

# Build and use of an LGD Benchmark Model

-using Machine Learning  
-based on GCD data template

Occasion:	GCD Conference at HSBC London
Presenters:	Jeroen Berends HSBC Thomas Aldheimer FCG
Moderator:	Philip Winckle FCG
Date:	2022-11-08

# Important Cautions:

**Disclaimer:** Any views expressed in this presentation are those of the presenters and do not necessarily represent the views of HSBC, FCG, Global Credit Data or any of its members.

**Anti-trust warning:** Participants are warned not to provide sensitive information about their bank or to engage in discussions which might encourage or lead to collusive behaviour. If in doubt then please seek guidance from your own bank's policies or legal counsel.

**Data shown:** Data used in this presentation has been modified to ensure that it does not expose the portfolio of the bank involved, while still being a good representation of the industry

## Story Line



- 👍 Background to the exercise
- 👍 Machine Learning methods used
- 👍 Segmentation and Model Structure
- 👍 Model performance
- 👍 Insights and outcomes
- 👍 Future use of ML in Credit Risk Modelling

## Discussion Points (Breakout session)



1. **Incomplete data fields:** remove, impute or model?
2. **High loss segments:** can this help us achieve a **wider spread** of LGDs in implementation?
3. **Segmentation and model structure:** treat important collateral and industry segments specially?
4. **dev/train/test split:** How to interpret the inevitable drop in out of time performance.
5. **Non-linear drivers:** Assumption of linearity not required in ML; good for modelling or an over-fit?



- Advanced IRB models being re-developed
- GCD Member bank
- Wanted to explore more options for Corporate LGD
- Historical LGD data in GCD 8 table data format
- Interested in Machine Learning for LGD modelling



- Started in 2005 to help banks model better by pooling data
- 55 plus members world-wide
- Created industry standard LGD data structure (8 table)
- Promoting banks' use of data in modelling



- Specialist in helping banks with Credit Risk Modelling
- Long term GCD partner
- 2019 built Machine Learning LGD model on GCD Large Corp data set as demonstration on pro-bono basis
- Interested to develop ML for banks more widely

Full scale LGD challenger model using Machine Learning

## Summary of Aims and Scope

**Deliverable:** **“Challenger” or “Benchmark” LGD Model** to explore alternatives in:

- Model design
- Data handling

**Purpose:** **Contrast with a bank’s “champion” model** to find:

- Improvements to champion model
- Demonstrate a wide range of alternative methods
- Find new risk drivers
- Compare performance
- find max performance ceiling

**Methods:** **More methods incorporated:**

- Segmented multi-step model (cure, real estate collateral, high/low LGD elements)
- Alternative data transformations
- Machine Learning (ML) methods using many more drivers, non-linear, multi-factor optimised
- Special focus on “real estate collateral heavy” facilities



## Target Definitions and Model Structures

LGD is difficult to model as most cases are full recovery or low loss, with other cases bunching around full loss. This is very different to the average outcome.

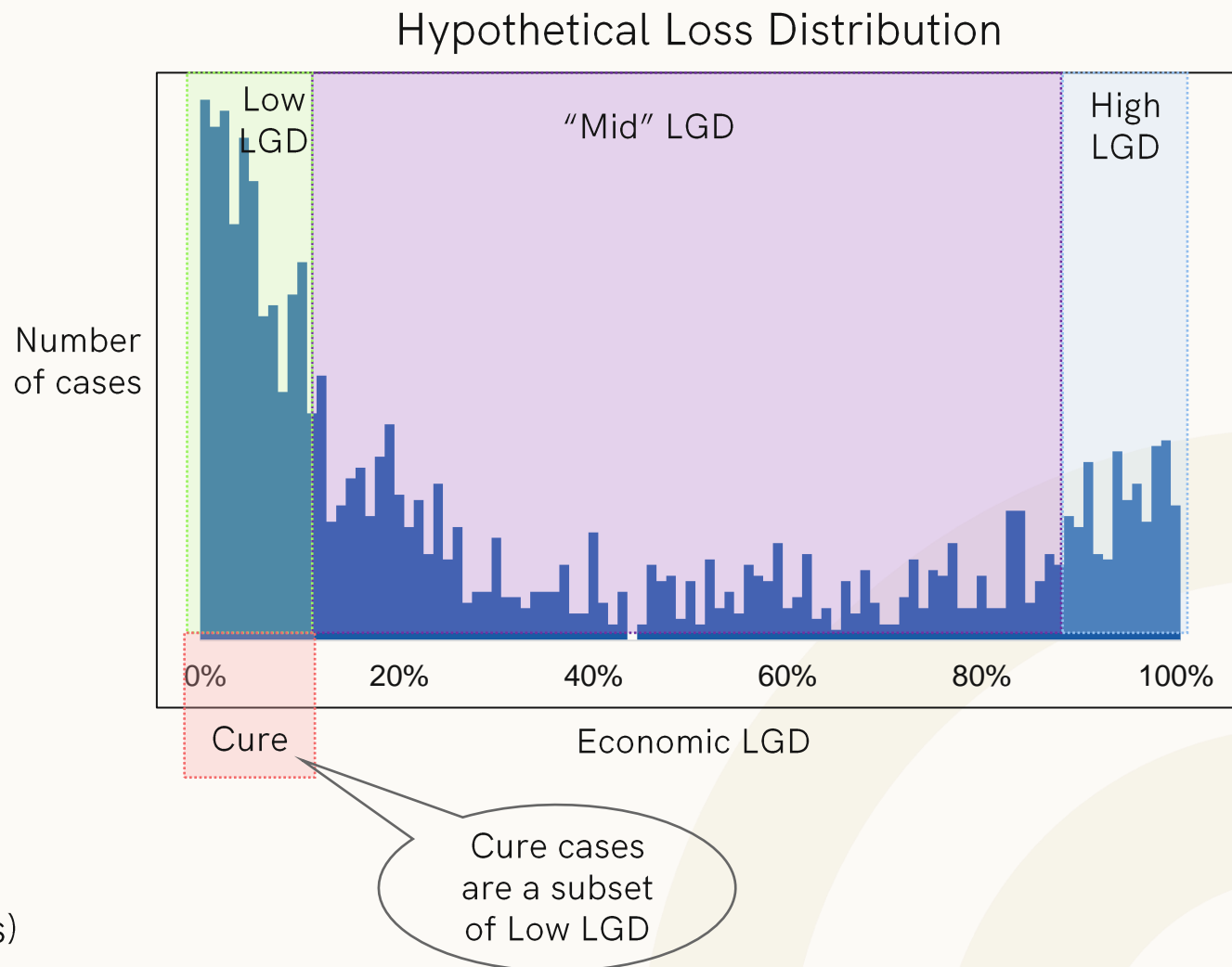
- Modelling "cure" cases separately is already regarded as good practice as it handles some of the low mode
- We tried to find drivers for the other modes of "Low LGD" and "High LGD" as well as for "Cure"

The 3 targets for "classifier" (yes/no) models were:

- **Cure** = zero nominal loss and recovery within 1 year and no sale of collateral (GCD Definition).
- **Low** = not cure and LGD < 10%
- **High** = not cure and LGD > 90%

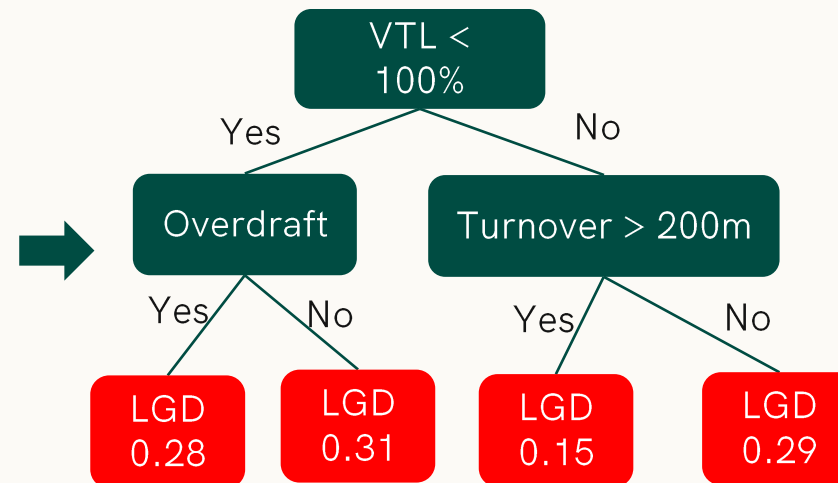
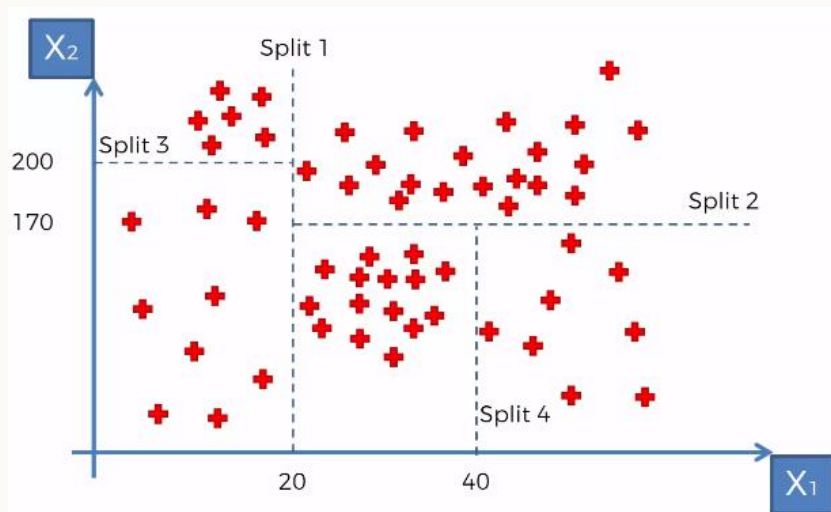
The remainder of cases were targeted in a regression model:

- **LGD** = Observed LGD (full range or remainder of cases)

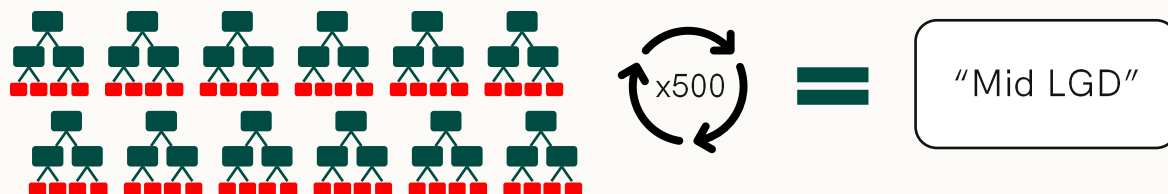


8 different techniques explored, including neural networks, decision trees and linear models. The best performing was "XGBoost" based on a decision tree structure.  
An illustrative example of training and application of such a model:

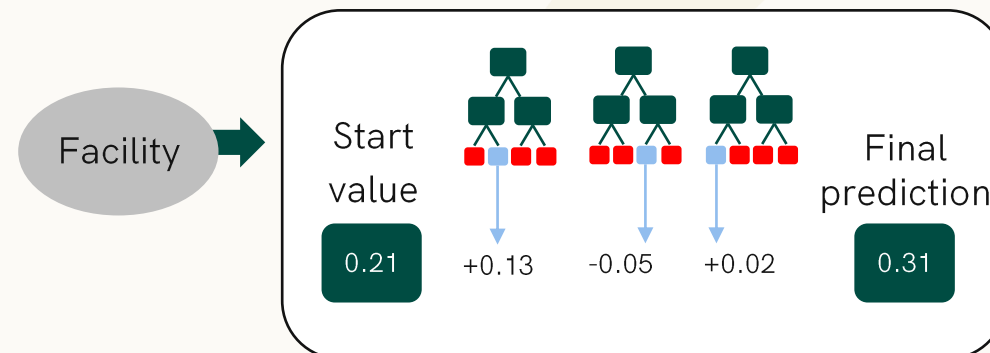
## 1 Split the data and build a decision tree



## 2 Create x100 different decision trees and iterate the process 500 times until an optimal structure is found:

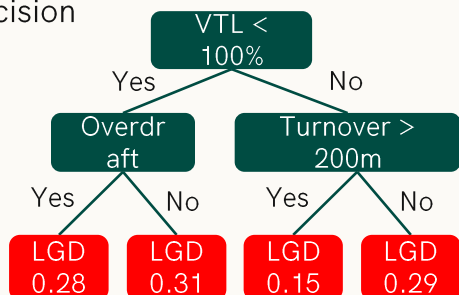
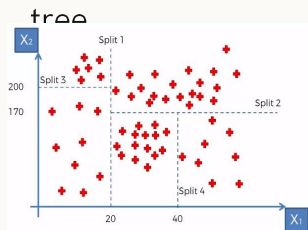


## 3 Use the optimal trees to make a prediction on actual data so that we can test performance:

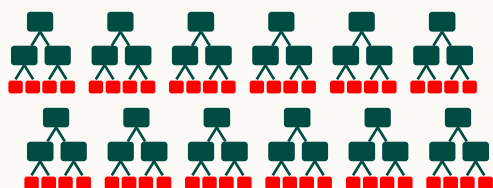




## 1 Split the data and build a decision



## 2 Create x100 different decision trees and iterate the process 500 times until an optimal structure is found:

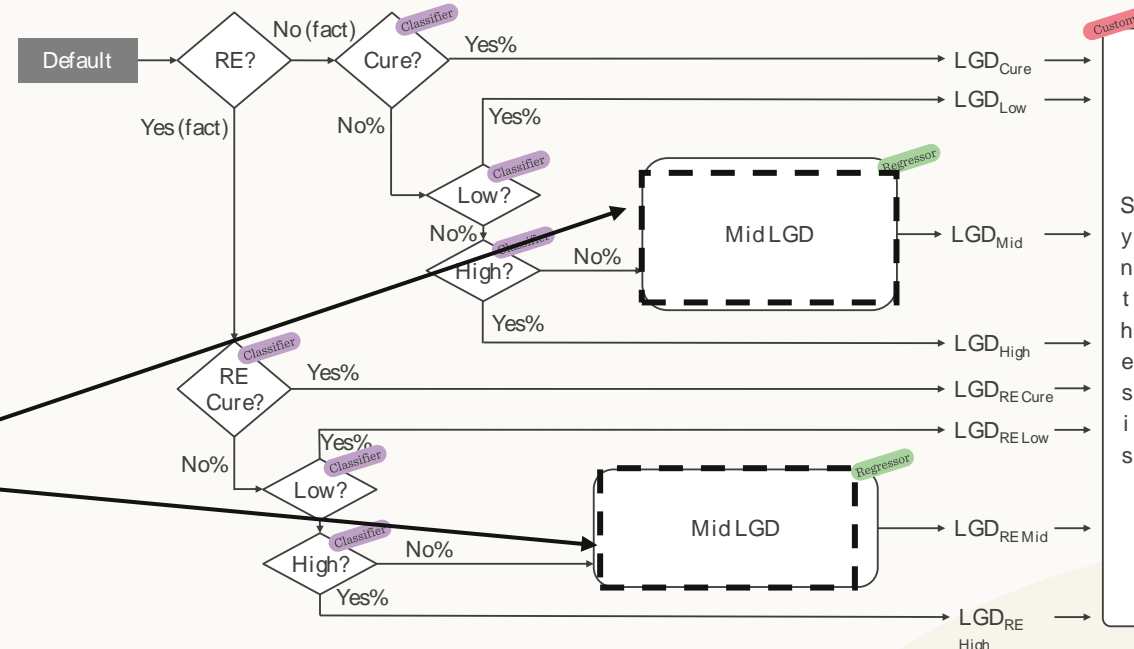
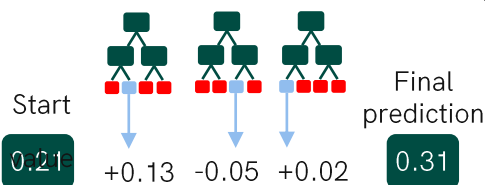


x500

"Mid LGD"

## 3 Use the optimal trees to make a prediction on actual data so that we can test performance

Facility



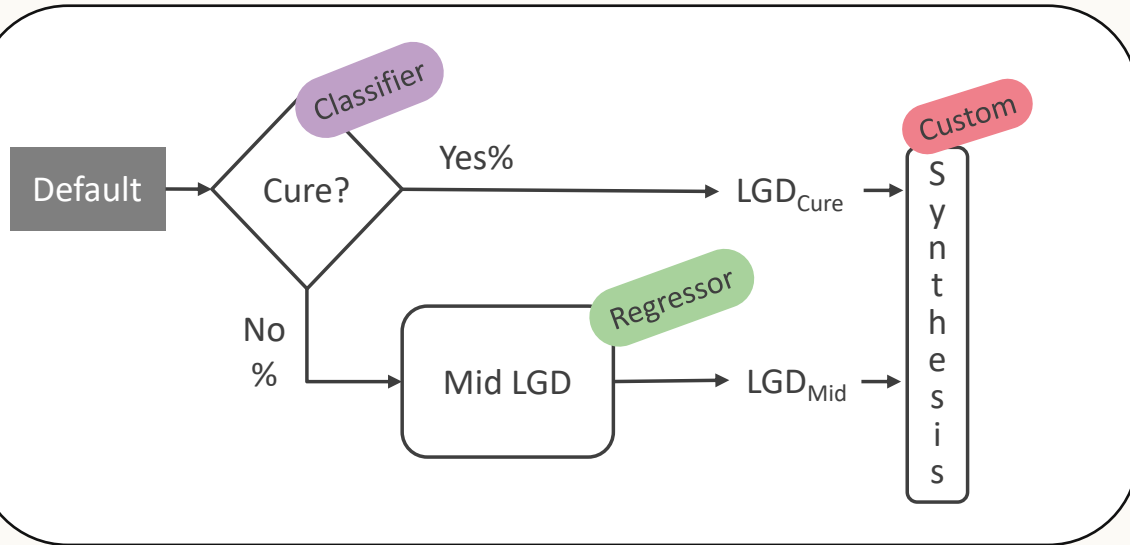
## 4 Repeat 1 - 3 for each sub-model

- Mid LGD predicts an LGD
- Cure, Low and High predicts a probability

## 5 Combine the probabilities and LGD predictions from all the sub-models into a final LGD prediction

## 6 Apply final model by running the live data through the tree structures which produces a non-linear outcome

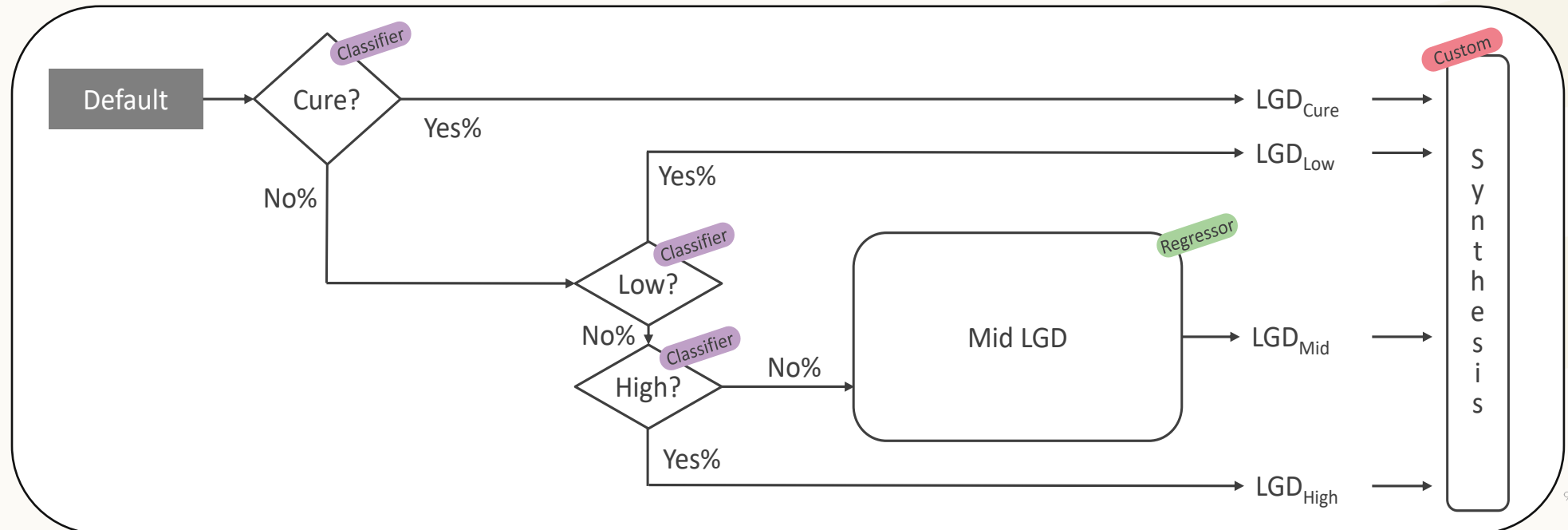




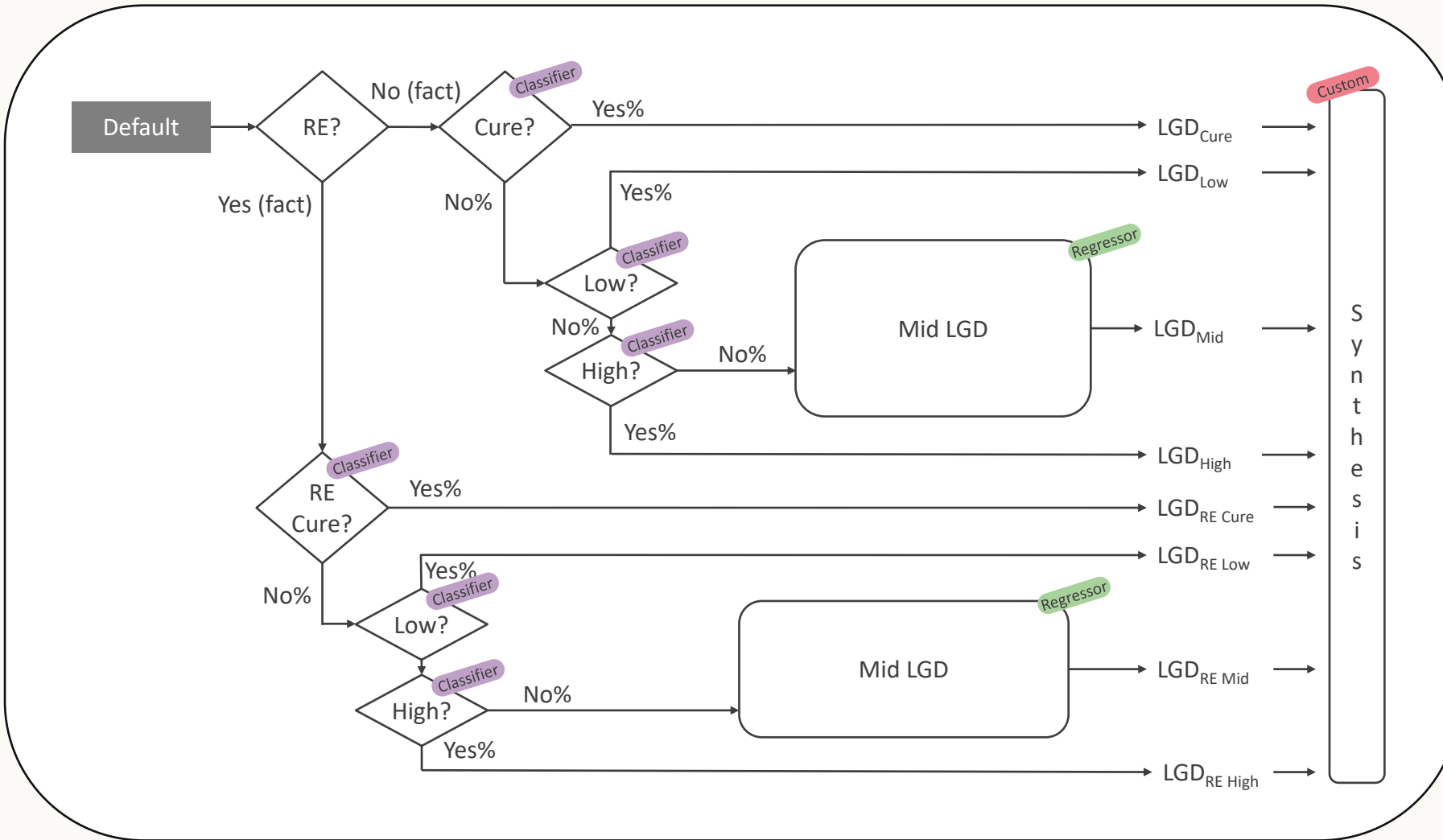
Over 20 different model structures were repeatedly tested:

← Simplest, with only a cure sub-model

Adding high and low classifier sub-models

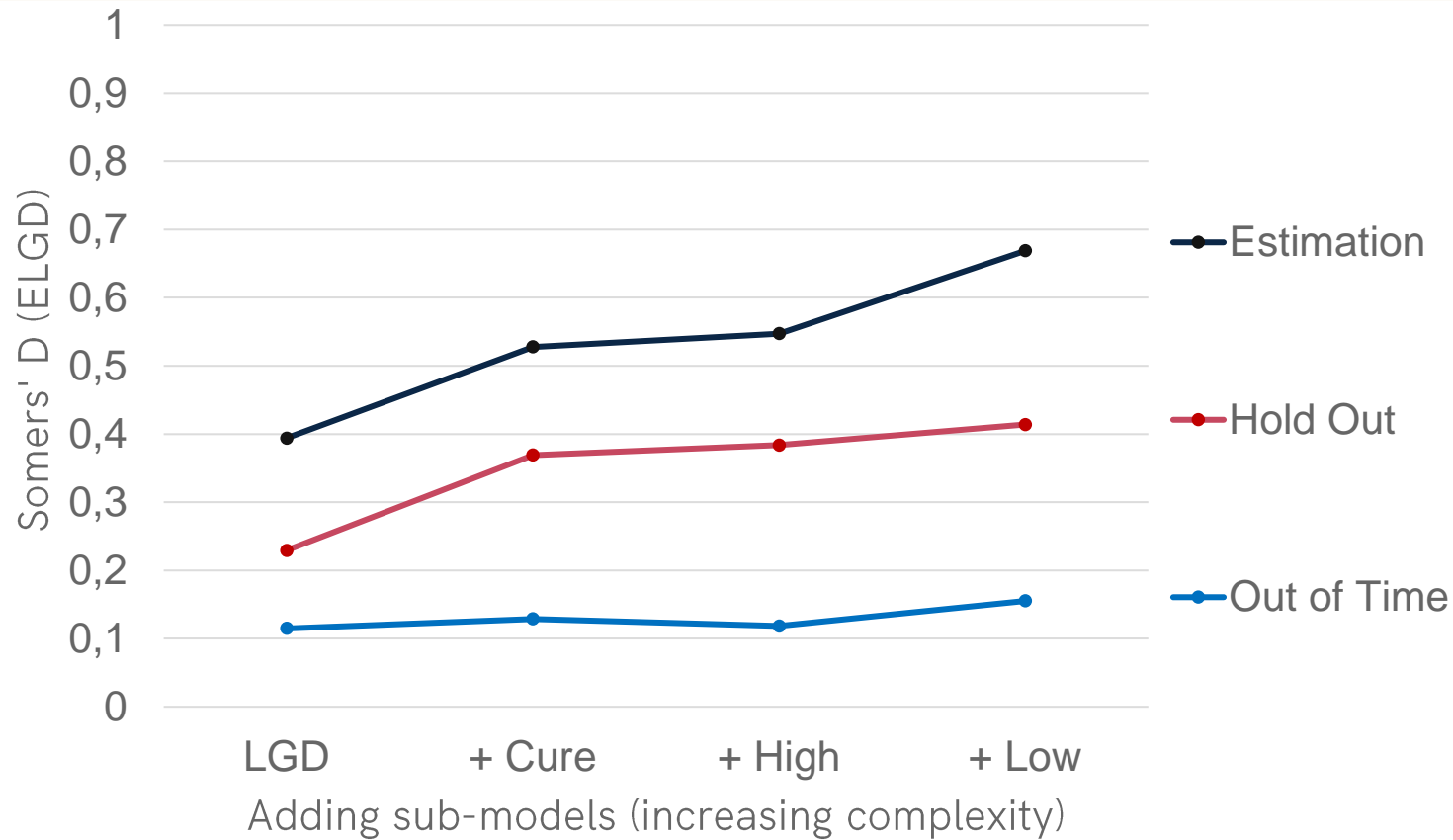


Most complex model tested, with real-estate-heavy-collateral cases broken into sub-model group:



$$LGD = P(Cure) * LGD_{Cure} + (1 - P(Cure)) * (P(Low) * LGD_{Low} + (1 - P(Low)) * (P(High) * LGD_{High} + (1 - P(High)) * LGD))$$

Effect on predictive power of adding complexity to the model structure?

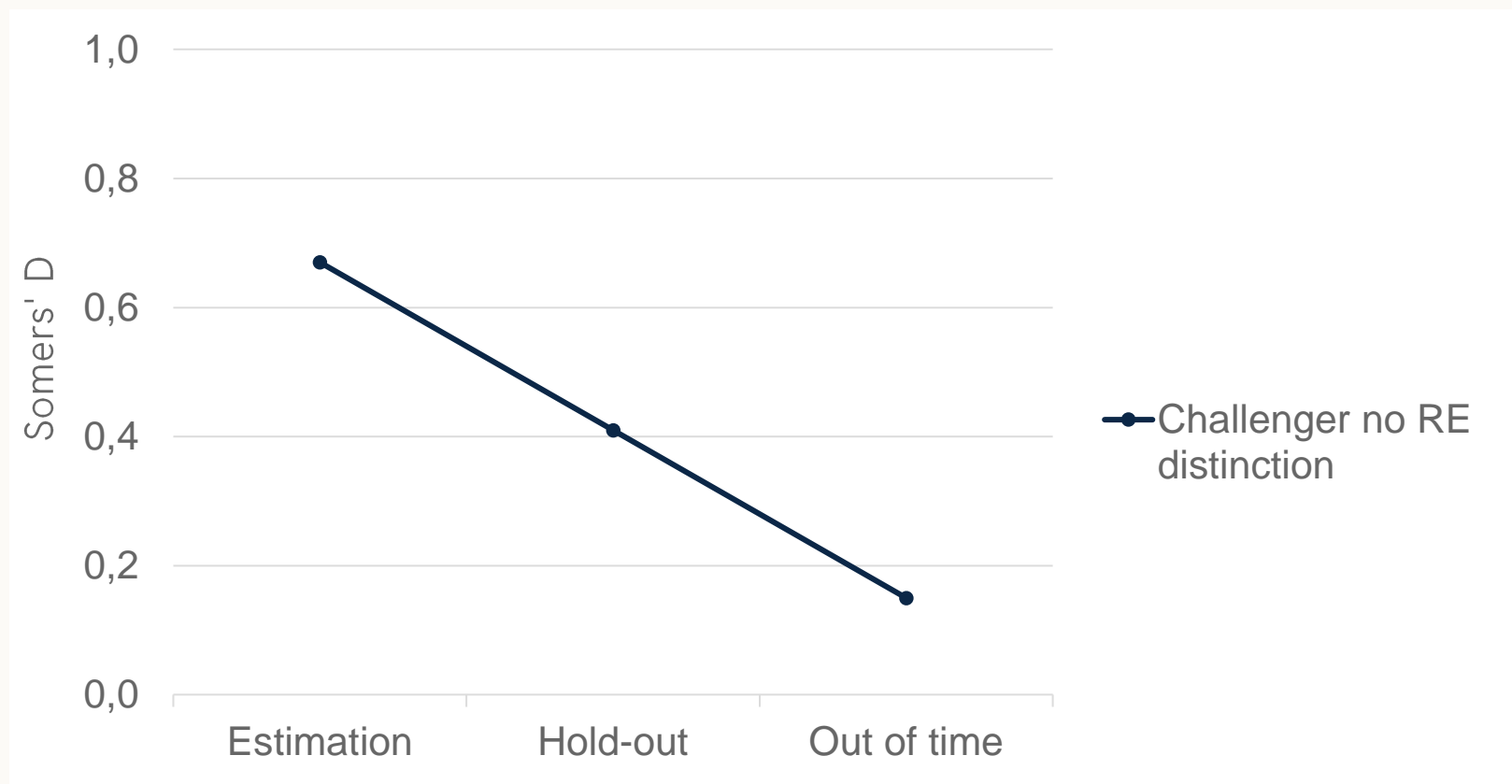


## Key Points

- Adding a cure phase increases performance strongly
- The “low” model addition adds predictive power in and out of sample and time
- The “high” model (which has less observations) adds marginal value

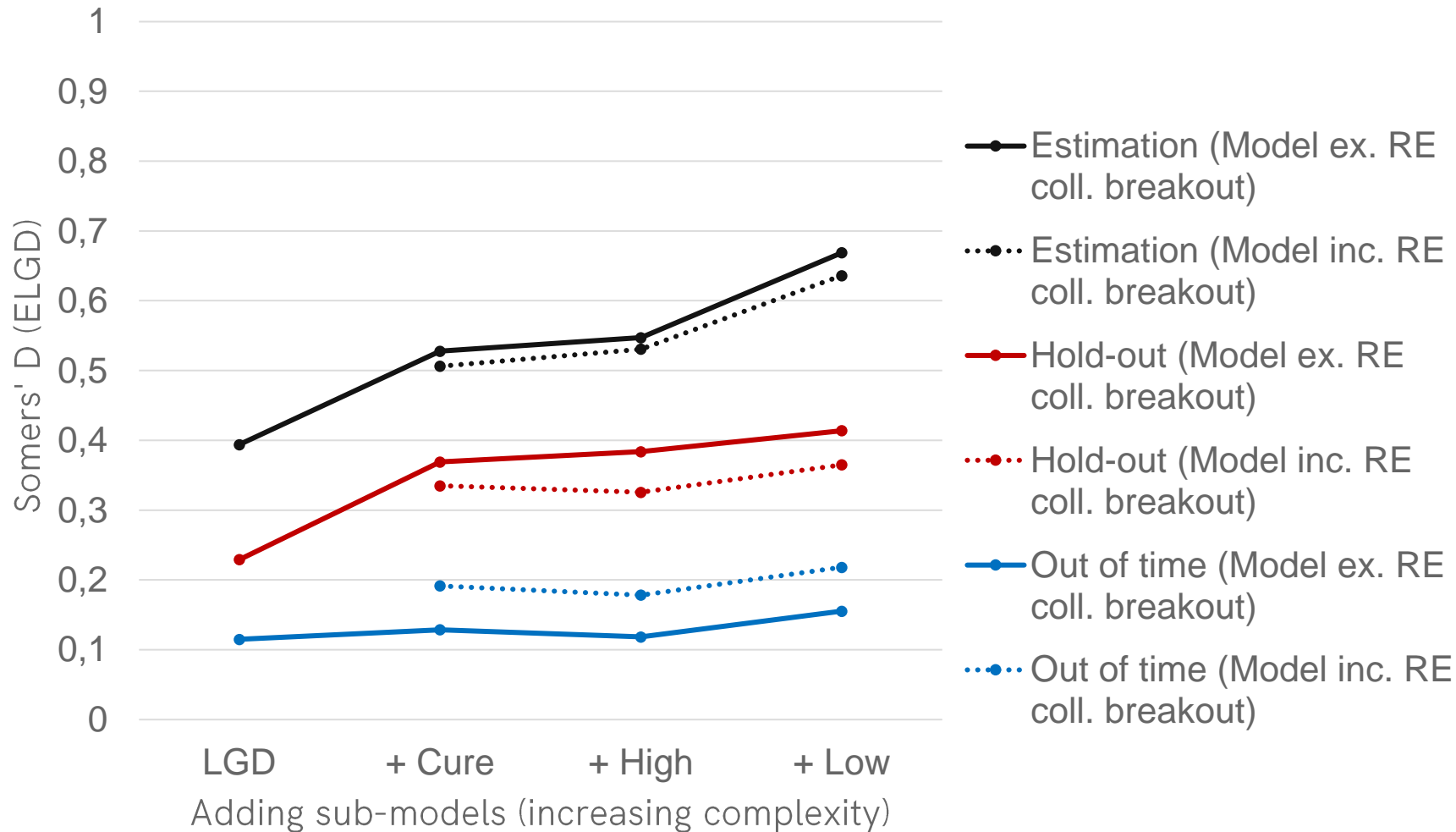
## Model Performance

What's the best performing challenger model over time?



## Key Points

- This challenger model details:
  - Including Cure, Low and High
  - 50+ risk drivers
- The levels of performance of the challenger model can be used as a benchmark to explain the ceiling of possible model performance in discussions with a regulator
- Estimation -> Hold-out/Out of time fall off in performance (overfitting tendency) a known challenge for ML
  - Can be reduced further by being more conservative hyperparameter optimization and reduce the number of risk drivers

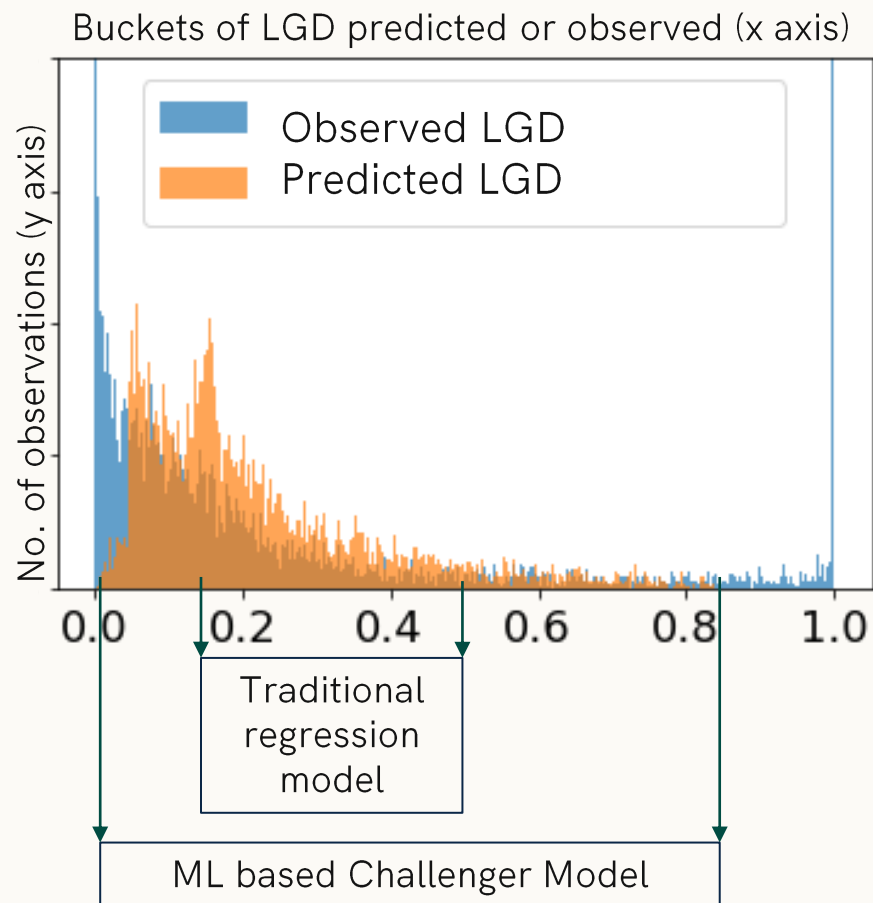


Our experience with the “real estate collateral heavy” sub-model group was mixed:

- Worse performance in estimation and hold-out
- but strong improvement in out of time.

## Model Performance

### Dispersion of Predicted LGD

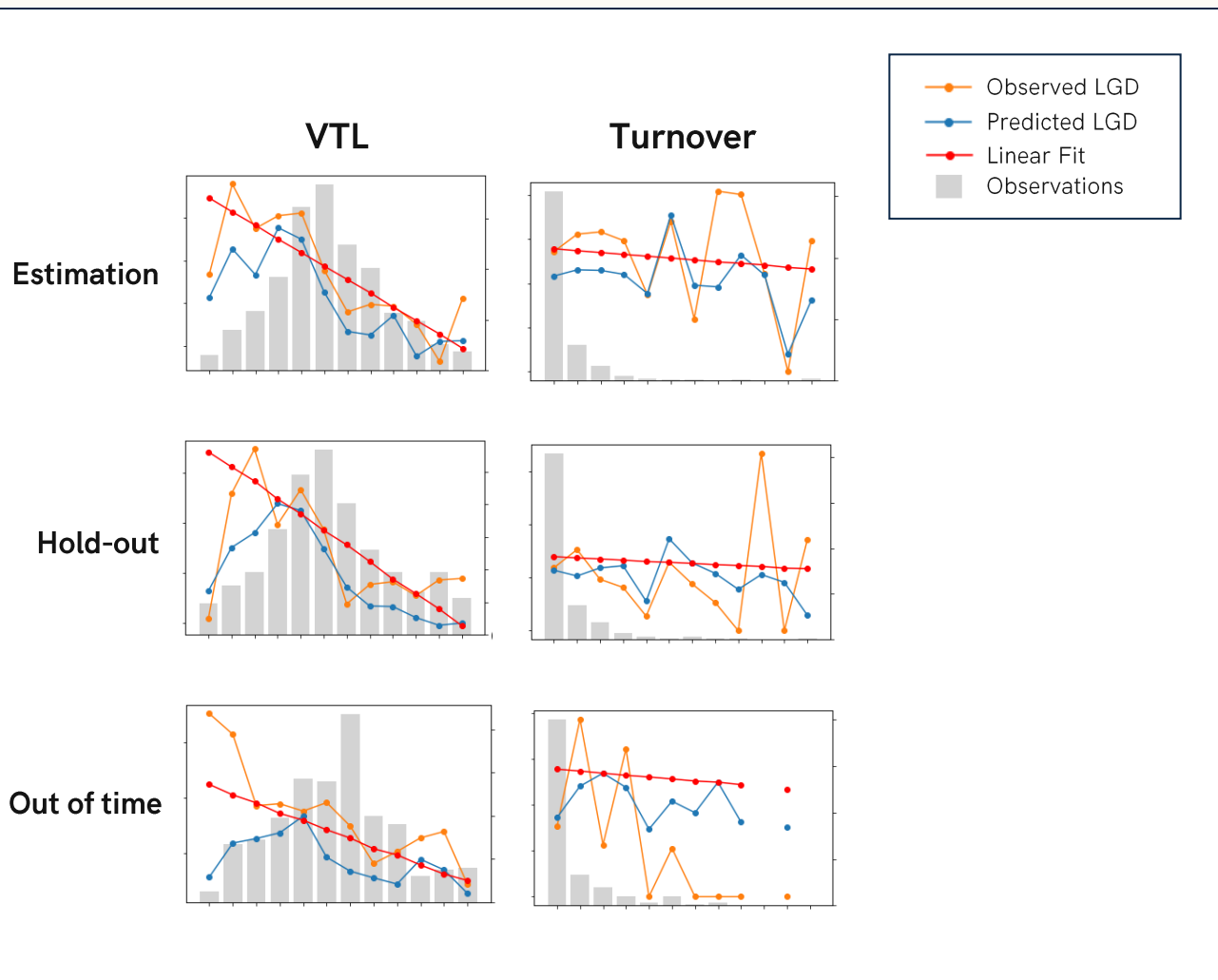


## Key Points

- The ML based challenger model distributes LGD predictions in a wider range than most traditional models, which is useful for business, pricing and helpful in discussions with regulators
- Dispersion is better for model structures with Low and High sub-models compared to those without any of these components
- Larger number of drivers and no linearity assumptions also assist the dispersion

## Model Performance

### Non-Linear Relationships



## Key Points

- In most cases the specific drivers chosen by the ML method perform as we intuitively expect (which is most often linear)
- Due to using ML, there can exist a noticeable break (non-linearity) in the relationship between driver and predicted LGD. A relationship difficult to model using linear models
- The figures are an example of non-linear relationships in a risk driver. As can be seen, the ML model fit tightly to the estimation sample. Note how the behavior in the observed LGD differs in the low VTL buckets between the estimation and out of time samples
- The modeller and risk experts can assess whether the non-linear relationships are real factors of the population or just artefacts from over-fitting (an important distinction)



## ML Techniques and Data Transformations

### Best performing ML techniques

- Classifiers: Brier-score, AUC, AUCPR are considered in ranking

Classifier ML Technique	Overall	Cure (non-RE)	High (non-RE)	Low (non-RE)	Cure (RE)	High (RE)	Low (RE)
xgbclassifier	1	1	2	1	2	1	2
histgradientboostingclassifier	2	3	1	2	4	2	4
randomforestclassifier	3	6	3	4	1	4	3
xgbrfclassifier	4	2	4	3	2	5	6
extratreesclassifier	5	4	4	6	8	2	1
sgdclassifier	6	7	6	5	5	7	5
ridgeclassifier	7	5	8	7	6	8	7
mlpclassifier	8	8	6	8	7	6	8

- Regressors: MAE, RMSE and Somers'D are considered in ranking

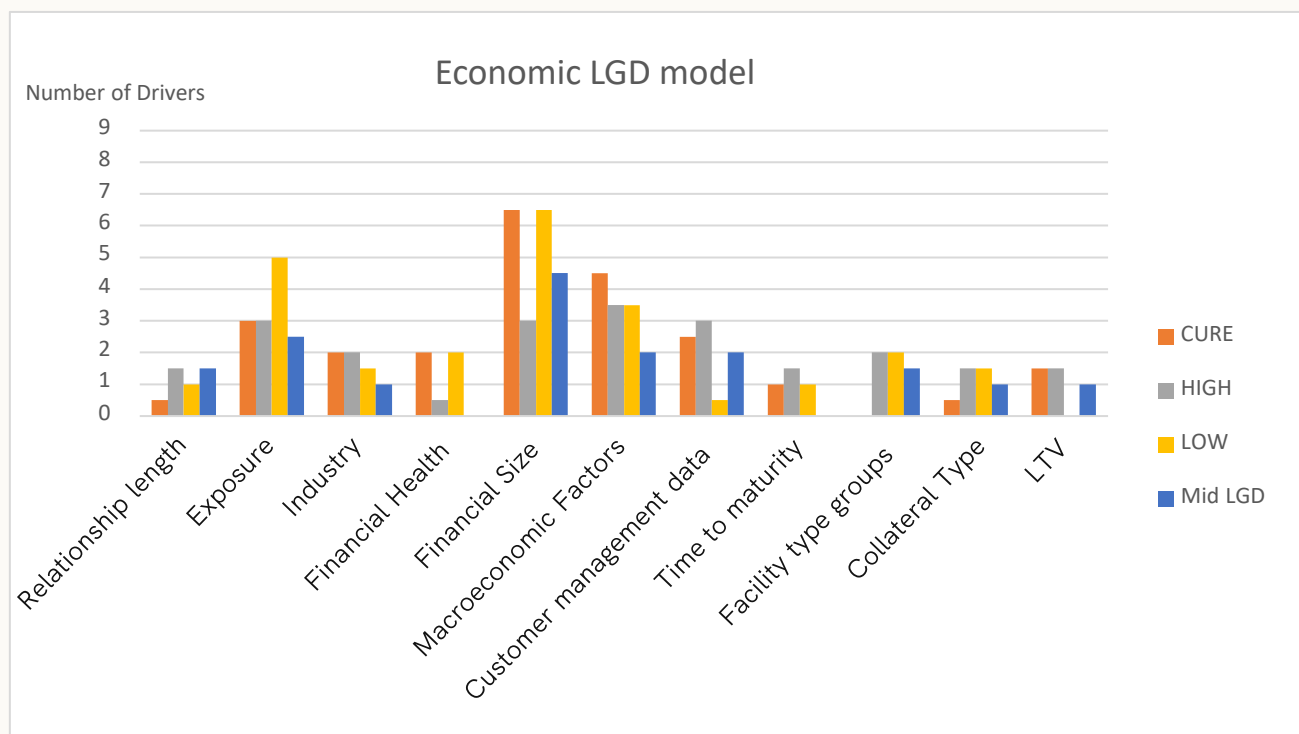
Regressor ML Technique	Overall	Error		Dispersion	
		LGD (non-re)	LGD (RE)	LGD (non-re)	LGD (RE)
extratreesregressor	2	1	2	6	5
histgradientboostingregressor	4	3	1	7	4
randomforestregressor	5	2	3	4	6
mlpregressor	6	6	7	3	2
xgbrfregressor	3	4	5	2	3
xgbregressor	1	5	4	1	1
sgdregressor	7	7	8	5	8
ridgeregressor	8	8	6	8	7

## Deep Dive into Sub-models

### Risk Driver Fingerprint

Overall, the challenger (ELGD) model uses 50+ different drivers. To help understanding we counted the number of drivers per category.

The comparison should mainly focus on the difference between the sub-models, not the absolute number of drivers in each category.



**Each one of the sub-models has a different set of drivers.**

### Cure:

- Relatively strong use of financials size drivers.
- Little influence from facility type groups

### High:

- Uses drivers from all categories.

### Low:

- Relatively strong use of exposure and financials size drivers.

### Mid LGD:

- Uses drivers from all categories.

### NLGD vs ELGD

- LTV, macroeconomic and financials health are stronger on NLGD than ELGD.
- Collateral type is slightly stronger on ELGD than NLGD. A possible reason is that different collaterals take longer or shorter time to realize.

## Sensitivity to Number of Drivers

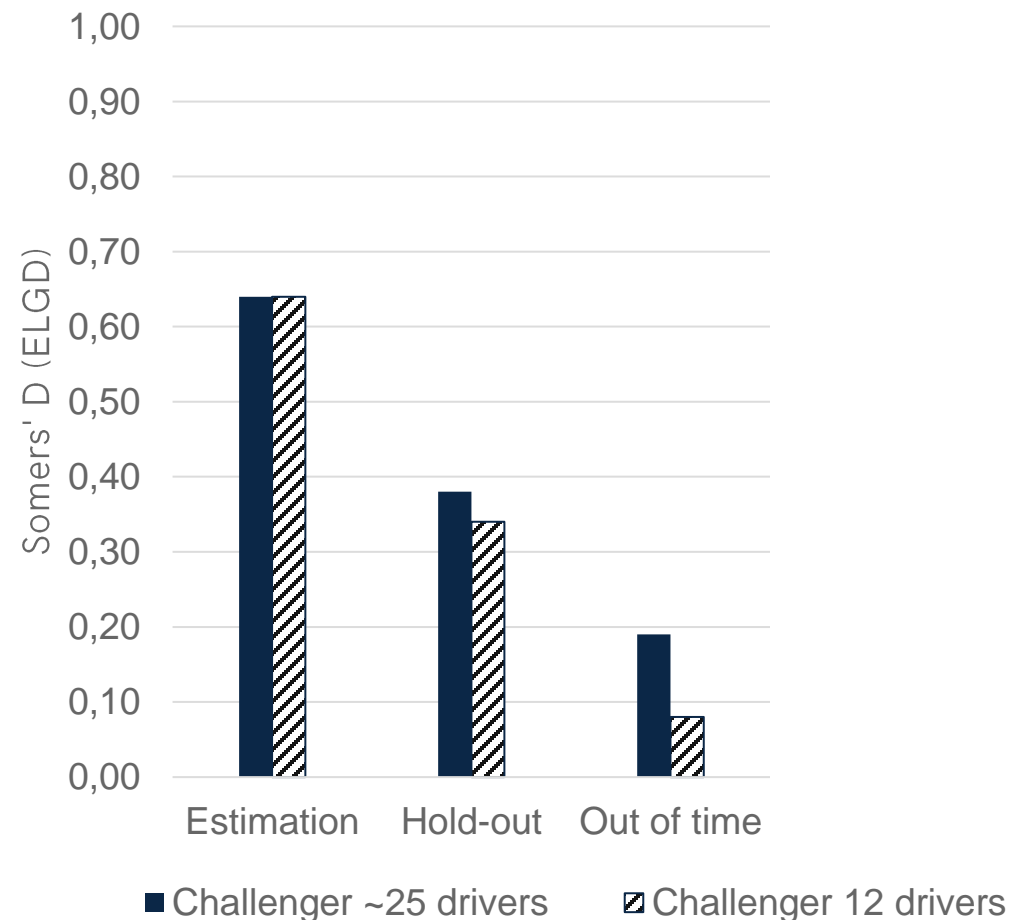
### Impact from reducing the number of risk drivers

Reducing the number of risk drivers is a robustness check of the model, fewer drivers = less information

The best model was ultimately reduced to 12 drivers per sub-model

#### Economic LGD (ELGD):

- Estimation and Hold-out only slightly affected by the drop in drivers
- Most significant advantage from more drivers in out of time
- Drop in performance:
  - Estimation: 0%
  - Hold-out: -11%
  - Out of time: -58%



*Possibilities*

- + Produces a theoretical “ceiling” of best performance (good benchmarking)
- + Chooses risk drivers objectively, giving comparison with expert models
- + Increasing the number of drivers above 8-10 drivers adds predictiveness and stability
- + Adding model complexity can help predict the bimodality of LGD
- + Predicts non-linear relationships, if they exist (see VTL Estimation example in previous slide 12)
- + Objective production of “best model” to test new or improved data

*Limitations*

- Overfitting  
(see VTL Out of time example in previous slide 12)
- Lower transparency: it’s not a black box, rather a grey box
- Regulatory approval of ML lacks precedent  
(e.g. difficult to quantify model risk margin of conservatism)

# Appendix 1: Benchmarking Tool on GCD Data Set

## Tool Overview



### Use of framework for investigating 10,000x different LGD models

- No installations, no integration, just deliver data – return output
- Combining different data transformations, model structures, ML techniques, risk drivers
- Deliveries focus on performance comparison between combinations



### Pick best performing models

- Top 5 best performing models along:
    - Rank correlation (Somers' D, Spearman)
    - Error (MAE, RMSE)
    - (Classifiers sub-models optimized on Brier score and also evaluated on AUC)
    - LGD prediction distribution (max-min diff)
    - Test samples (Hold-out, Out of time)
- = near-theoretical maximum predictive performance a model could produce on this data set, a performance ceiling



### Time-stability analysis

- More robust time stability analysis across multiple out of time samples or specific time periods within hold-out sample
- Top 5 best performing models along all metrics are “stressed” by being re-trained (with identical parameters and drivers) on a redefined development and out of time sets

## Tool Overview



### Leverage your *Global Credit Data* membership with machine learning

- Quick-start way to use GCD data (total data set or own bank subset).
  - For banks already having data in GCD's standard format (best practice for data), this adds a companion "off the shelf" model which builds on GCD member discussion of best practice.
- Provides much deeper risk and modelling insights directly from anonymized internal data of member banks - *an ideal "quick benchmark" tool for IRB banks*, without the need of significant resources
- Promotes state-of-the-art technology for your modelling capabilities



### Machine learning for IRB models

- Increased attention from regulators (e.g. EBA discussion paper on machine learning)
- The time is right to start making more informed decisions on the future use of ML in the organization's risk modelling - we can help with that
- Possible future use in reg cap models, driven by the capital relief from superior risk models

### Multiple application areas other than IRB

- IFRS9
- Stress testing
- ICAP models
- Pricing models
- Reg cap *challenger* model
- Validation



### Model performance boost

- Quantifiable results, for example:
  - Improved rank correlation +30%
  - Decreased model error -10%
- Improvements to champion model (or even adoption of some parts of challenger model), for example:
  - +60% rank correlation adding a *Cure* component
  - +10% rank correlation adding a *Low* component





## Your input data



1

### Use of framework for investigating 10,000x different LGD models

- No installations, no integration, just deliver data – return output
  - Built in Python, computing power by AWS
- It's "brute force" investigation of all possible models generated from the combinations of (with the weakest performing ones sorted out):
- ~200 data transformation combinations
  - ~100 model structures with 2-8 sub-models
  - ~10 machine learning algorithms / sub-model
  - Iterative key driver selection process using three different feature importance evaluations (replacing traditional SFA analysis)

#### Deliverables from stage 1:

Fully detailed output and high:

- Performance comparison of alternative model structures
- Performance comparison of key drivers by segment and overall
- Performance comparison of data transformations
- Performance comparison to a linear benchmark model

#### Possible custom add-ons:

- Your own idea of model structure / sub-model
- Your own idea of data transformations
- Your own idea of performance metrics
- Your own idea of specific drivers



2

### Best performing models

Top 5 best performing models along:

- Rank correlation (Somers' D, Spearman)
  - Error (MAE, RMSE)
  - (Classifiers sub-models optimized on Brier score and also evaluated on AUC)
  - LGD prediction distribution (max-min diff)
  - Test sample (Hold-out, Out of time)
  - (Custom metric)
- = near-theoretical maximum predictive performance a model could produce on this data set, a performance ceiling

A candidate model must fulfill:

- Sufficient minimum performance
- Sufficient stability between development and hold-out/out of time sets
- Better performance than linear benchmark

#### Deliverables from stage 2 and 3:

- Predictive performance and stability analysis
- Driver analysis
- Model behavior analysis (why it predicts as it does)
- Conclusions and suggestions, drawn by the results, by FCG experts
- Access to best performing models (pre-trained, no code)

#### Possible custom add-ons:

- Full model documentation along one/several of the best performing models (deep-dive)



3

### Time-stability analysis

Top 5 best performing models along all metrics are "stressed" by being re-trained (with identical parameters from stage 2) on redefined development and out of time sets. Redefined new out of time sets are:

- Consecutive 2-year intervals
- Consecutive 3-year intervals
- Known downturn periods

With existing pre-trained model: deepen analysis on specific time periods that might lie within hold-out sample

# Appendix 2: about FCG

## FCG

is a leading governance, risk and compliance firm, offering best-in-class services and tech solutions to the European financial industry.

We help navigate a changing and complex regulatory environment, supporting our clients in every step from analysis and advice to implementation and outsourcing



## Founded 2008

FCG was founded in 2008 in Stockholm and has grown to become the leading Nordic advisor to businesses in Europe, having supported +700 clients of various size and business models.



## Experience

We are a Governance, Risk & Compliance Advisory / Services & Technology firm offering standard- or customized solutions depending on the client needs. FCG has a profound understanding of the challenges that our clients meet.



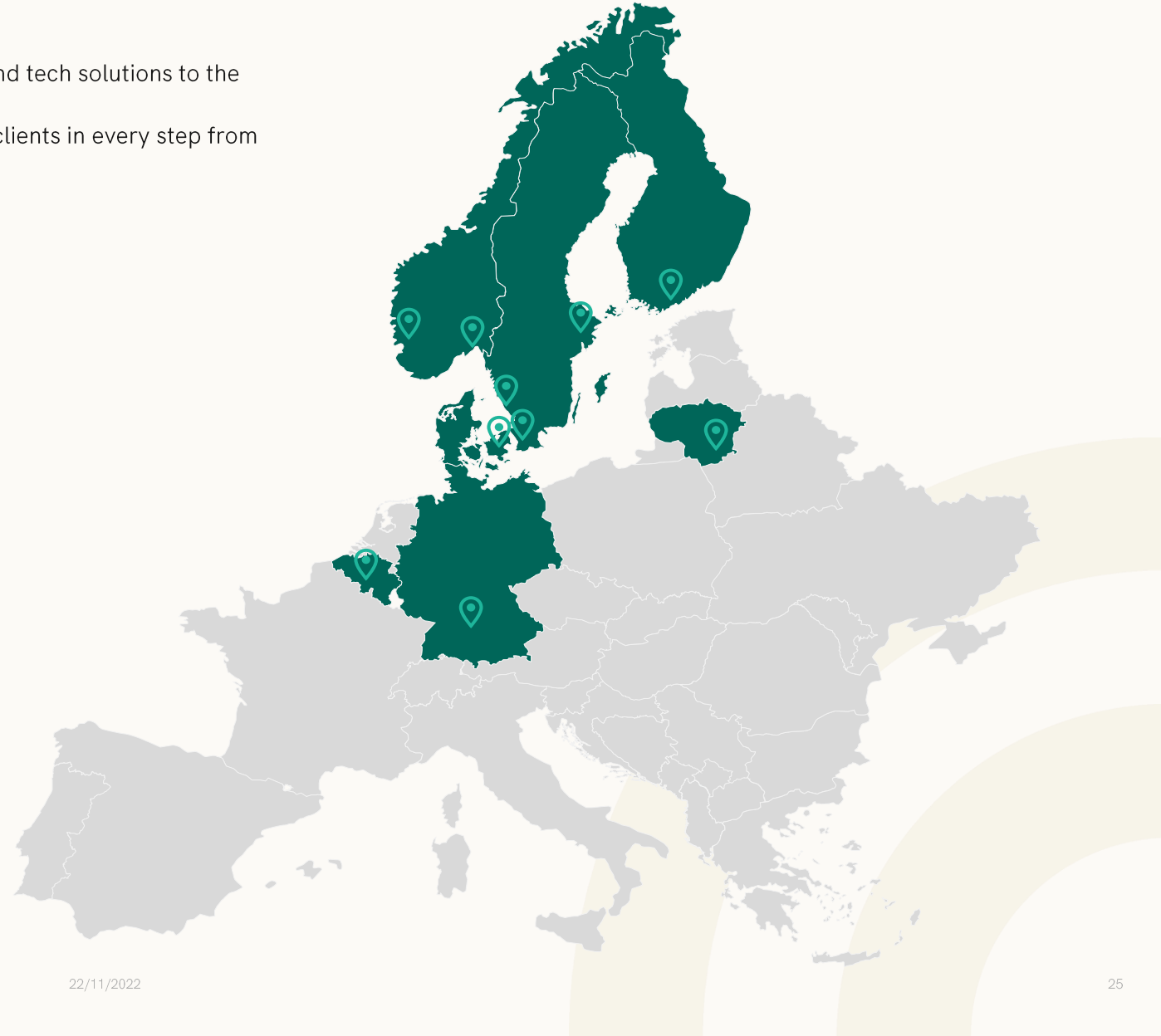
## +450 Employees

FCG has more than 450 employees and grows continuously.



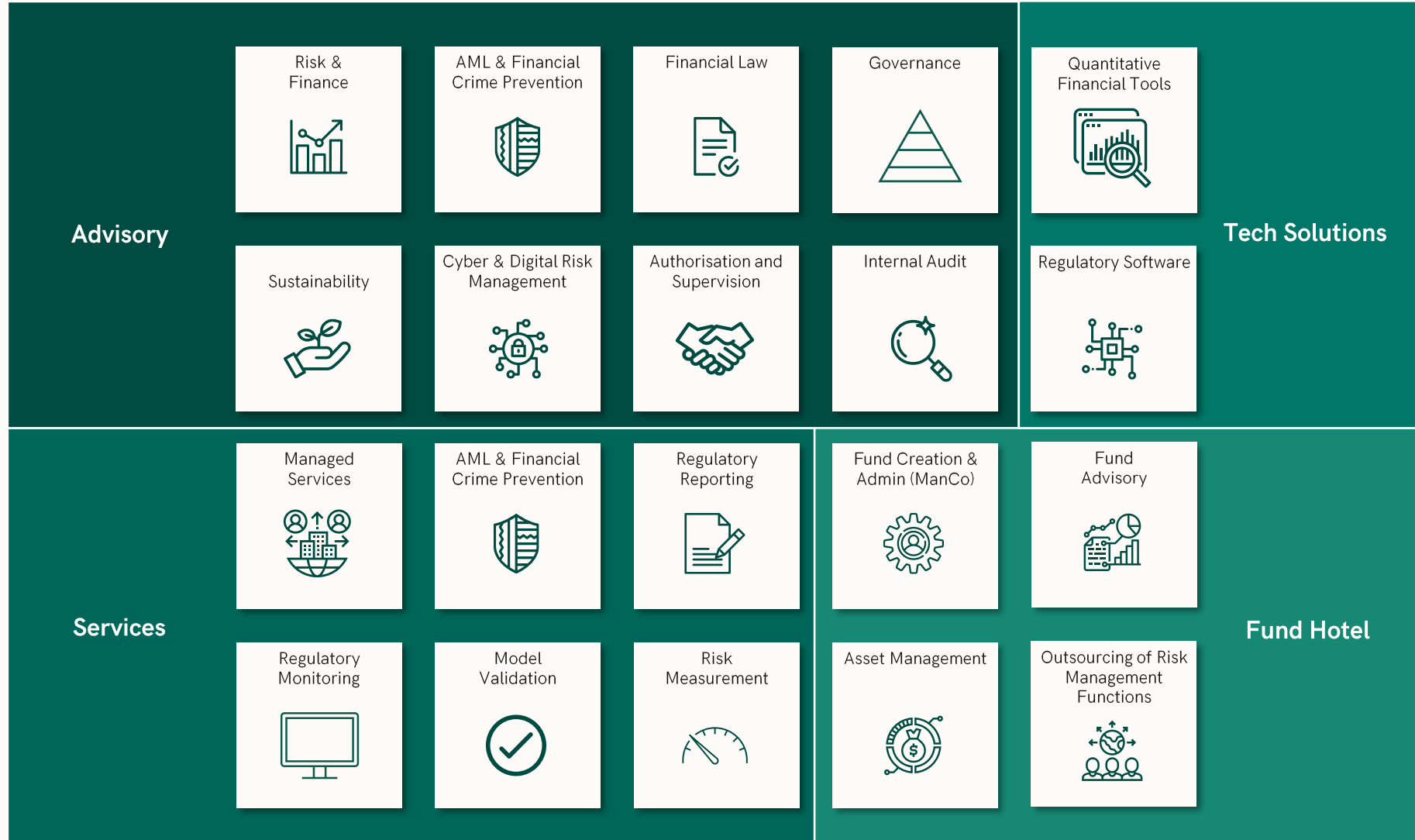
## 10 Locations

Headquarter in Stockholm and offices in Gothenburg, Malmo, Copenhagen, Oslo, Bergen, Helsinki, Frankfurt, Brussels and Vilnius.



# One-stop-shop for mission-critical services across the entire GRC spectrum

*FCG has a unique position as a one-stop-shop for GRC solutions and has become the go-to firm for the Nordic financial sector:*





## Risk Expert

Jimmi Brink

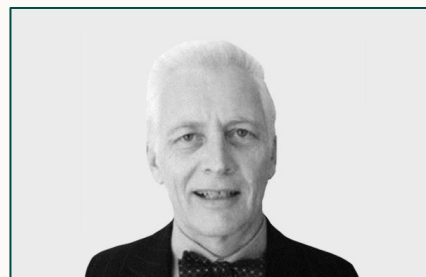
Partner

+46 73 253 11 19

✉ jimmi.brink@fcg.se

### Competence:

Former regulator and banker with 20+ year of experience within risk management. Extensive know how covering risk strategy & steering as well as more technical aspects of financial risk modeling including IRB and IFRS9.



## LGD Expert

Philip Winckle

Senior Adviser

+46 70 784 7654

✉ Philip@winckle.com

### Competence:

Many years experience in bank lending, credit approval, loan workout and credit risk modelling. Top level risk management experience including Basel models, risk appetite and stress testing. Former CEO of industry group The Global Credit Data Consortium.



## Risk Expert

Jonas Ljungqvist

Partner & Head of FCG Germany

+49 174 152 42 27

✉ jonas.ljungqvist@fcg-international.com

### Competence:

20+ years experience from senior positions in Risk Control and Risk Management in banks with significant hands-on credit risk modelling and validation experience from IFRS9 and IRB.



## Model & Architecture Development

Thomas Aldheimer

Data Scientist

+46 72 179 49 79

✉ thomas.aldheimer@fcg.se

### Competence:

Several years' experience with machine learning in credit risk modelling. Other areas of machine learning expertise includes natural language processing, regulatory monitoring & horizon scanning, AML, payments performance, fraud and user behavior analytics. Academic background from Accounting and Financial Management (M.Sc.), Astro physics (B.Sc.) and Mathematics (B.Sc.).



## Model Development

Christoffer Eduards

Senior Associate

+46 73 230 26 56

✉ christoffer.eduards@fcg.se

### Competence:

Credit risk experience including development and validation of models (IRB & IFRS9) encompassing all risk parameters, retail/non-retail and domestic/global portfolios. Background within industrial engineering, financial mathematics and computer science.