Machine Learning in Life Insurance

Searching for Patterns in Cash Flow Models

Pierre Joos

Hannover, 16. May 2019

Issue: Life insurance valuation is extremely complex

External data

- Capital market scenarios

Internal data

- Balance sheet data
- Company strategy

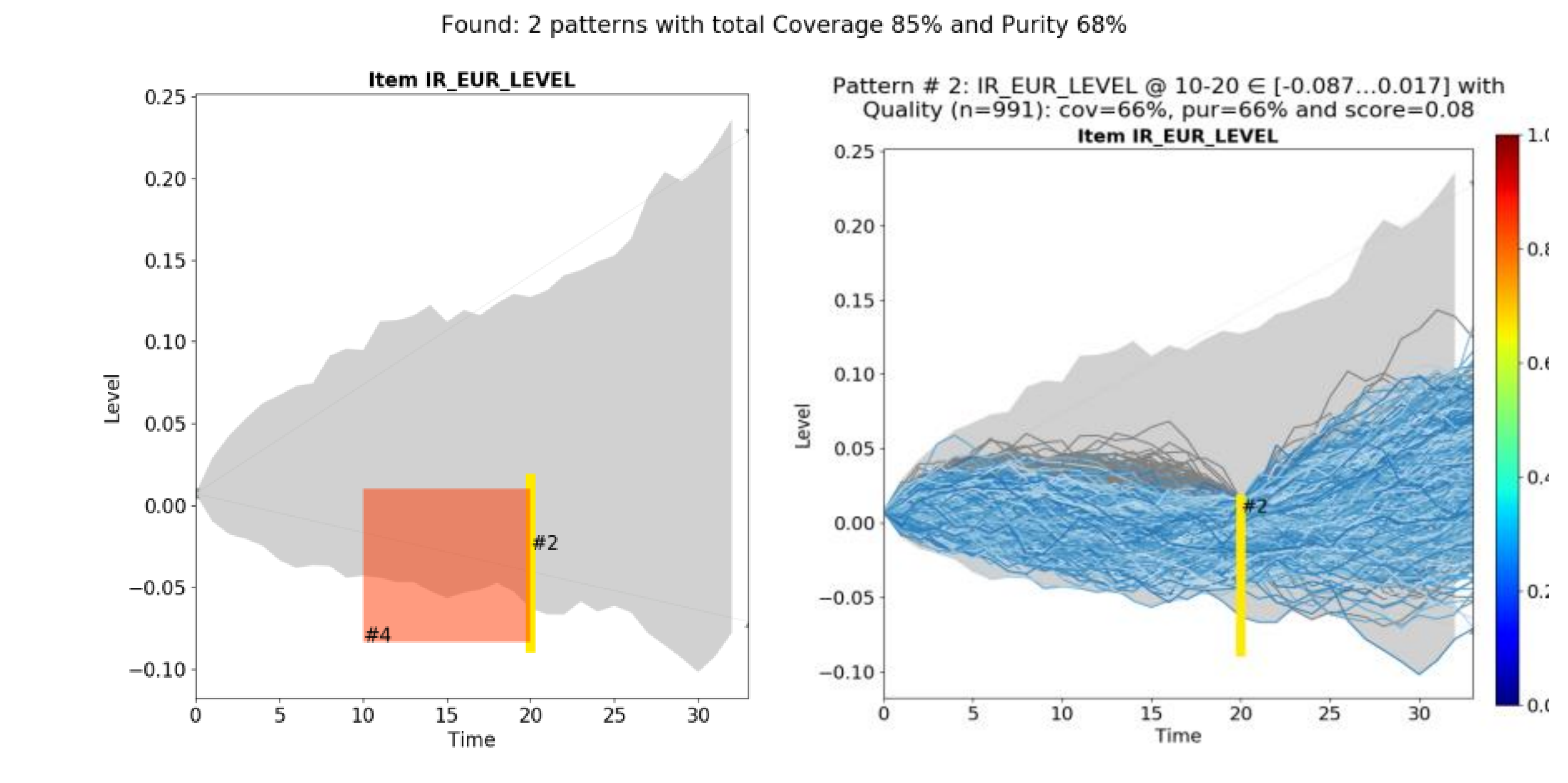
Cash Flow Model

Valuation

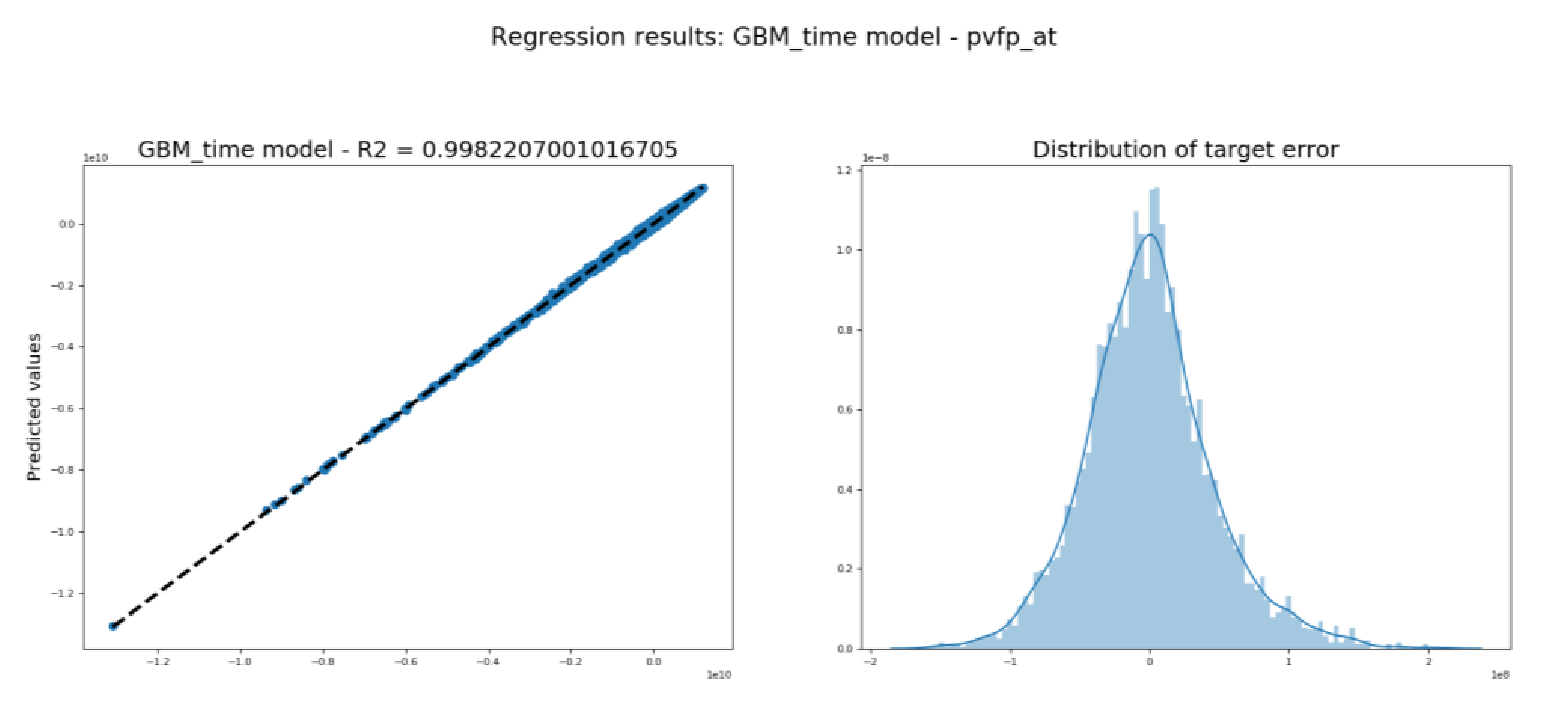
2 **Prediction:**
Can we predict the results?

1 **Analysis:**
What drives the results?

Analysis:
Long-term low interest rates are worst case for insurer



Prediction:
GBMs can predict the PVFP for a given market scenario quite good



Agenda

Life insurance policy

(Permanent) Life insurance:

Financial contract between policy holder and insurer where premiums are accumulated and paid out later

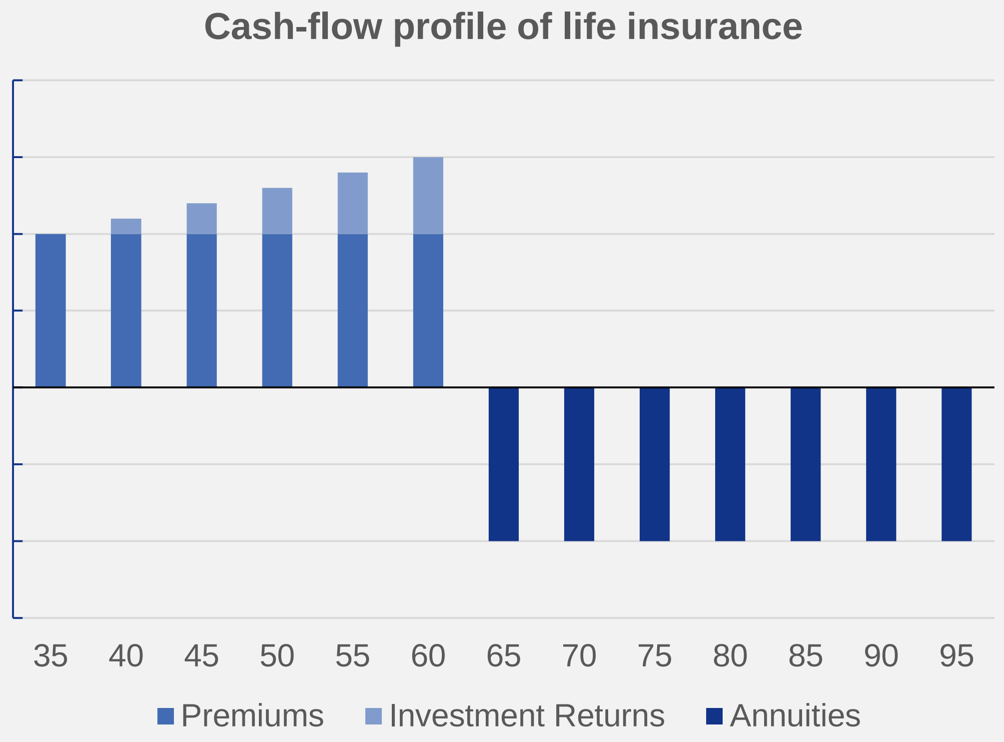
Cash-flows:

- Policy holder pays premiums until retirement
- Returns from investments are accumulated
- Insurer pays annuity after retirement

Uncertainties:

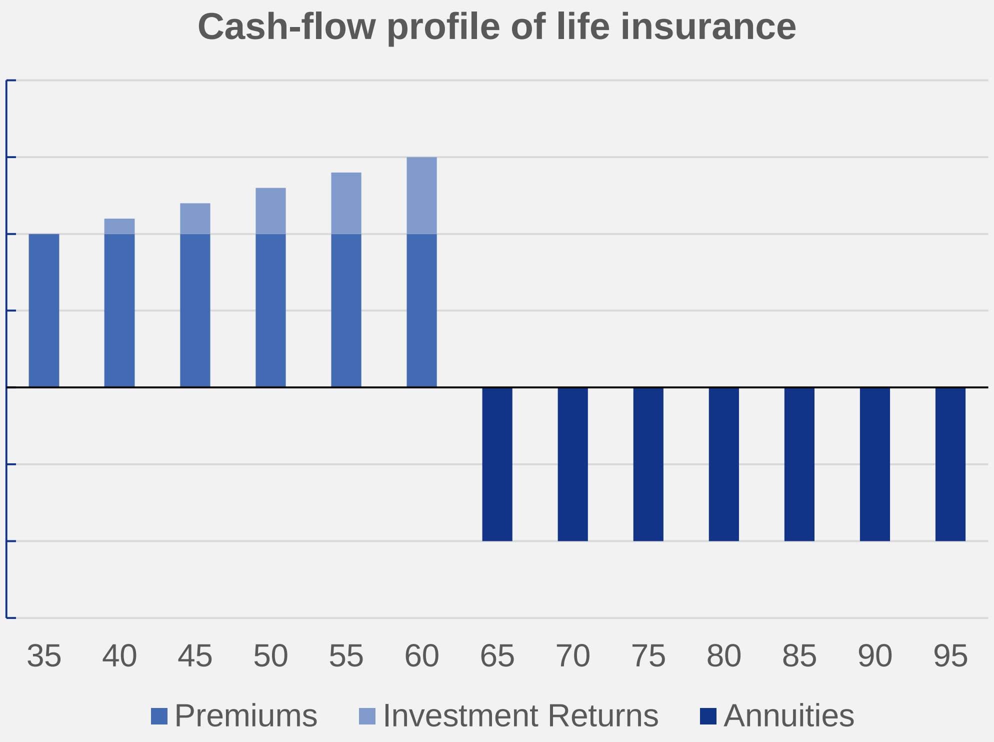
- Investment returns
- Longevity of policy holder

➔ **What is the total profit / loss of this policy?**

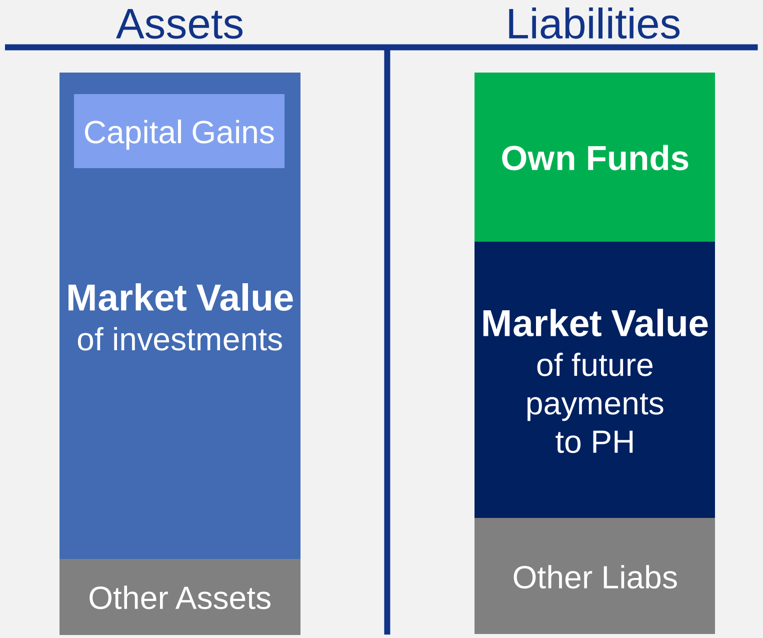


Life insurance balance sheet in Solvency II

Policy view



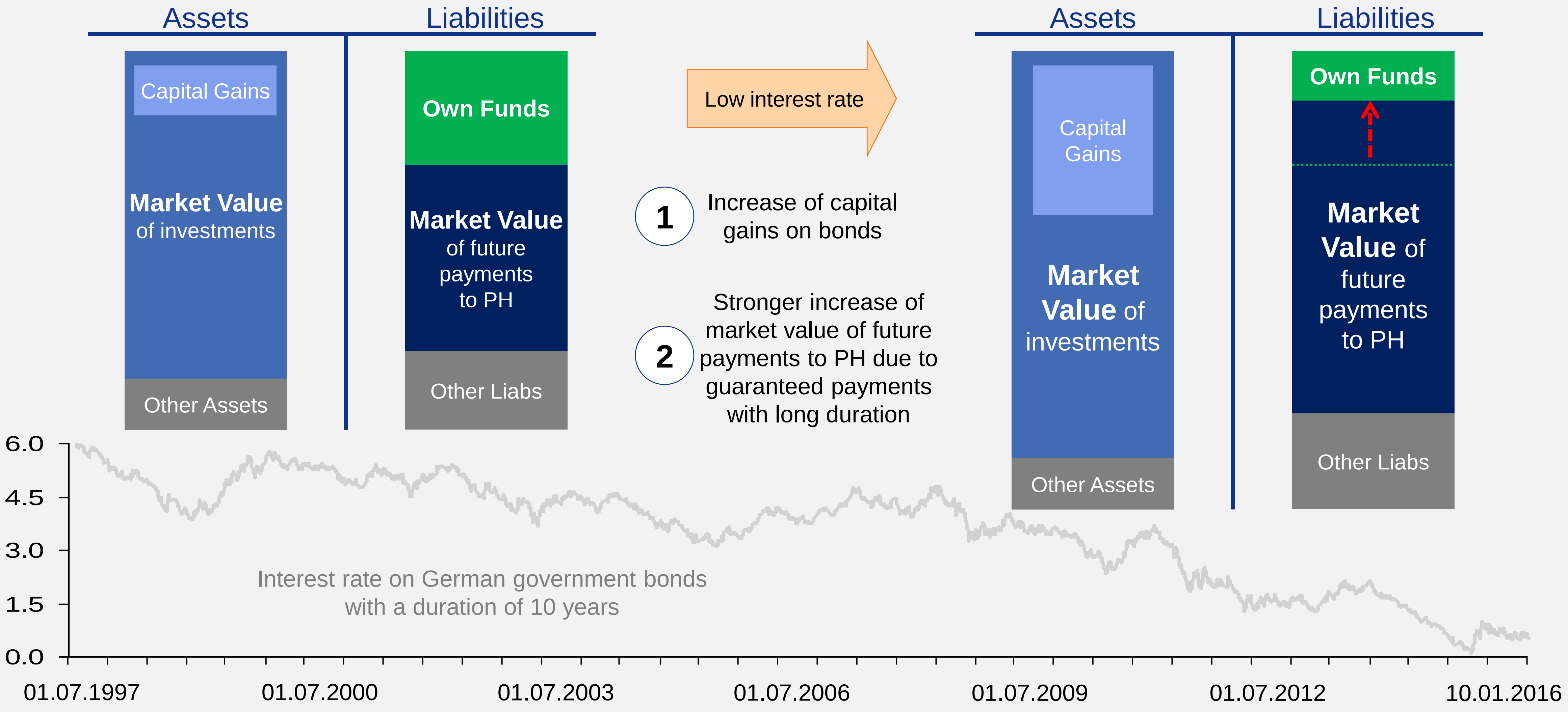
Portfolio view



1 **What is the value of all assets?**

2 **Who will get this value?**

Market changes impact balance sheet



Cash flow models (CFMs) are central to valuation

External data

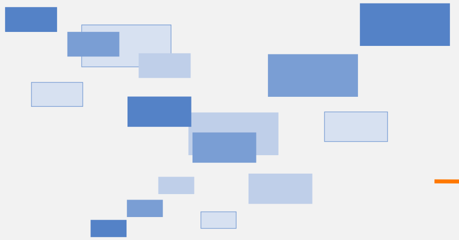
- Capital market scenarios

Internal data

- Balance sheet data
- Company strategy

Cash-Flow Model

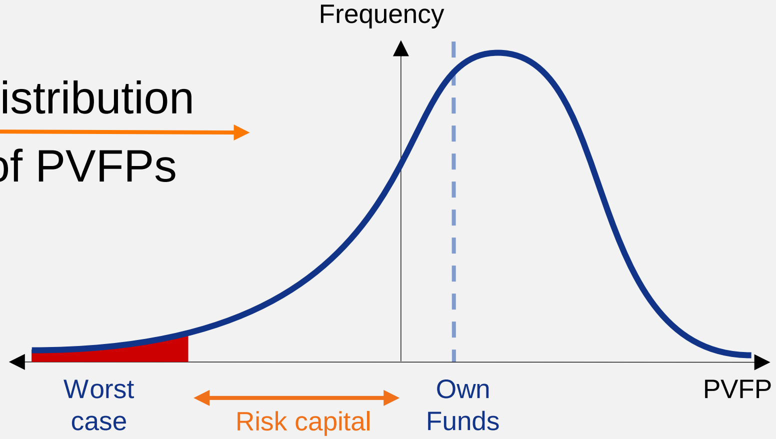
Monte-Carlo Simulation



- Assets $A_1,\ldots,A_N$
- Reserves $R_1,\ldots,R_N$
- Cash flows $C_1,\ldots,C_N$
- Profits $P_1,\ldots,P_N$

→ Present Value of Future Profits (PVFP)

Distribution of PVFPs

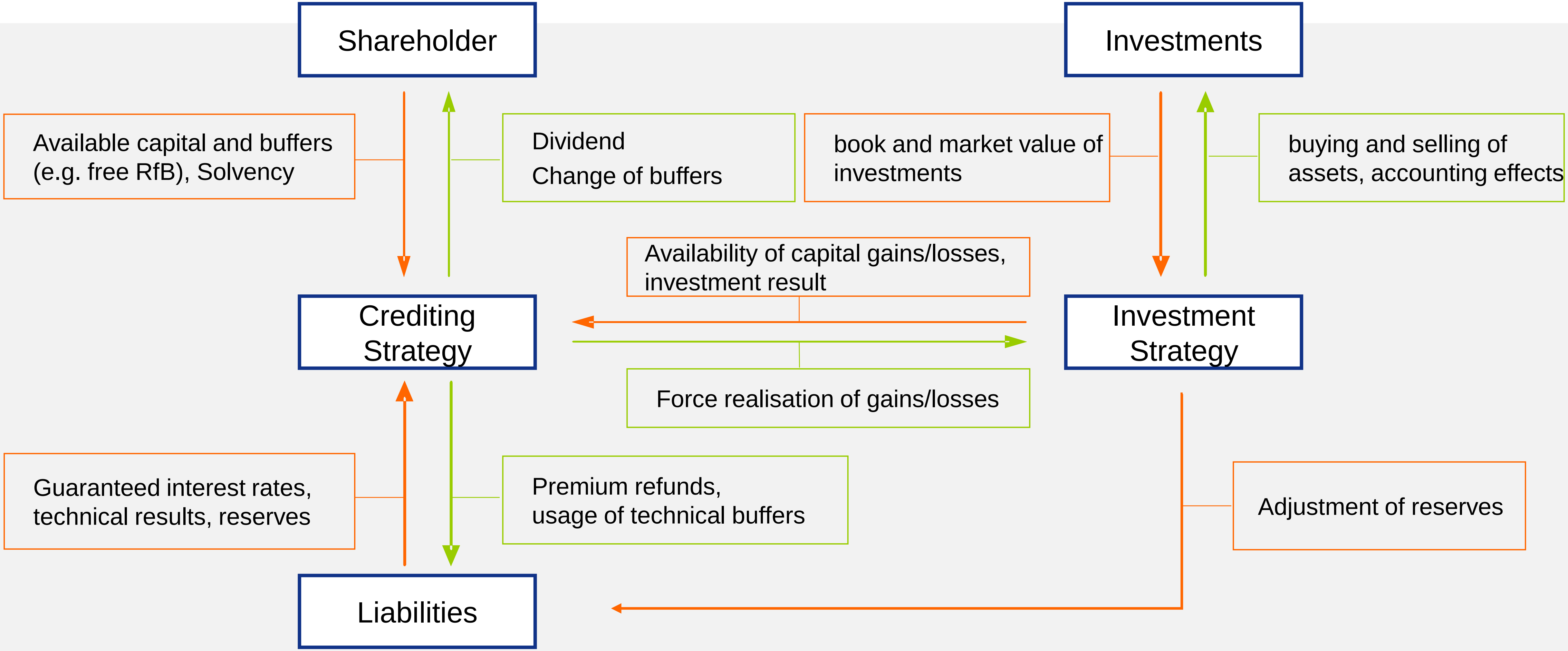


Detailed representation of investments and insurance products

Future management behaviour included (e.g. profit sharing, future asset allocation)

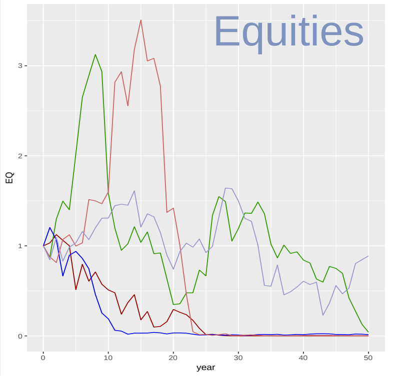
Options & guarantees are explicitly valued using financial math concepts

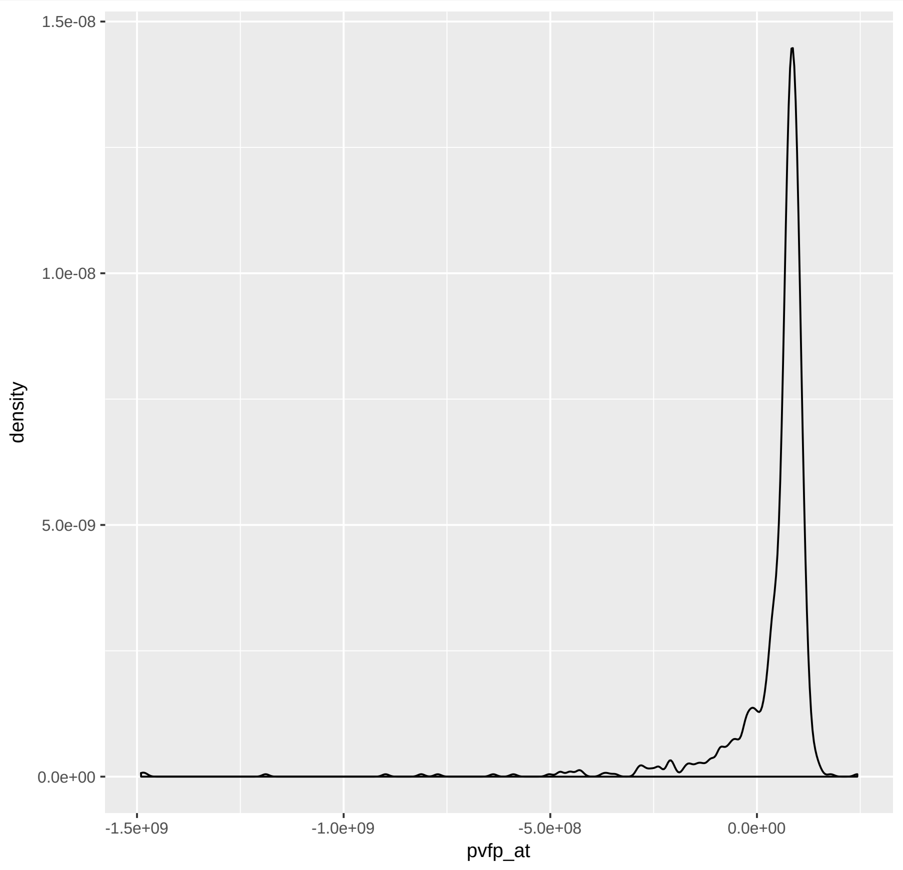
CFMs project complex annual decision processes…

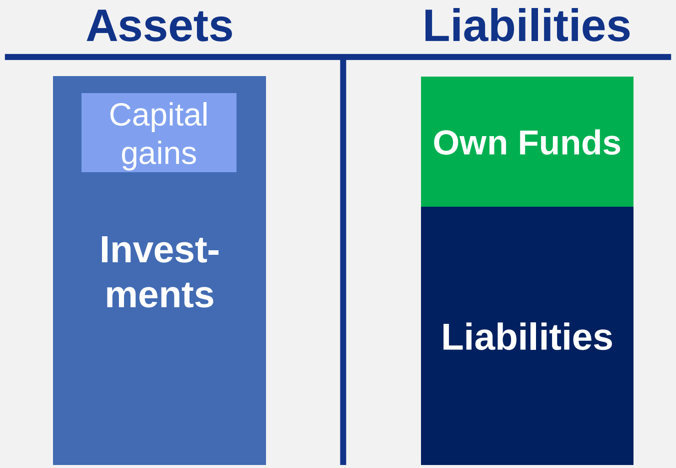


- Projection horizon typically over 50 years

… and are applied to a variety of different market scenarios







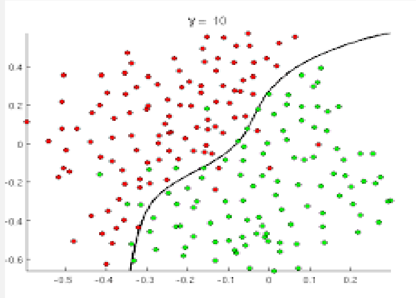
- 1000 – 5000 market scenarios
- Typically 10+ market factors

- Cash flow model of Willis Towers Watson
- Average German insurer

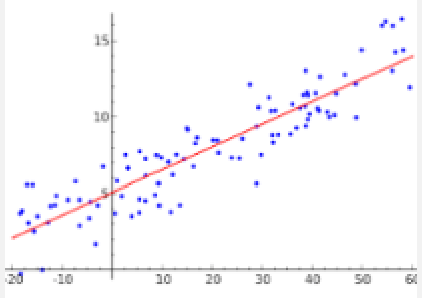
- 1000 – 5000 PVFPs as output
- Stochastic nature of results

Machine learning is now able to tackle complex problems…

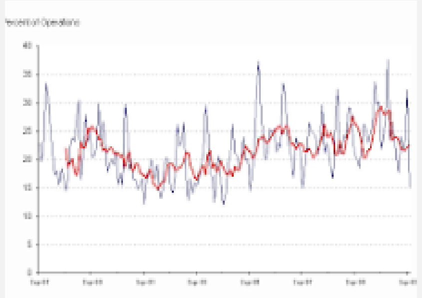
Classification



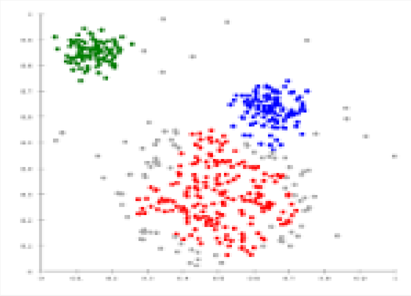
Regression



Time series analysis



Clustering



Dimension reduction

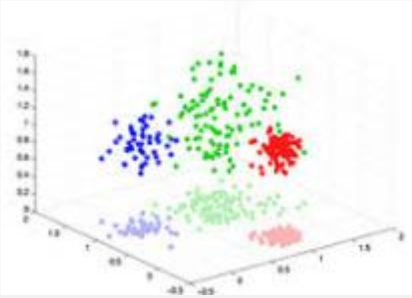
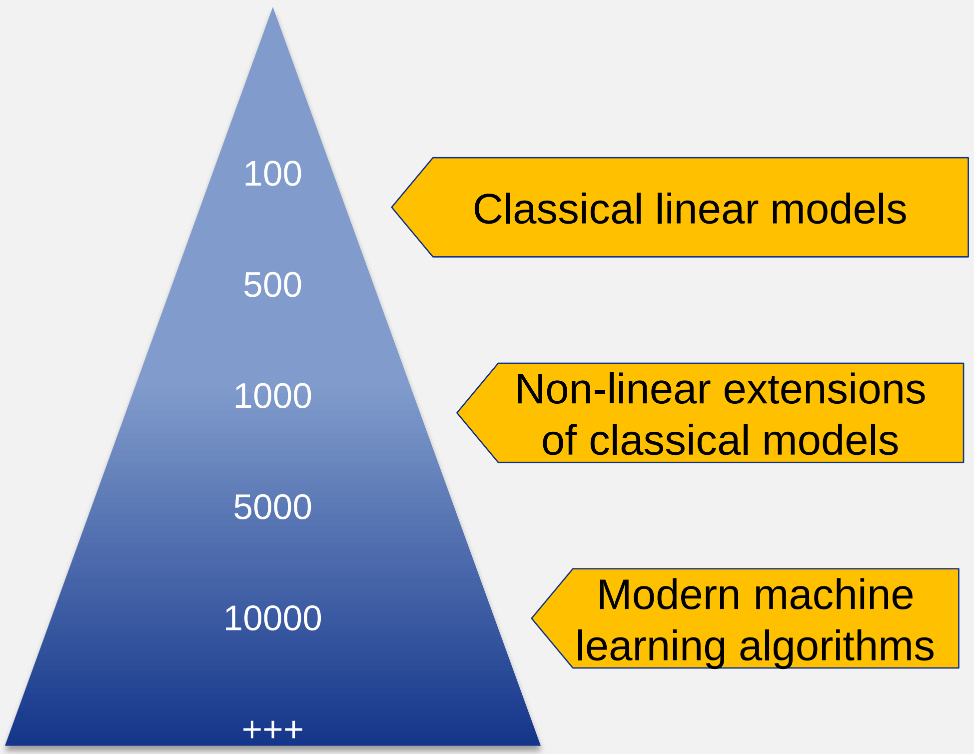


Image analysis

Other

Text & speech analysis

…depending on the quantity and quality of the available data



Number of observations required for a good model fit

Classical (linear) methods have advantages

- Better interpretability
- Faster calibration

Data quantity and quality is critical for the model choice

- Representative period of time
- Covering relevant special cases and not only the „average“
- Target variable as objectively as possible

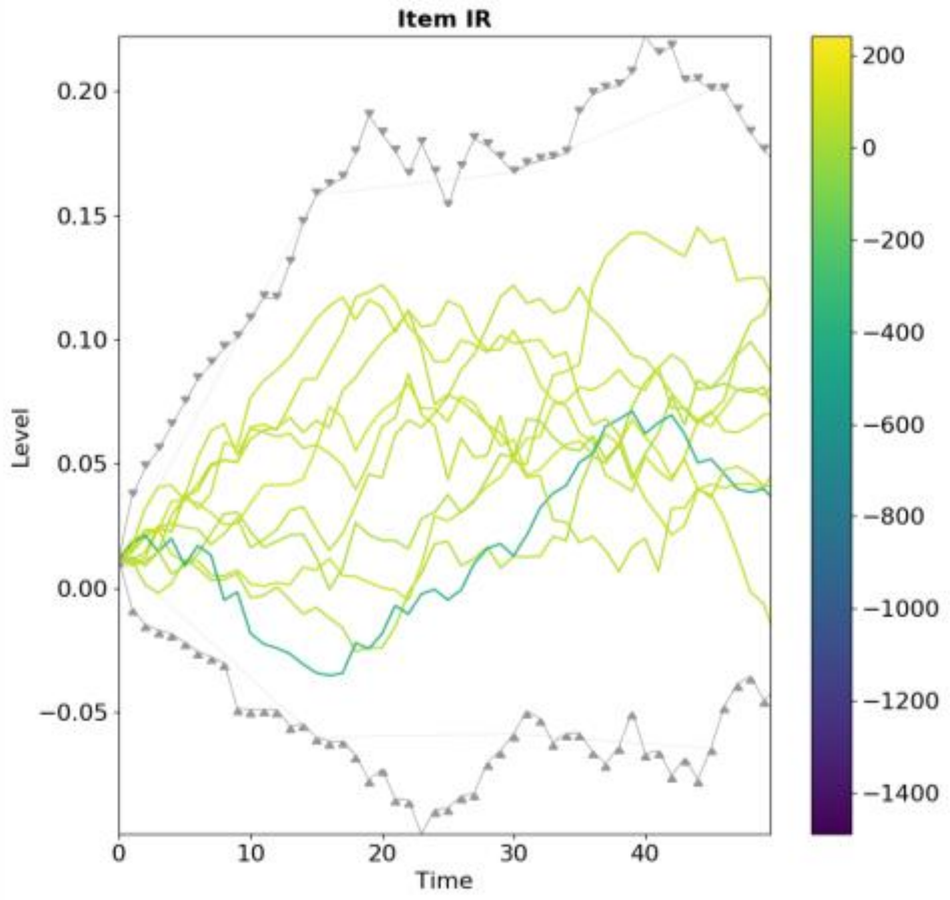
Risk of overfitting with too complex ML models

➔ **Simpler models are always preferable**

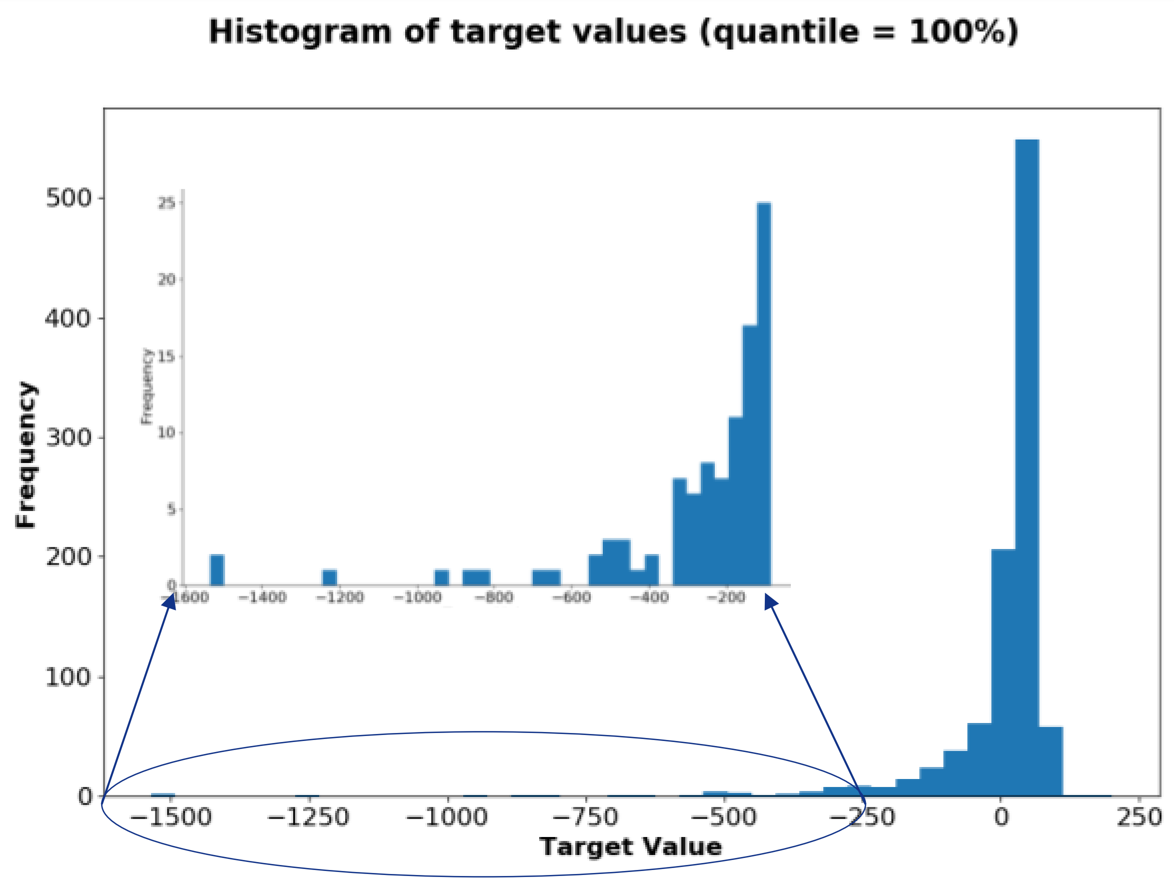
Agenda

Which scenarios are responsible for extreme losses?

Input

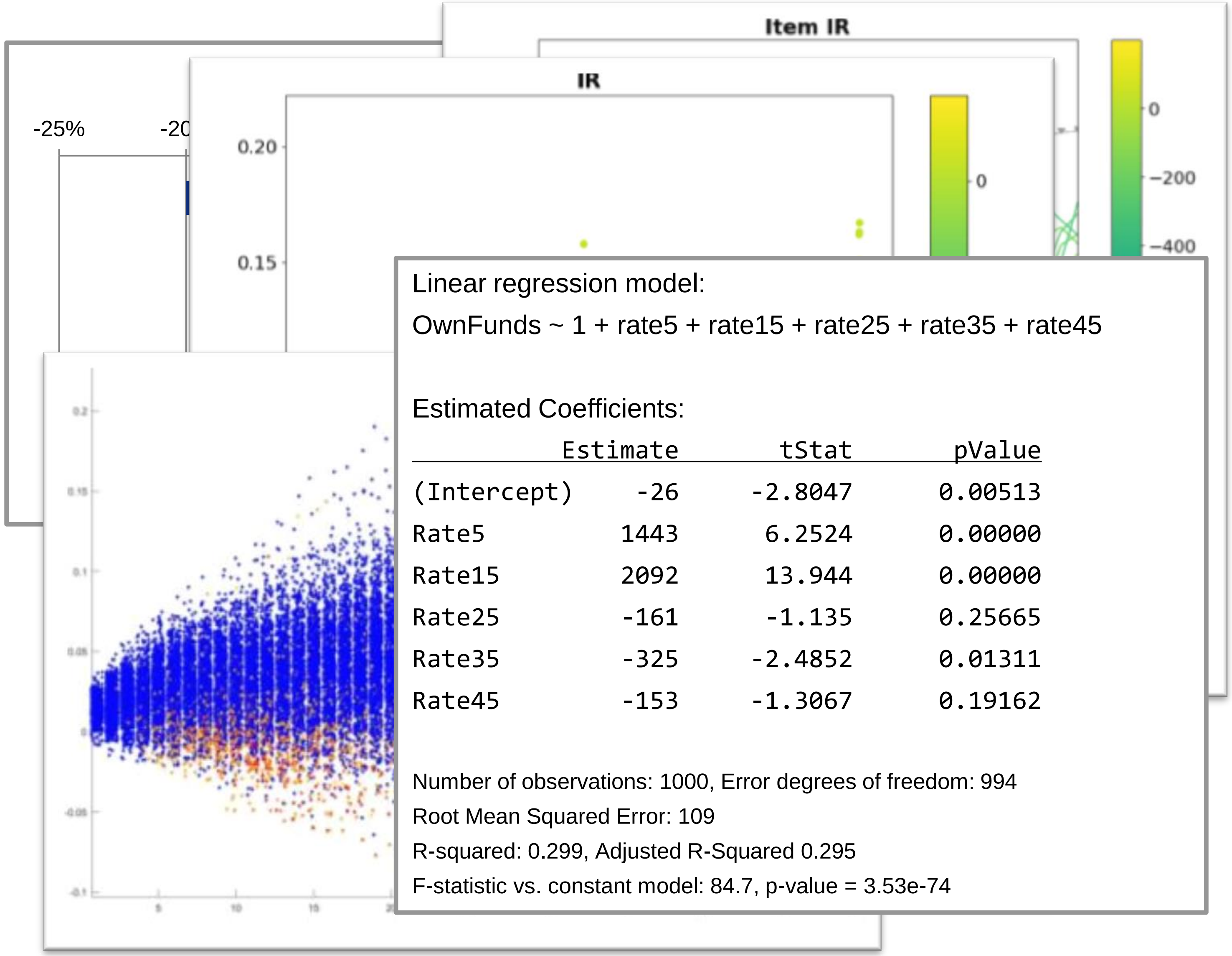


Output



© Allianz SE 2017

Traditional approaches provide limited insights



Linear regression model:

OwnFunds ~ 1 + rate5 + rate15 + rate25 + rate35 + rate45

Estimated Coefficients:

	Estimate	tStat	pValue
(Intercept)	-26	-2.8047	0.00513
Rate5	1443	6.2524	0.00000
Rate15	2092	13.944	0.00000
Rate25	-161	-1.135	0.25665
Rate35	-325	-2.4852	0.01311
Rate45	-153	-1.3067	0.19162

Number of observations: 1000, Error degrees of freedom: 994

Root Mean Squared Error: 109

R-squared: 0.299, Adjusted R-Squared 0.295

F-statistic vs. constant model: 84.7, p-value = 3.53e-74

Traditional approaches:

- Sensitivity analysis
- Plot worst scenarios
- Mulit-dimensional plots
- (Linear) regression

Issues:

- Only linear or 1-dim dependencies are identified
- Analysis often requires additional runs of the CFM
- Time-consuming manual analysis

Potential for Machine Learning ?

The evolution of data analytics

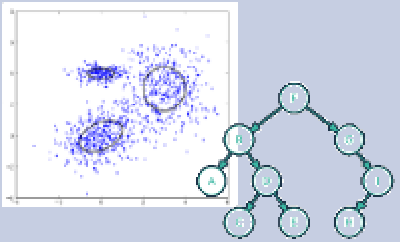
Statistics

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$$

Analyzing quantitative information

- Descriptive analysis and hypotheses testing
- Scientific sub-topic of mathematics
- “Data generating process”

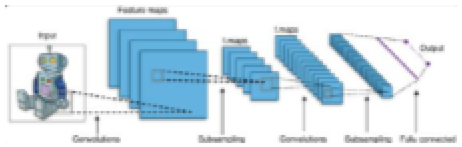
Data Mining



Explaining patterns in the data

- Information extraction from large data sets
- Visualization and structuring
- “Patterns”

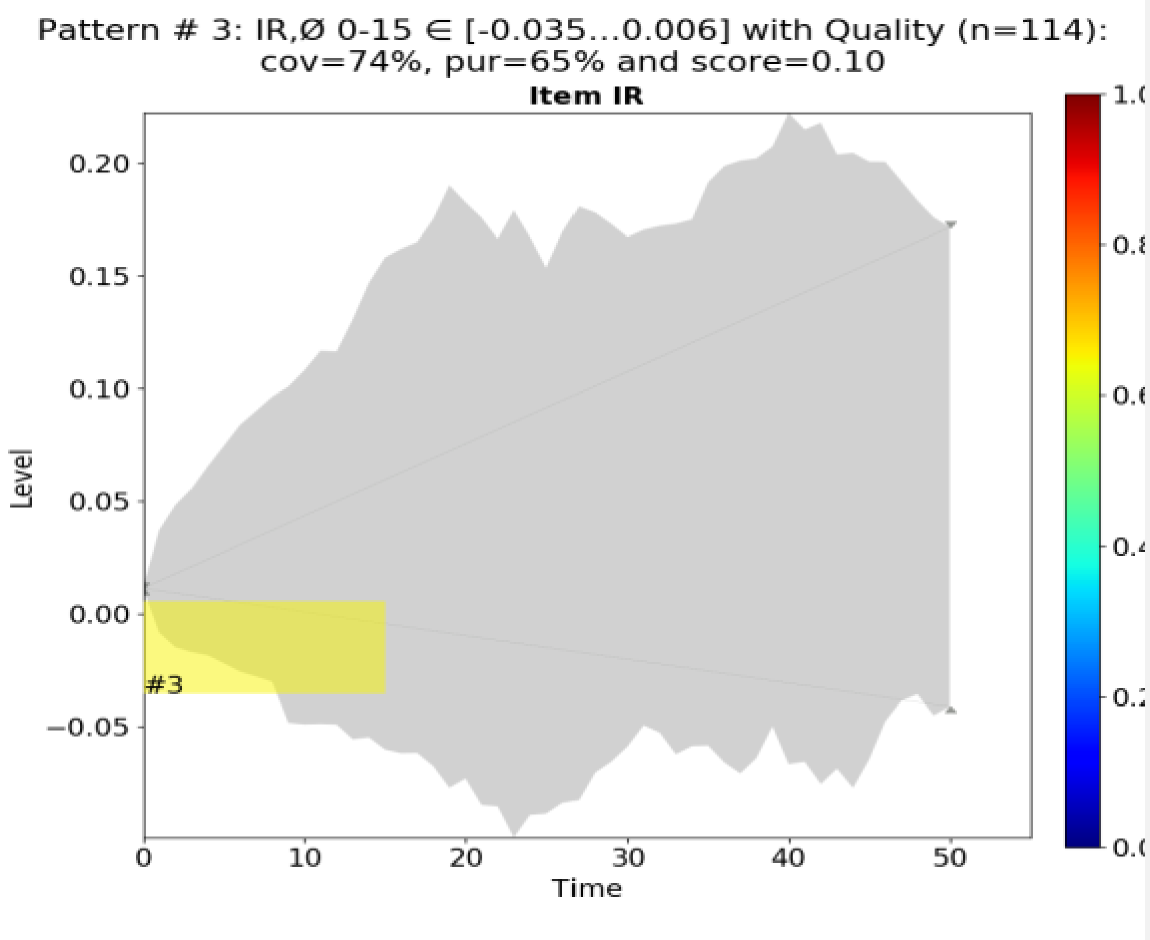
Machine Learning



Prediction of future based on experience

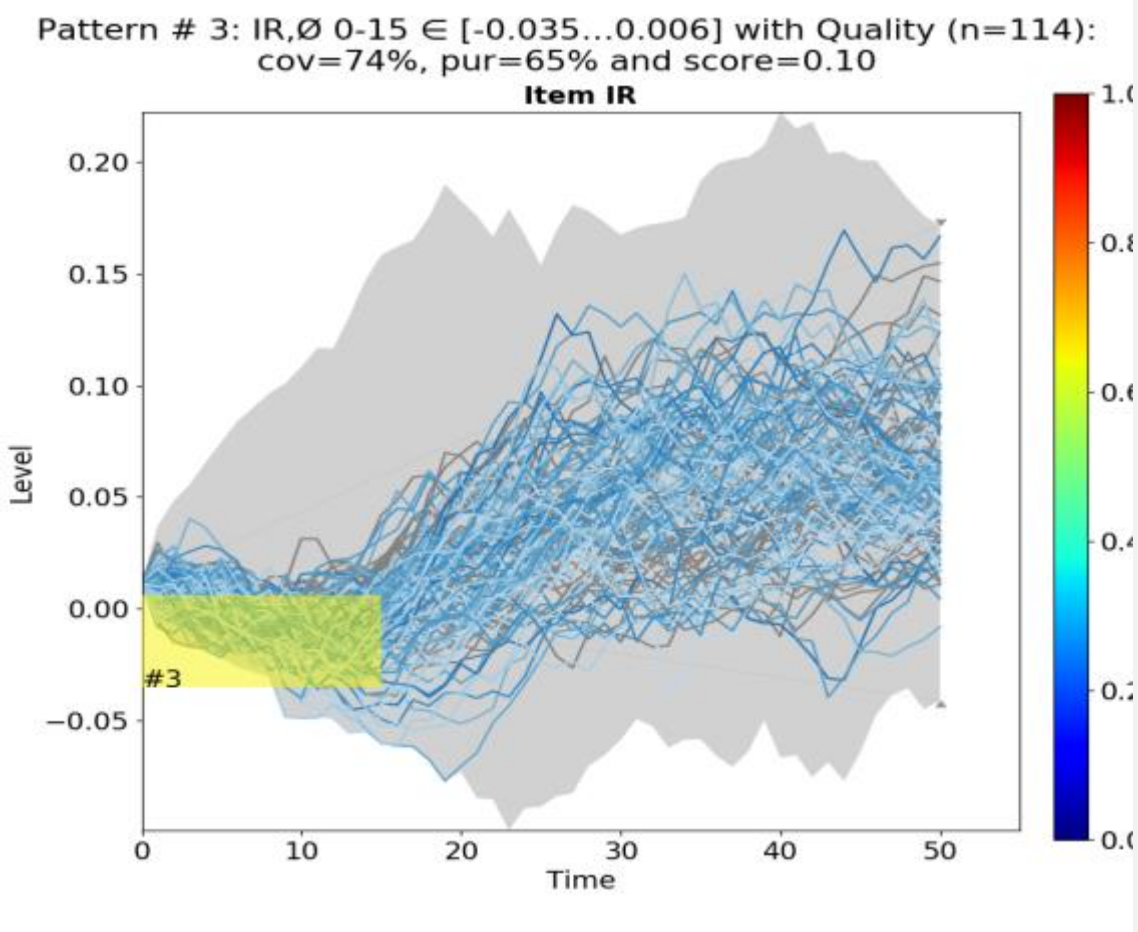
- Flexible models for complex data sets
- Model learn from data / experience
- “Prediction”

What is a pattern?



- The algorithm searches for patterns which are in the scenarios of certain target (e.g. lowest OFs)
- A pattern was found for interest rates (Item IR)
- Colored horizontal bar indicates time range, gauge and value range:
  - Horizontal bar = average over that time within a certain range
  - Width: time range between year 0 and 15
  - Height: value range between –3.5% and 0%
- Quality of this pattern is shown visually (color of bar) and as text in heading:
  - Coverage (cov=74%): fraction of targeted scenarios covered by this pattern
  - Purity (pur=65%): fraction of scenarios in this pattern belonging to the target

How to read the detailed output



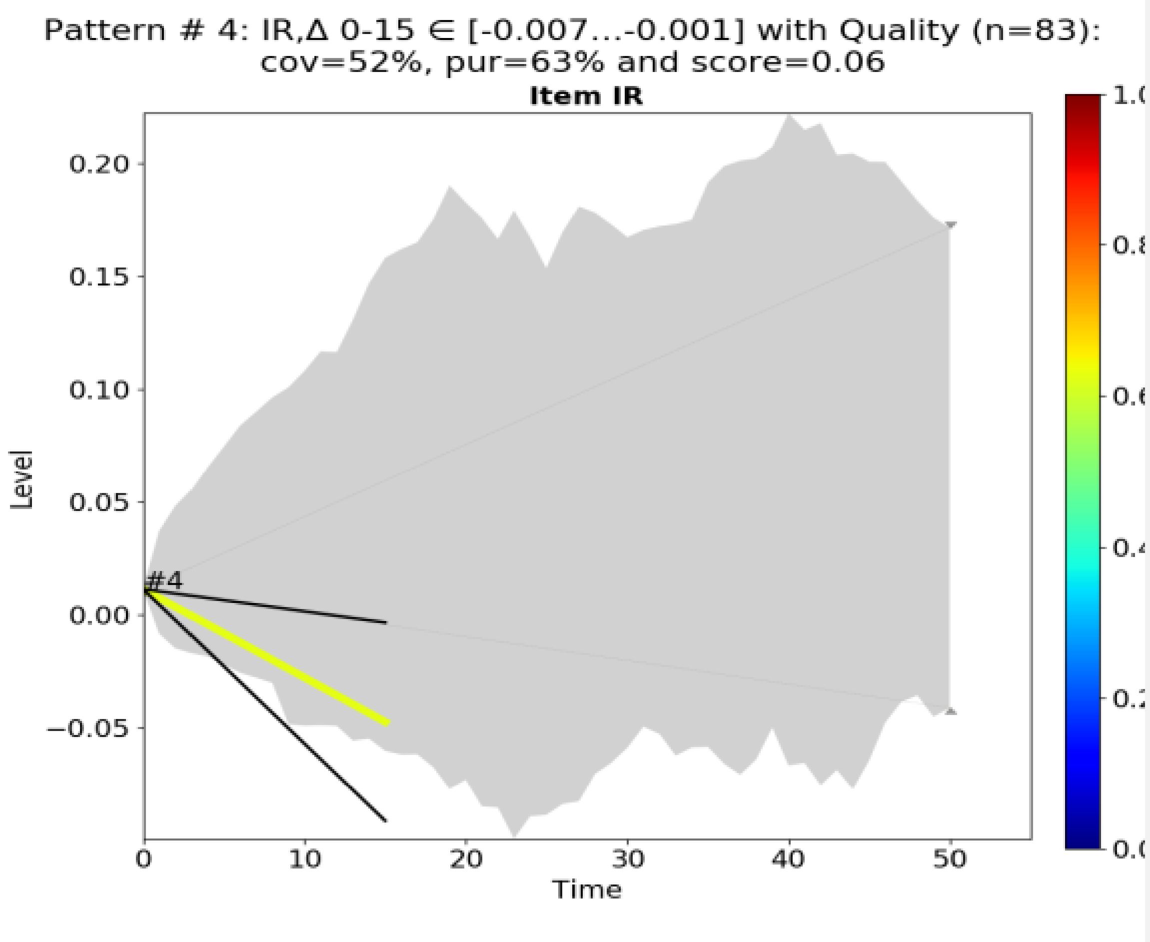
- All scenarios matching the pattern are shown as lines
  - Blue lines = scenario in target
  - Grey lines = scenario not in target
- Total range of all scenario is shown as light grey background

- 65% of all scenarios matching this pattern are in our target (purity)
  - Visually represented by color of the bar
- 74% of all targeted scenarios are covered by this pattern (coverage)

Interpretation:

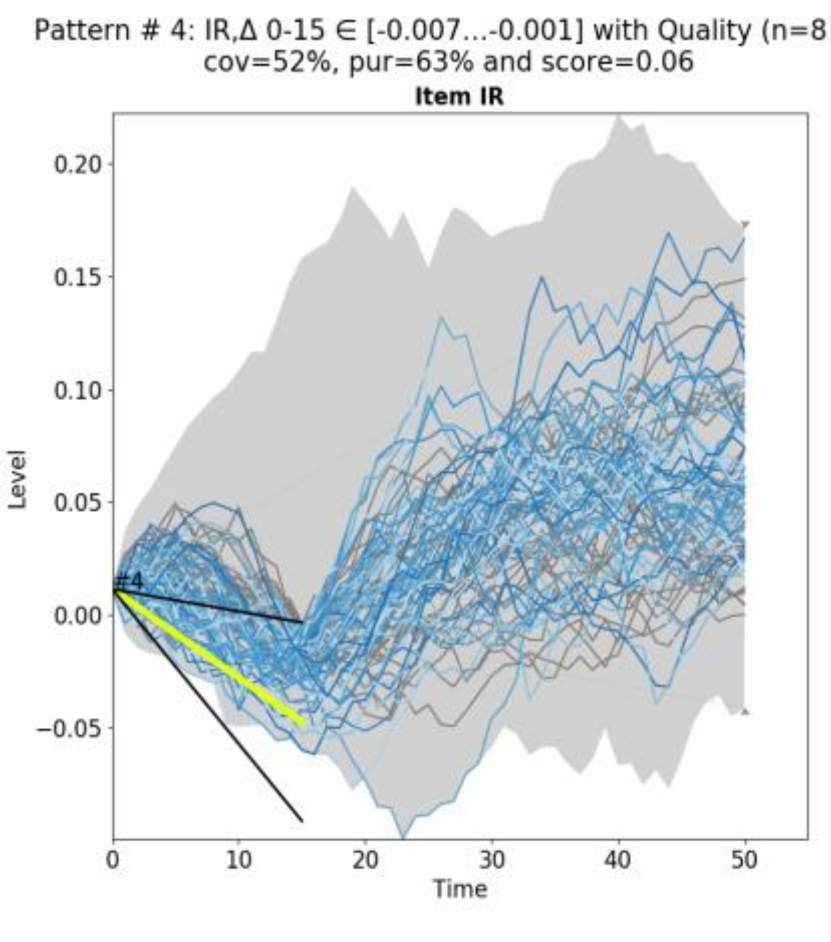
Falling interest rates are driver for bad OFs

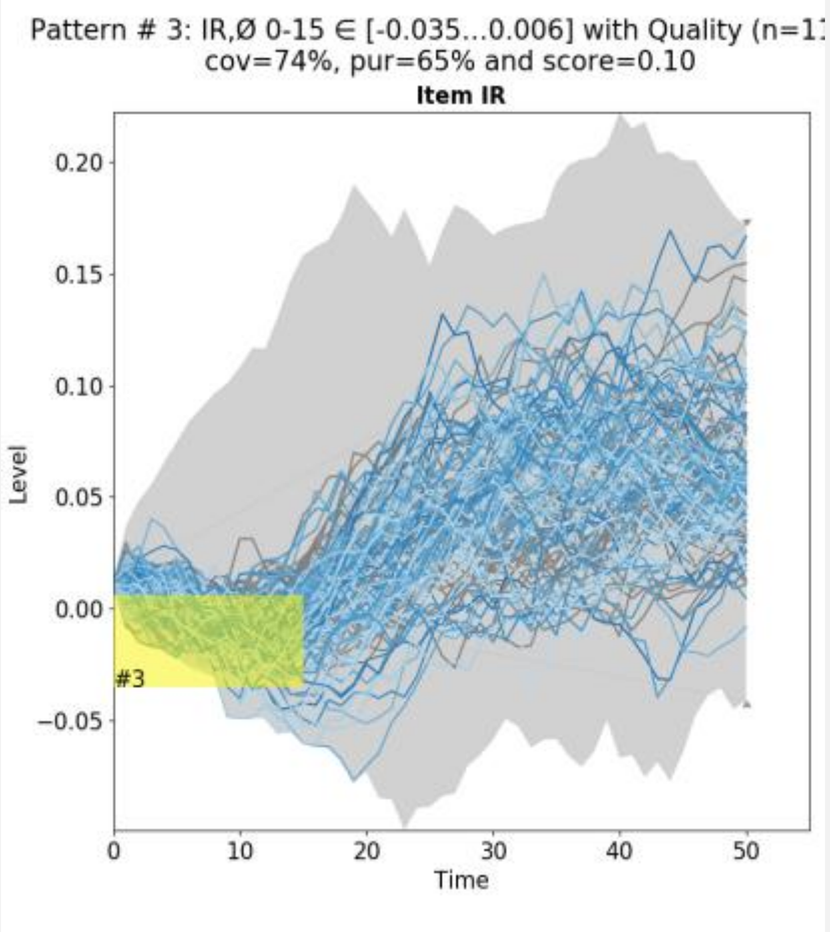
Second pattern describes bad OFs



- Line indicates that interest rates change (Δ) over a the first 15 years
- Scenarios match this pattern if interest rates fall by 0.4% per year
  - Interest rates are down by approx. 6% after 15 y
  - Bandwidth is at ±5%
  - In this case this is equal to „IR at year 15 between –10% and 0%“ (due to fixed rate at t=0)
- A scenario meeting this requirement has a probability of 63% to be in the target (purity)
- This pattern covers 52% of the targeted scenarios (coverage)
- Pattern #4 seems to describe same scenarios as pattern #3 before

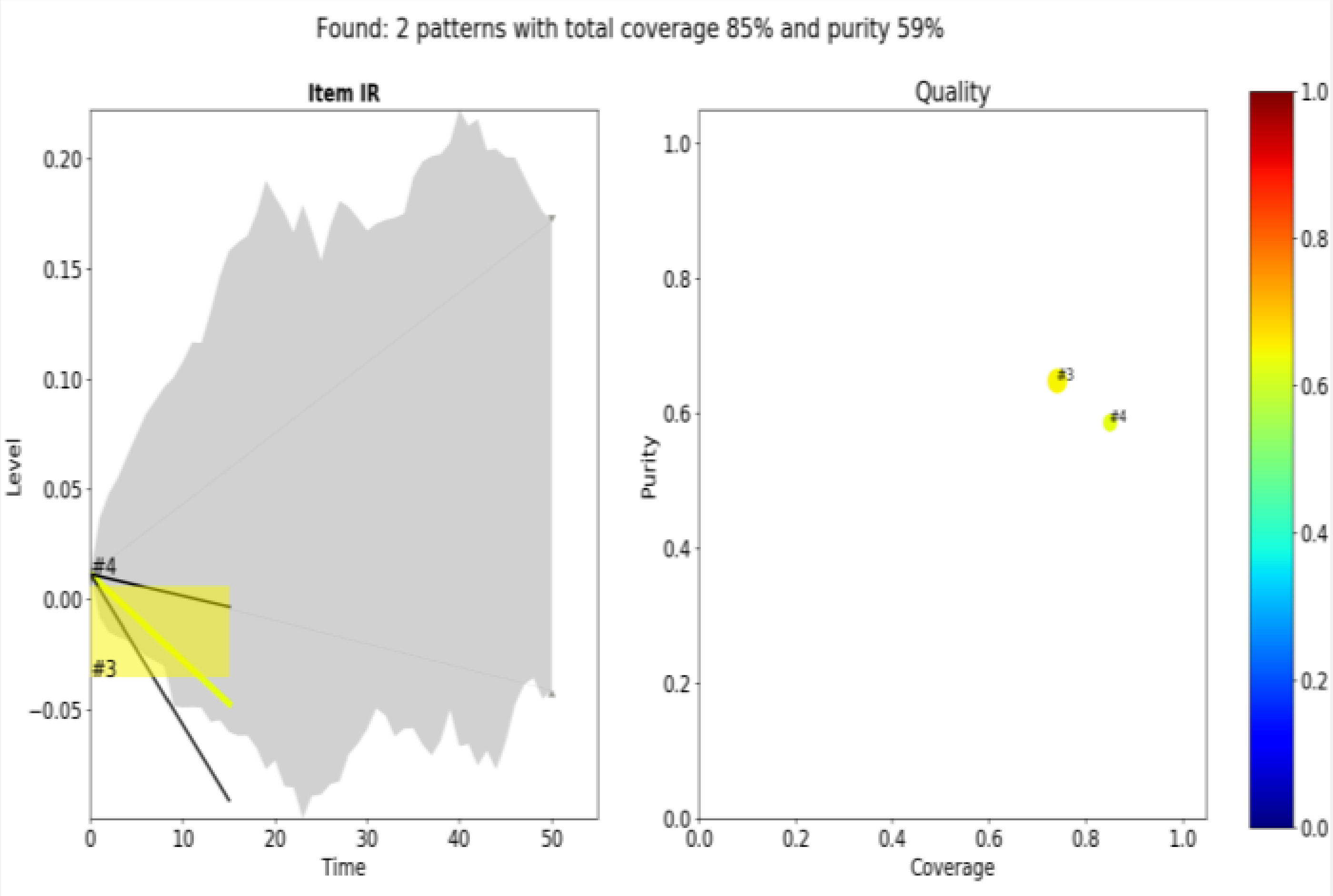
Additional scenarios covered by new pattern





- Pattern #4 covers partly different scenarios than pattern #3
- E.g. IR scenarios, that
  - Rise first
  - Then fall abruptly
- Pattern #4 covers approx. 11% of scenarios in target that are not covered by pattern #3
- Marginal coverage can be only seen when sorting the found patterns

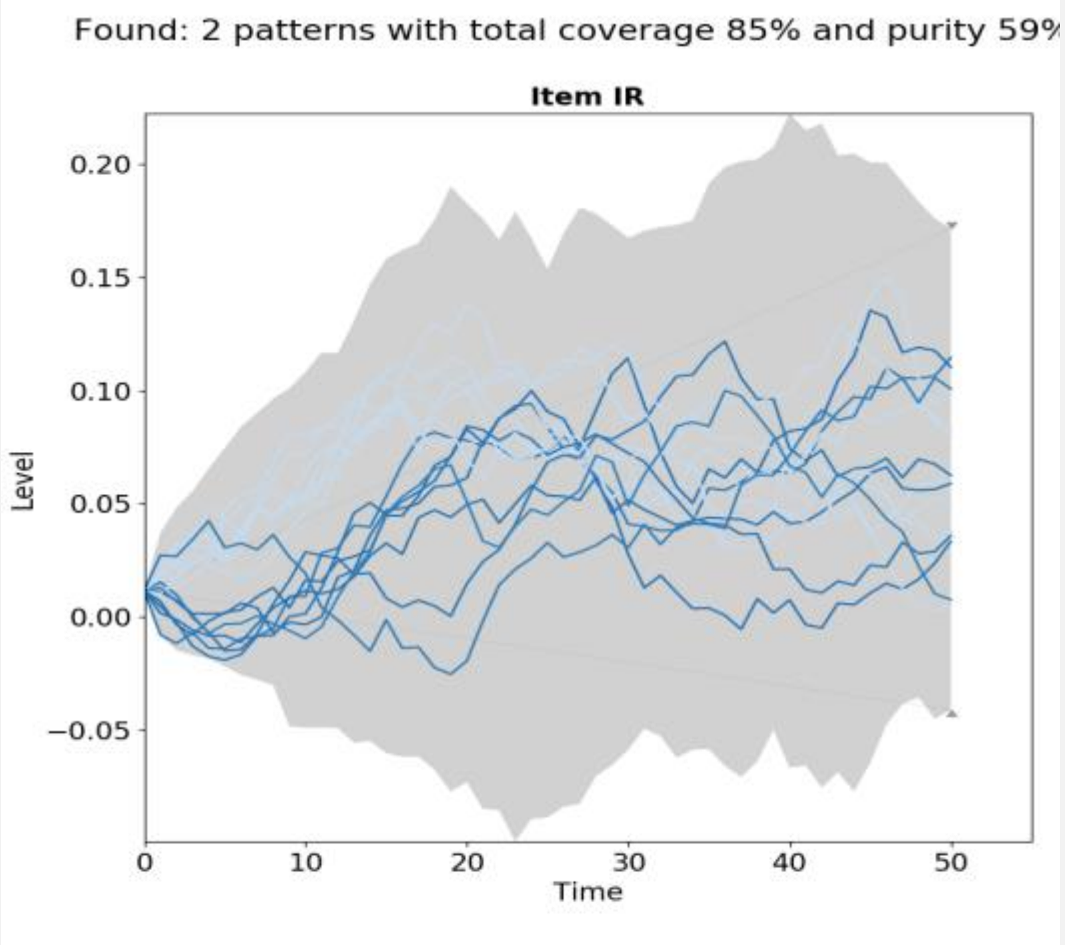
Overview plot shows all found patterns



- Projection horizon was segmented into three buckets (1–15, 16–30 and 31–50 years) to reduce computational workload
- Search tries to find a combination of pattern with best coverage and purity
  - Each pattern has an identifier (#)
  - Different patterns are alternative explanations (OR)
- Quality plot on the right hand shows coverage und purity of all patterns together
  - Size and color of individual points show coverage and purity of individual pattern
  - Pattern #4 increases total coverage but decreases total purity
- Interpretation: Worst OFs are driven by declining interest rates

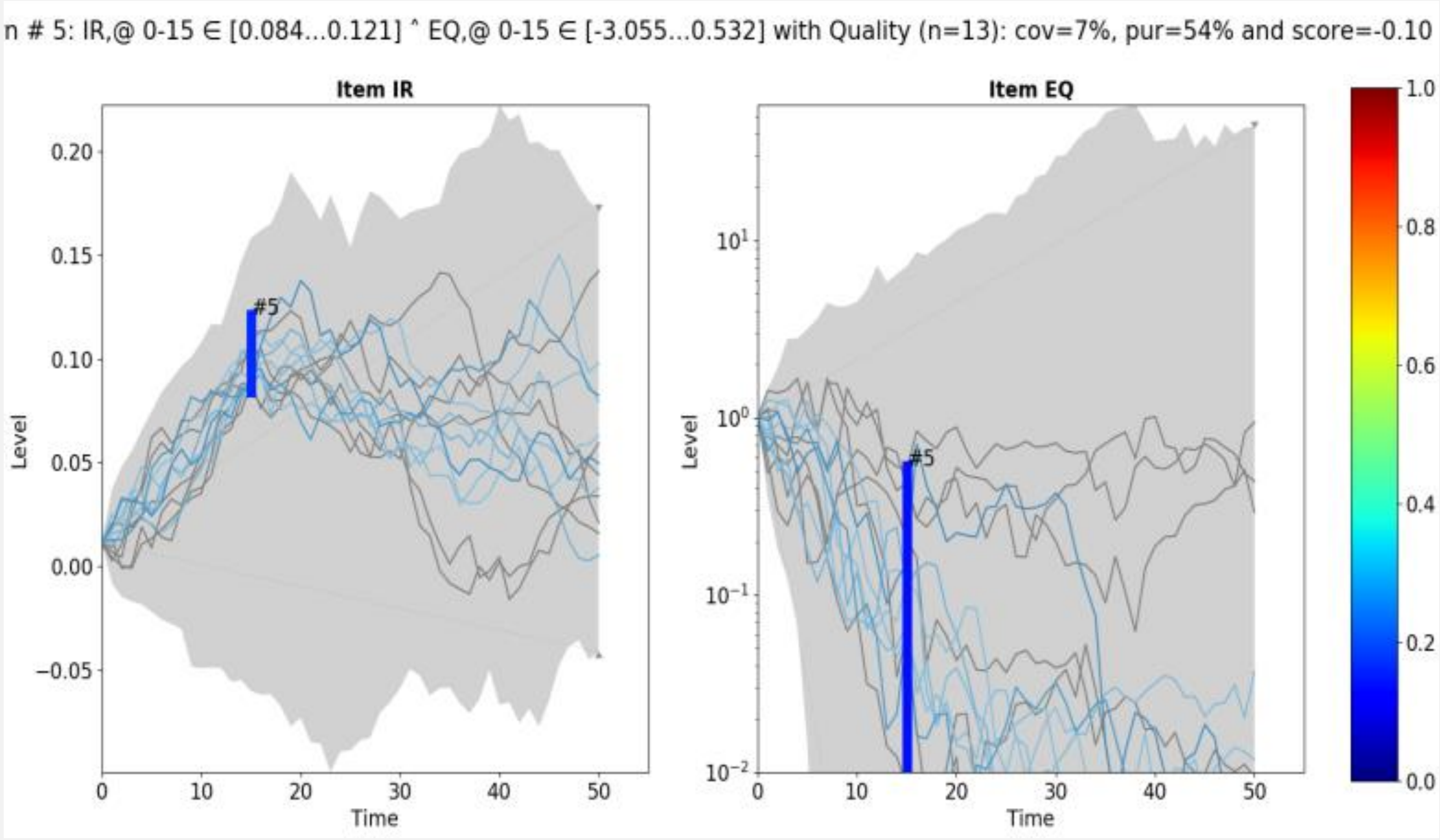
→ Can we improve coverage further?

Residual plot shows scenarios that did not match a pattern



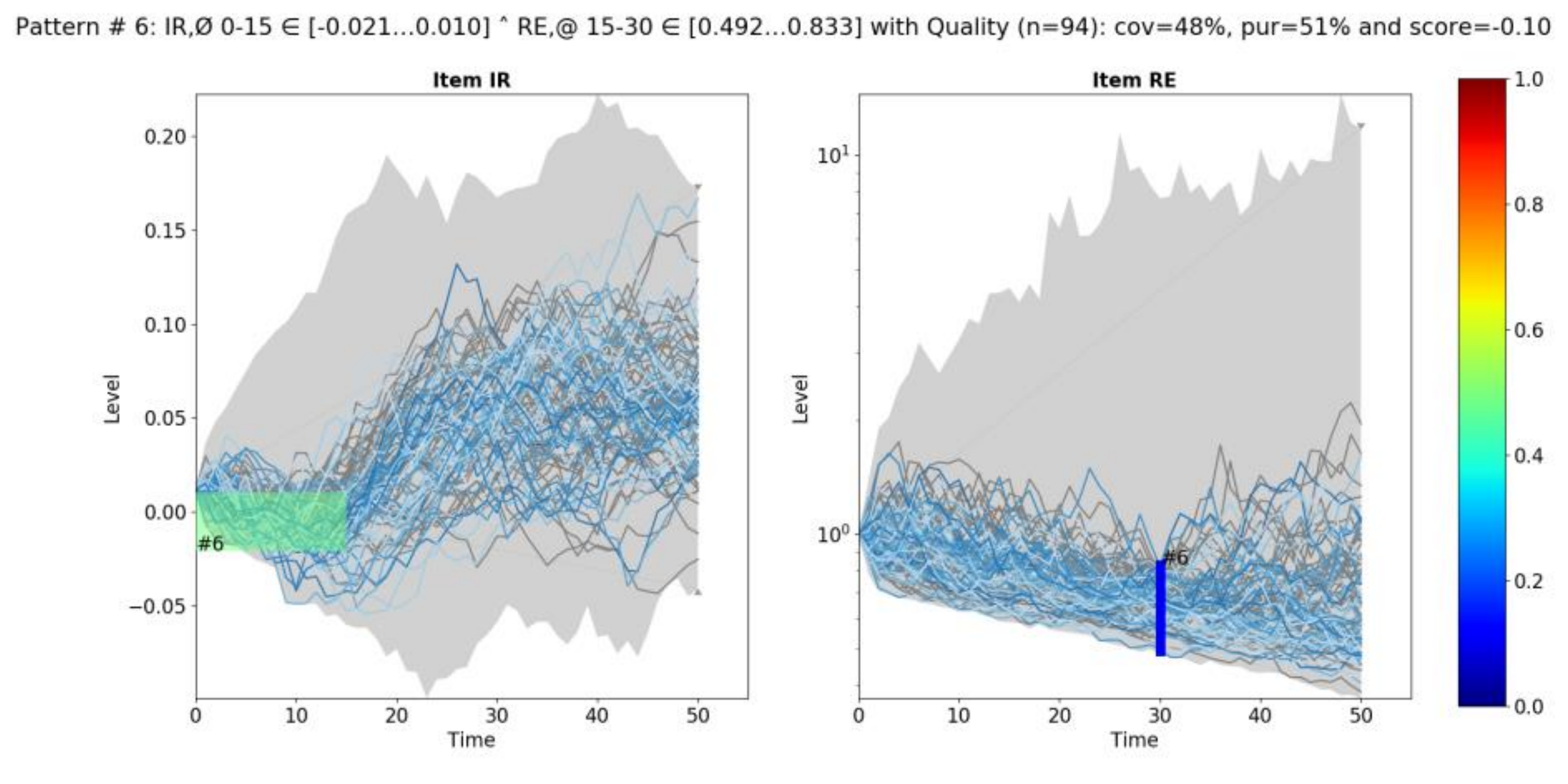
- Two groups of interest rate developments can be identified visually
  - Scenarios in the middle
  - Rising interest rates
- Search for further patterns should include combinations of variables and has to improve total coverage without diluting total purity too much

Algorithm can find more complex patterns

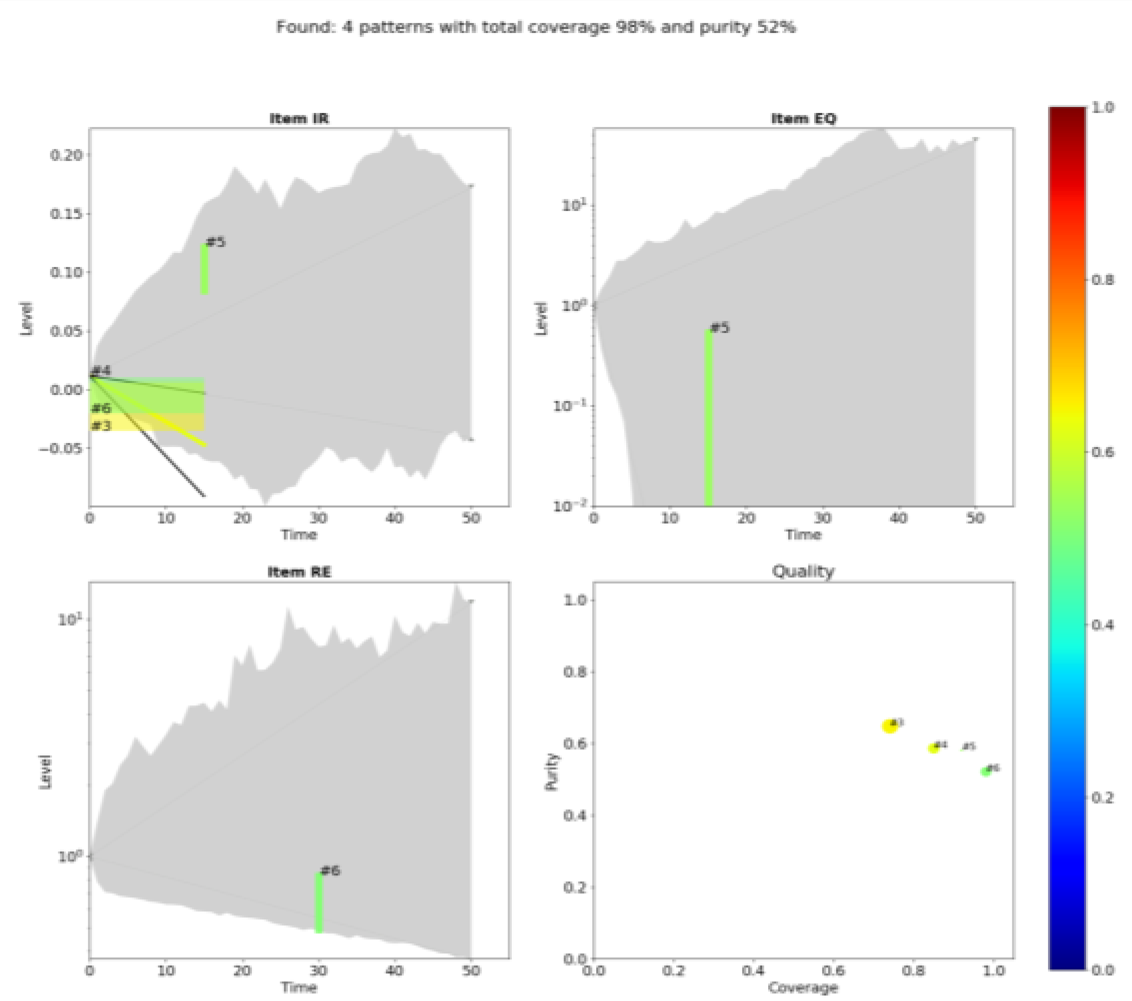


- Pattern #5 has two conditions:
  - Interest rates have to be between 8% and 12% in year 15 (vertical line)
  - Equity index has to be less than 50% after 15 years (logarithmic scale!)
- Only scenarios meeting both conditions (AND) are included in this pattern
- This mechanism allows to identify interactions between variables
- Individual purities of both conditions (ca. 20%) are weaker than joint purity (54%)
- This pattern covers only a small part of our targets (coverage 7%)

Other combination found: RE decrease relevant when IR flat

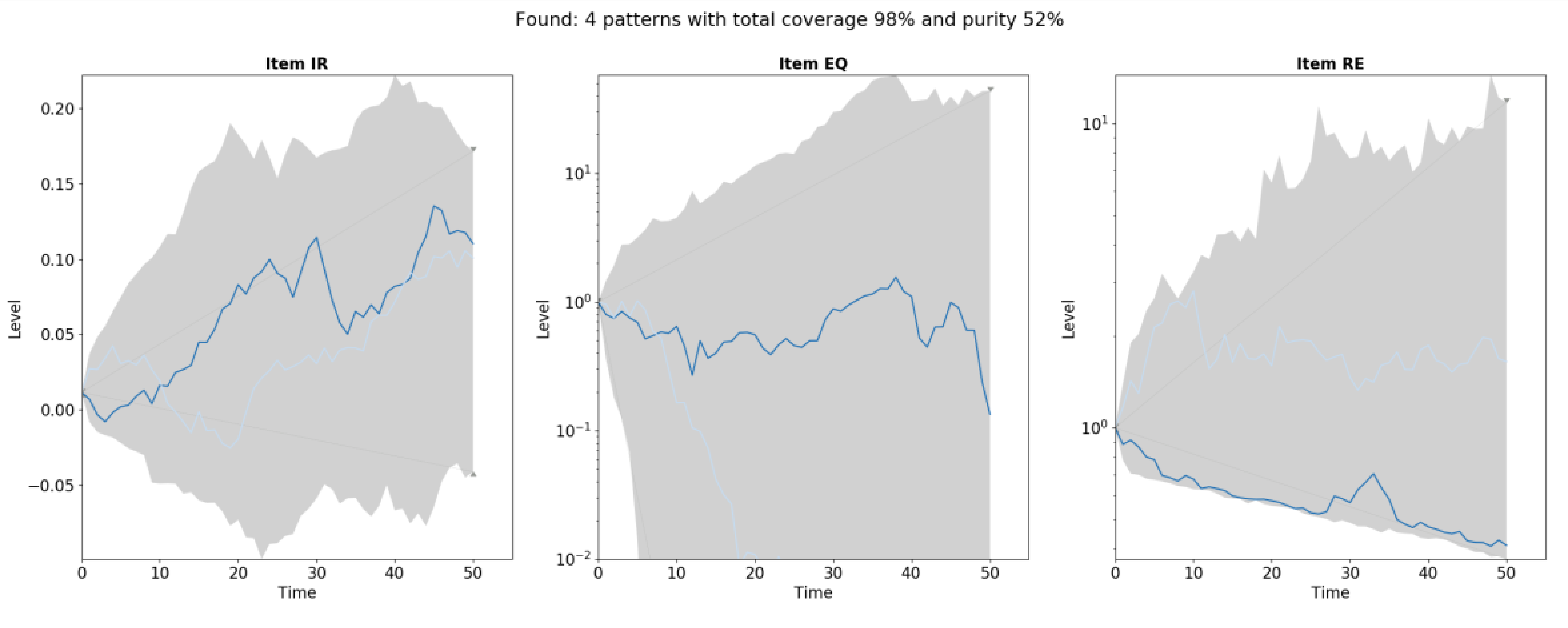


Overview of results

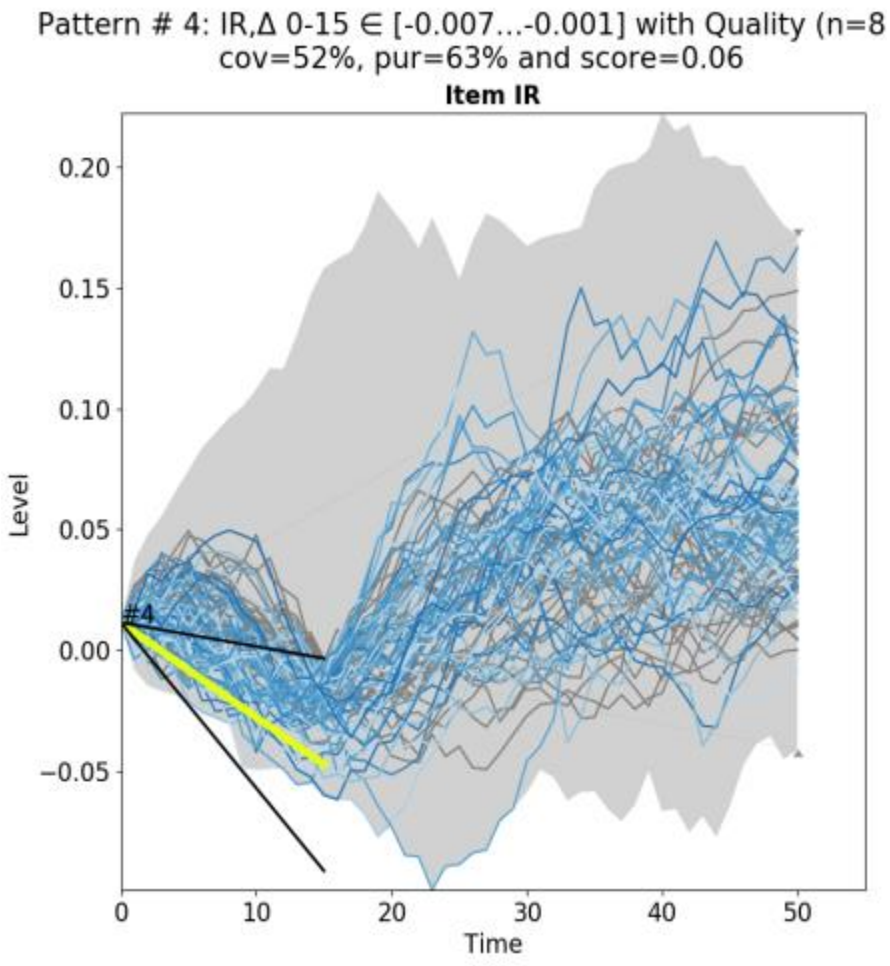


- Analysis of worst OFs
  - Data: Towers Watson, internal industry model
  - Closing: 2015 Q4
  - Scenarios: 1000 iterations (with neg. interest rates)
  - Target: 10% lowest quantile of OFs
  - Patterns with max 2 conditions
  - Patterns with min 50% purity
- Plot shows all found patterns
  - Falling interest rate as main risk driver visible
  - Negative real estate relevant if interest rates are flat
  - High interest rates only relevant if there is a significant equity shock at the same time (crafted scenario)

Only 2 unexplained scenarios remain



How does it work?



1. Simplify data

- Segment time data into buckets (here 0–15, 15–30 and 30–50)
- Define general patterns of scenarios within a bucket, e.g.
  - IR drops from 1% to around – 5% within first bucket
  - IR in first bucket on average around –3%

2. Find optimal parameters of patterns

- Optimal w.r.t. coverage and purity to a given target (worst PVFPs)
- Using classical optimisation algorithm

3. Find optimal combination of patterns

- Test all combinations of patterns in order to find combination effects, e.g. low interest rates together with losses on real estate
- Criteria needed which combinations are preferred, e.g. as few conditions as possible (simple is better)

Applications

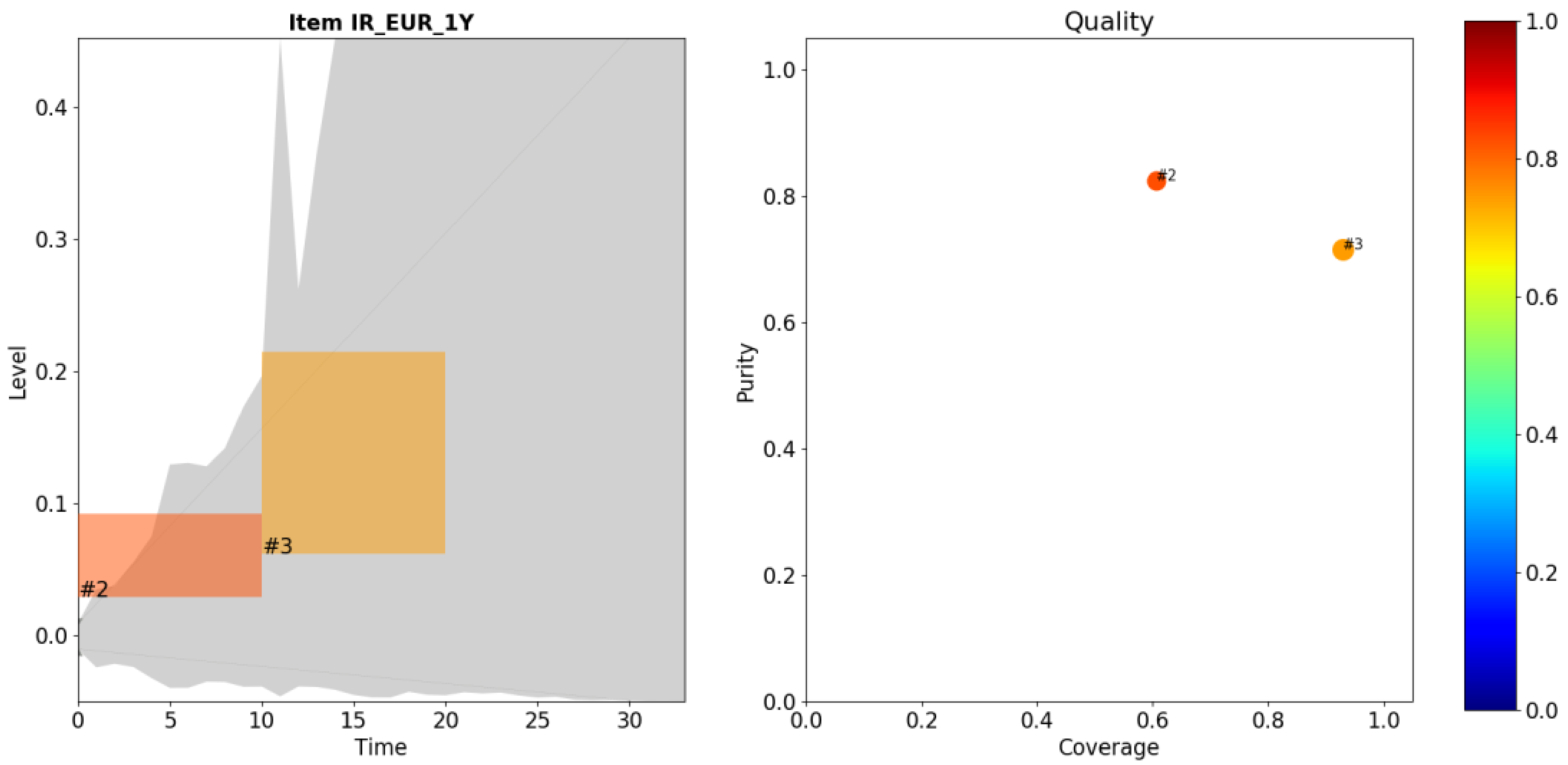
1. **Static analysis**
   - All funds are analysed based on closing runs
     - Model runs automatically in batch mode
     - Saves all graphics onto hard-drive
   - Allows quick first analysis of sensitivities

2. **Analysis of changes**
   - Analyse the changes from one quarter to the next
     - Run analysis on all funds for previous and current quarter
     - Analyse the changes in PVFP by scenario (if scenarios have same seed)
   - Can also be applied to what-if-calculations

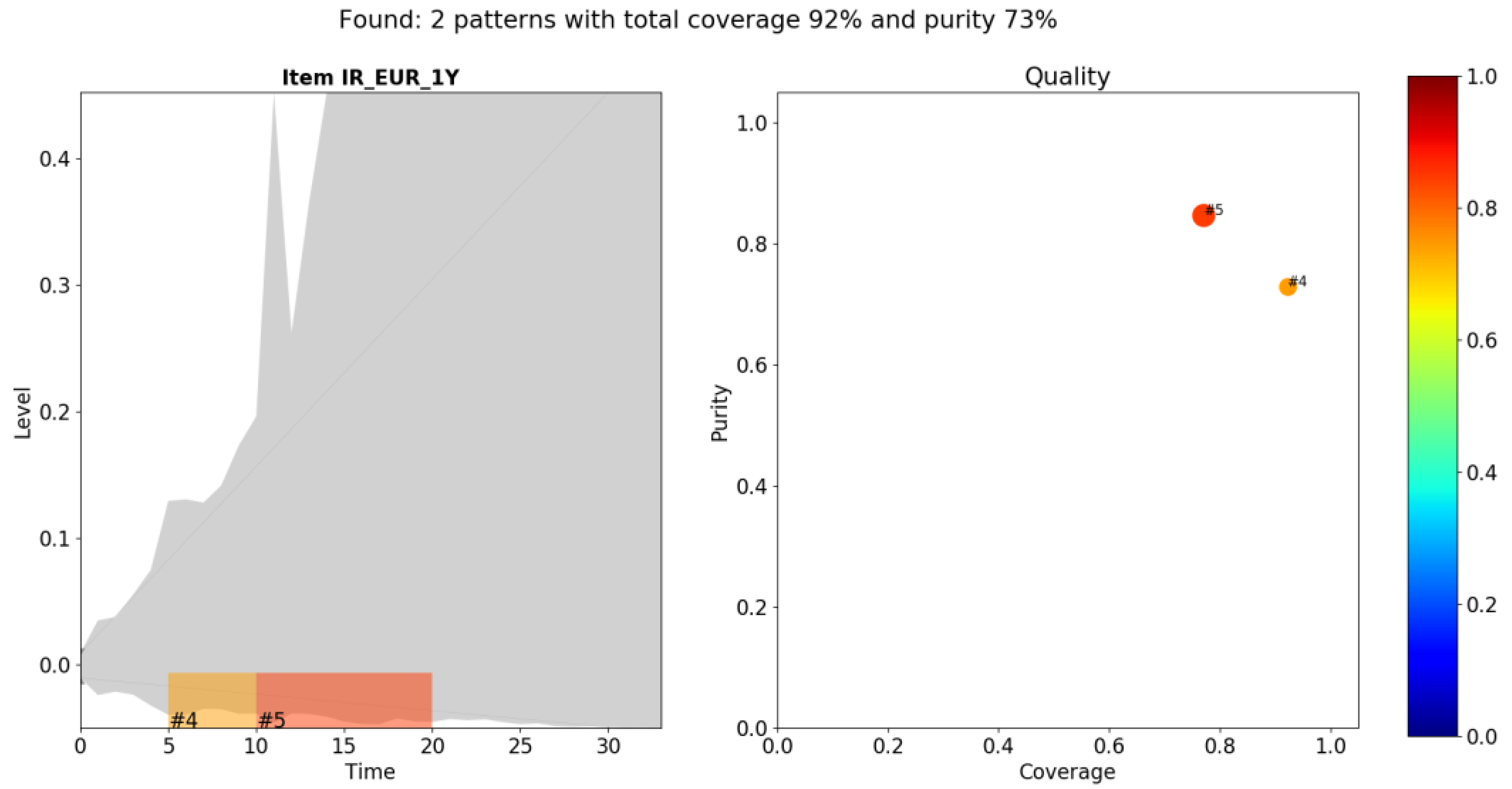
3. **Model validation**
   - Using the analysis tool to validate cash flow model after model change
     - Run analysis on all funds and on multiple targets (low PVFP, high PVFP, …)
   - Results of analysis gives hints for further validation steps

Example 1: Protection product with IR up sensitivity

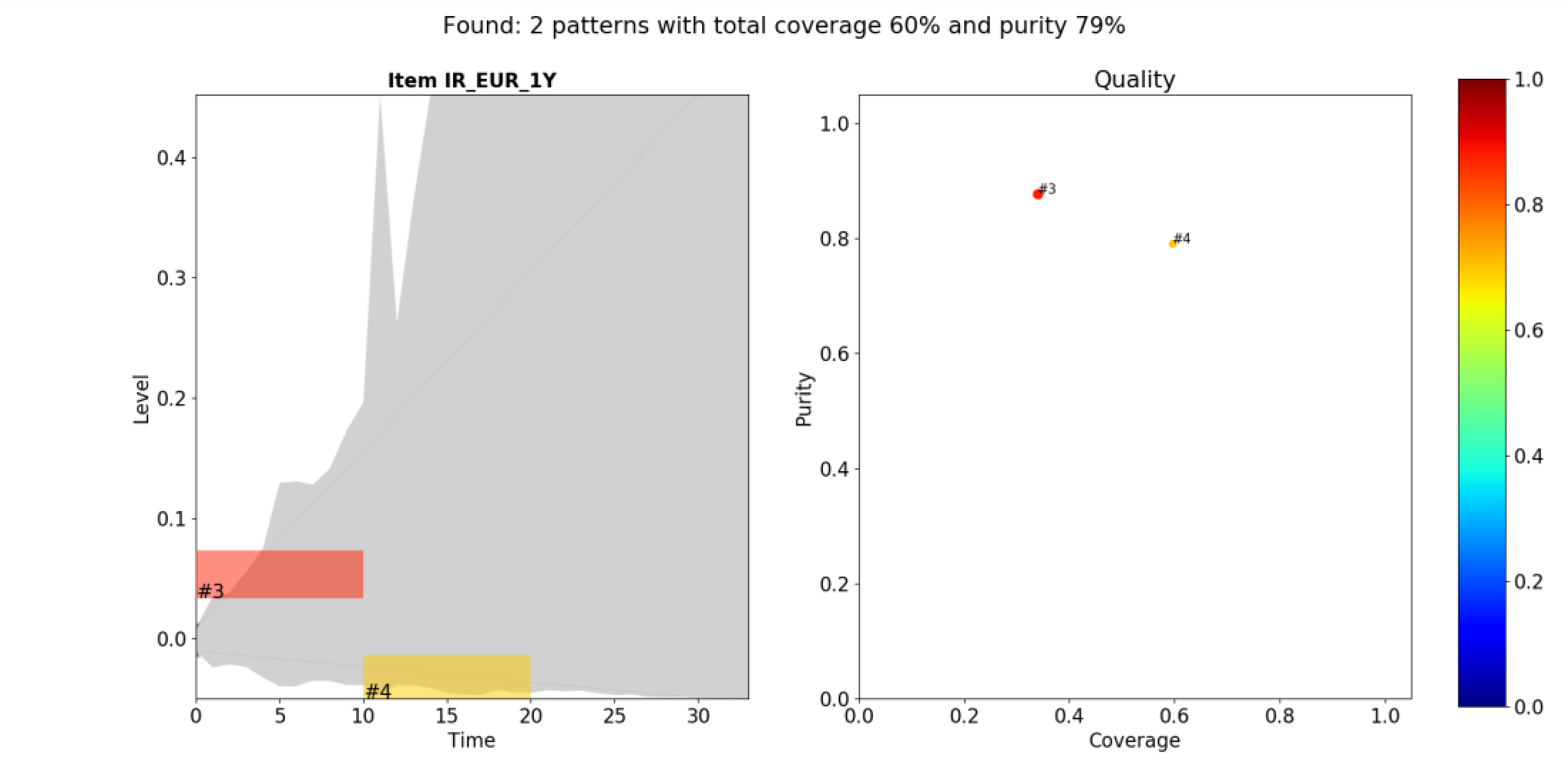
Found: 2 patterns with total coverage 93% and purity 72%



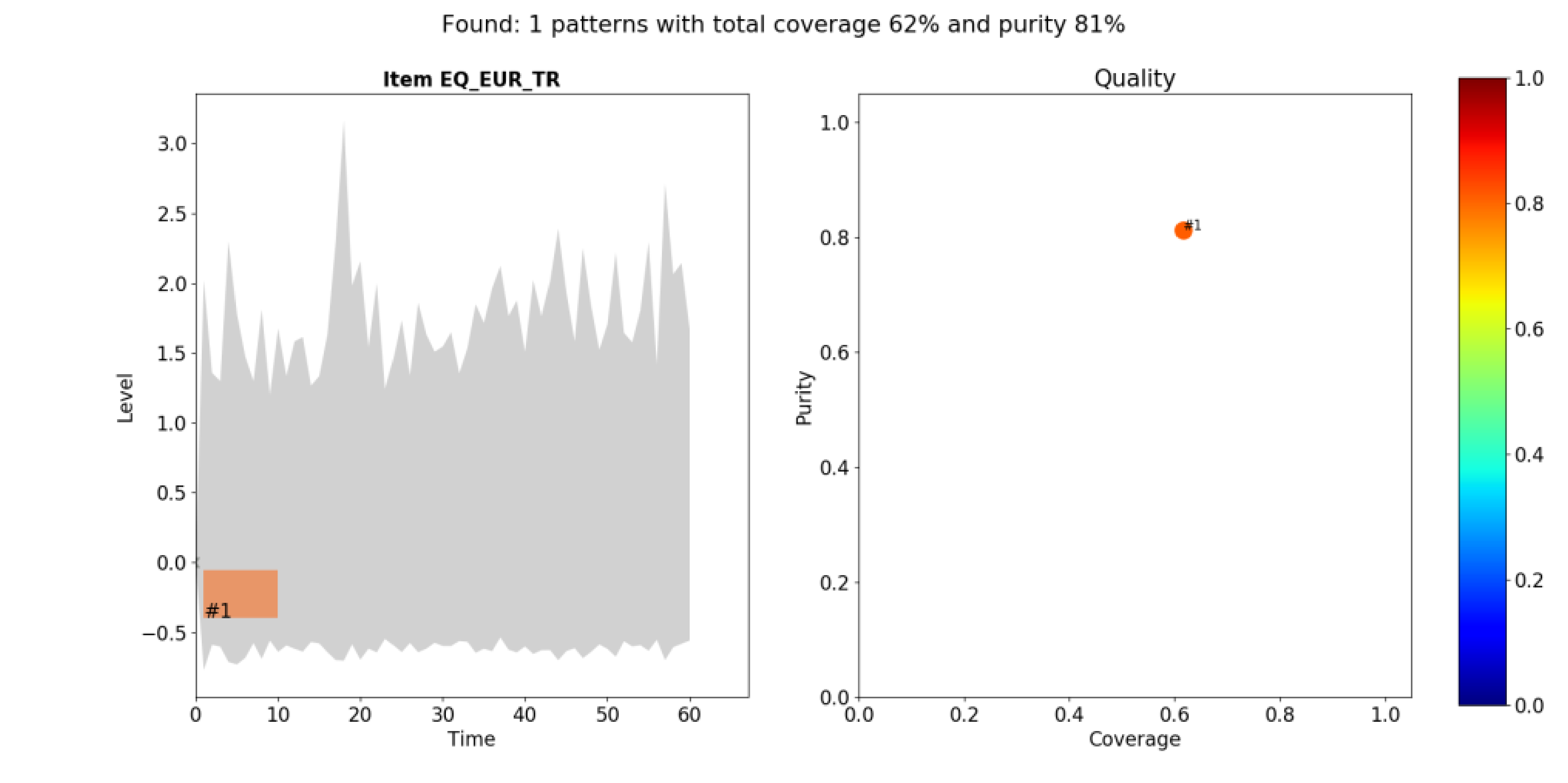
Example 2: Classic product with high guarantee



Example 3: Classic product with guarantee and fixed surrender value



Example 4: Fund with significant equity investments

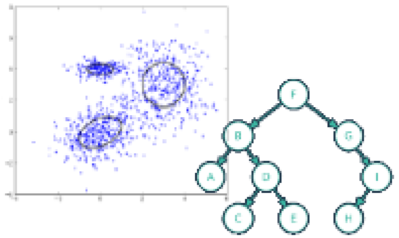


Agenda

The evolution of data analytics

Statistics	Data Mining	Machine Learning

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$$



Statistics

Analyzing quantitative information

- Descriptive analysis and hypotheses testing
- Scientific sub-topic of mathematics
- “Data generating process”

Data Mining

Explaining patterns in the data

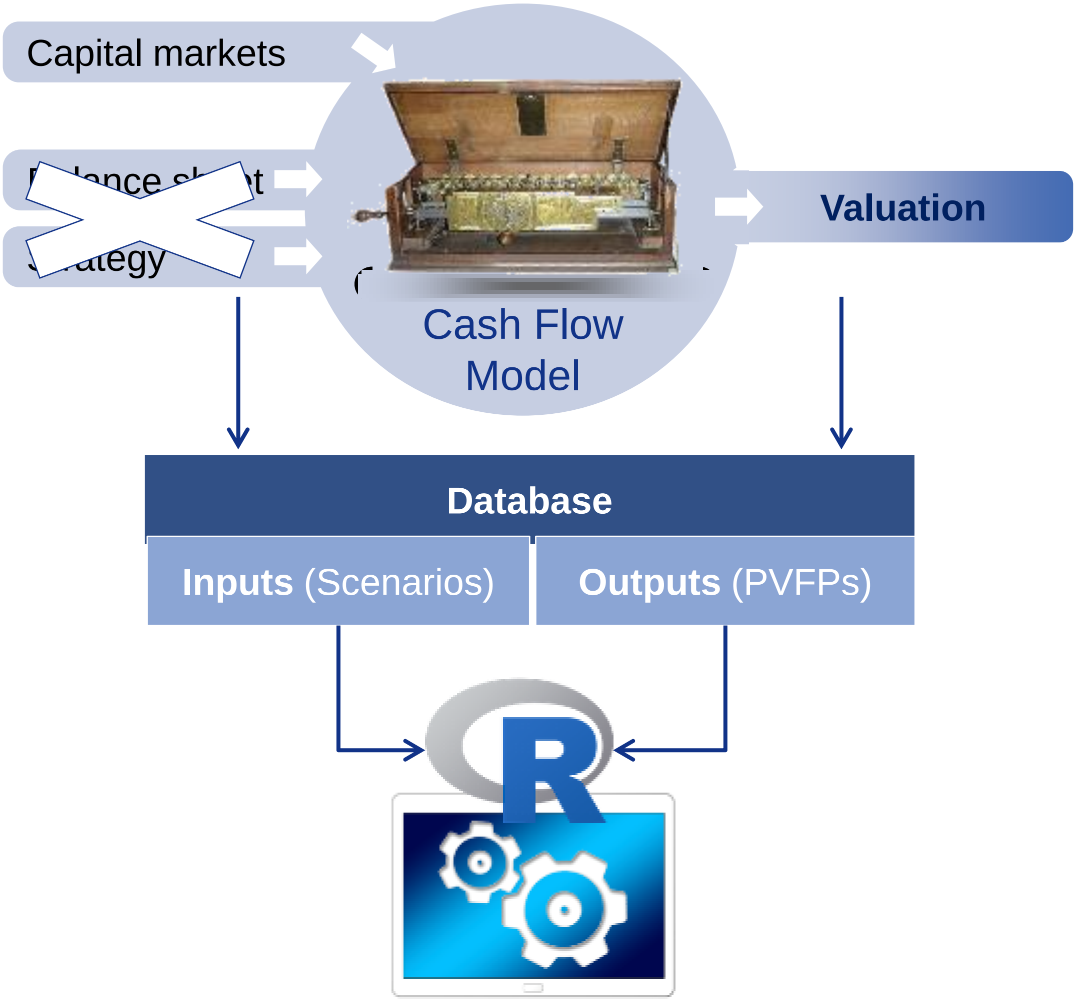
- Information extraction from large data sets
- Visualization and structuring
- “Patterns”

Machine Learning

Prediction of future based on experience

- Flexible models for complex data sets
- Model learn from data / experience
- “Prediction”

What do we want to achieve



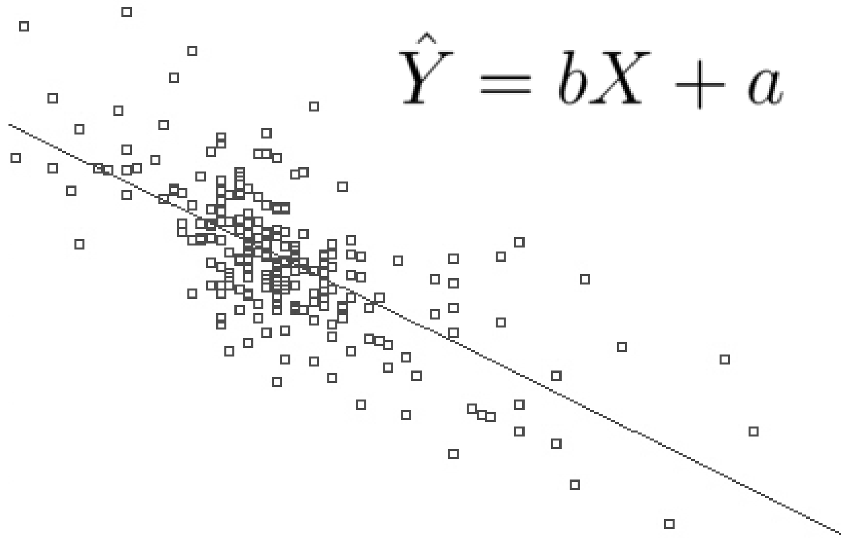
- Run time of cash flow models is very high
  - Projection of huge number of internal fields
  - Depends on granularity of assets and liabilities (model points)
- Store results from all calculations
- Apply Machine Learning algorithms to predict PVFP for a given capital markets scenario
- Use these models for quick calculations of PVFPs within the same quarter

Huge speed enhancement

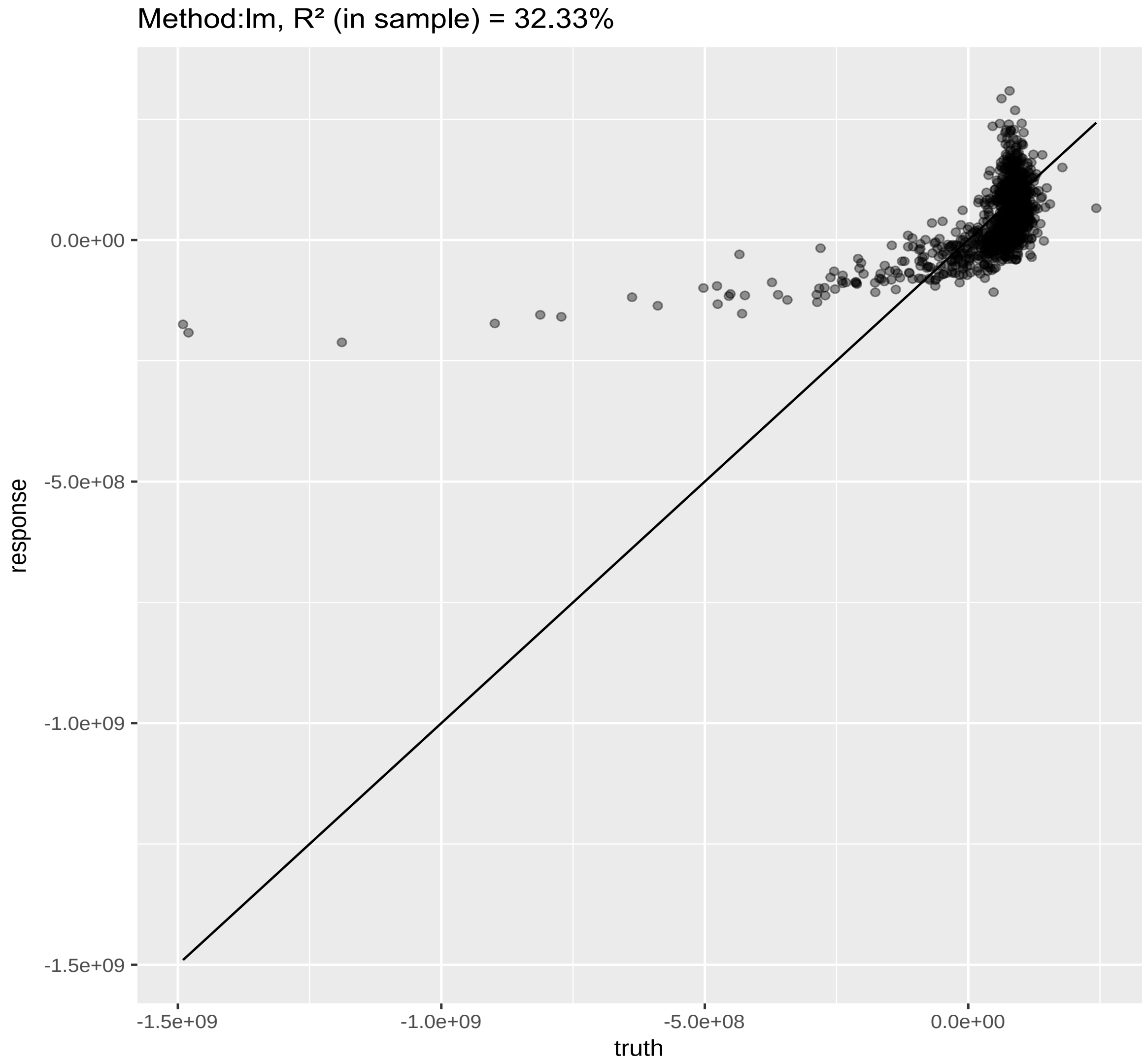
Linear regression

Model assumes linear dependency between

PVFPs and the capital market factors

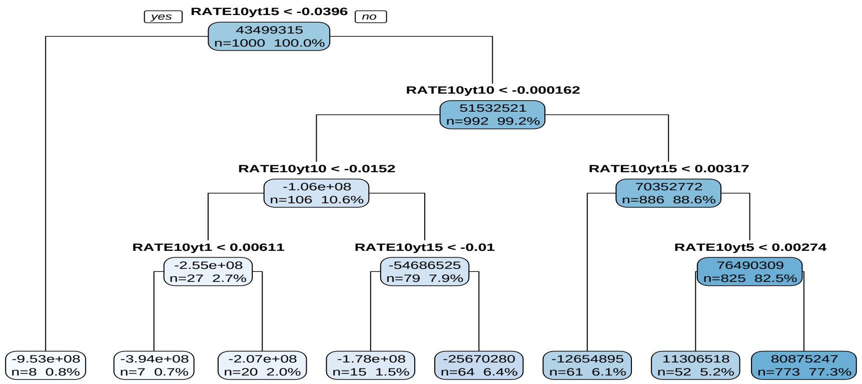


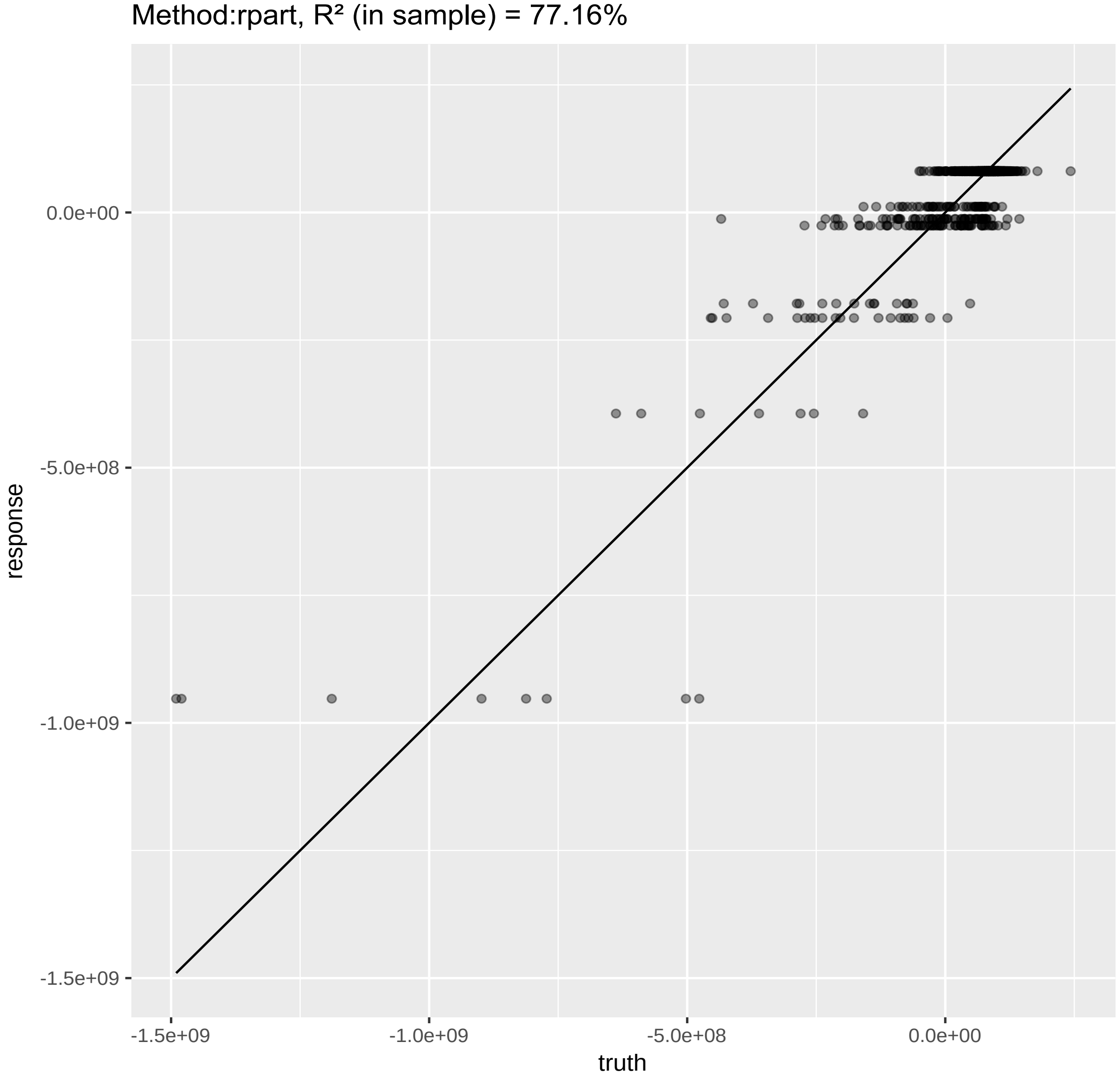
https://commons.wikimedia.org/wiki/File%3ALinear_regression_scatterplot_with_generic_formula.png



Decision trees

A decision tree defines a hierarchical sequence of rules (decisions) on the capital market factors which branches out to a predicted value

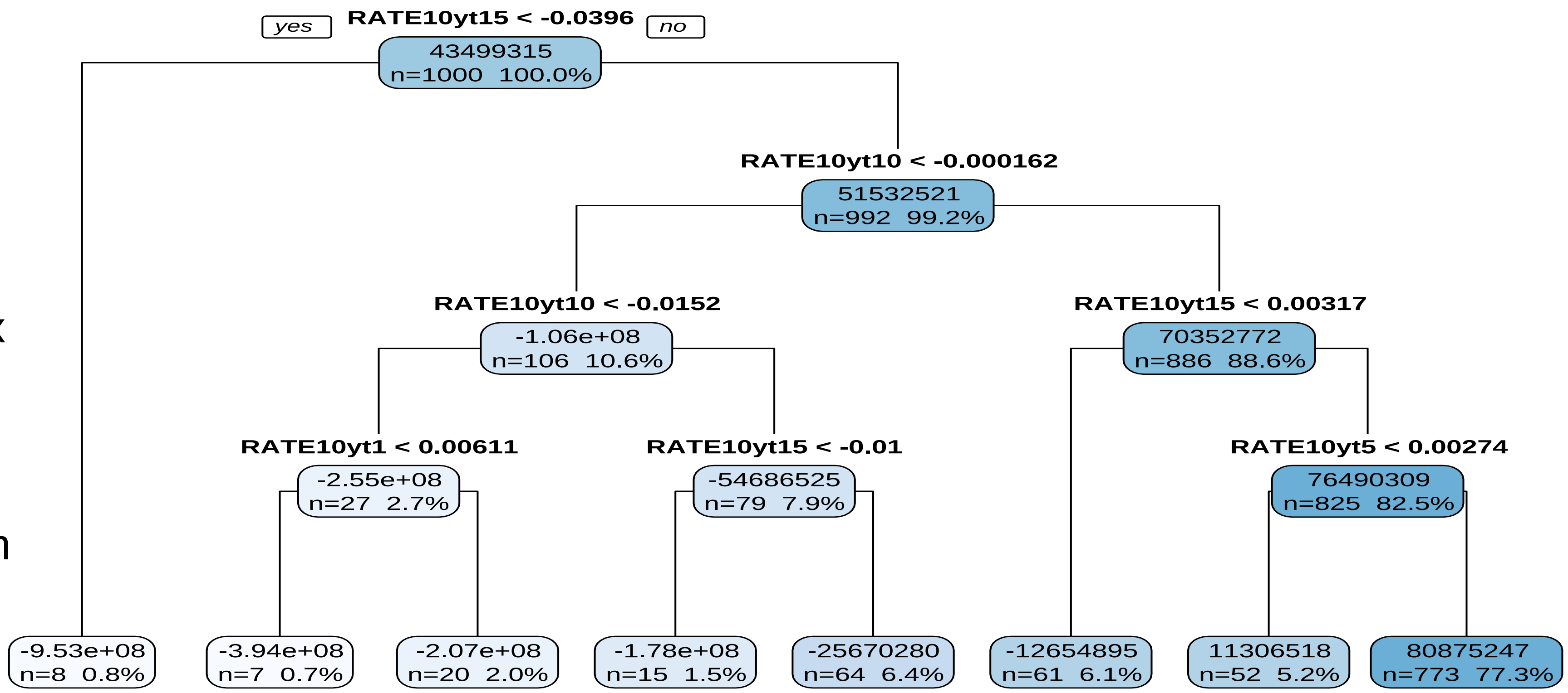




Decision trees are quite flexible but weak

Properties

- Decision trees are generated very fast
- Typically not too complex (binary splits and few leaves)
- Are readable for a human

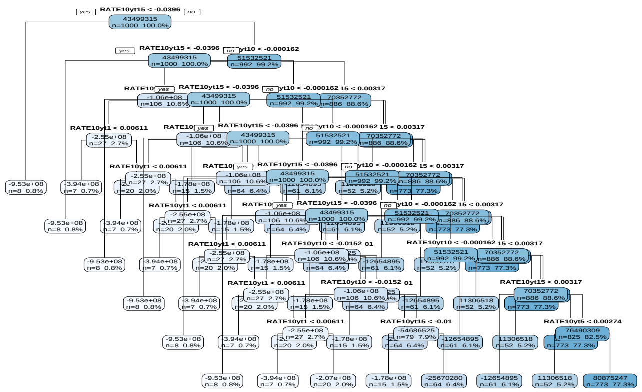


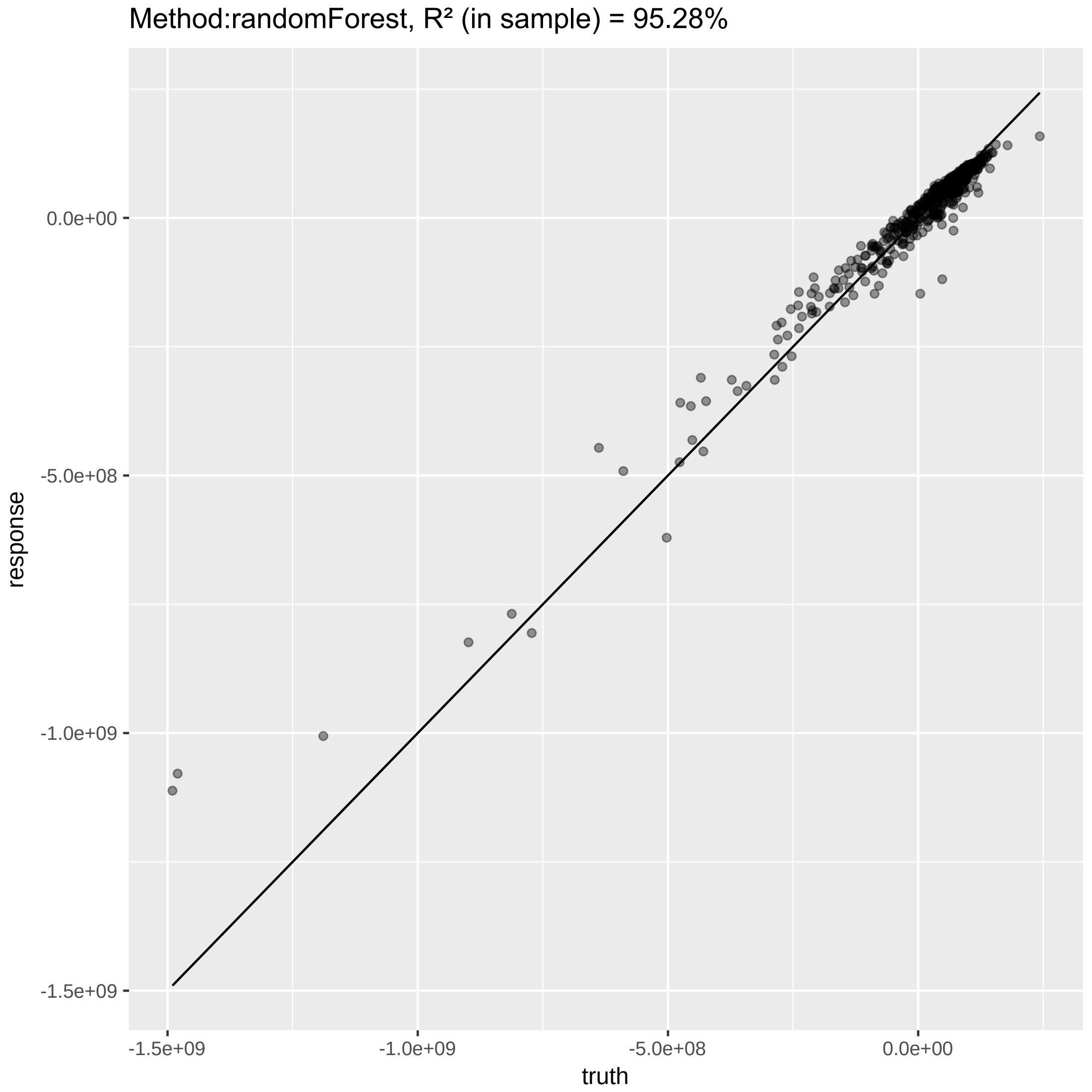
Random Forest

Random = Trees based on random sub-sets of features and data

Forest = Many trees (ensamble)

Prediction of a Random Forest is average of predictions of the individual trees

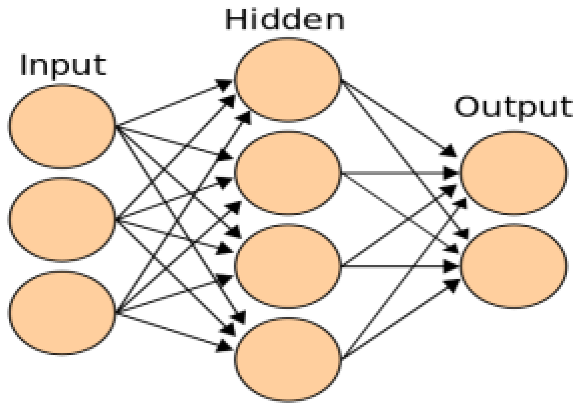




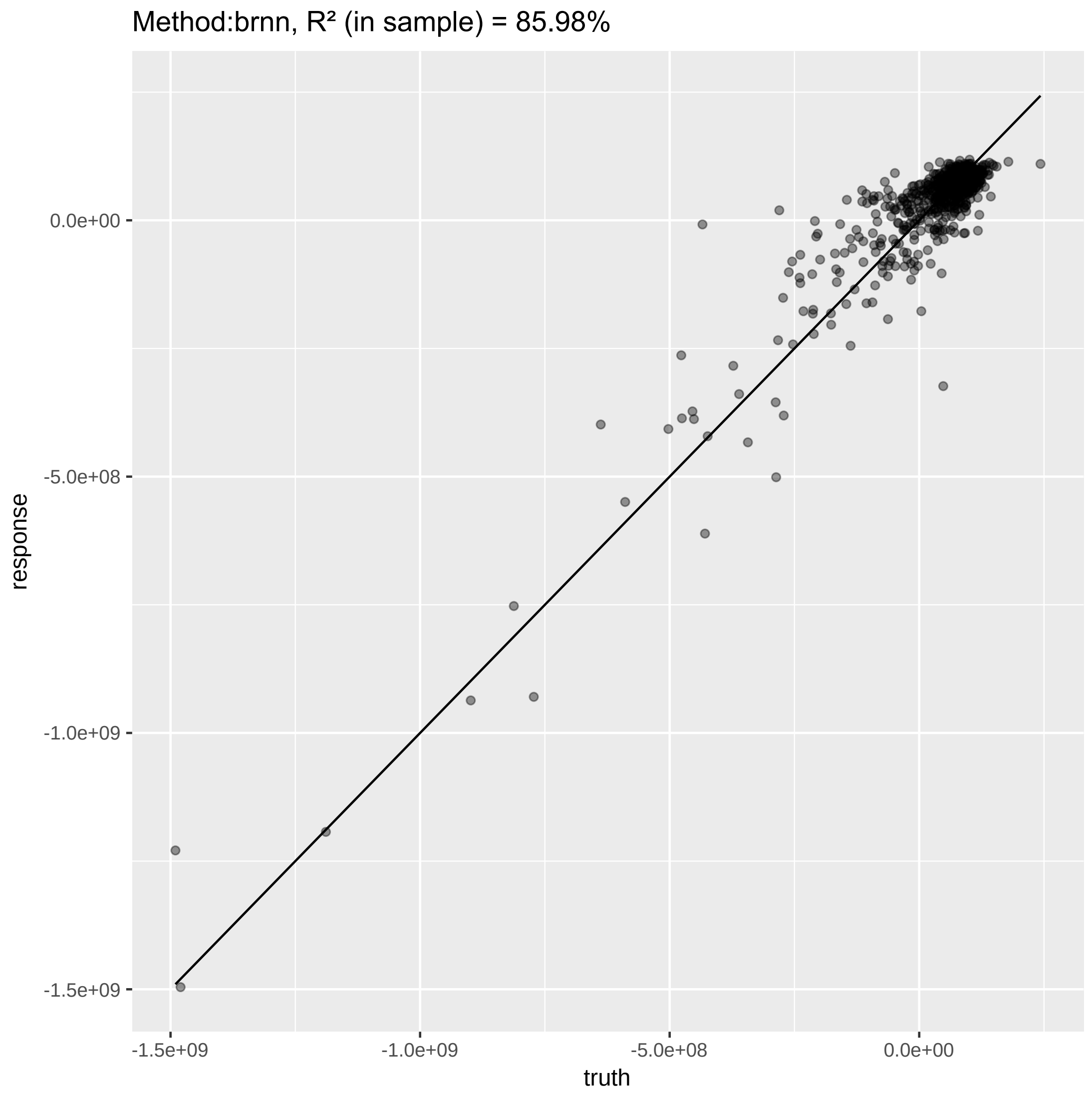
Neuronal network

Structured data processing

- Input layer has nodes (neurons) for each feature at each time step
- Output layers represents prediction
- Hidden layers react to patterns in the input
  - Exact pattern cannot be prescribed
  - Number of hidden layers determines predictive power → deep neural networks

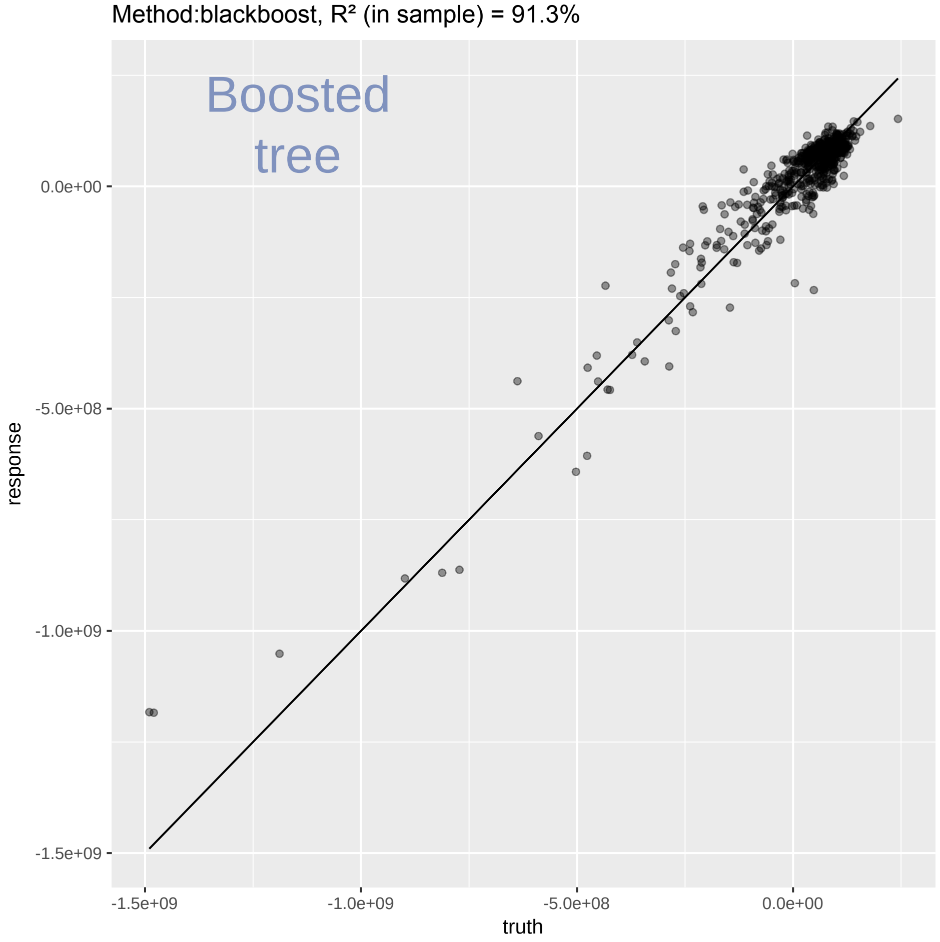


https://commons.wikimedia.org/wiki/File%3AArtificial_neural_network.svg

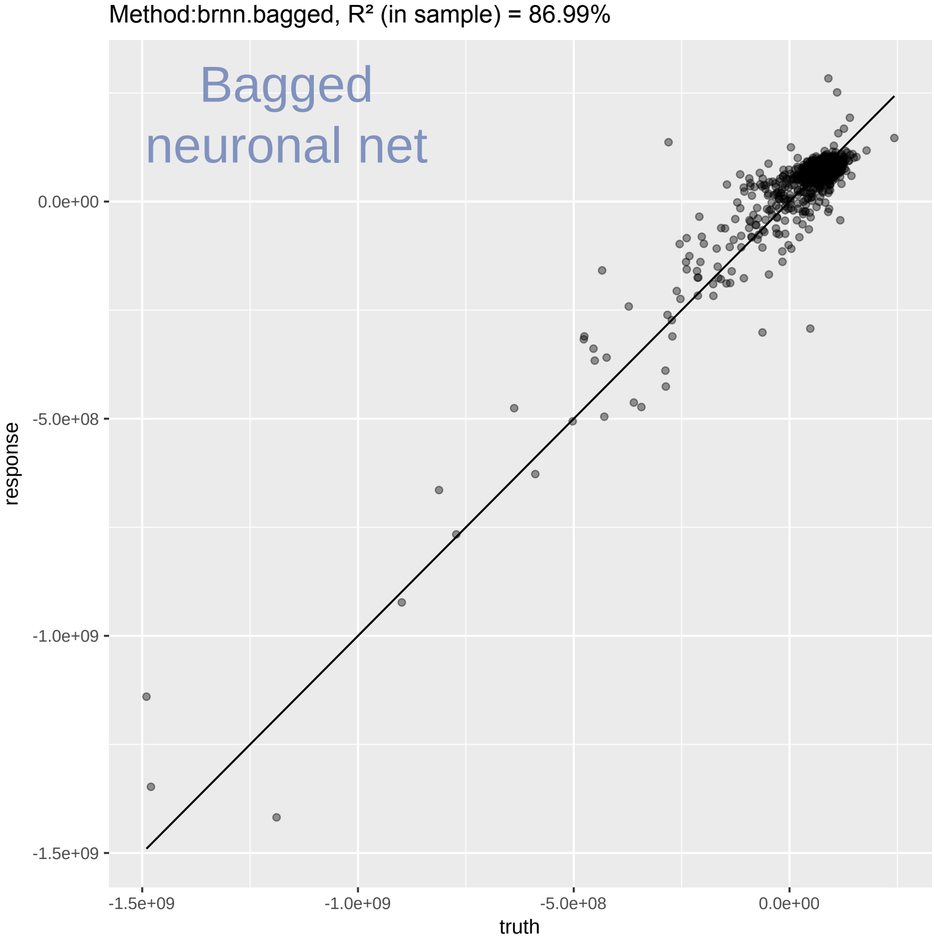


Boosting + Bagging

Boosting (forward stagewise modelling)

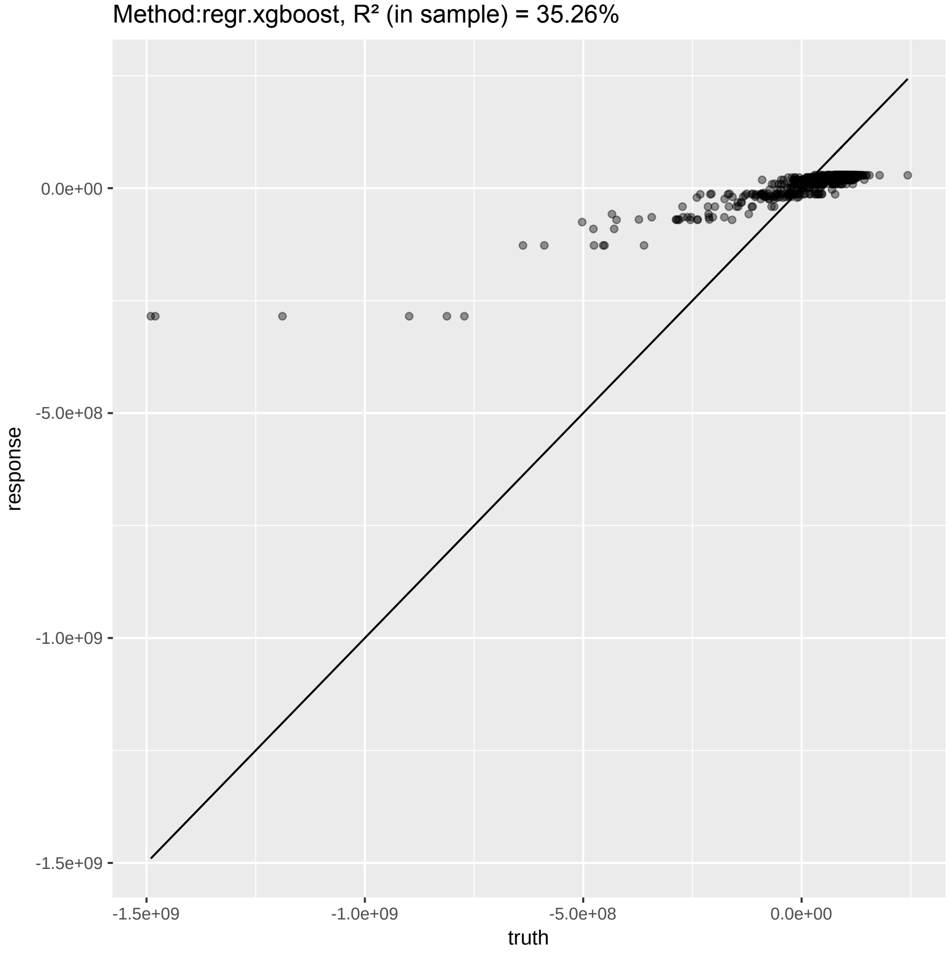


Bagging (Subsampling)

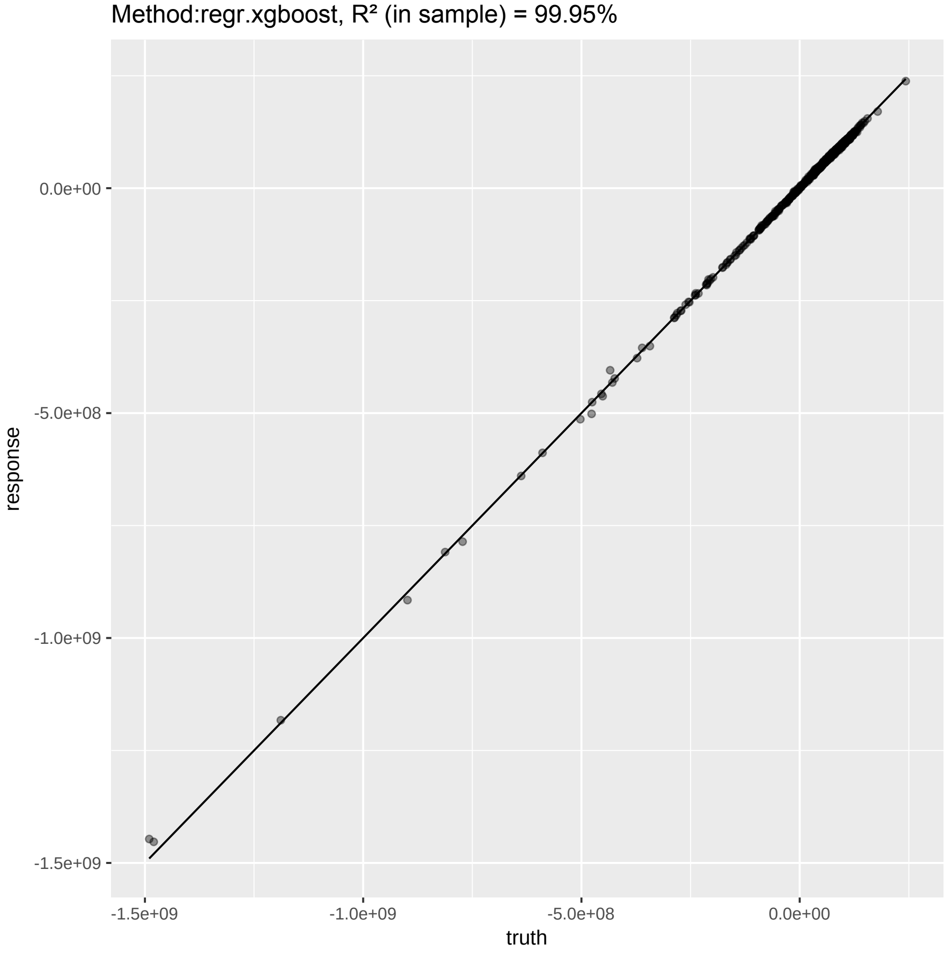


Hyper parameter tuning

Without tuning



With tuning



Machine learning is not complicated

#1. Define data

mydata = *CF-Model-Output-Data combined with Scenario Information*

#2. Define tasks

tasks = list(makeRegrTask(data= mydata, target=„PVFP"),…)

#3. Define methods

learners = list(makeLearner("regr.rpart"), …)

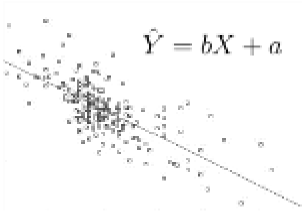
#4. Make the analysis (combine all tasks with all methods, do cross validation, compare e.g. MSE, MAPE, R^2,…)

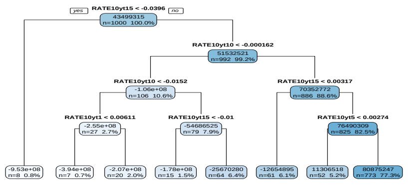
(bmr = benchmark(learners, tasks, cv10, measure= list(mse, mape, rsq)))

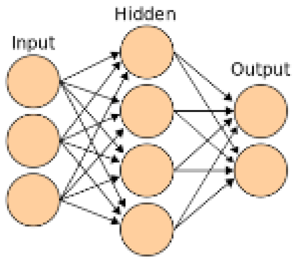
	task.id	learner.id	mse.test.mean	mape.test.mean	rsq.test.mean
1	mydata	lm	1.183934e+16	0.9909817	0.2960641
2	mydata	rpart	6.879358e+15	1.1763703	0.6037440
3	mydata	randomForest	4.693410e+15	0.8590164	0.7381586
4	mydata	brnn	3.775355e+15	0.8320113	0.7650239

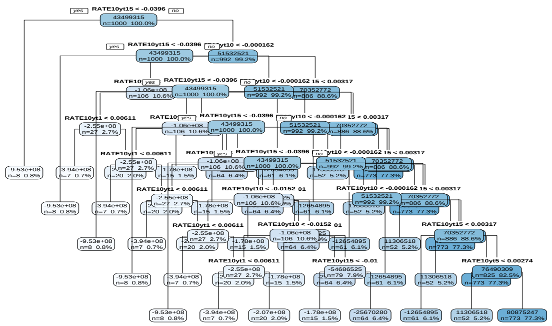
Comparison of approaches

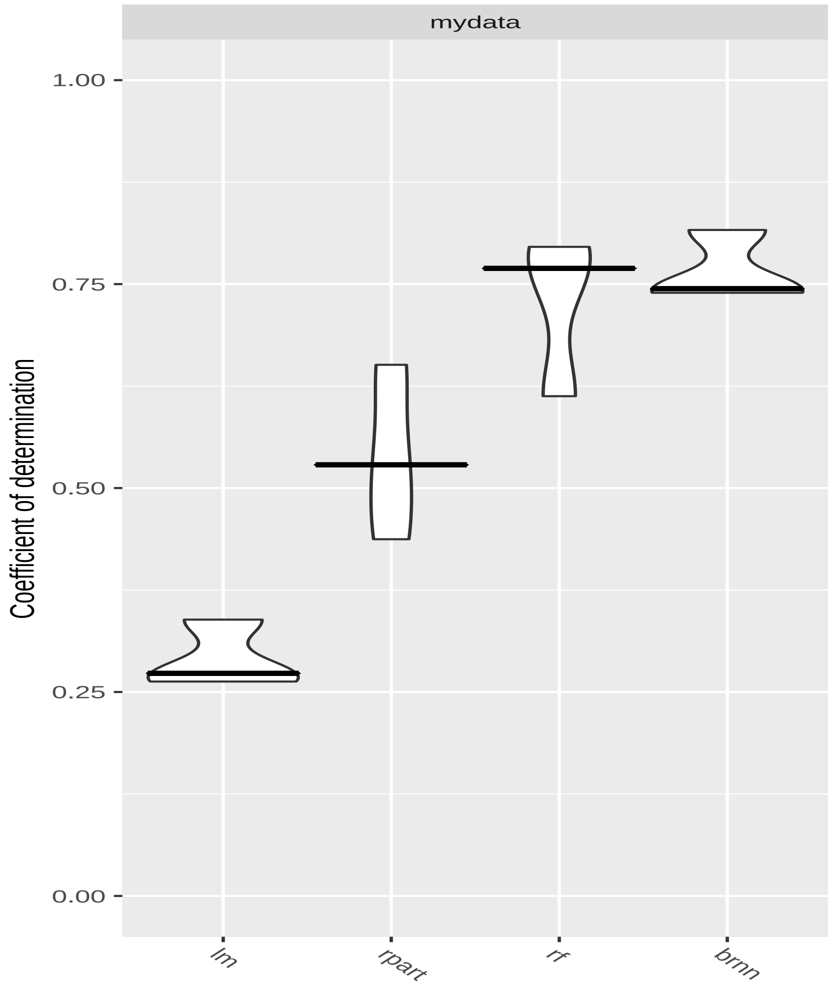
- Modern ML models are often easier to use and yield as good or better results than classical methods
- Overfitting is an issue which has to be addressed (e.g. using cross-validation, bagging, randomisation)
- ML models are often black boxes



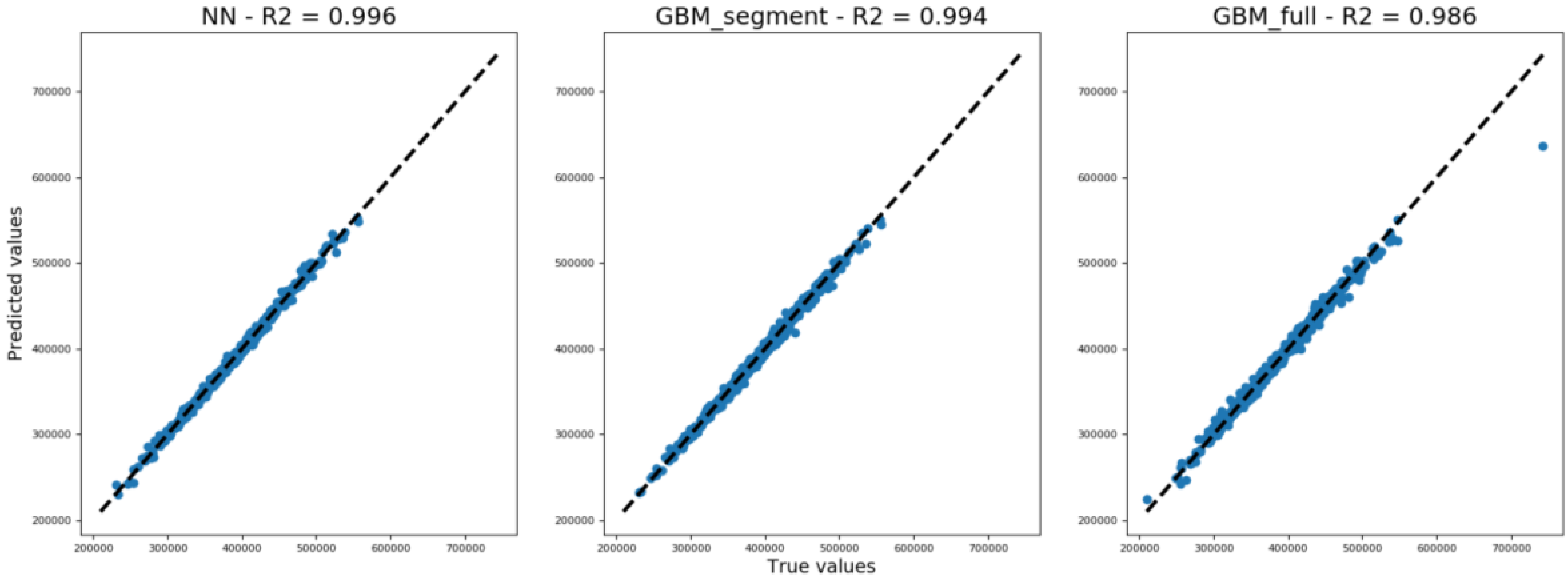




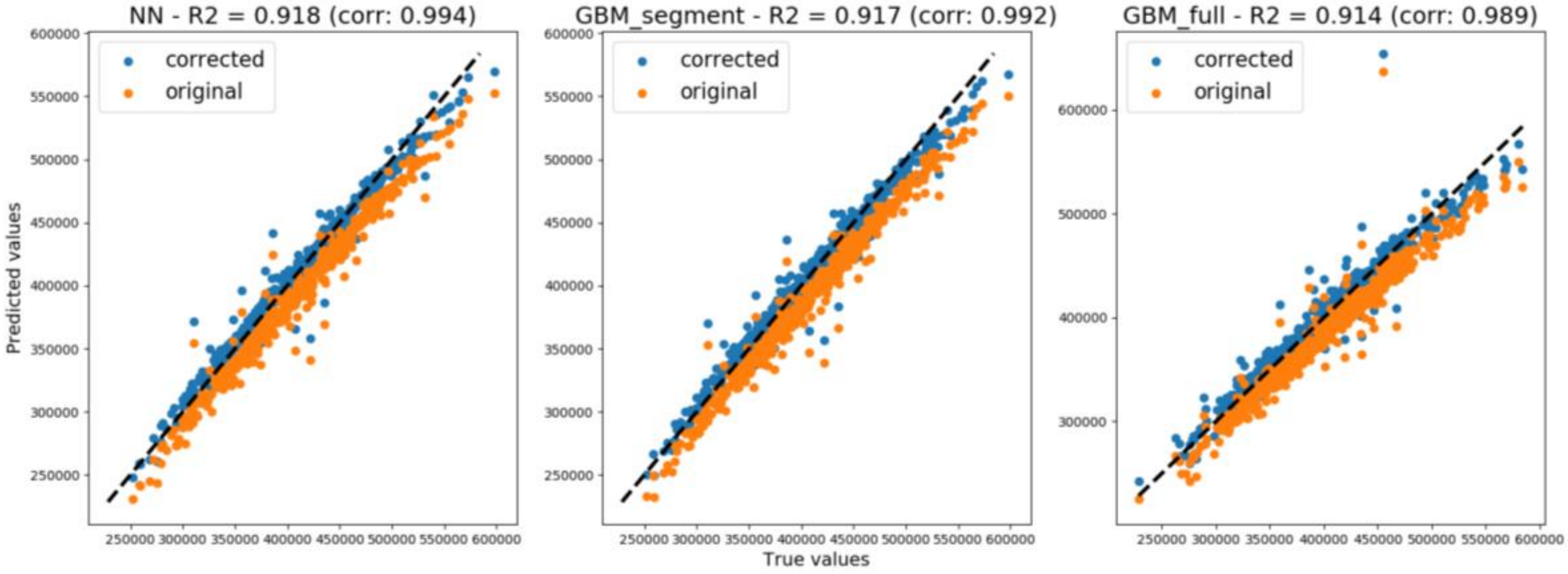




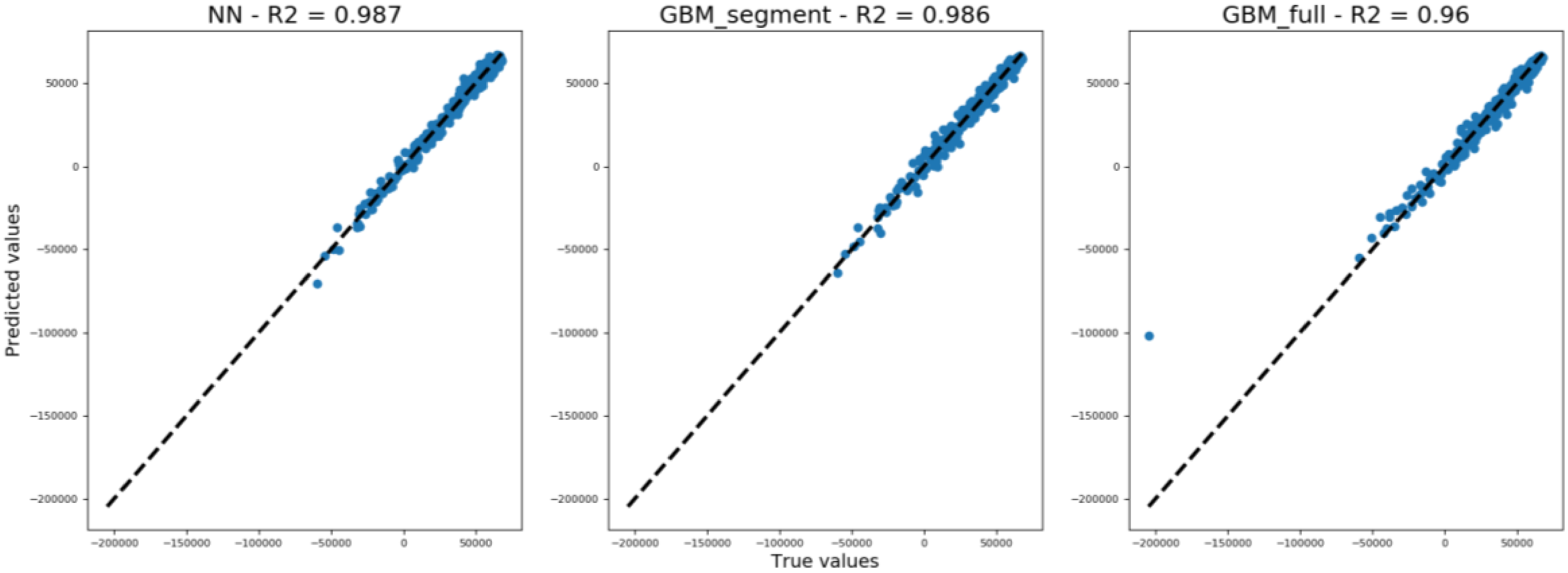
Fund 1: protection product



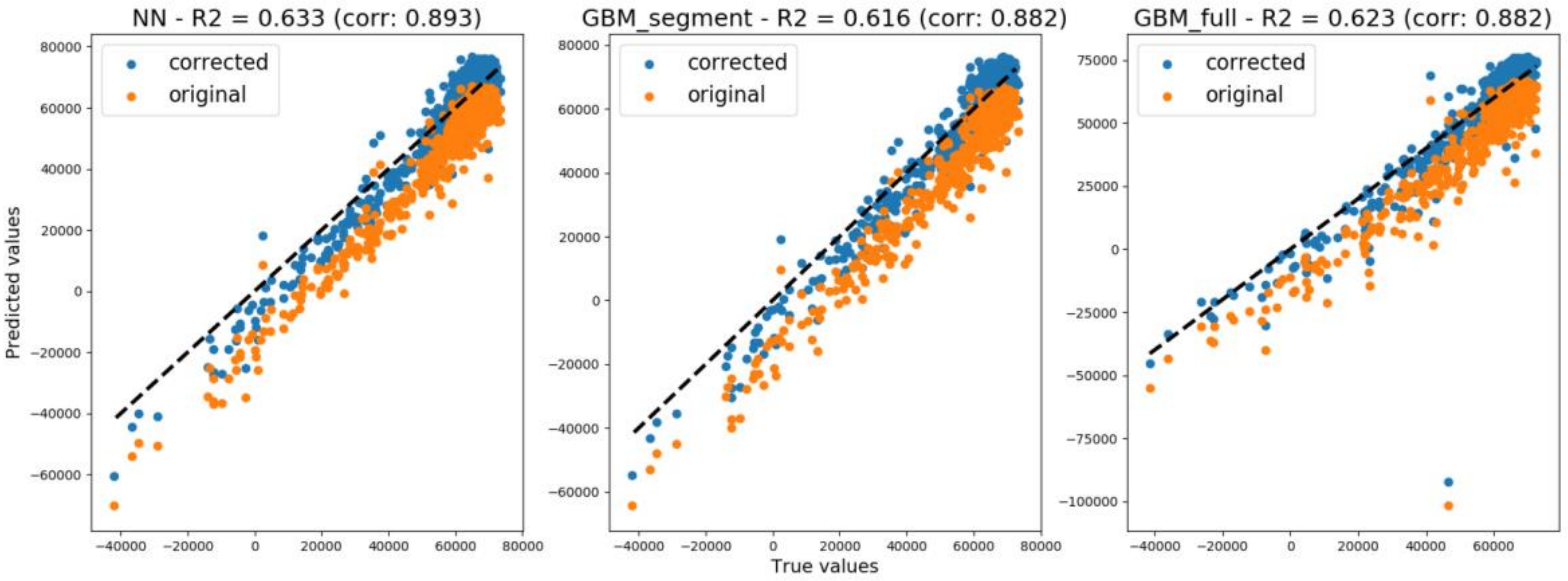
Fund 1: next quarter prediction



Fund 2: high guarantee product

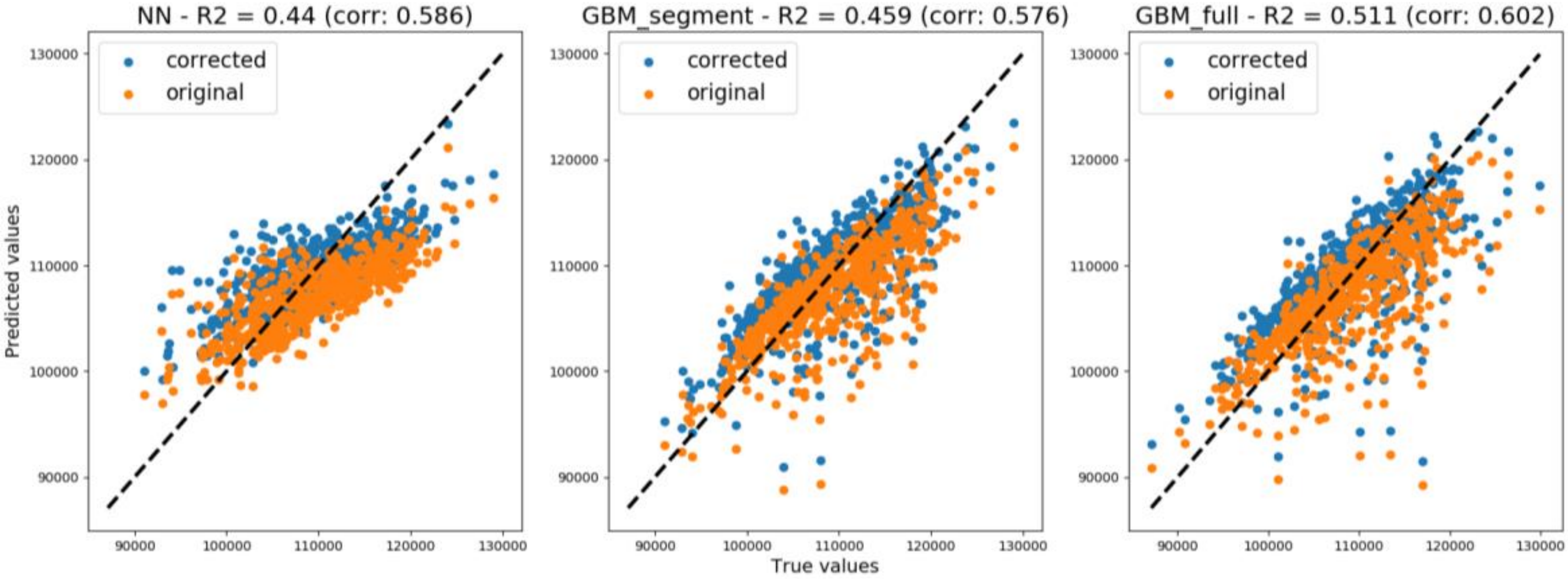


Fund 2: next quarter prediction

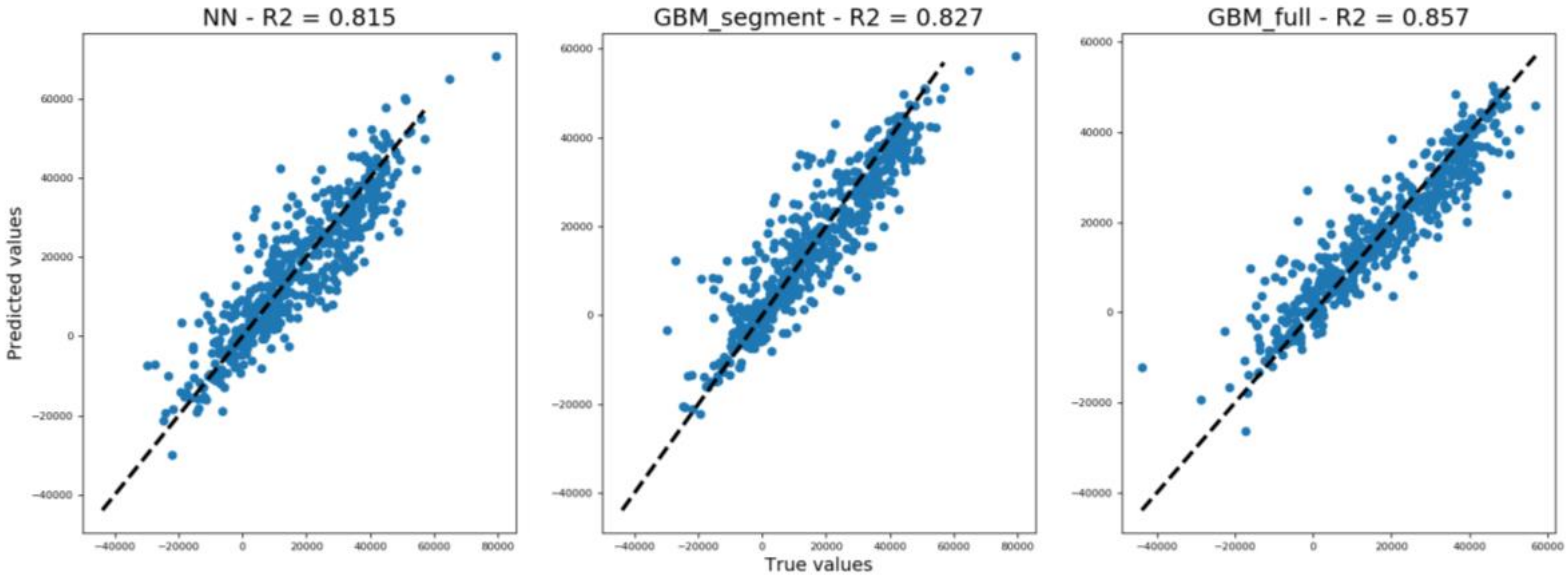


Fund 3: guarantee and fixed surrender value

Fund 3: next quarter prediction



Fund 4: equity focused product



Fund 4: next quarter prediction