2021 CAS Annual Meeting

GBMs: The Real Impact of a Rate Change

Michael McPhail, FCAS
Serhat Guven, FCAS

November 8, 2021

Willis Towers Watson

Agenda

- Understanding Customer Demand
- Model Form
- Applications

What is insurance customer demand?

- Customer demand in insurance generally reflects the following:
  - Conversion of new policies
  - Renewals of existing policies
- To model the conversion of new policies, we build an acquisition model
- To model the renewals of existing policies, we build a retention model
- In addition to the models above, insurers might be interested in modeling mid-term cancellations on existing policies

What impacts customer demand?

Attributes & Attitudes	What is the customer like?
Influences	What you have done to the customer?
Environmental	What are the external influences?
Status Changes & Triggers	What has changed and when?

Why consider customer demand?

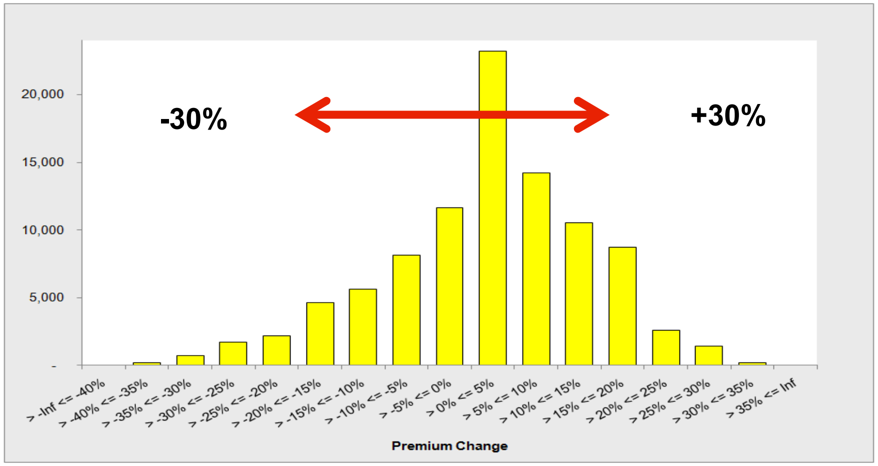
- Forecasting profitability
- Studying retention & acquisition
- Understanding the future mix of business

Why consider customer demand?

Revenue Neutrality Example

- Revenue neutrality
  - Overall effect on premium volume is revenue neutral
  - Individual policies could still see large swings in rate

$$\frac{\sum_{CurrentPolicies} \text{ProposedPremium}}{\sum_{CurrentPolicies} \text{CurrentPremium}} = 1.000$$



- Disadvantage of this view of revenue neutrality is that it fails to consider the future shape of the book (need to reflect retention and conversion effects)
- Off-balance to achieve desired demand-weighted loss ratio

How to model customer demand?

- Any modeling approach that produces a probability can be used to model customer demand
  - GLM
  - Decision trees
  - GBM
  - And so on…

Modeling Customer Demand with a GLM

Advantages & Disadvantages

- Advantages
  - A known commodity
  - Relatively easy to interpret and explain
  - Reponse variable does not have to be normally distributed

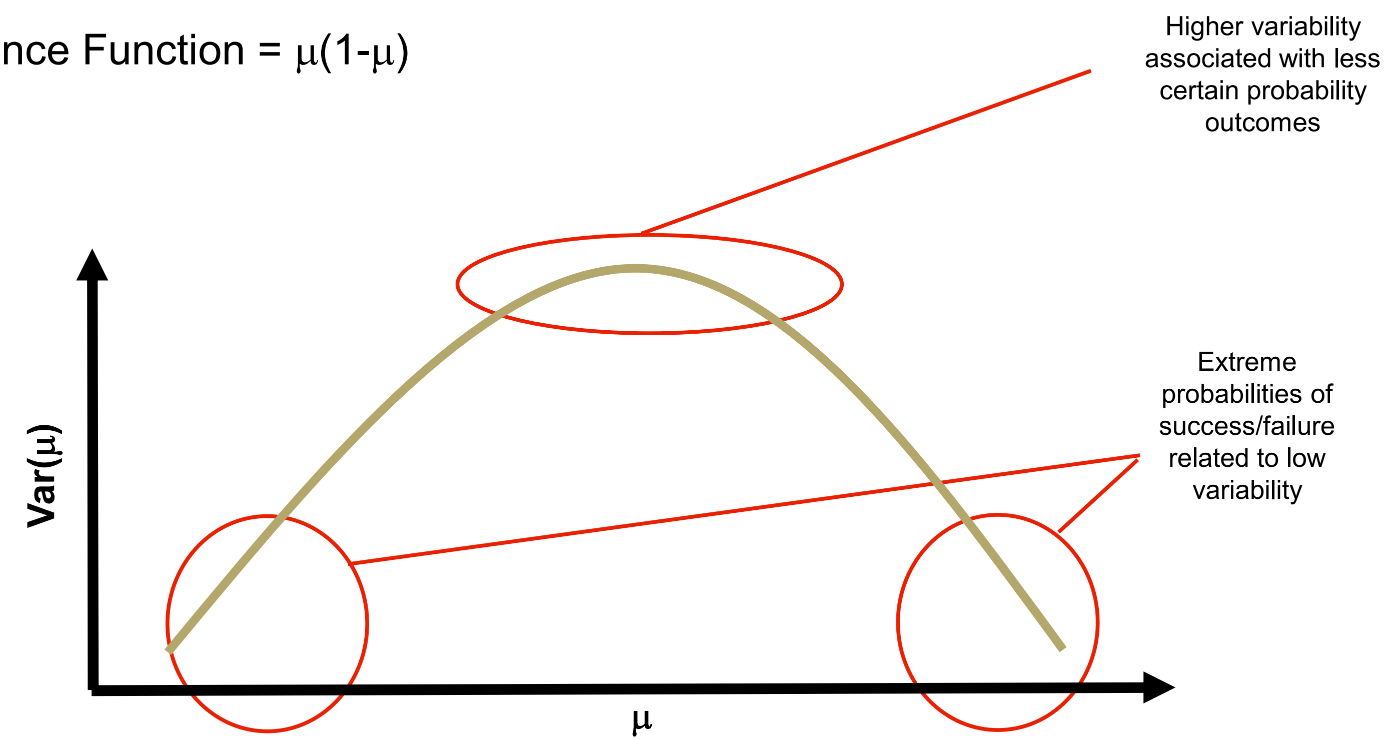
- Disadvantages
  - Does not handle non-linear relationships well
  - Modeling interactions often requires manual adjustments

Modeling Customer Demand with a GLM

Distribution Function

- Binomial
  - Basic functional form in decision modeling
  - Belongs to the exponential family of distributions

- Variance Function = $\mu(1-\mu)$



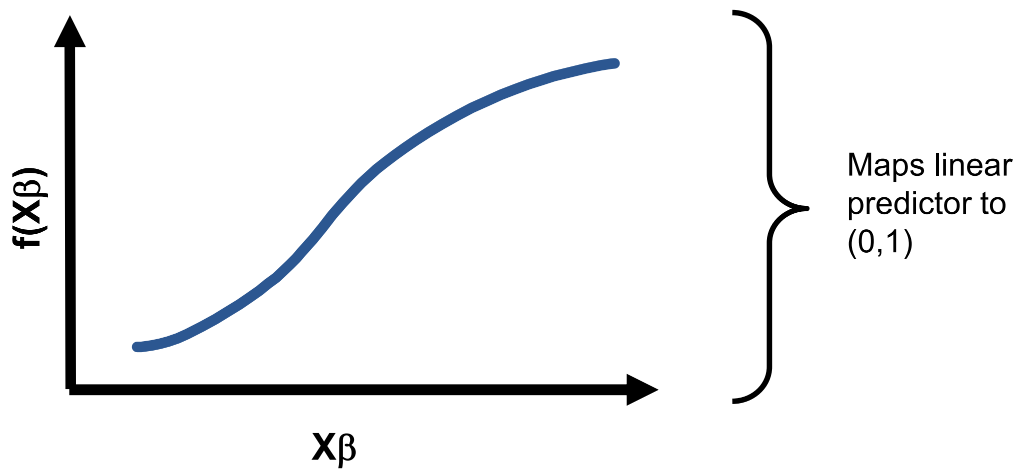
Modeling Customer Demand with a GLM

Link Function

- Logit link is canonical for the binomial distribution:

$$\frac{1}{1+\frac{1}{\exp(X\beta)}}$$

- Properties of the logit link function:



- S-shape curve “traps” the predictive value to the probability range

Modeling Customer Demand with a Gradient Boosted Machine (GBM)

Advantages & Disadvantages

- Advantages
  - Can handle non-linear relationships
  - Naturally models interactions without manual adjustments
  - Tends to produce better model fits

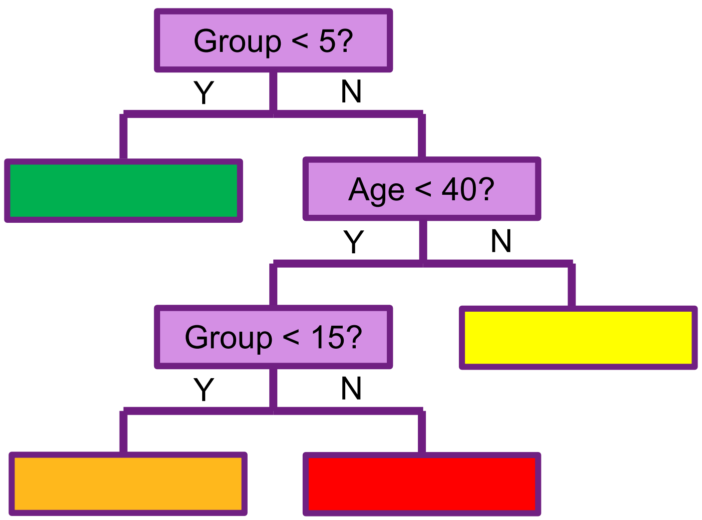
- Disadvantages
  - Not as common in the insurance industry
  - Can overfit the data without proper tuning
  - Large trees can be difficult to intepret

How Does a GBM Work?

What is a GBM?

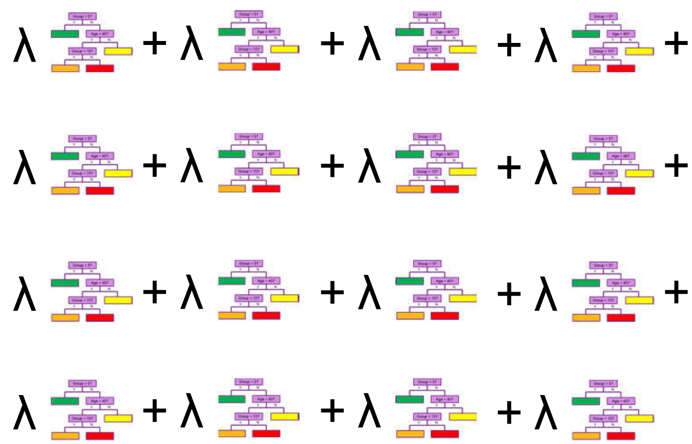
A tree

$$f_i(x)$$



A GBM

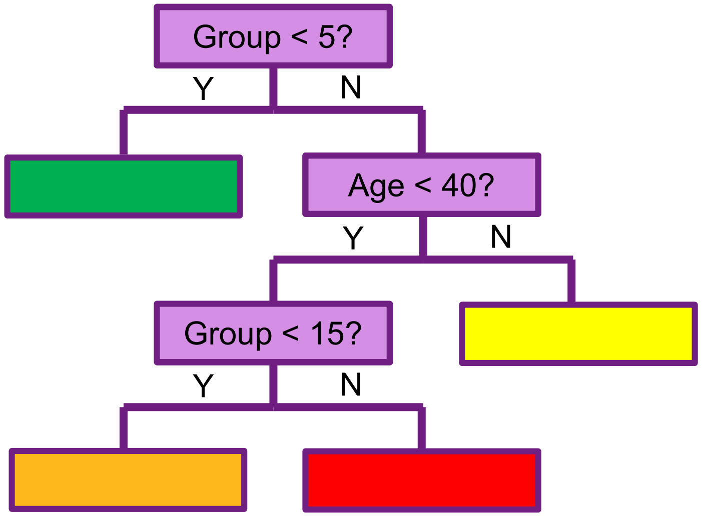
$$f(x) = \lambda \sum_{n=1}^{N} f_n(x)$$



How Does a GBM Work?

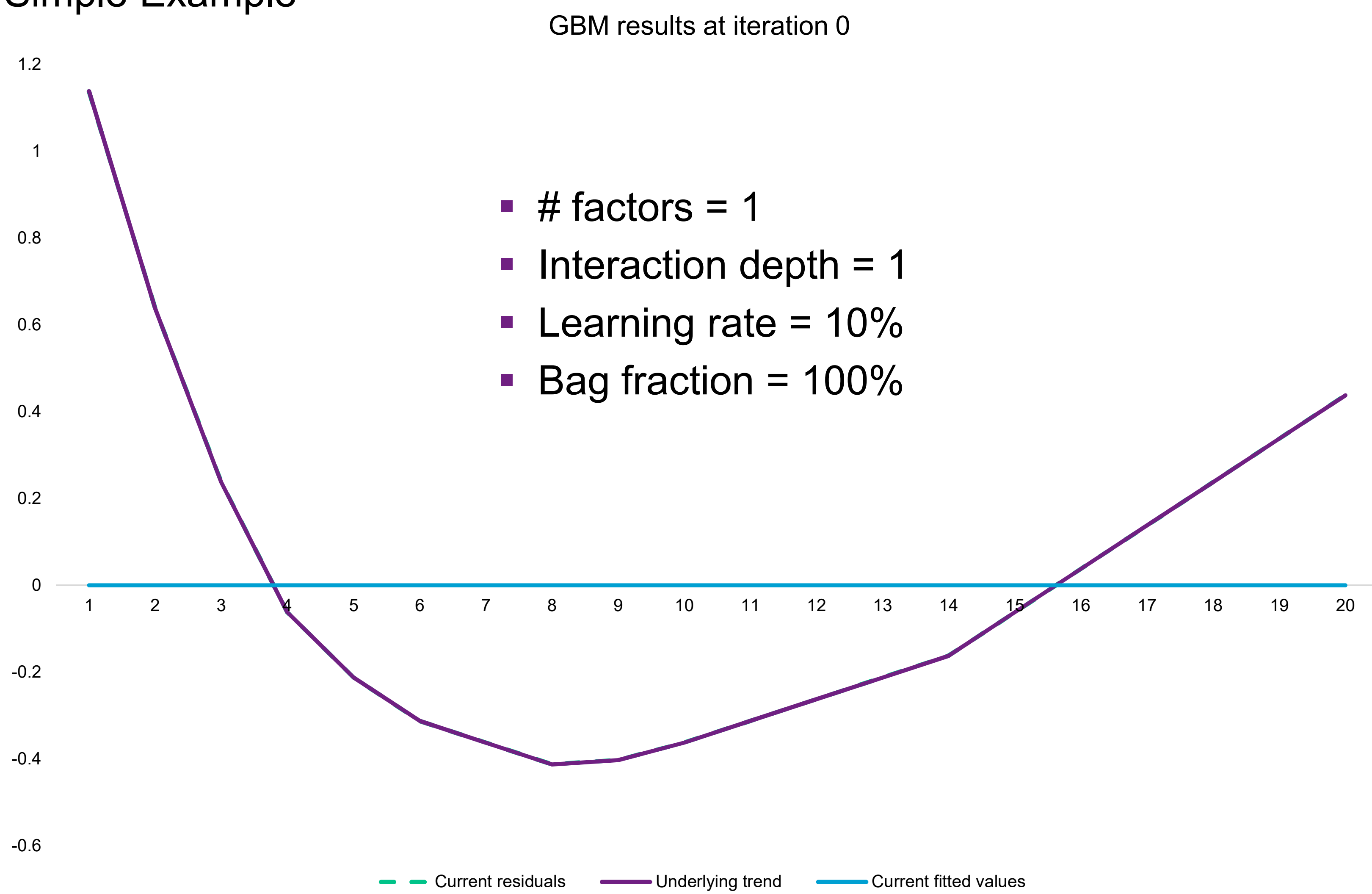
Four Main Assumptions

- 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



How Does a GBM Work?

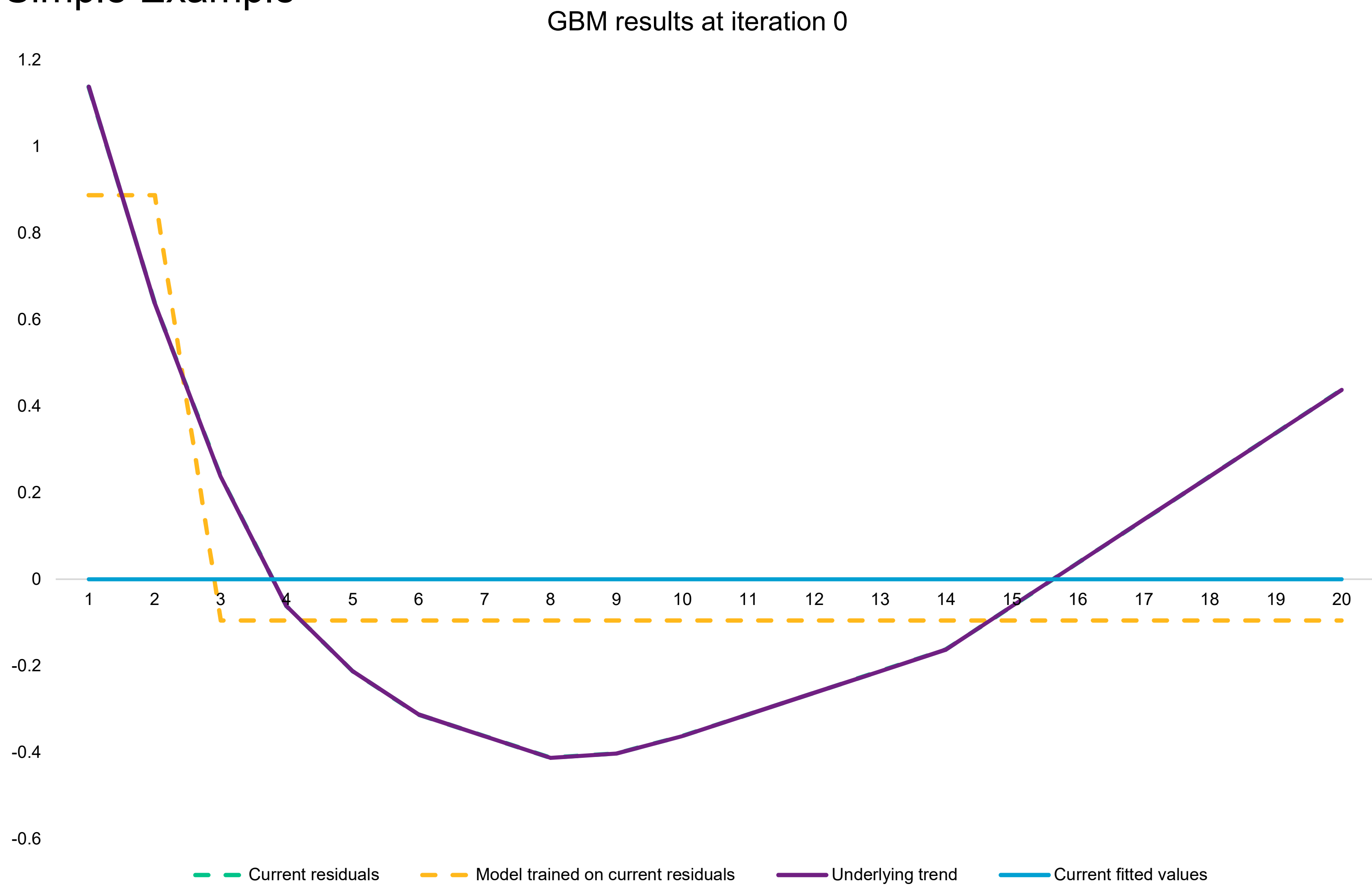
A Simple Example



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

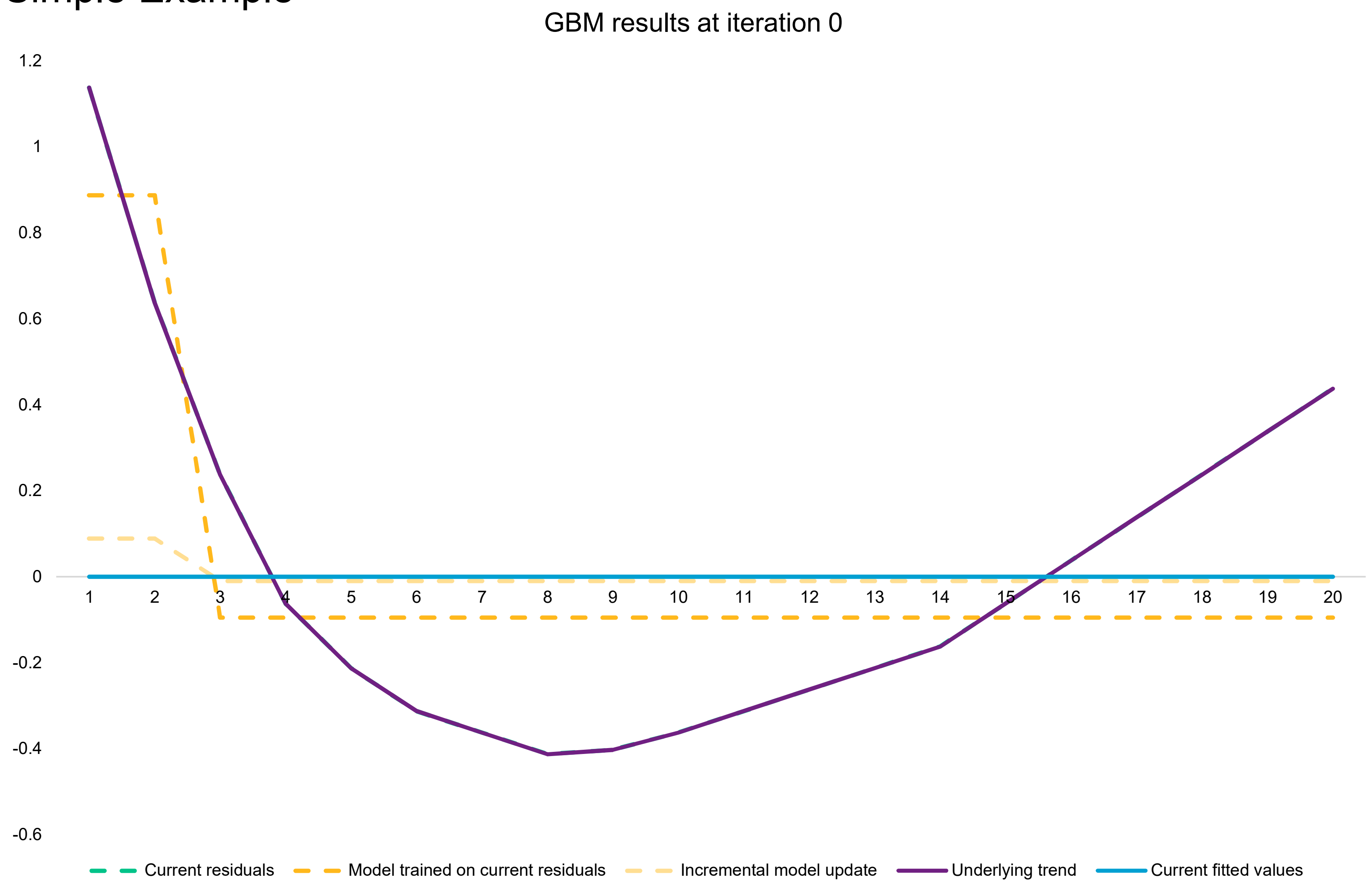
How Does a GBM Work?

A Simple Example



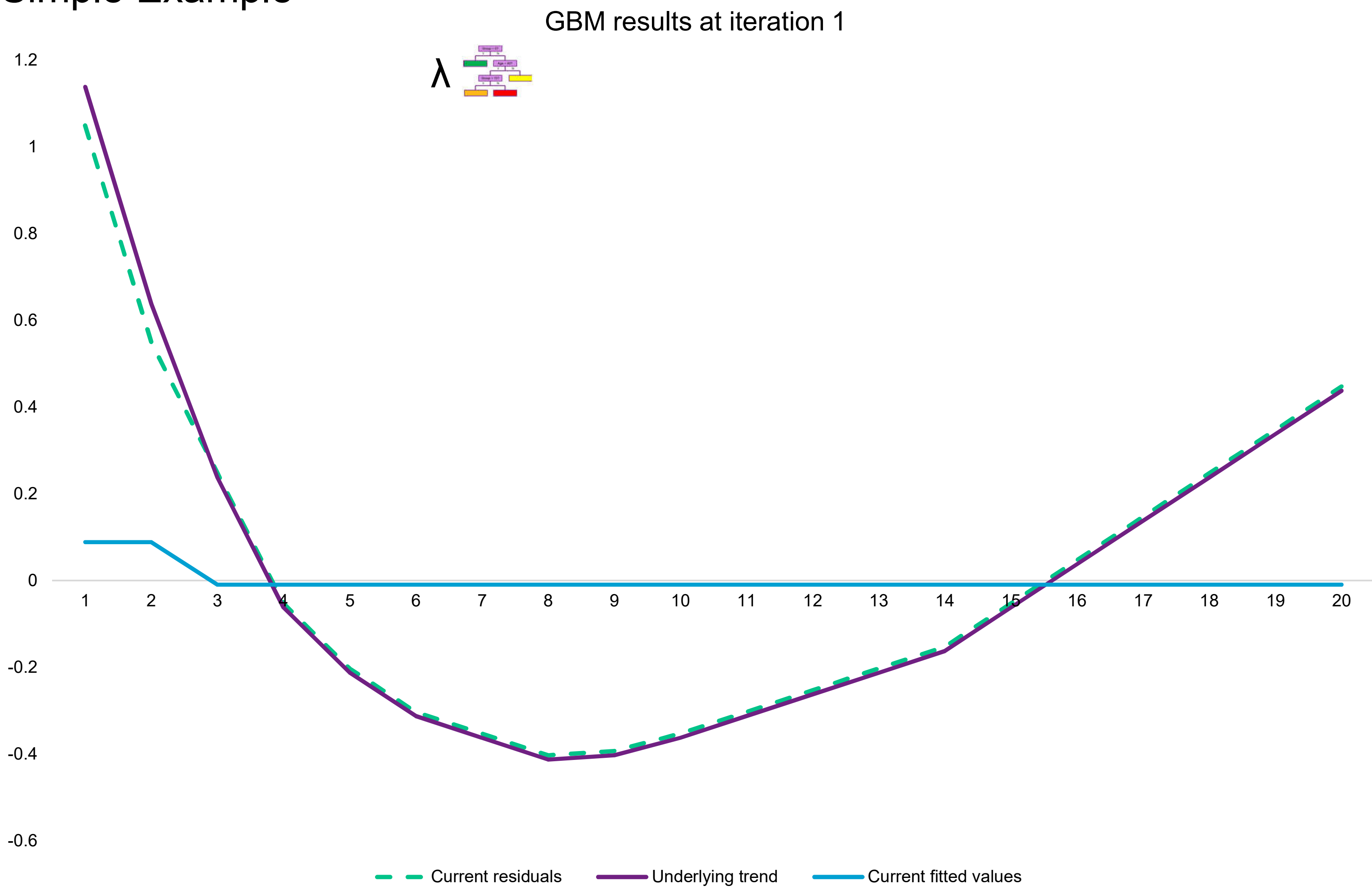
How Does a GBM Work?

A Simple Example



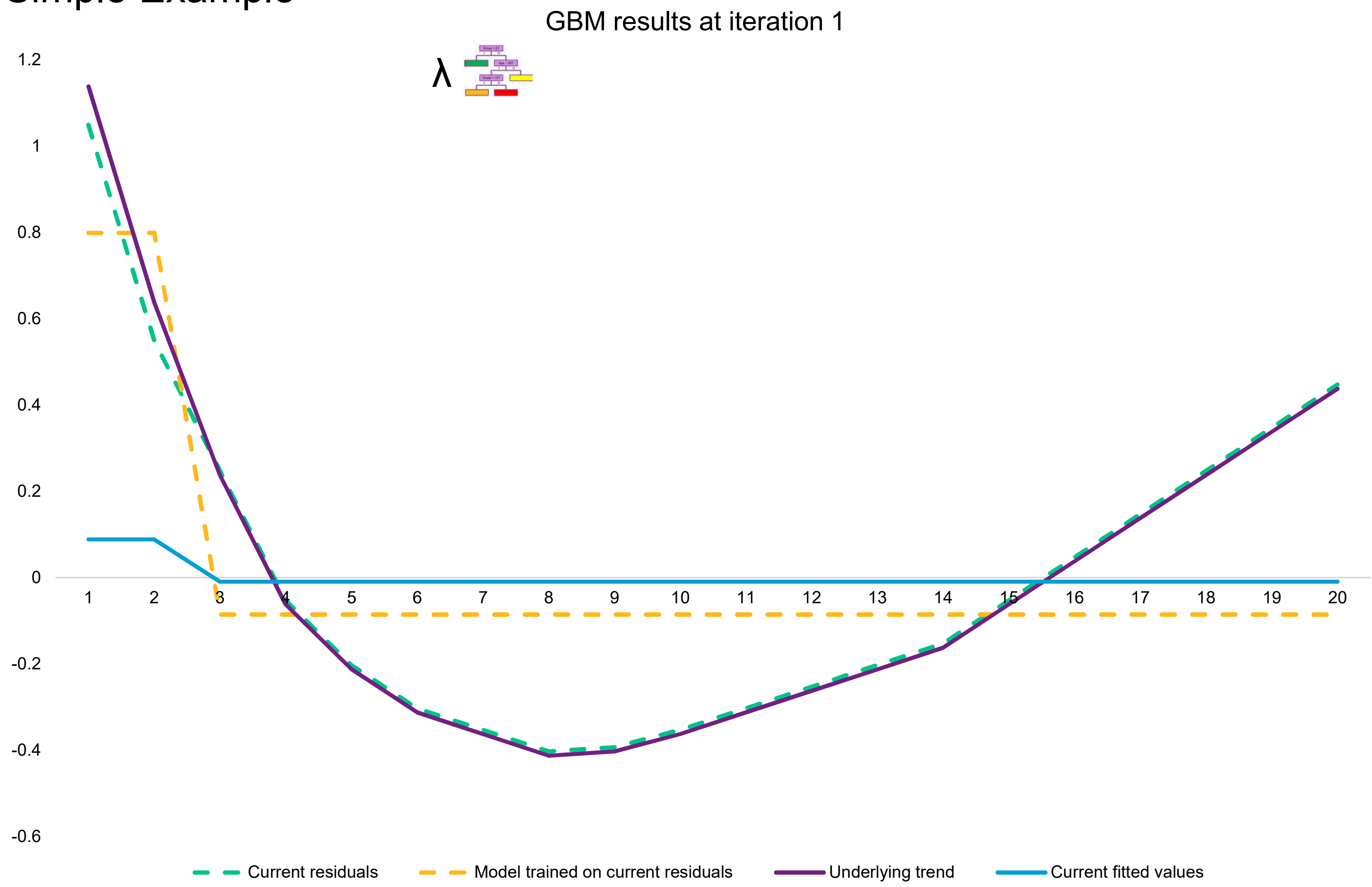
How Does a GBM Work?

A Simple Example



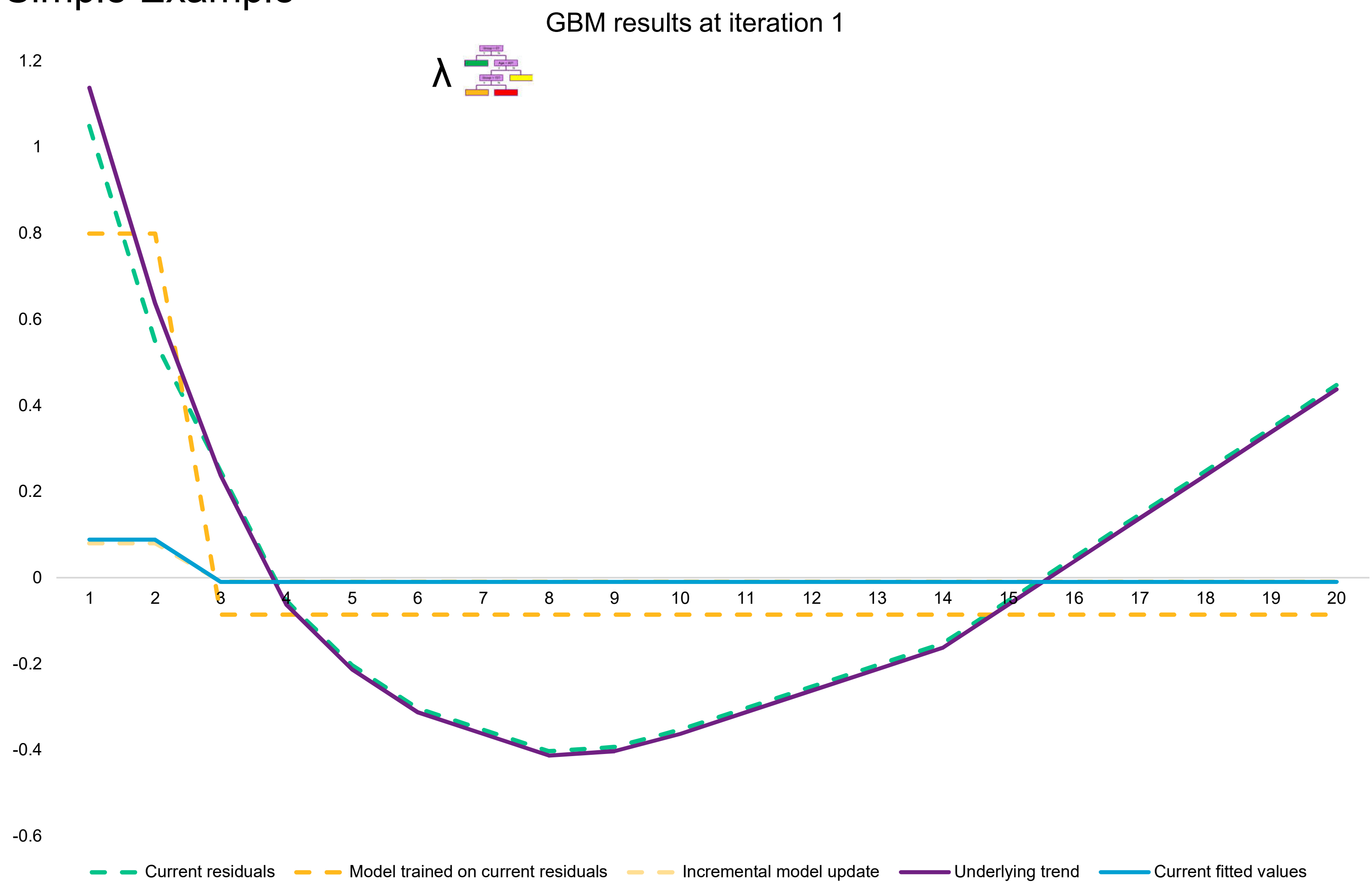
How Does a GBM Work?

A Simple Example



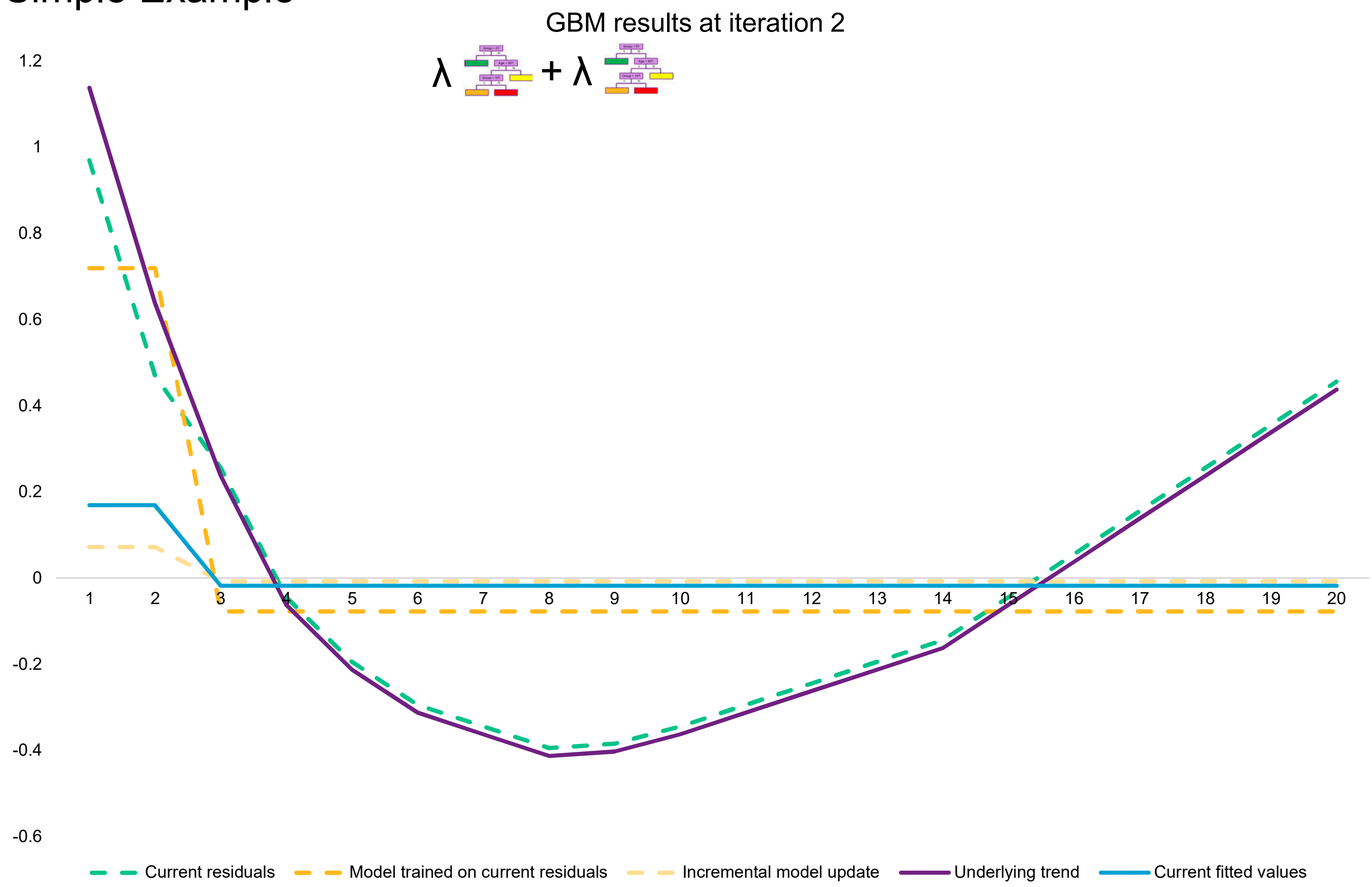
How Does a GBM Work?

A Simple Example



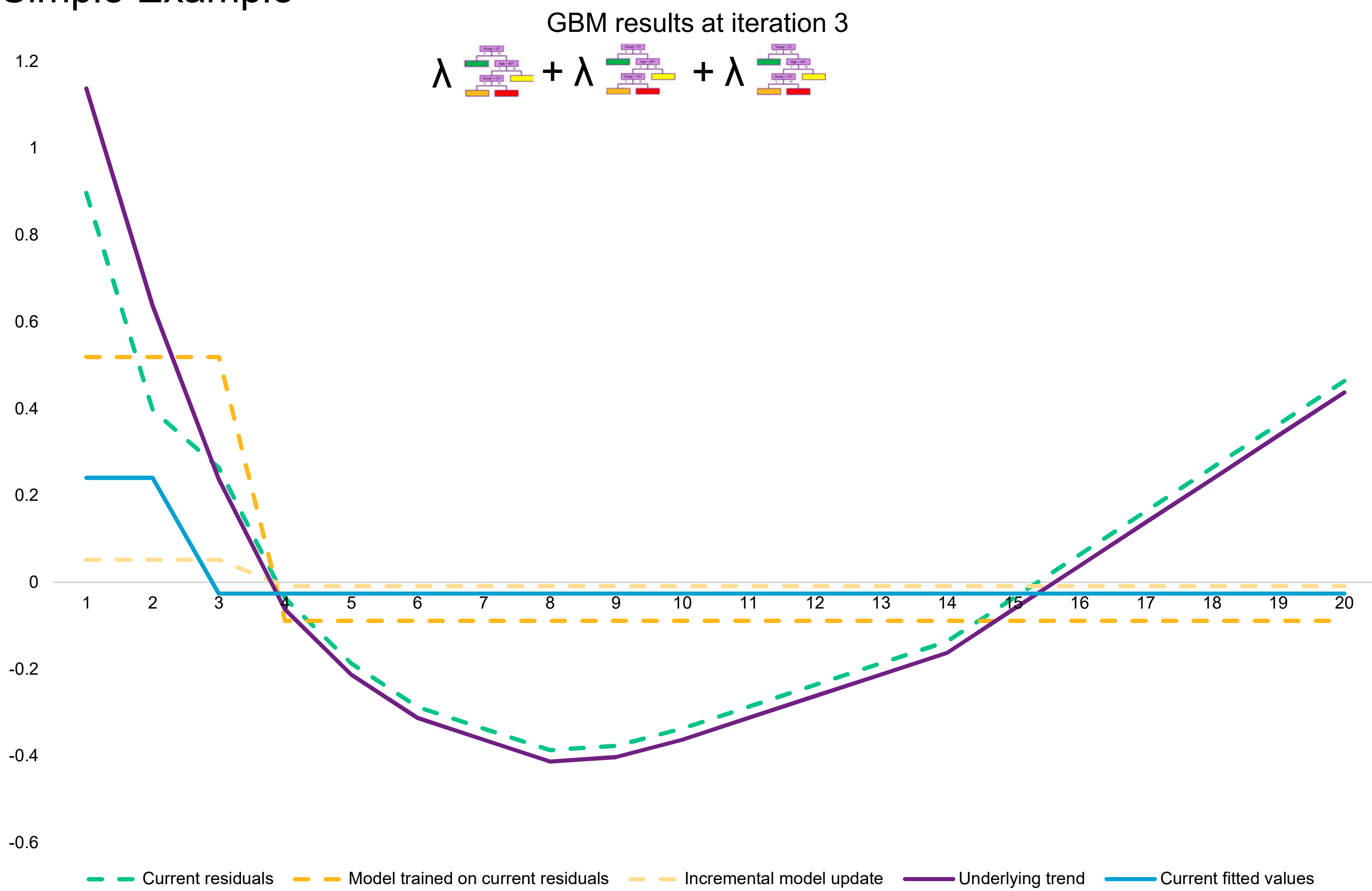
How Does a GBM Work?

A Simple Example



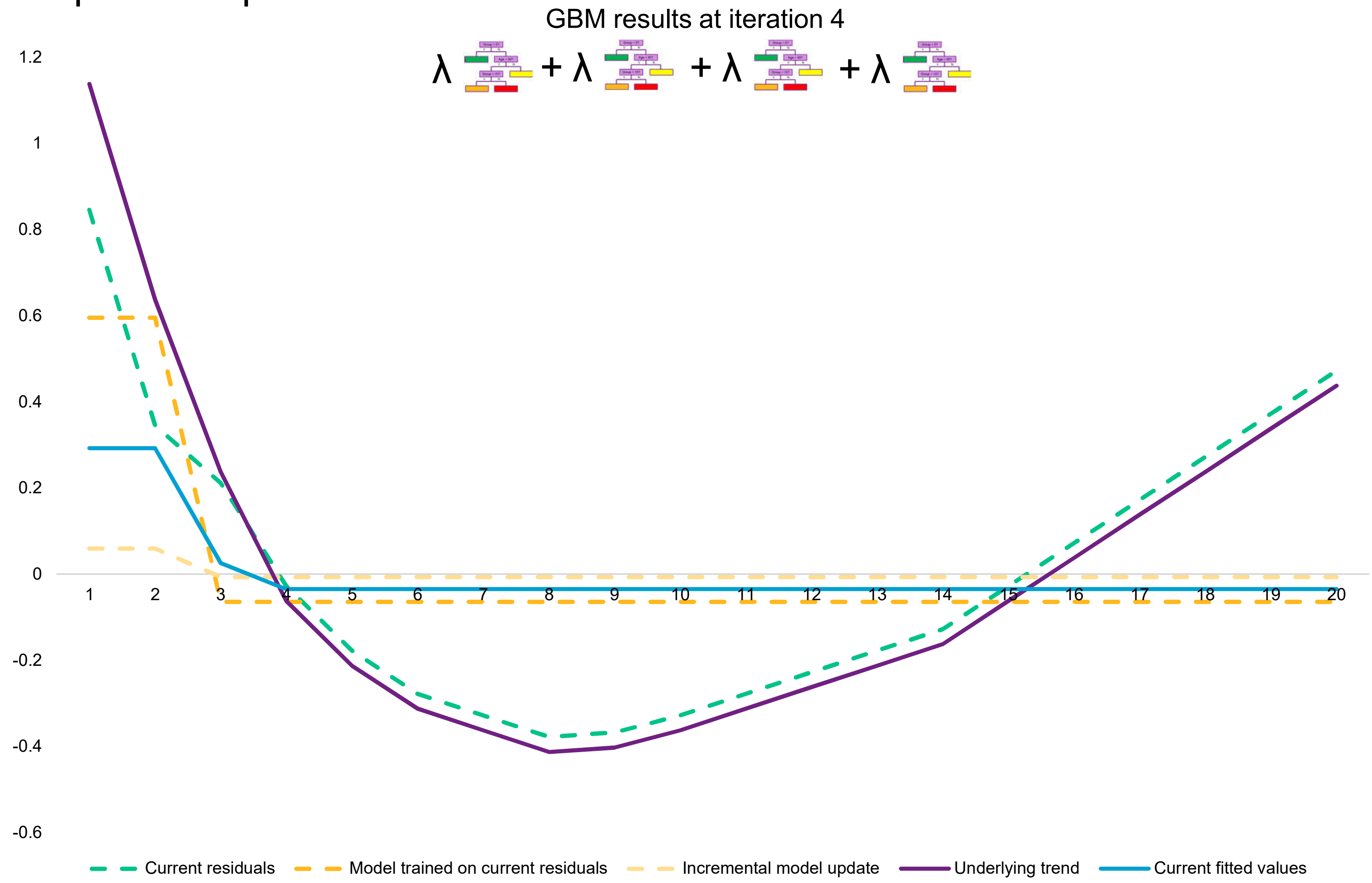
How Does a GBM Work?

A Simple Example



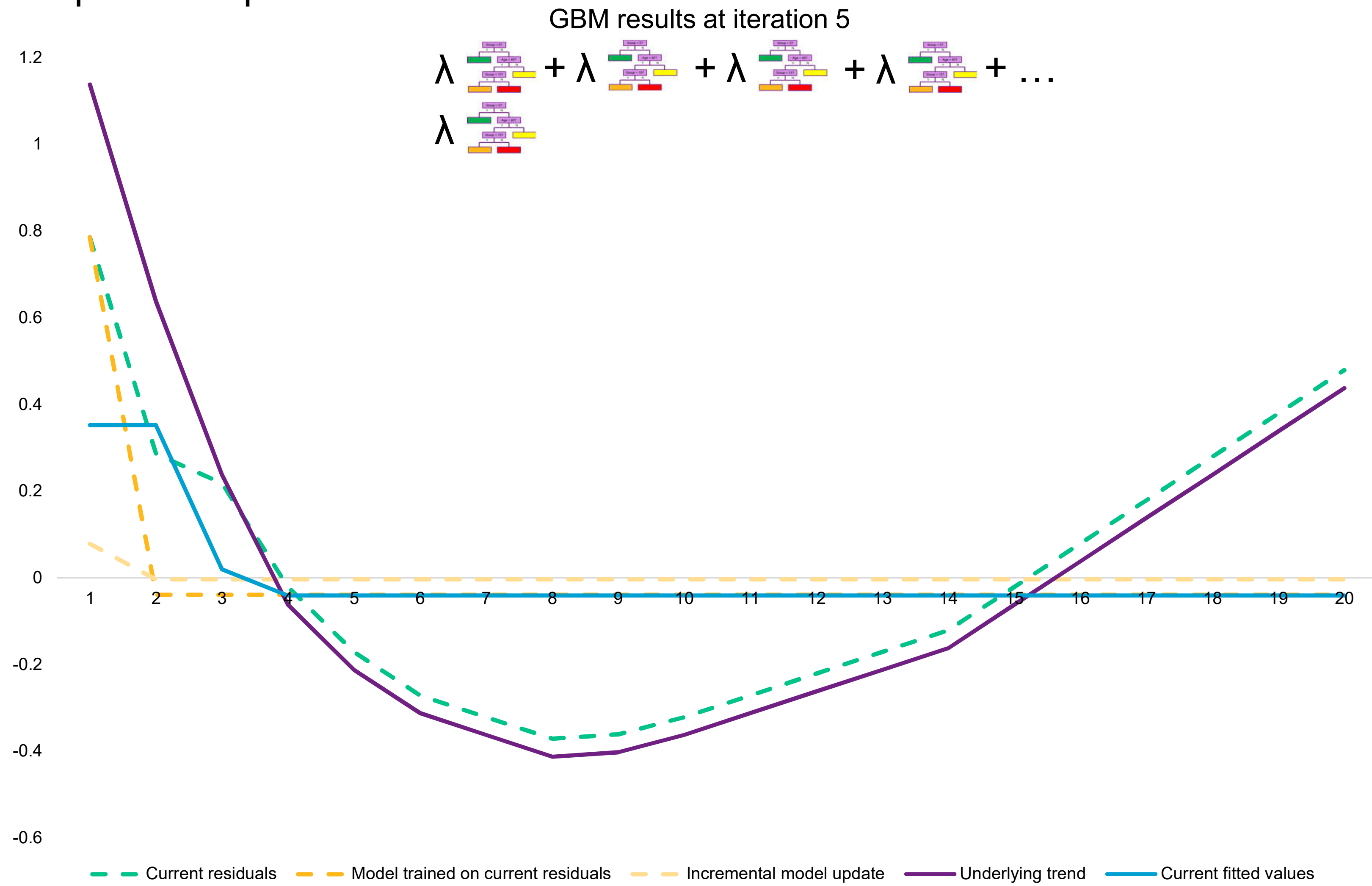
How Does a GBM Work?

A Simple Example



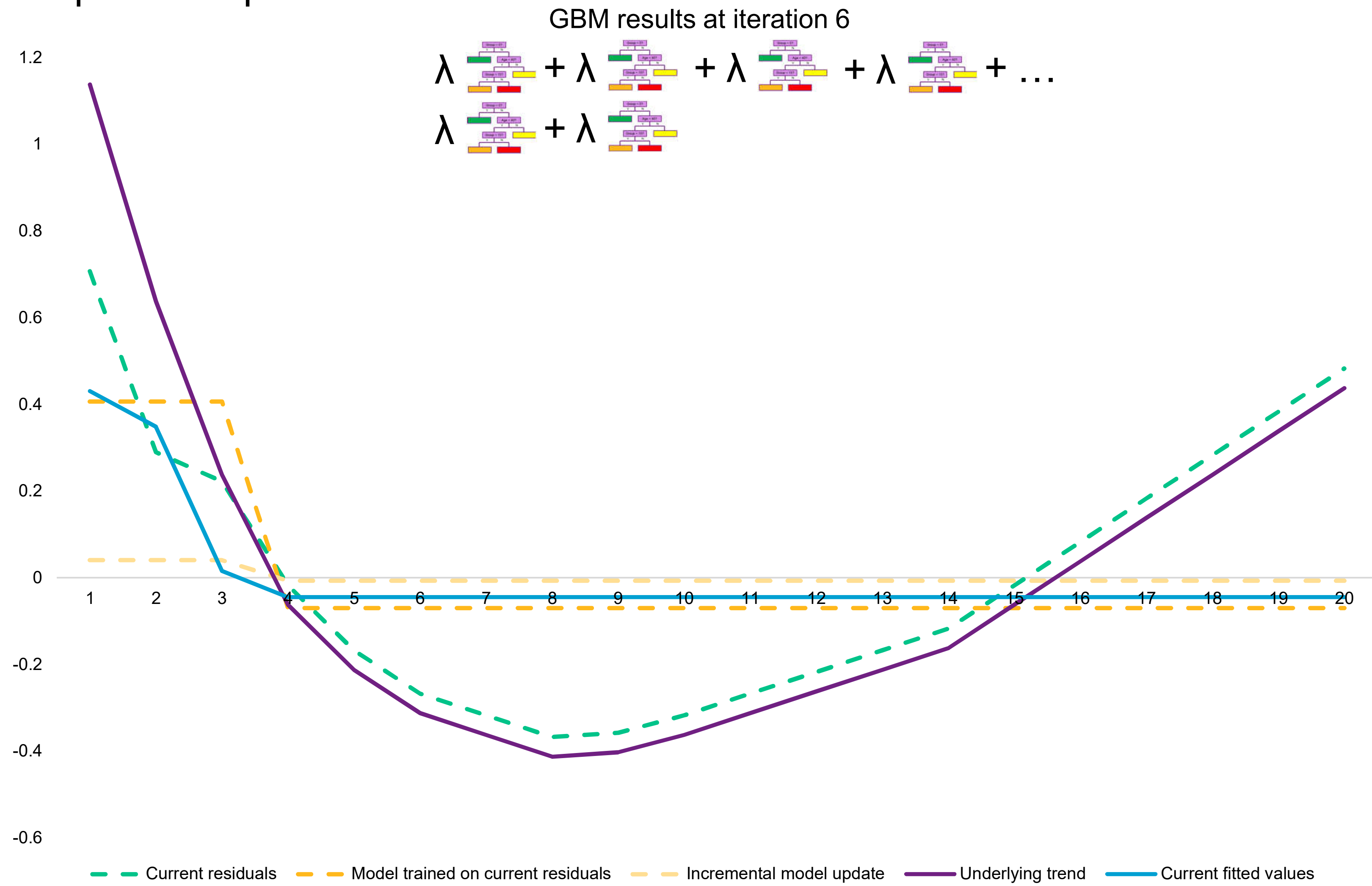
How Does a GBM Work?

A Simple Example



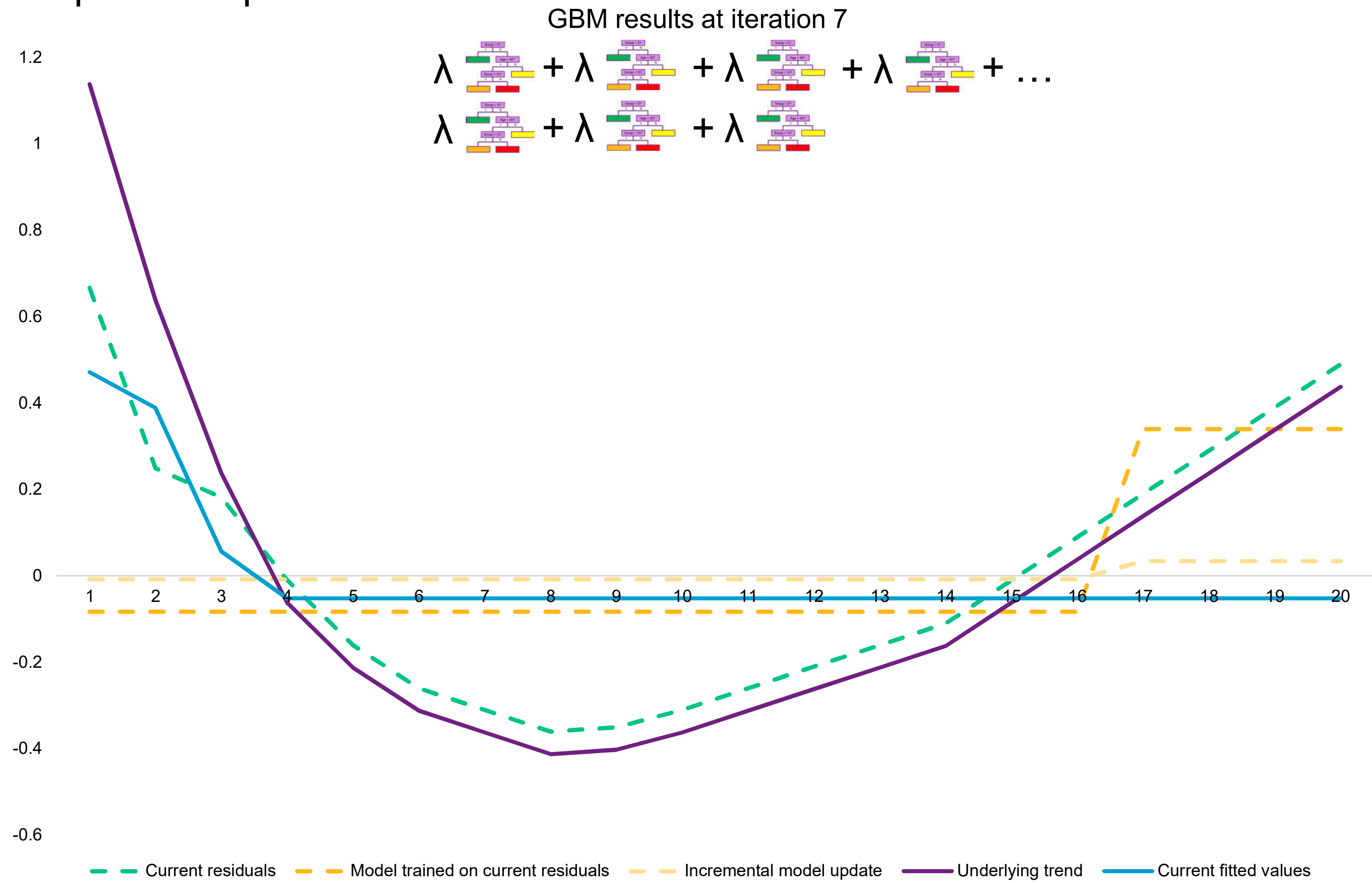
How Does a GBM Work?

A Simple Example



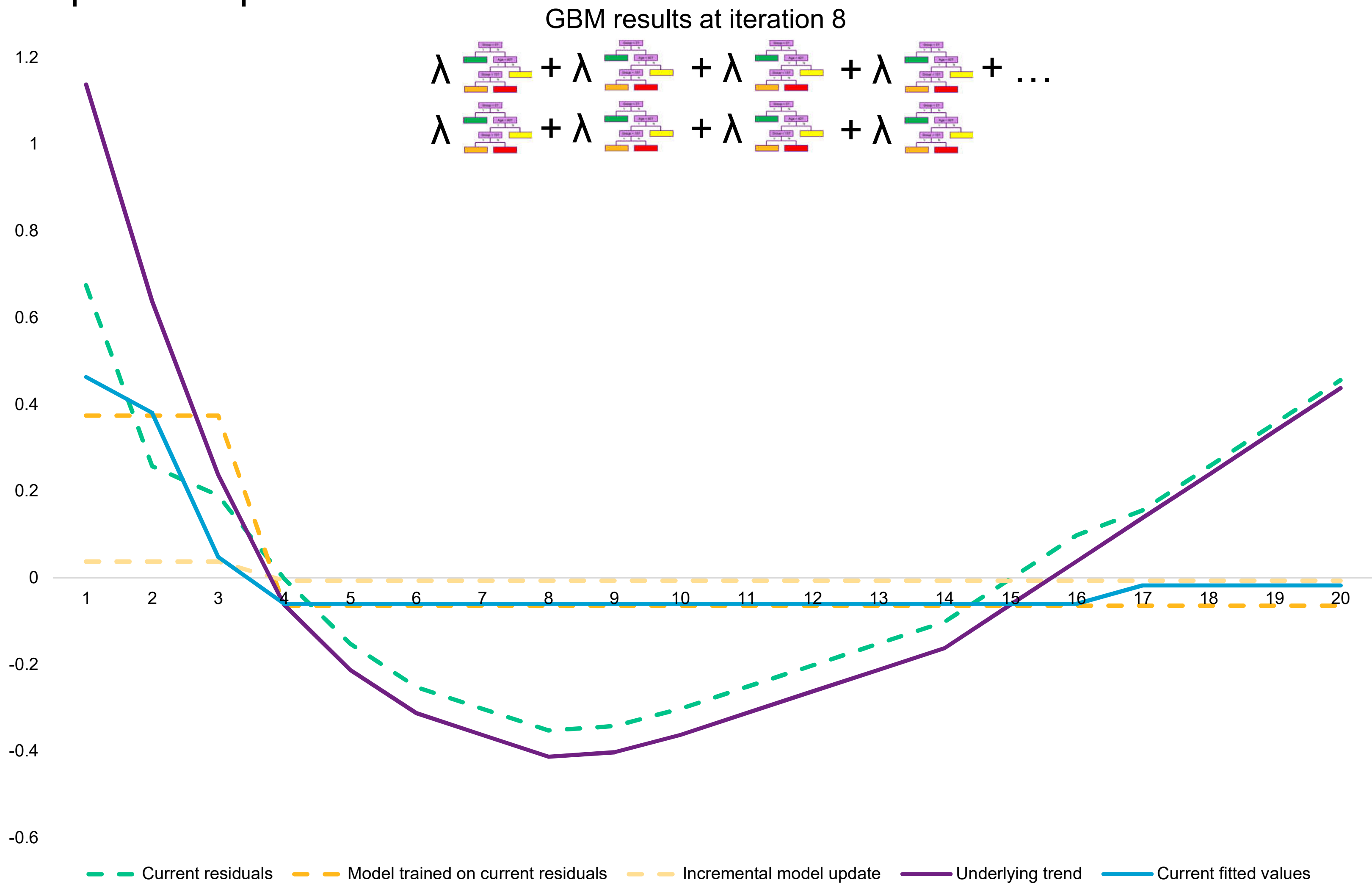
How Does a GBM Work?

A Simple Example



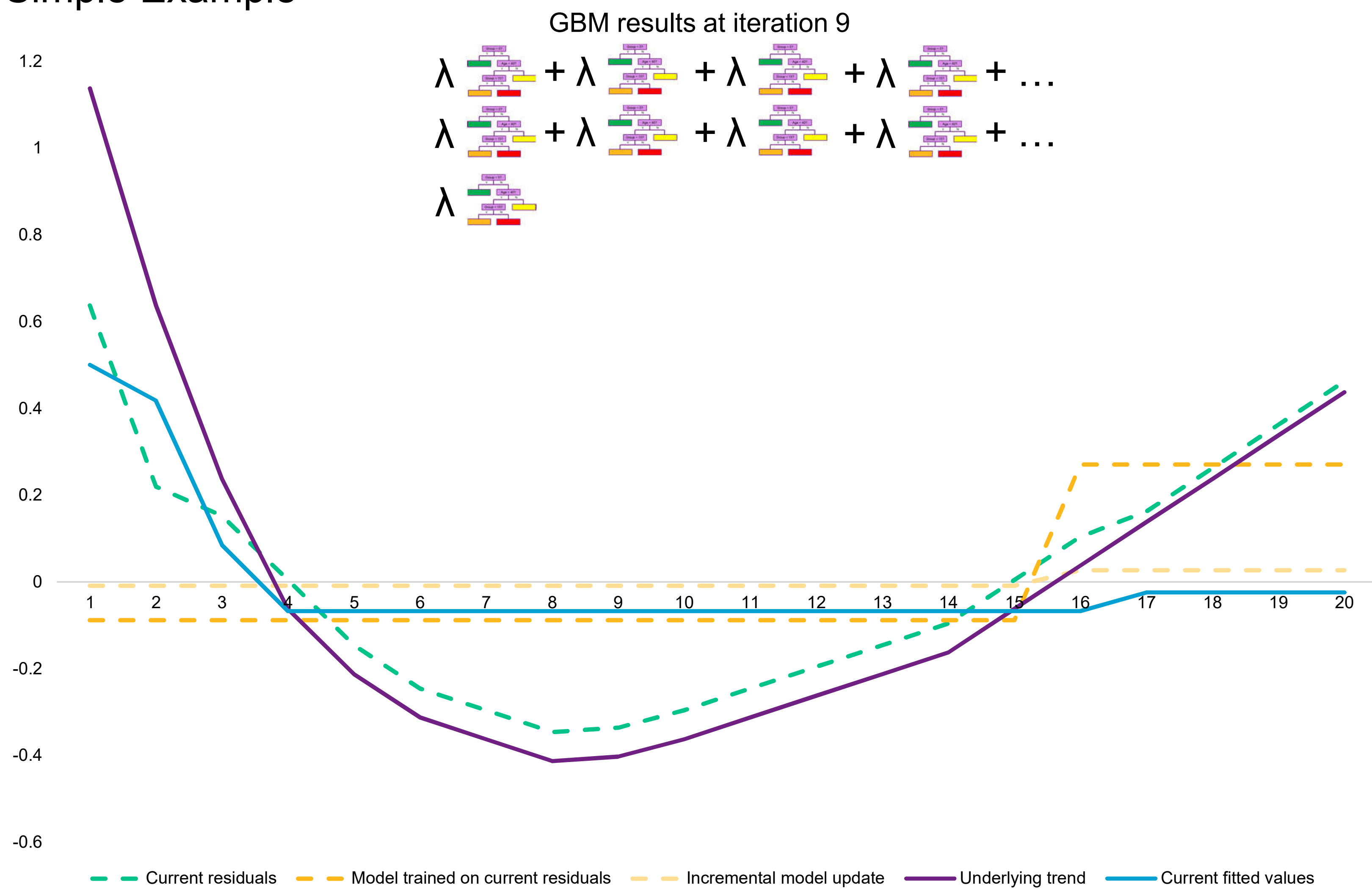
How Does a GBM Work?

A Simple Example



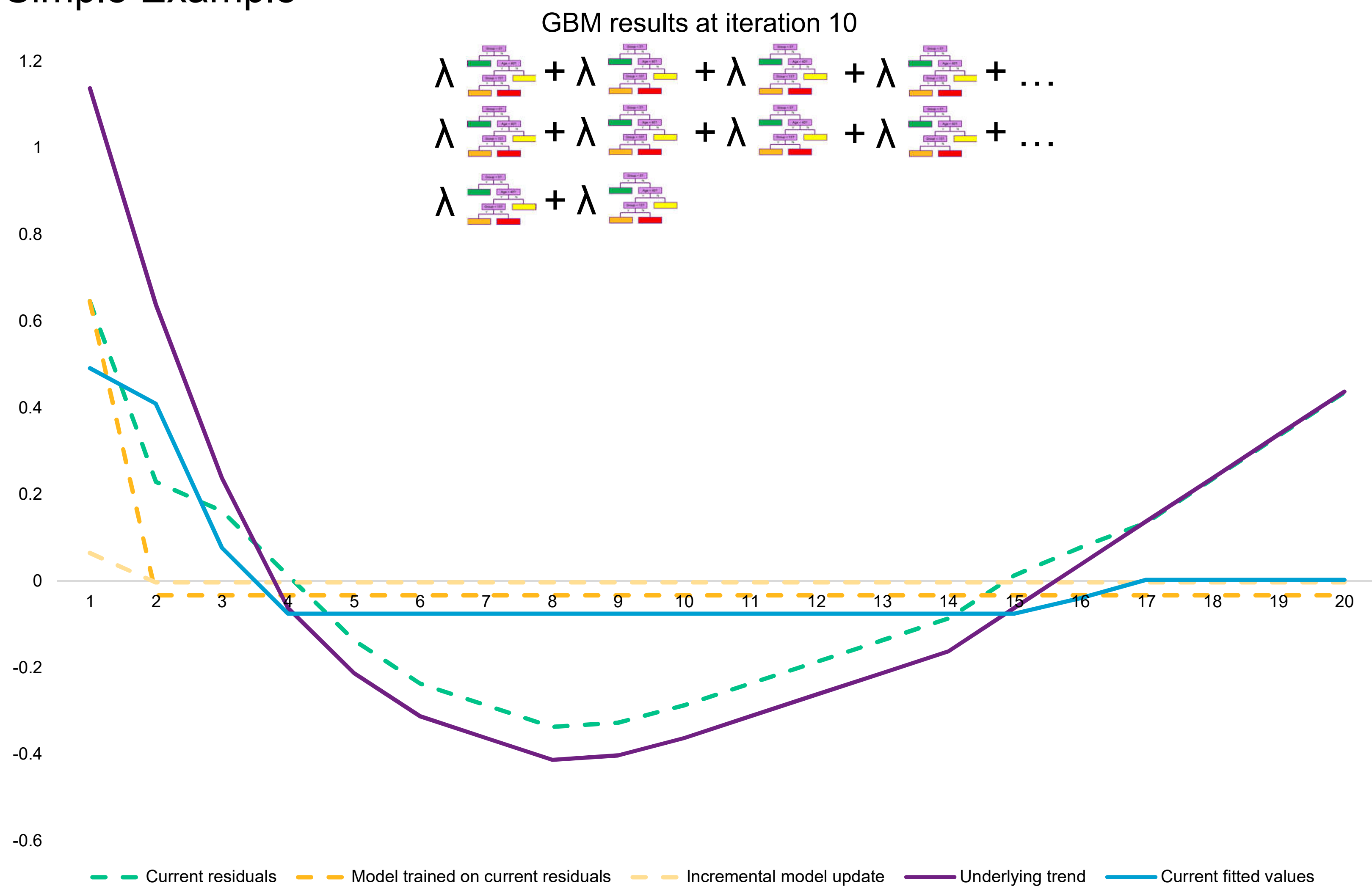
How Does a GBM Work?

A Simple Example



How Does a GBM Work?

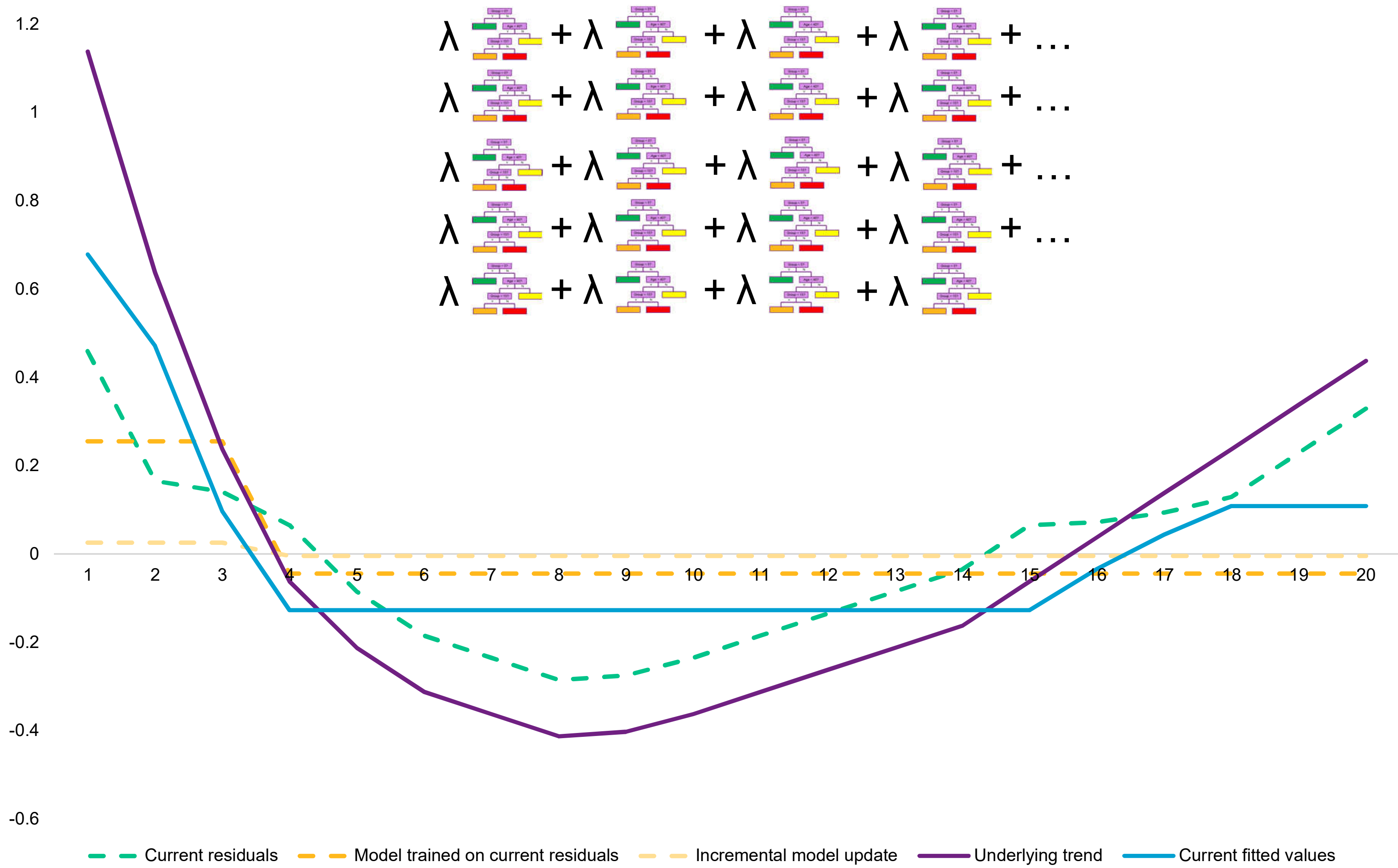
A Simple Example



How Does a GBM Work?

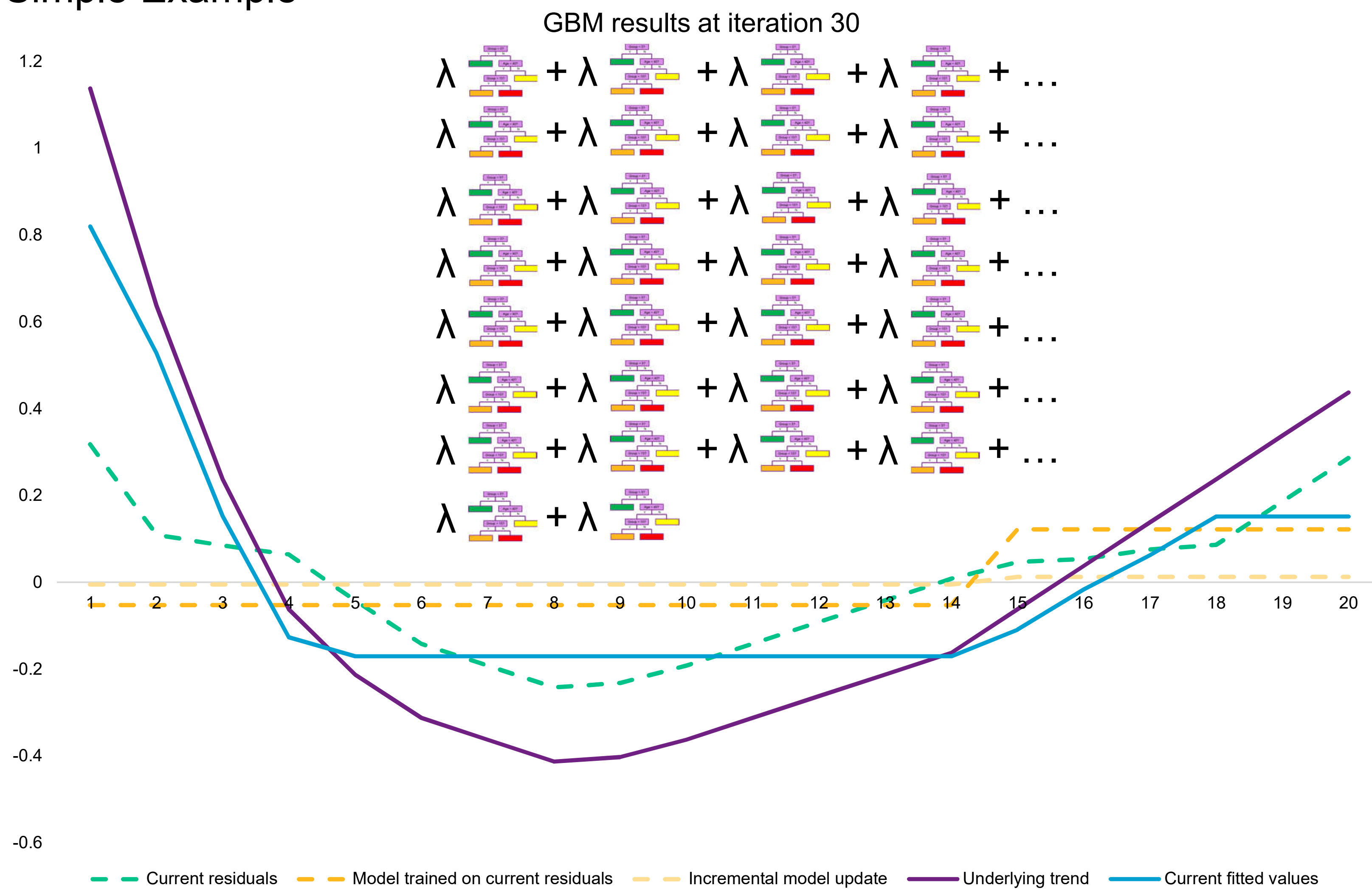
A Simple Example

GBM results at iteration 20



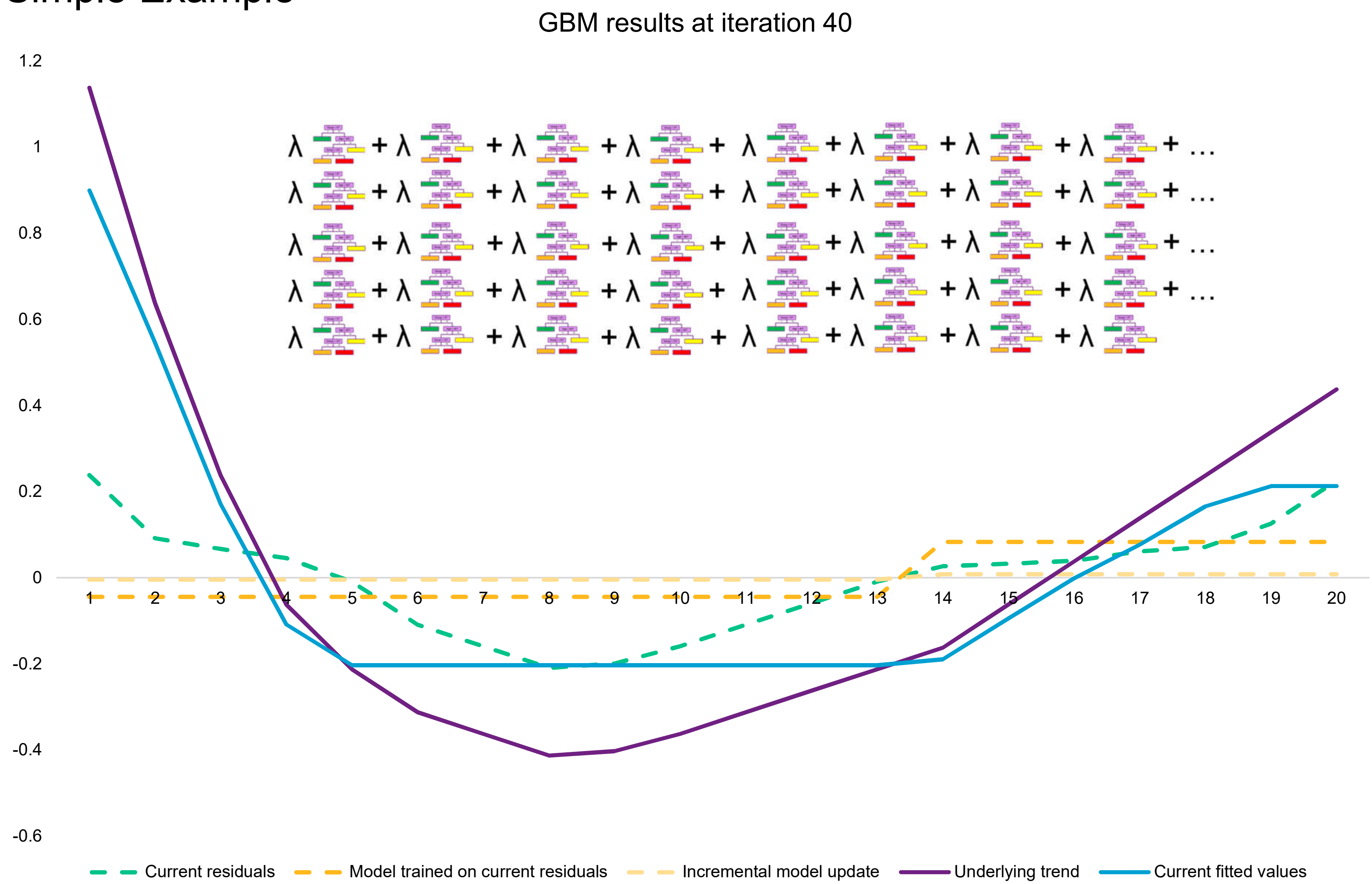
How Does a GBM Work?

A Simple Example



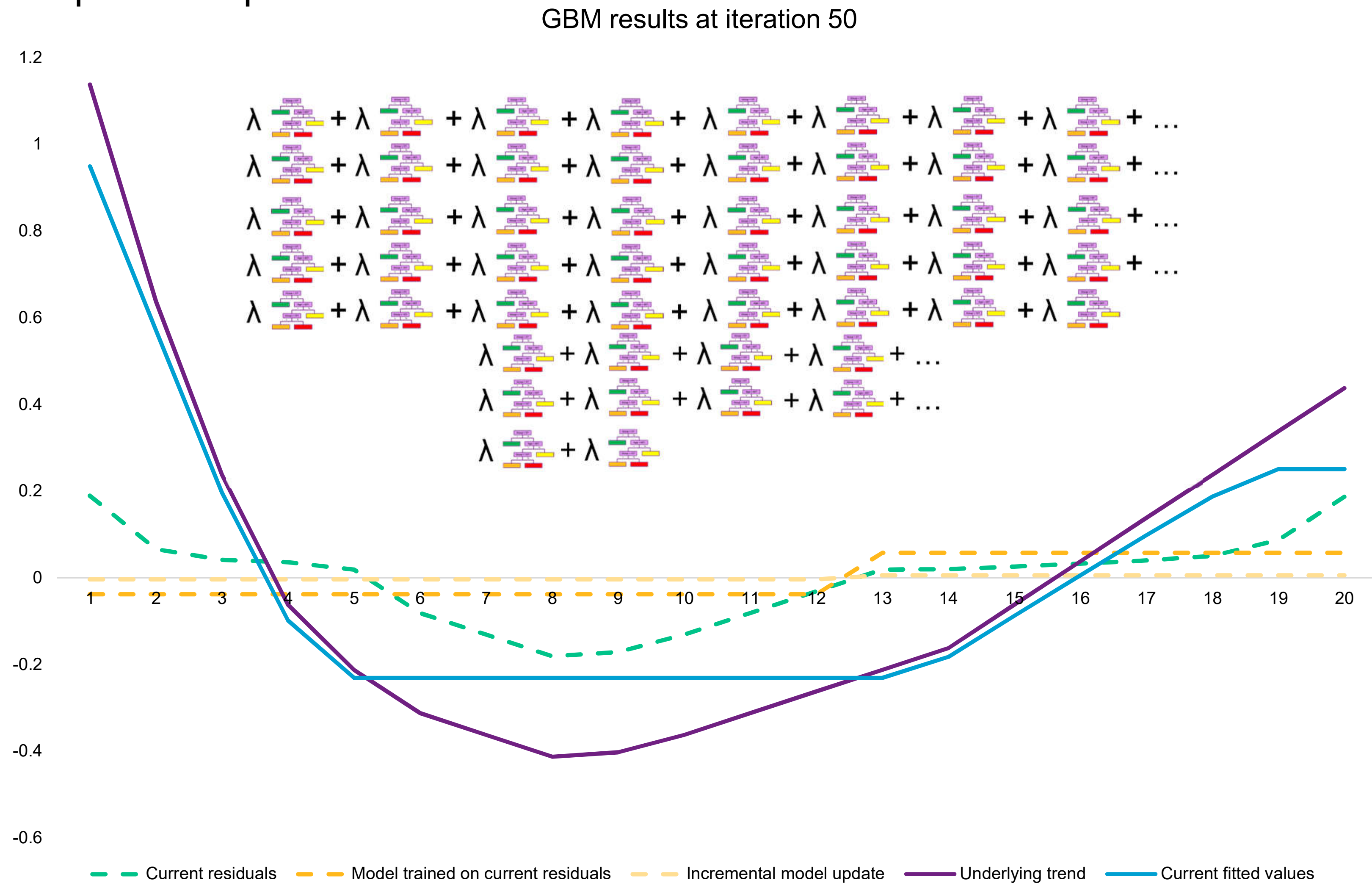
How Does a GBM Work?

A Simple Example



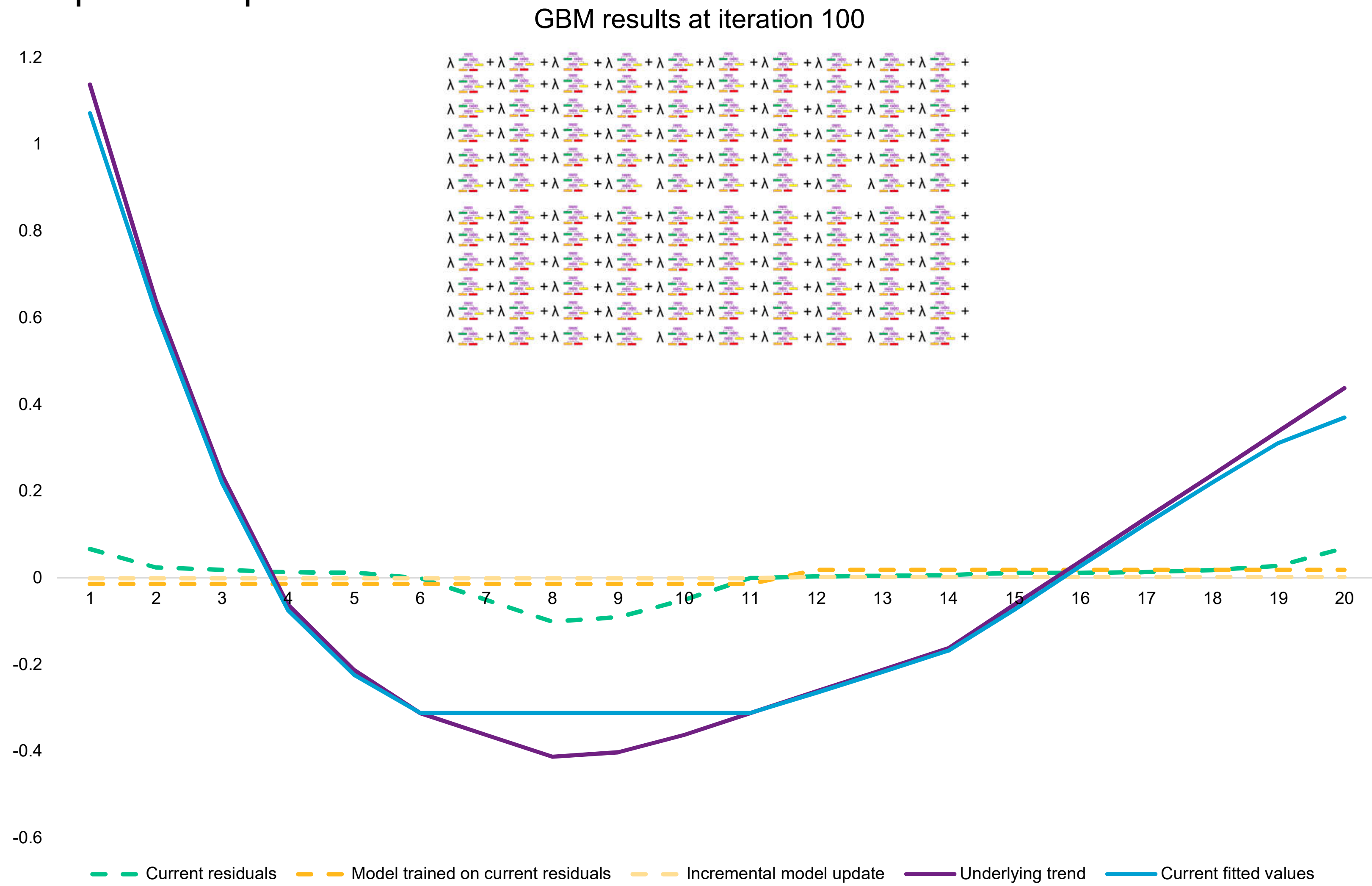
How Does a GBM Work?

A Simple Example



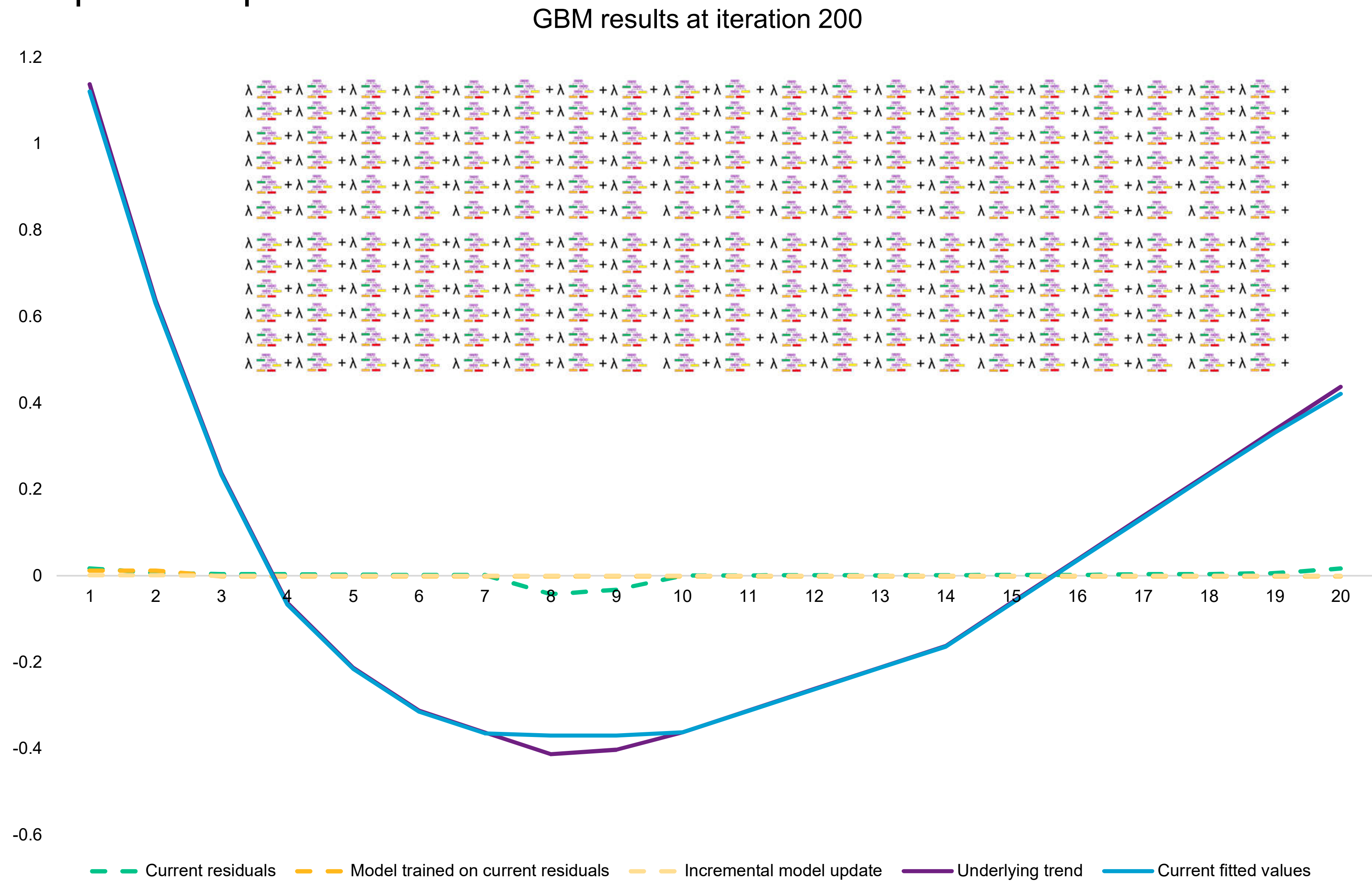
How Does a GBM Work?

A Simple Example



How Does a GBM Work?

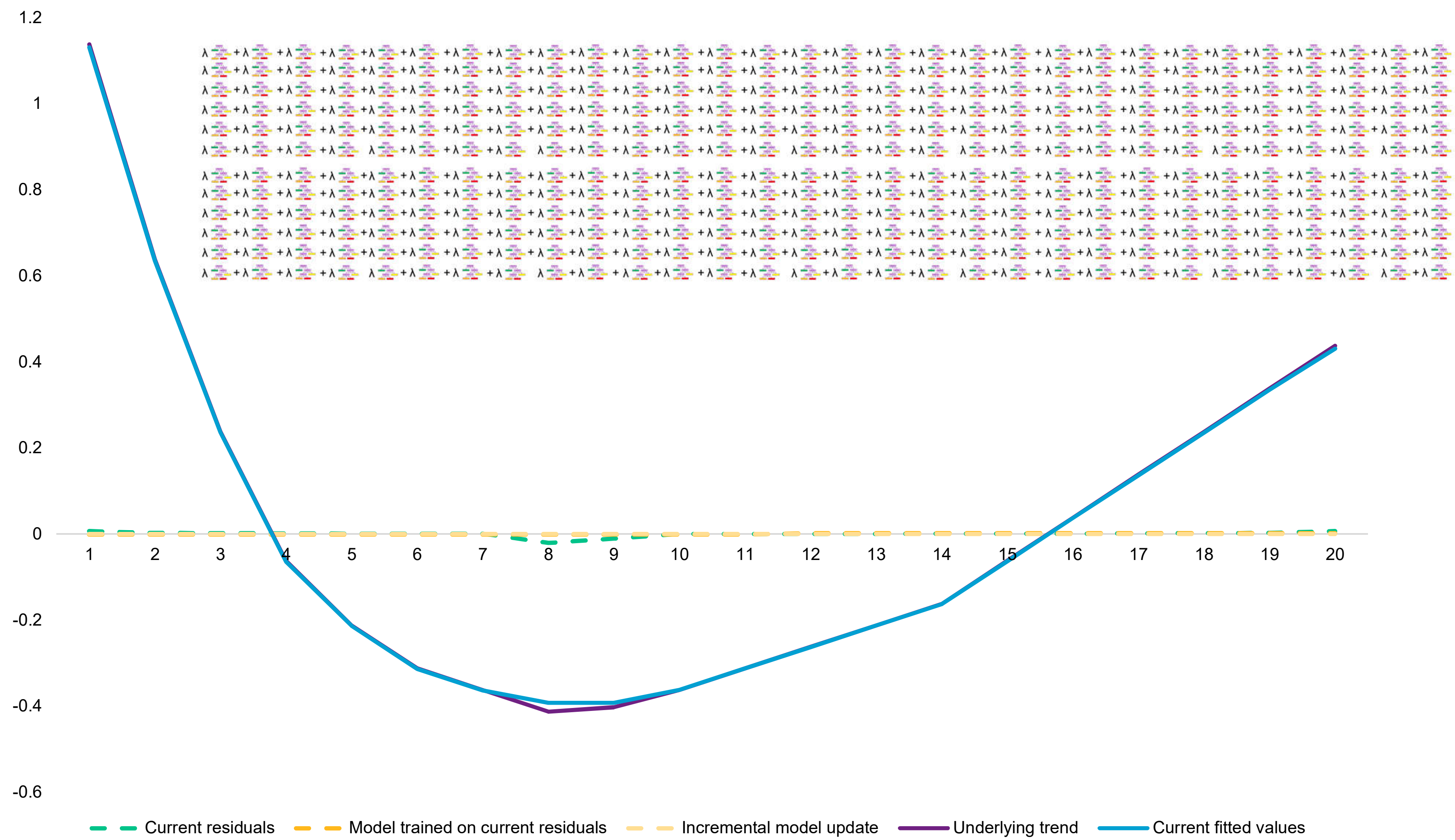
A Simple Example



How Does a GBM Work?

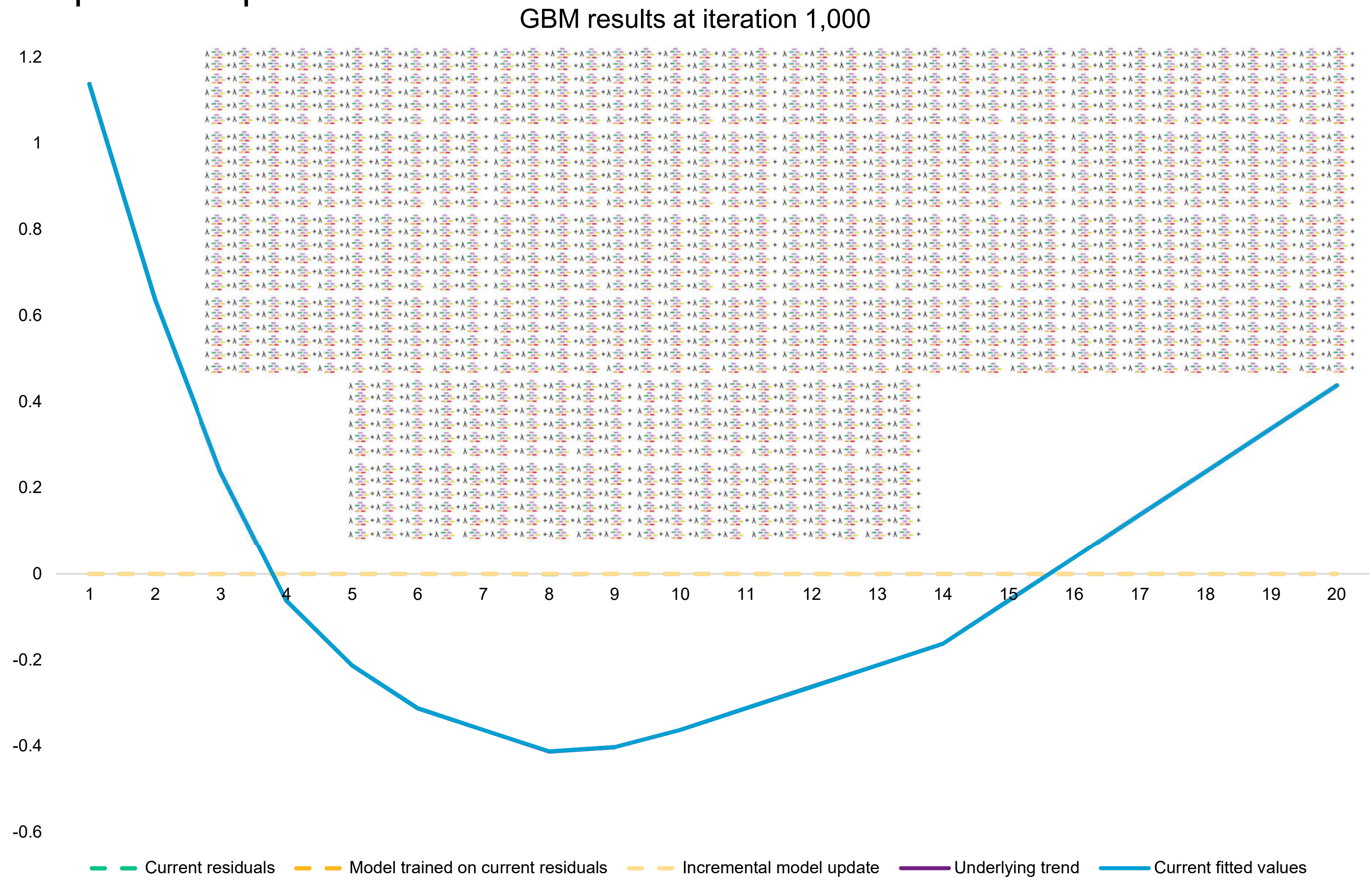
A Simple Example

GBM results at iteration 300



How Does a GBM Work?

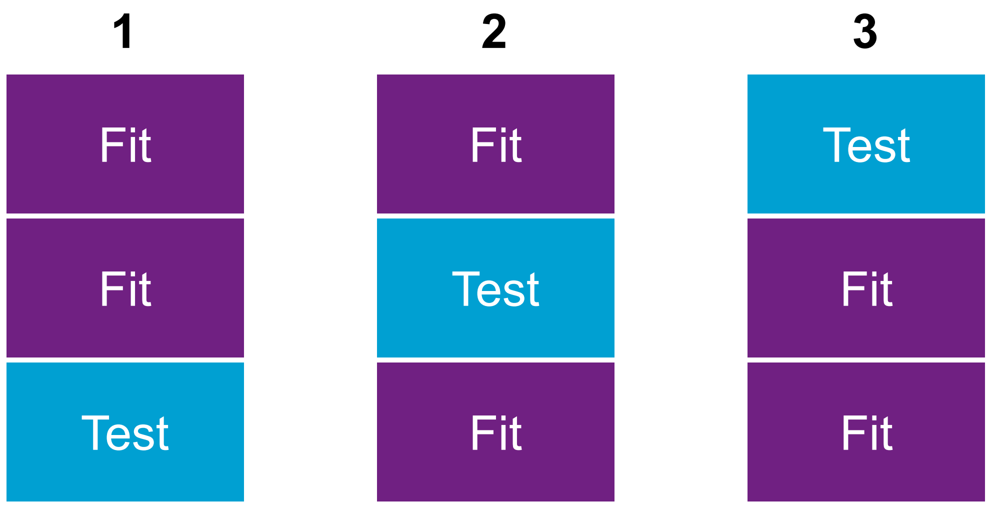
A Simple Example



How Does a GBM Work?

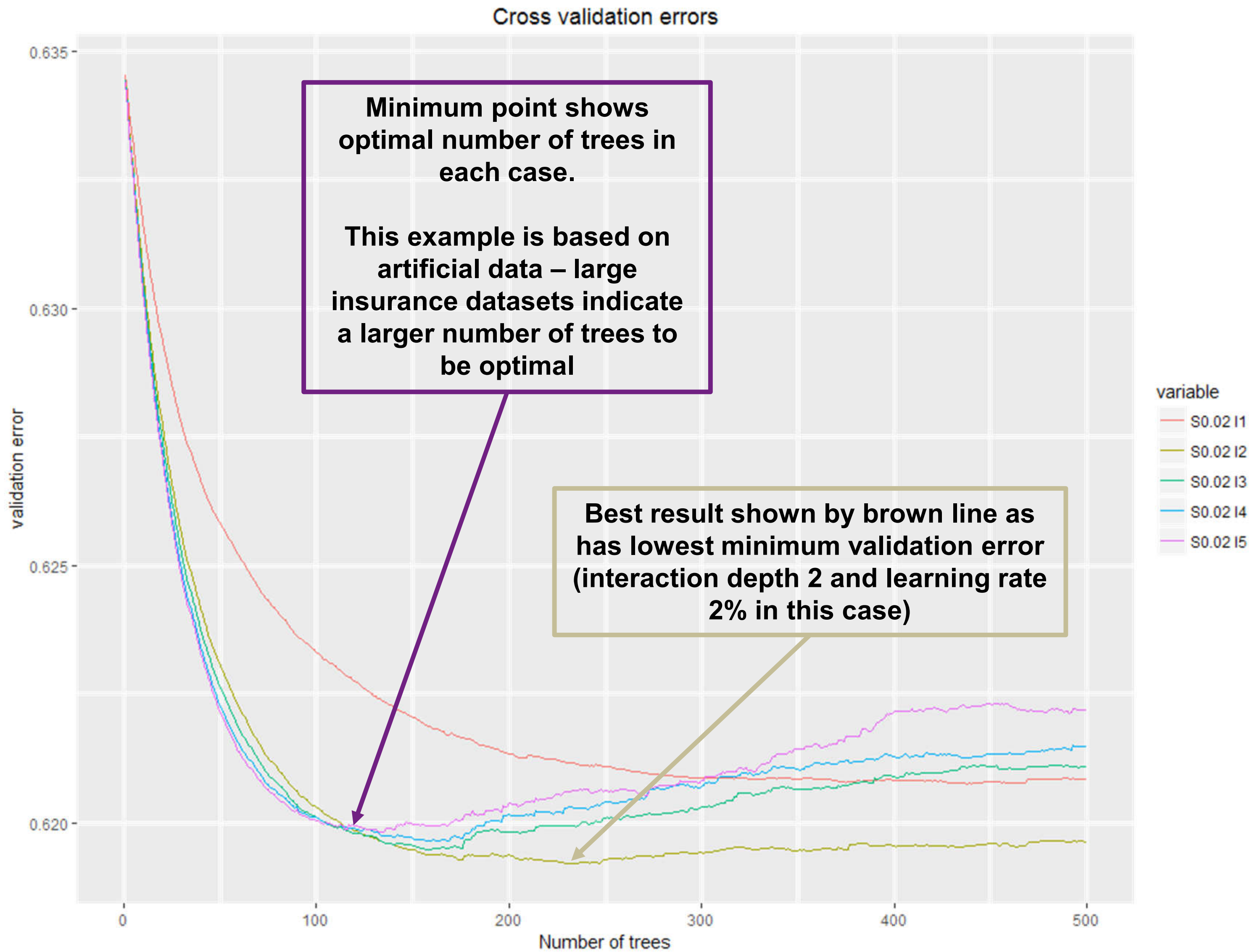
Calibrating the Assumptions

- *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

Example 5-fold cross validation



Application #1

Off-balancing (5% rate increase)

- Inforce Dataset

Inforce Policy	Current Premium	Proposed Premium	Expected Losses	Retention
1	974	1,023	682	0.88
2	958	1,006	680	0.89
3	950	998	684	0.90
4	968	1,016	707	0.91
5	986	1,035	730	0.92
6	955	1,003	716	0.93
7	965	1,013	733	0.94
8	963	1,011	742	0.95
9	973	1,022	759	0.96
10	961	1,009	807	0.97
Total	9,653	10,136	7,240	

Application #1

Off-balancing (5% rate increase)

- Quote Dataset

Quote	Proposed Premium	Expected Losses	Conversion	Retention
1	1,044	835	0.20	0.80
2	1,048	891	0.22	0.82
3	1,063	914	0.24	0.85
4	1,079	950	0.26	0.88
5	1,095	986	0.28	0.92
Total	5,329	4,575		

Application #1

Off-balancing (5% rate increase)

- Traditional view focuses on inforce dataset
  - Current loss ratio = 7,240/9,653 = 0.75
  - Proposed loss ratio = 7,240/10,136 = 0.714

- Alternative view focuses on inforce and quote datasets AND considers demand
  - Demand-weighted proposed premium on inforce dataset = 1,023(0.88) + … + 1,009(0.97) = 9,376
  - Demand-weighted expected losses on inforce dataset = 6,707
  - Demand-weighted proposed premium on quote dataset = 1,044(0.20)(0.80) + … + 1,095(0.28)(0.92) = 1,102
  - Demand-weighted expected loss on quote dataset = 952
  - Current loss ratio = 7,240/9,653 = 0.75
  - Demand-weighted proposed loss ratio = (6,707 + 952)/(9,376 + 1,102) = 0.731

- **Key point: Traditional off-balance approach may lead to insufficient rate**

Application #2

Multi-period Simulation

- Personal auto
- Renewal dataset & quote dataset
- Time Horizon – four periods, each lasting six months
- Quote growth rate – 5% each period
- Quotes do not enter simulation until new rates go into effect at the beginning of period 1
- Quote distribution constant over time
- Aging assumptions
  - Operators age by 1 every other period
  - Vehicles age by 1 every other period
- Current loss ratio is 75%
- Ignore trend
- Scenarios
  - 5% base rate decrease
  - 15% decrease operators aged 25-30 off-balanced to an overall 5%

Application #2

Multi-period Simulation

Quotes							
	Period	**Policies Offered**	**Policies Written**	**Conversion**	**Policies Retained**	**Retention**	**Profit Margin**
Scenario 1	0	N/A	N/A	N/A	N/A	N/A	N/A
	1	20,000	5,493	27.5%	4,669	85.0%	1.9%
	2	21,000	5,767	27.5%	4,902	85.0%	1.9%
	3	22,050	6,058	27.5%	5,150	85.0%	1.9%
	4	23,153	6,360	27.5%	5,406	85.0%	1.9%
Scenario 2	0	N/A	N/A	N/A	N/A	N/A	N/A
	1	20,000	5,646	28.2%	4,743	84.0%	1.8%
	2	21,000	5,928	28.2%	4,980	84.0%	1.8%
	3	22,050	6,228	28.2%	5,231	84.0%	1.8%
	4	23,153	6,538	28.2%	5,492	84.0%	1.8%

Application #2

Multi-period Simulation

Renewals					
	Period	**Policies Offered**	**Policies Retained**	**Retention**	**Profit Margin**
Scenario 1	0	50,000	44,000	88.0%	2.5%
	1	44,000	41,287	93.8%	2.4%
	2	45,956	44,162	96.1%	2.3%
	3	49,064	47,147	96.1%	2.2%
	4	52,296	49,315	94.3%	2.2%
Scenario 2	0	50,000	44,000	88.0%	2.5%
	1	44,000	41,287	93.8%	2.4%
	2	46,030	44,155	95.9%	2.5%
	3	49,135	47,121	95.9%	2.6%
	4	52,352	49,263	94.1%	2.7%

Application #2

Multi-period Simulation

Quotes + Renewals							
	Period	Policies Offered	Policies Written	Policies Retained	Earned Premium	Profit Margin	Absolute Profit
Scenario 1	0	50,000	50,000	44,000	$35,250,000	2.5%	$881,250
	1	64,000	49,493	45,956	$34,486,258	2.3%	$810,152
	2	66,956	51,723	49,064	$36,412,258	2.3%	$822,930
	3	71,114	55,122	52,296	$38,800,399	2.2%	$842,146
	4	75,449	58,657	54,722	$40,949,798	2.2%	$888,759
Scenario 2	0	50,000	50,000	44,000	$35,250,000	2.5%	$881,250
	1	64,000	49,646	46,030	$34,729,064	2.3%	$812,026
	2	67,030	51,958	49,135	$36,692,114	2.4%	$891,271
	3	71,185	55,363	52,352	$39,087,466	2.5%	$985,029
	4	75,505	58,890	54,755	$41,236,423	2.6%	$1,076,159

- **Key point: A multi-period view is often needed to make the best rate decision**

Conclusions

- GBMs provide a robust alternative to GLMs for modeling demand
- Traditional off-balance approach may lead to insufficient rate
- A multi-period view is often needed to make the best rate decision

Thank You!

Michael McPhail

Director

Willis Towers Watson

Michael.mcphail@willistowerswatson.com

Serhat Guven

Managing Director

Willis Towers Watson

Serhat.guven@willistowerswatson.com