



# Global Credit Data

*by banks for banks*

## Unresolved LGD Methodology and its application in long-run average calculation



Nina Brumma & Ben Galow (GCD),  
Mohamed Sirfouq (HSBC)



# Agenda

- ❑ Why is the LGD for unresolved defaults important?
- ❑ GCD's unresolved LGD methodology – an overview
- ❑ The unresolved LGD methodology
  - Results
  - Backtesting results
  - Monitoring results
- ❑ How to use it – RR Dashboards 2022
- ❑ Outlook
- ❑ Appendix



[www.globalcreditdata.org/library/unresolved-defaults-lgd-study-2020](http://www.globalcreditdata.org/library/unresolved-defaults-lgd-study-2020)

# Why is an LGD for unresolved defaults important?

## Regulatory compliance

Incorporation of unresolved defaults required (e.g. European EBA GL 2017/16 or US BCC Bulletin 13-5)



## Crises

Increased number of defaults after Covid19 & energy crisis



## Overcoming the resolution bias

avoiding a resolution bias when using most recent default cases

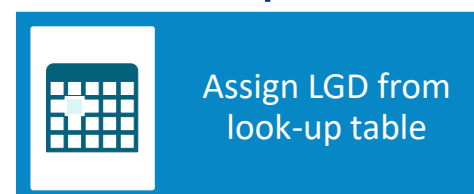
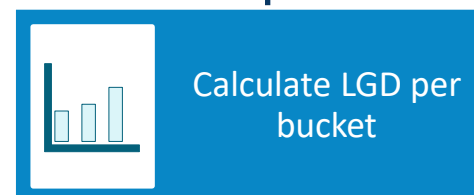
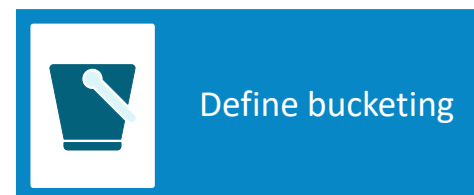
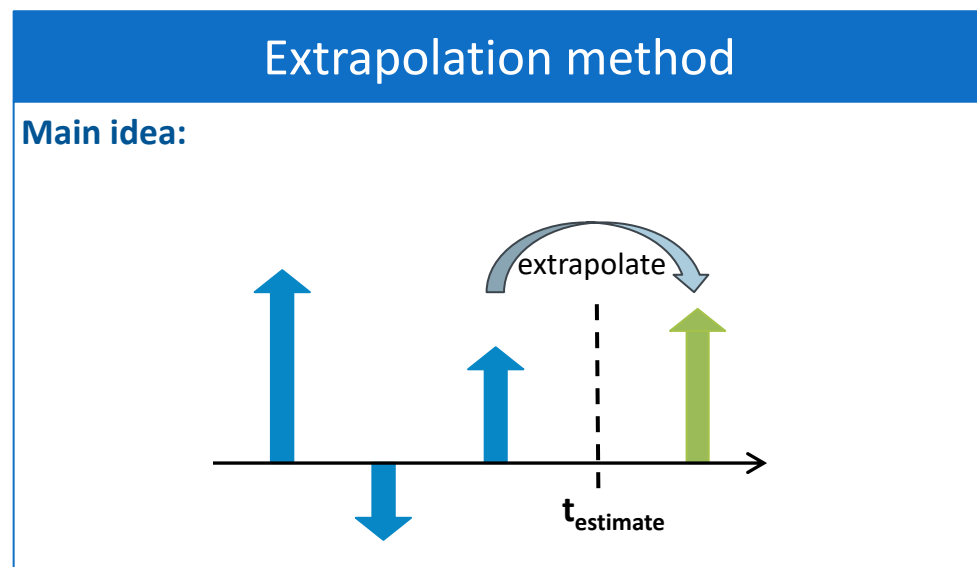


## NPL pricing

Unresolved LGD estimates can be used for pricing of non-performing loans




# Overview of GCD's unresolved LGD methodology



**Determination of parametrization** on RDS based on resolved cases and unresolved cases exceeding the maximum recovery period

**Assign unresolved LGDs** to individual loans via time-in-default & already recovered amount

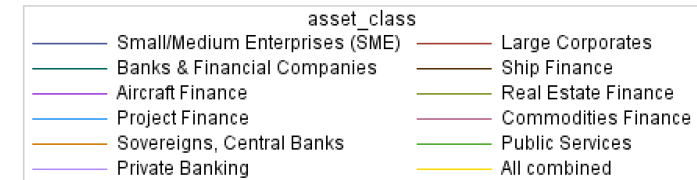
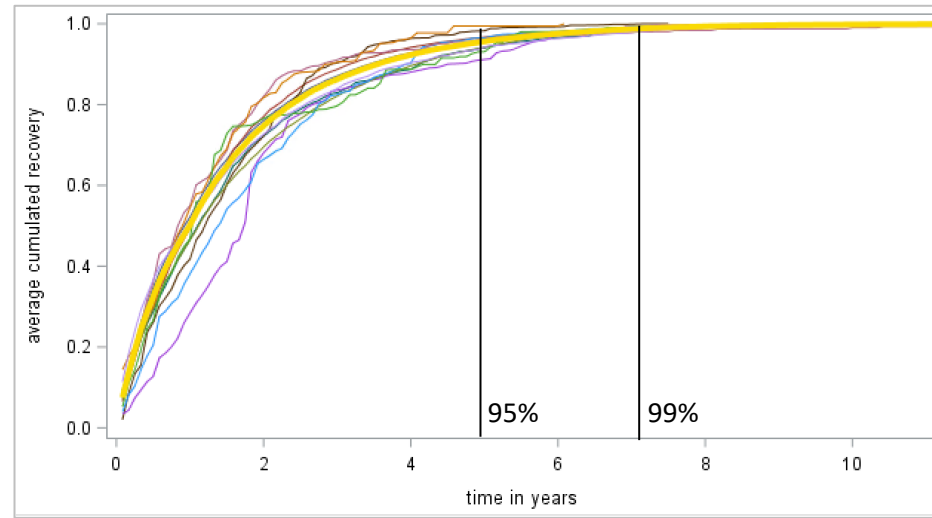
 The methodology has been approved by the GCD Methodology Committee (MethCom)

# Maximum recovery period

## 7 years

Maximum Recovery Period used in calibration

	Maximum Recovery Period [in years]	Stdev [in %]
95% Percentile	4.8	20.2
99% Percentile	7.3	9.2



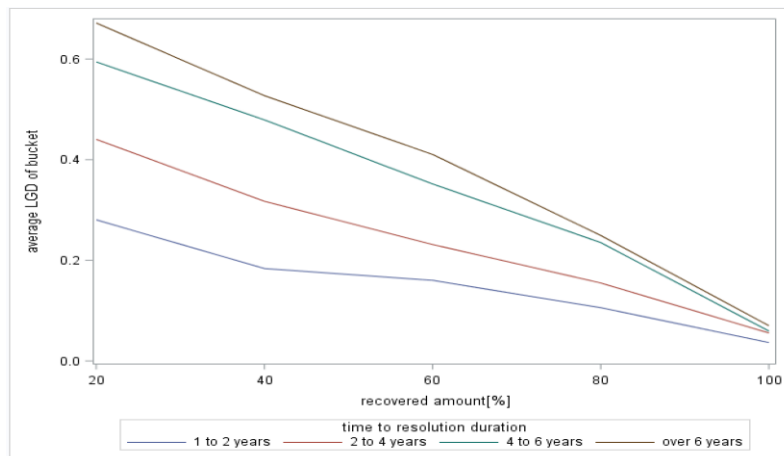
## Rationale

- ❑ Different Percentiles have been tested: The full year close to 99% has been chosen in order to be in the flat region of cumulated recovery curves (cf. European Central Bank guide)
- ❑ Facility Asset Class does not appear to be a strong driver for the maximum recovery period (see graph)
- ❑ The parameter's impact on the outcome is relatively small as it is only used for the cut-off for the development sample

# Structure of parametrization

Parametrization Example Large Corporates

time to resolution [years]	already recovered amount	average LGD of bucket
1 - 2	0 - 20%	39.3%
1 - 2	20 - 40%	30.1%
1 - 2	40 - 60%	23.4%
1 - 2	60 - 80%	16.0%
1 - 2	80 - 100%	4.7%
2 - 4	0 - 20%	49.6%
2 - 4	20 - 40%	37.5%
...	...	...



- ❑ Besides facility asset class, two main dimensions/ risk drivers are used:
  - With increasing **time to resolution** the average LGD is higher
  - With increasing **received recovery amount** the realized average LGD is lower.
- ❑ The received recovery amount is determined at the left boundary of the respective time-to-resolution bucket
- ❑ The **average LGD of bucket** is determined by averaging over all observations with time-to-resolution  $\geq$  the time of the current bucket and already recovered amount  $\geq$  the amount of the current bucket

# Assignment of LGD values for unresolved defaults

LGD calculated with extrapolation method (the loan is in default for less than 7 years)

Results				
Loan	Default Date	Default amount	Recovery Rate	LGD
A	04-06-2017	5,000,000	<b>0.6</b>	<b>0.4</b>

Loan	Default Date	Default amount	Time in Default Unresolved [days]
Loan A	04/06/2017	5,000,000	1,122

Transaction date	Transaction amount
01/04/2018	€ 1,000,000
01/03/2019	€ 250,000

Facility asset class	TTRes bucket	RR bucket	RR
...	...	...	...
<b>2</b>	<b>[720;1440[</b>	<b>[20;40[</b>	<b>0.6</b>
...	...	...	...

Bucket structure makes use of the LGD drivers:

- a. **time-in-default;**
- b. **recovered amount;**

- ❑ Loan A three years in default → **TTRes bucket [720;1440[**
- ❑ Recovery received 25% → **RR bucket [20;40[**

# Facility asset class grouping and results

Grouped asset class	avg LGD resolved	avg LGD unresolved with model
Small/Medium Enterprises (SME)	23%	<b>40.1%</b>
Large Corporates	23.8%	<b>41.1%</b>
Banks. Sovereigns. Public Services	25.2%	<b>37.7%</b>
Spezialized Lendings	16.9%	<b>29.9%</b>
Real Estate Finance	22.5%	<b>40.2%</b>
Private Banking	19.3%	<b>44.1%</b>

- ❑ Facility asset classes have been consolidated for robustness (Spezialized Lending: Ship Finance, Aircraft Finance, Commodity Finance and Project Finance)
- ❑ The values for the unresolved cases are conservative particularly due to long time-in-default durations that result in higher prognosed LGDs



# Backtesting & Monitoring

## Backtesting

During model development forecasted values have been compared with realized values



## Monitoring

On an annual basis the performance of the model is monitored



# Overview backtesting results



## Backtesting methods

- ❑ On portfolio level: Calibration of forecasted average LGD.  
*“Does the forecasted average LGD match the realized average LGD?”*
- ❑ On individual case level: Comparison of forecasted and realized LGD via. e.g. via Kernel Density plots  
*“Does the forecasted individual LGD match the realized LGD?”*

### In-Sample Backtesting

PASSED

- ❑ Use all cases which have been in default in. e.g. 2012. and resolved later. i.e. a final resolved LGD is available
- ❑ Throw away all cash flow data after the observation time point (e.g. 31.12.2012) & calculate LGD using the method
- ❑ Perform Backtesting
- ❑ Repeat for several

### Out-of-Sample Backtesting

PASSED

- ❑ Omit a sample part for method development
- ❑ Perform backtesting on the omitted sample

# Annual monitoring (on 2021 data\*)



The monitoring metrics are structured along EBA's dimensions:

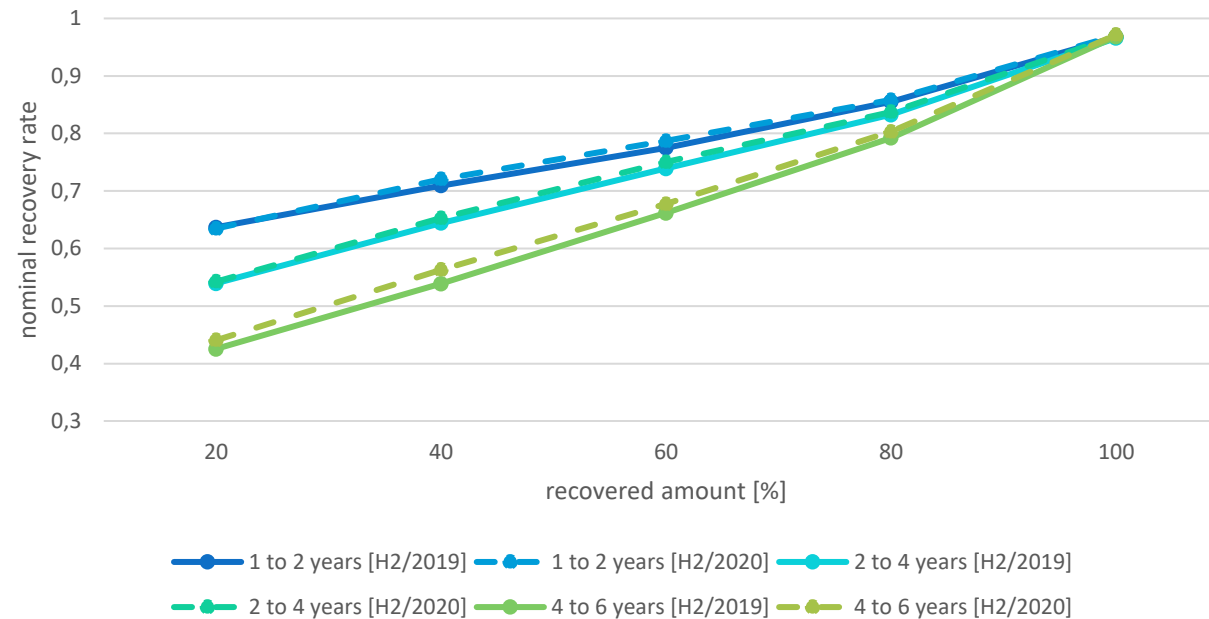
- ❑ risk classification
  - *“Do the risk drivers time-to-resolution and already-recovered-amount still show the same behavior?”*
- ❑ risk calibration
  - *“Is the methodology calibrated correctly?”*
- ❑ accuracy of risk projection
  - *“Is the distribution of the projected UR LGDs based on 2020 data similar with the distribution of UR LGDs based on 2021 data?”*

\*The validation with 2022 data is under preparation.

# Annual monitoring – risk classification



- “Do the risk drivers time-to-resolution and already-recovered-amount still show the same behavior?”



▶ The risk drivers time-to-recovery and already-recovered-amount still show the same behavior based on the updated calibration.

# Annual monitoring – risk calibration



- ❑ “Does the parametrization change strongly compared with current parametrization?”

grouped asset class	avg LGD unres H2/2019 Param	avg LGD unres H2/2020 Param
Small/Medium Enterprises (SME)	37.86%	37.04%
Large Corporates	36.23%	35.65%
Banks. Sovereigns. Public Services	32.26%	31.82%
Spezialized Lendings	32.04%	31.09%
Real Estate Finance	38.70%	38.14%
Private Banking	33.75%	34.52%

- ❑ Further, out-of-sample backtesting on the H2/2020 data using the current calibration is successful for most asset classes and elsewhere the deviation can be explained.
- ❑ The current calibration also **did not show signs of instability** regarding **sensitivity analysis**.

▶ The current parametrization (based on H2/2019 data) is valid

# Annual monitoring – accuracy of risk projection



- “Is the distribution of the projected UR LGDs based on H2/2019 data similar with the distribution of UR LGDs based on H2/2020 data?”

grouped asset class	total loans unresolved in H2/2019	Hellinger distance Passed: < 0.25
Small/Medium Enterprises (SME)	1370	0.2097
Large Corporates	663	0.2581*
Banks. Sovereigns. Public Services	63	0.2174
Spezialized Lendings	216	0.2736*
Real Estate Finance	512	0.2202
Private Banking	276	0.2789*

\*The reason for not passing the test, is, that a significant portion of loans either experienced a shift to a higher time-after-default bucket, or, even exceeded the MRP and, hence, no further projected cashflows for the latter cases are assumed leading to a significantly more conservative LGD value.

▶ The **accuracy assessment is successful** for most asset classes and elsewhere **the deviation can be explained**.

# RR Dashboards 2022: Include unresolved defaults in time series graph



## Observed Recovery Rates Dashboard Corporates

April 2021

2021 version

### Banks recover 78% from Corporate defaults

#### Key Findings

- When corporates default lenders eventually recover on average almost 78% of the amount lent.
- Recovery Rates were 7% lower in the last downturn compared to non downturn times.
- It takes between 1 and 2 years to recover the cash flows.

#### Seniority and Collateral

Seniority and collateral are observed as affecting recoveries. Secured recovery is higher than unsecured, particularly where a strong (primary) collateral is held.

#### Region

GCD data has its strongest database in Europe and North America, which register similar figures. GCD members receive country level data for deeper analysis.

#### Industry

Granular industry groups are available to GCD member banks and are condensed here into key industry groupings. The differences are mainly explained by the availability of real estate collateral.

#### Recoveries and Losses in Crisis Times

The long timespan of the GCD database and the detailed cash flow data allow for dedicated analysis by year of default. Higher numbers of defaults and lower recoveries are observed during the financial crisis starting in 2008. The peak of defaults were in year 2009 which was the worst year of the crisis. As workouts take more than one year the 2008 defaults were being recovered in that worst year and hence show the lowest recovery. The final result of the defaults from the current Covid-19 economic crisis will be reported by banks in the years 2021 to 2025.

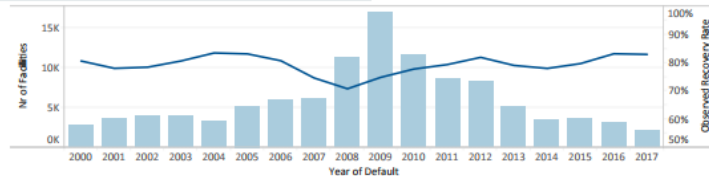
#### Note on Terms Used (see [Appendix](#) for more details)

Observed Recovery Rate refers to the historically observed nominal average recovery cash flows divided by outstanding at default.

Time to Peak Recovery is calculated as the center point of recovered cash flow.

Primary Collateral is Cash, Marketable Securities, Ship, Airplanes, Real Estate and Other Objects. Secondary Collateral is all other collaterals.

	Nr of Facilities	Observed Recovery Rate	Time to Peak Recovery
Grand Total	108,502	78%	1.4
Secured	Primary	84%	1.5
	Secondary	74%	1.3
	Total	79%	1.4
	Senior	76%	1.2
Unsecured	Subordinated	60%	1.7
	Other	77%	1.8
	Total	76%	1.3
Region	Africa & Middle East	87%	1.4
	Asia & Oceania	78%	0.8
	Europe	79%	1.5
	Latin America	71%	1.5
	North America	78%	1.2
	Other	62%	1.5
Industry	Agriculture	83%	1.2
	Communications	75%	1.4
	Construction	77%	1.5
	Hotels and Restaurants	79%	1.5
	Manufacturing	79%	1.2
	Mining	83%	1.0
	Real Estate	82%	1.5
	Social/Health Services	75%	1.7
	Other Services	76%	1.3
	Transportation	79%	1.2
	Utilities	80%	1.3
	Wholesale/Retail Trade	76%	1.2
	Other	78%	2.0
Downturn 2007-2009		73%	1.6
Non Downturn		80%	1.2



#### About Global Credit Data

Global Credit Data (GCD) is a non-profit association owned by 50+ member banks. GCD operates pooled databases on a "give to get" basis, meaning that members who supply high quality data and receive detailed data in return. The robustness of GCD's data collection infrastructure place the GCD databases as the global standard for credit risk data pooling.

#### About this report

This report summarizes Recovery Rates for Corporates, in which numerical evidence of recoveries and losses is presented. See other GCD Reports for Banks and Sovereigns.

#### Contact

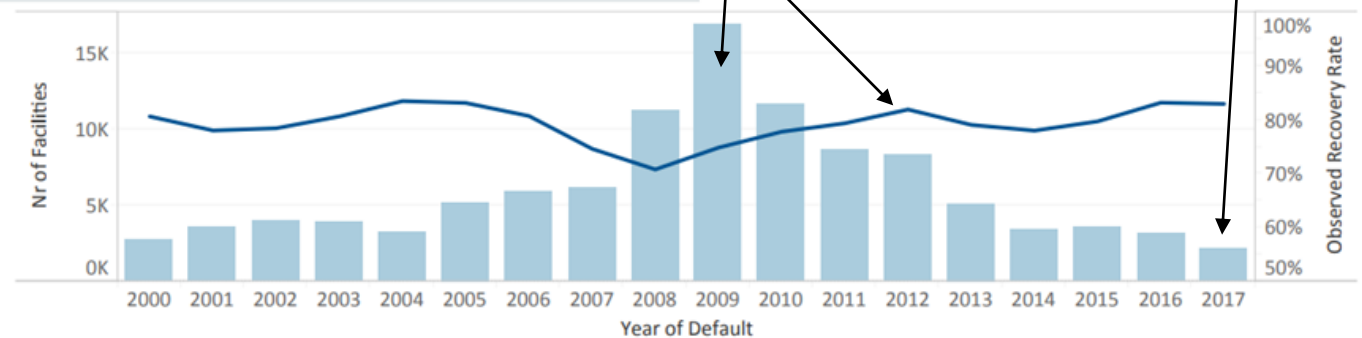
Nina Brumma  
Head of Analytics and Research  
[nina.brumma@globalcreditdata.org](mailto:nina.brumma@globalcreditdata.org)

Nunzia Rainone  
Analytics and Research  
[nunzia.rainone@globalcreditdata.org](mailto:nunzia.rainone@globalcreditdata.org)  
[www.globalcreditdata.org](http://www.globalcreditdata.org)

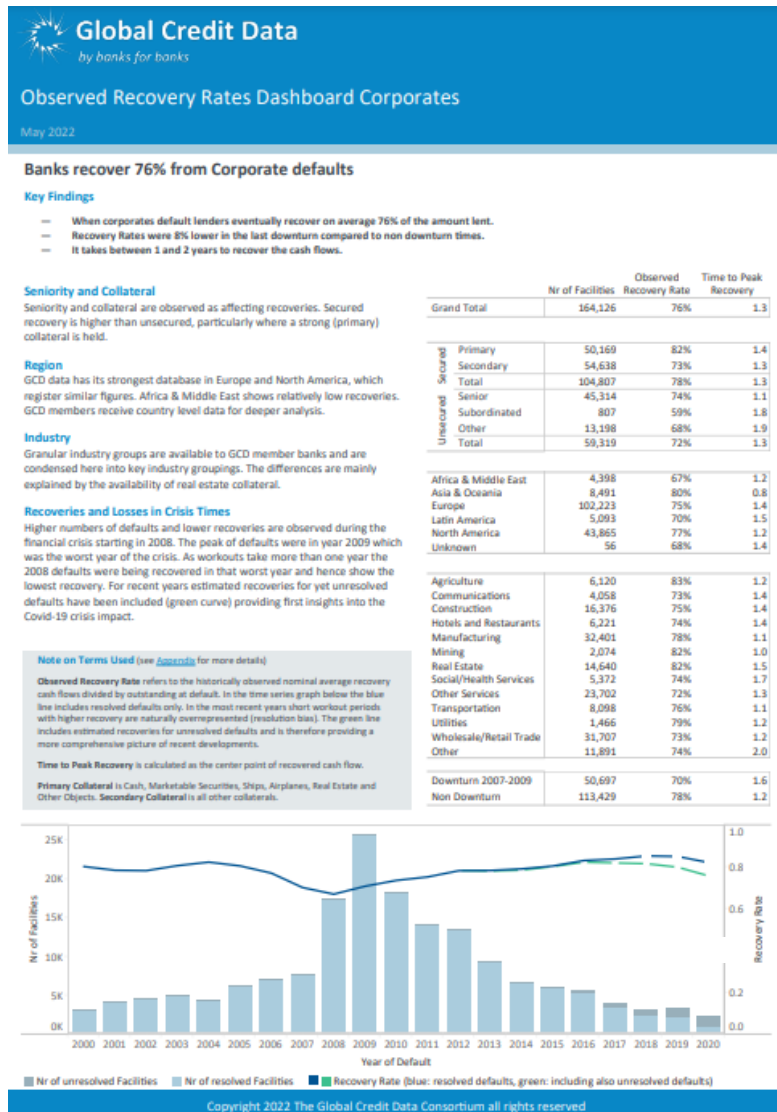
Copyright 2021 The Global Credit Data Consortium all rights reserved

Includes **only** resolved loans

**Stops** in 2017

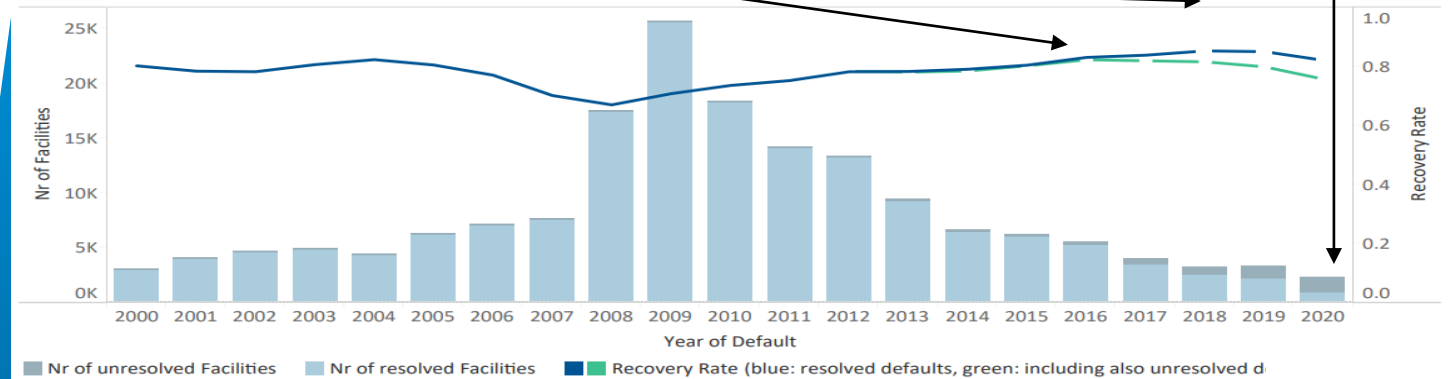


# RR Dashboards 2022: Include unresolved defaults in time series graph



Display resolved for recent years (dotted blue line) and resolved and unresolved loans (dotted green line)\*

Until 2020



## 2022 version change in graph

- ✓ include recent years (up to 2020)
- ✓ include unresolved defaults in recovery rates using the same methodology as in Unresolved LGD Study (separate line)\*  
[https://globalcreditdata.org/wp-content/uploads/public/unresolved\\_defaults\\_lgd\\_study-large\\_corporates.pdf](https://globalcreditdata.org/wp-content/uploads/public/unresolved_defaults_lgd_study-large_corporates.pdf)
- ✓ Explain/reference methodology in Appendix

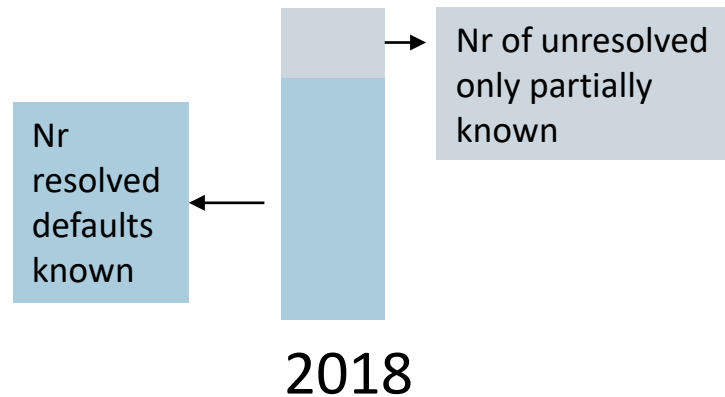
\*for segments with enough unresolved data points.





# RR Dashboards 2022: What we need to create the green line

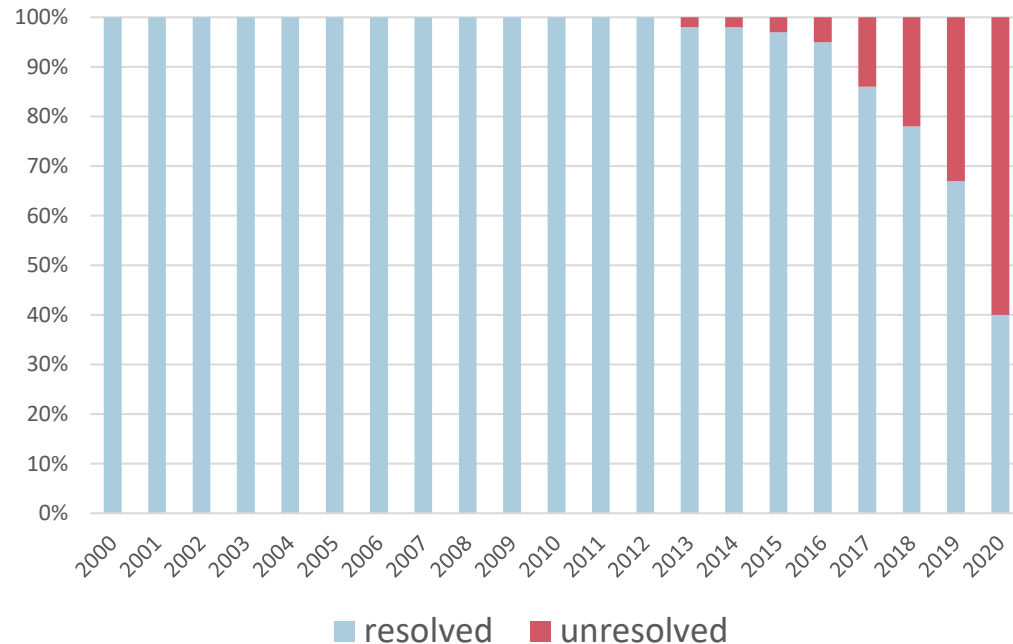
- ❑ LGD unresolved methodology (✓ approved by Methcom)
- ❑ Number of unresolved defaults by default year (✓ approved by Methcom)
  - Unresolved default information in GCD is incomplete as data submission is optional and not all banks provide unresolved. Situation slightly improved since aggregated unresolved data is collected.
- ❑ Update for H2 2021 data (for Dashboards)



## methodology

- ✓ Calculate completion rate based on subsample of unresolved defaults in database (unaggregated and aggregated)
- ✓ Use completion rate and number of resolved in total RDS to estimate number of unresolved

# RR Dashboards 2022:



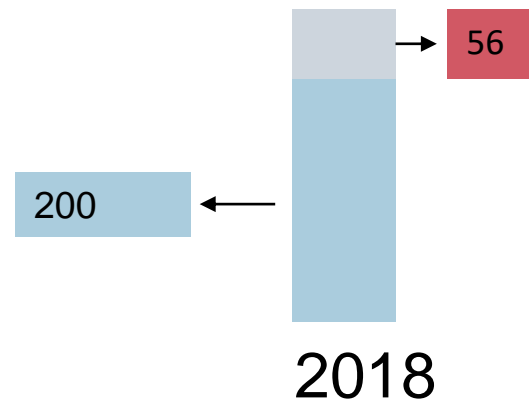
Year of Default	Completion Rate
2000	100%
2001	100%
2002	100%
2003	100%
2004	100%
2005	100%
2006	100%
2007	100%
2008	100%
2009	100%
2010	100%
2011	100%
2012	100%
2013	98%
2014	98%
2015	97%
2016	95%
2017	86%
2018	78%
2019	67%
2020	40%

## Completion rate results

- The completion rate is calculated on the overall database. No distinction between Facility Asset Classes or any other segmentation (not enough data).
- We have no evidence to suggest a different completion rate for a specific segment. However, for low default segments not enough unresolved data points are available.

# RR Dashboards 2022: Number of Unresolved defaults

- ❑ Example: Nr resolved in overall RDS 2018: 200
- ❑ completion rate is 78%
- ❑ Number of extrapolated unresolved for 2018:  $200 * 22\% / 78\% = 56$  defaults

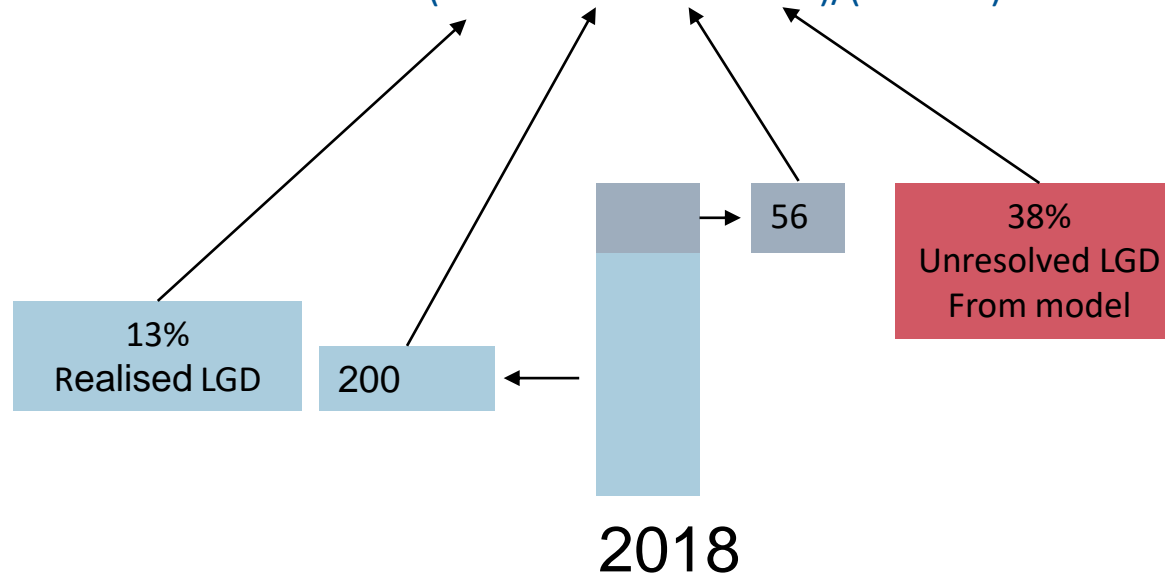


Year of Default	Completion Rate
2000	100%
2001	100%
2002	100%
2003	100%
2004	100%
2005	100%
2006	100%
2007	100%
2008	100%
2009	100%
2010	100%
2011	100%
2012	100%
2013	98%
2014	98%
2015	97%
2016	95%
2017	86%
2018	78%
2019	67%
2020	40%

# RR Dashboards 2022: LGD average including unresolved defaults

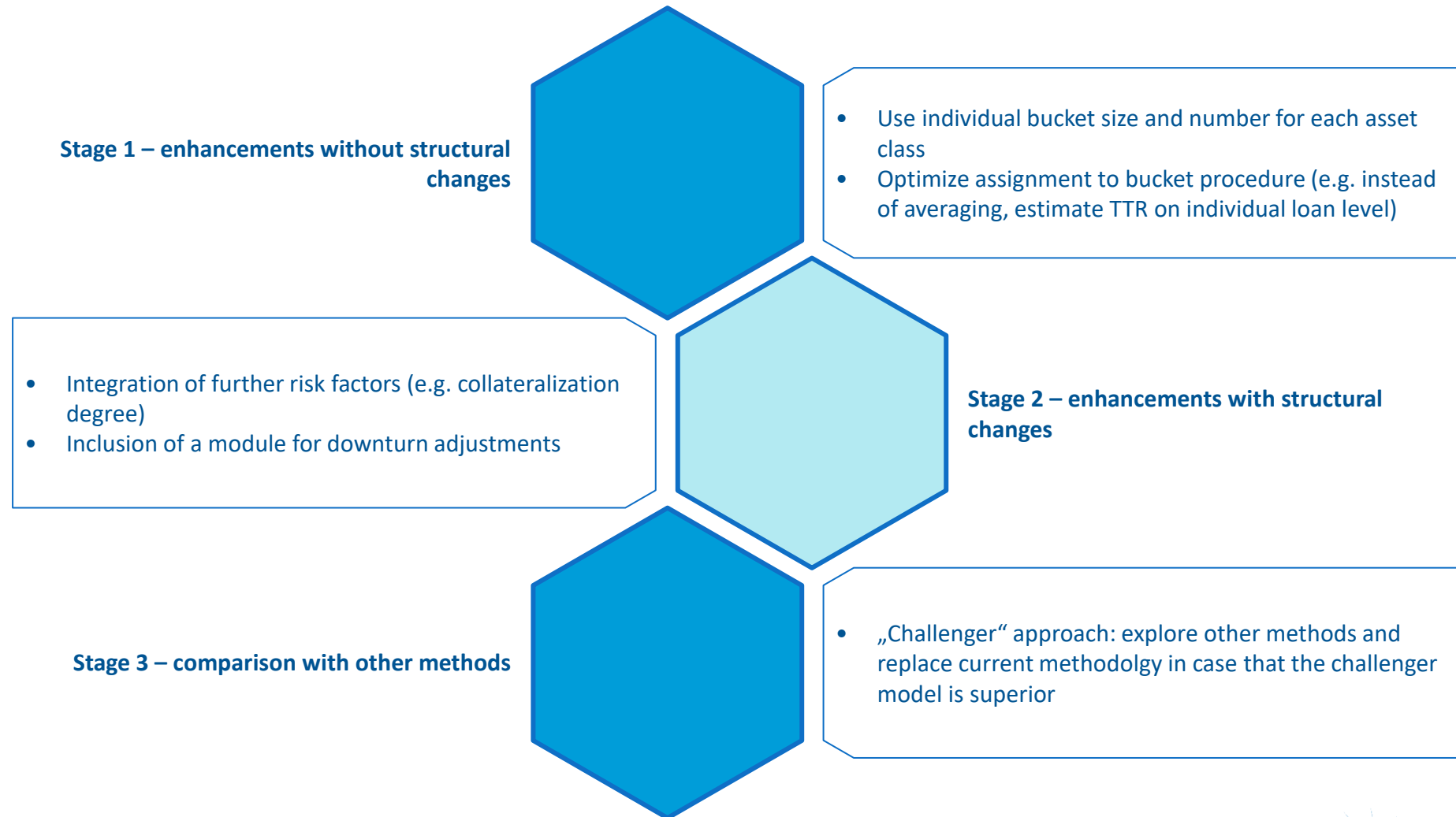
## □ Example LGD for 2018

□  $(13\% + 200 + 56 * 38\%) / (200 + 56) = 18\%$



- LGD for unresolved higher as we expect the realised LGD to be biased toward cures, easy quick workout processes

# Outlook (what could be done to improve the method?)



# Appendix

# In-sample backtesting results

(Example LC)

**PASSED**

Overall Backtesting Result

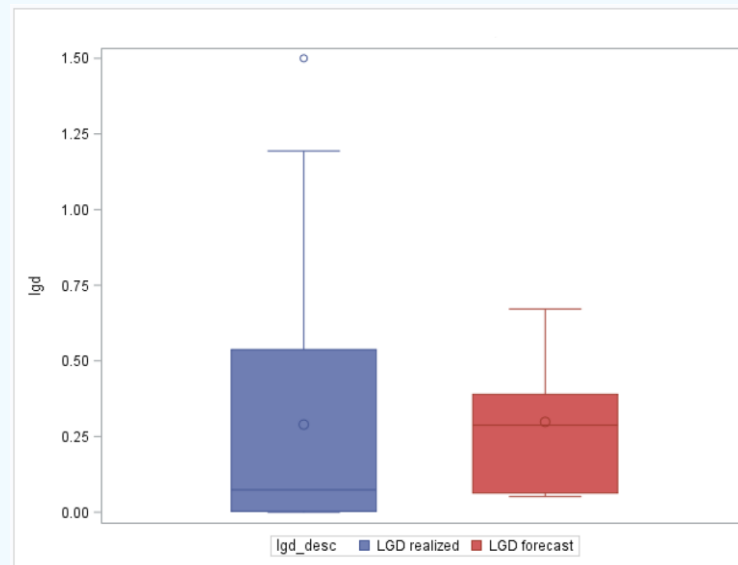
- ❑ Forecasted average LGD very close to realized LGD. T-Test requires the two mean values to be very close for large samples and should therefore not be the only value to consider.
- ❑ Bimodal nature of LGD is well reflected. The peak density at 0% is well modelled, the 100% peak is visible.
- ❑ The spread in the realized LGD is larger than in the forecasted due to averaging effects.
- ❑ The spread indicates the usual achievable accuracy of LGD estimates (best practice +/- 5% points)

## AVERAGE LGD and T-TEST

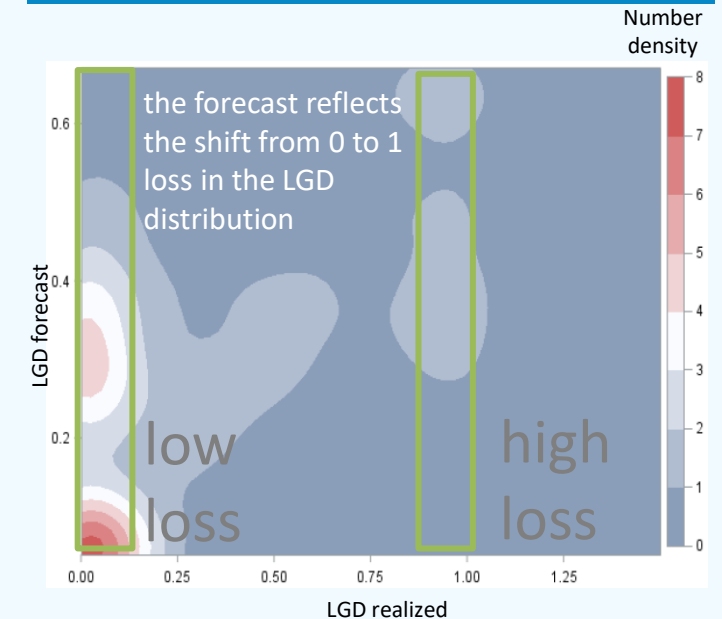
	LGD realized	Difference	LGD forecast	T-test 95%*
2010	27%	↗ 1%	28%	passed
2011	31%	↘ 2%	29%	failed
2012	28%	↗ 1%	29%	passed
2013	29%	↗ 1%	30%	passed

\* suboptimal due to normal assumption

## BOX PLOT



## KERNEL DENSITY PLOT



# Out of-sample backtesting results

(Example LC)

**PASSED**

Overall Backtesting Result

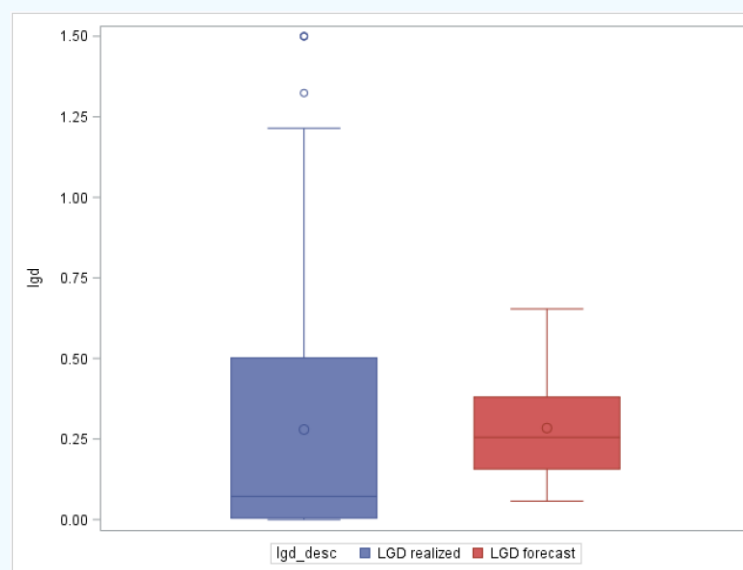
- Forecasted average LGD very close to realized LGD. T-Test requires the two mean values to be very close for large samples and should therefore not be the only value to consider.
- Bimodal nature of LGD is well reflected. The peak density at 0% is well modelled, the 100% peak is visible.
- The spread in the realized LGD is larger than in the forecasted due to averaging effects.
- The spread indicates the usual achievable accuracy of LGD estimates (best practice +/- 5% points)

## AVERAGE LGD and T-TEST

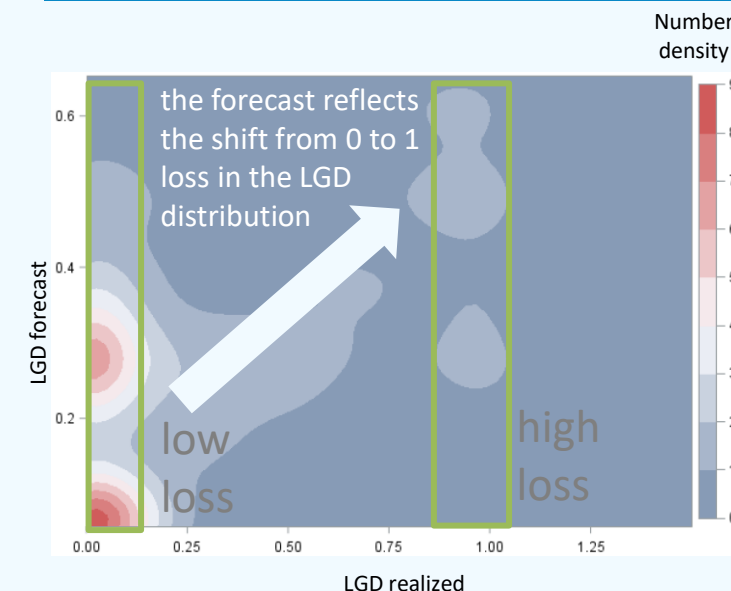
	LGD realized	Difference	LGD forecast	T-test 95%*
2010	27%	→ 0%	27%	passed
2011	31%	↘ 3%	28%	failed
2012	28%	→ 0%	28%	passed
2013	29%	↗ 1%	30%	passed

\* suboptimal due to normal assumption

## BOX PLOT



## KERNEL DENSITY PLOT





# Unresolved LGD: How to use it?

- ❑ Two flags have been introduced in the GCD database to identify unresolved cases on loan and borrower level respectively:
  - Standard\_RDS\_LGD\_Unresolved
  - BOR\_Standard\_RDS\_LGD\_Unresolved
  
- ❑ One additional field is delivered
  - Time\_in\_Default\_Unresolved

Values are in days and measures the difference time between the default date and the current date (the latest date of the data release). For unresolved loans with **time in default smaller than seven years** (2520 days), the **GCD unresolved LGD methodology is applied** for the calculation of LGD and recovery rate.

# Hellinger distance

## Hellinger distance measure

- ❑ Suppose you have two discrete probability distributions  $P=(p_1, p_2, \dots, p_k)$  and  $Q=(q_1, q_2, \dots, q_k)$  with relative frequencies  $p_i, q_i$  for the possible realizations  $i=1, 2, \dots, k$ .
- ❑ Test if the probability distribution of the dataset used for model calibration (pool),  $P$ , is representative for the probability distribution  $Q$  of the application portfolio (lender portfolio)

❑ The Hellinger Distance is defined as 
$$H(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^k (\sqrt{p_i} - \sqrt{q_i})^2}$$

- ❑ It gives values between 0 and 1, with  $H=0$  meaning that both distributions are identical and  $H = 1$  meaning that they are singular.
- ❑ To demonstrate representativeness values close to 0 are desirable, with values above 0.25 considered critical.