning in Pricing

Conclusions

- There are many forms of ML models
- New data and feature/response engineering generally add more value than new methods BUT we need to continuously explore which methods work on which problems
- Traditional measures of prediction value may not reflect applications in insurance
- And it’s not all about predictive power anyway – other criteria are important

- GBMs can provide predictive lift benefits by capturing higher order effects … BUT
  - Can you cope with not seeing the model and instead use broad diagnostics
  - Effort is required to expose/understand higher order effects in an expeditious manner
  - How will business leaders and regulators respond to this method?
  - Do you have the software and hardware to fit to large dataset
  - Do you have a rating engine that can implement a GBM

Machine learning in pricing

Conclusions

- Machine learning brings a proliferation of new methods
- Improving models is more than just finding the best method. Consider:
  - What data are available and how can data be transformed to give insight
  - What is the optimal model structure and target variable?
  - How can information be transferred between models?
- Earth is a fast, interpretable method that can improve overall lift by informing when/where to segment models
- Neural networks are complex and require numerous input decisions; analyzing unstructured data (e.g., imagery) is an intuitive application for this method … but where else may it be helpful?
- Penalized regression can aid in factor selection decisions and may in fact be a good method in its own right – particularly when the modeler has less of a "feel" for the data
- Machine learning in pricing is not all about improving predictive power. Consider:
  - Fast investigation of new data
  - Quick assessment and response of emerging experience