



**Global Credit Data**  
by banks for banks

# UNRESOLVED DEFAULTS LGD STUDY

LARGE CORPORATES

**NOVEMBER  
2020**

**A METHODOLOGY FOR INCORPORATING UNRESOLVED  
DEFAULTED LOANS INTO LGD CALIBRATION**

### **Global Credit Data Contributors:**

Nina Brumma, Head of Analytics and Research  
Nunzia Rainone, Analyst & Member Support Executive  
Richard Crecel, Executive Director

*Contact:* [secretary@globalcreditdata.org](mailto:secretary@globalcreditdata.org)

## CONTENTS

ABOUT GLOBAL CREDIT DATA.....	3
WHAT IS LOSS GIVEN DEFAULT? .....	4
EXECUTIVE SUMMARY .....	5
RESULTS: FROM OBSERVED TO LONG-RUN AVERAGE LGD .....	6
THE GCD METHODOLOGY FOR UNRESOLVED LGD.....	8
BACKTESTING.....	11
CONCLUSION.....	12
OUTLOOK.....	12
FAQS .....	13

## ABOUT GLOBAL CREDIT DATA

Global Credit Data (GCD) is a non-profit association owned by 50+ member banks with the simple mission to help banks better understand and measure their credit risks through data pooling and benchmarking activities. GCD's data pools support the key parameters of banks' credit risk modelling: Probability of Default (PD), Loss Given Default (LGD), and Exposure at Default (EAD).

GCD started collecting historical loss data in 2004, offering exclusive access to its member banks. These banks receive the detailed anonymised database and can therefore confirm results and test them on customised sub-sets of data. The LGD database now totals over 230,000 non-retail defaulted loan facilities from around the world to more than 130,000 borrowers covering 11 Basel asset classes.

In 2009, GCD introduced a PD database which now has over 15 years of default rates and PDs. GCD also runs a name and cluster benchmarking database to help banks calibrate and benchmark their PD, LGD and EAD models.

GCD operates all databases on a "give to get" basis, meaning that members must supply high-quality data to receive data in return. The robustness of GCD's data collection infrastructure helps place GCD's databases as the global standard for credit risk data pooling.

[www.globalcreditdata.org](http://www.globalcreditdata.org)

## WHAT IS LOSS GIVEN DEFAULT?

Loss given default (LGD) reflects how much money a bank or other financial institution loses when a borrower defaults on a loan, expressed as a percentage of total exposure at the time of default. LGD is one of the key factors used to calculate expected credit losses and Advanced Internal Ratings-Based (AIRB) regulatory capital along with probability of default (PD) and exposure at default (EAD).

### What's new with LGD?

Requirements for accurate credit loss and LGD modelling have been significantly increased by developments in regulation and standards over the last several years. Business usage of acute pricing information can also provide valuable insights in competitive markets. Both regulatory capital frameworks, impairment frameworks (such as IFRS 9 and CECL) and stress-testing frameworks (CCAR) created a need for detailed default and loss data. Investors, regulators and accountants require banks to be able to project expected and unexpected loss levels under different scenarios. Both banks' business as well as their capital holding strategies are significantly influenced by these calculations. Banks benefiting from GCD's consistent information exchange and wealth of data are also able to fine-tune their final estimates via benchmarking.



### GCD's LGD data set

[GCD's LGD data set](#) comprises over 230,000 non-retail defaulted loan facilities from around the world, encompassing more than 130,000 borrowers and 11 Basel asset classes. Collected since 2004, the data details the resolution of defaults from 2000 to 2017 the best part of two decades – taking in numerous sector shocks, as well two major crises, the dot-com crash and the global financial crisis.

The quality of this data set is ensured by GCD's high standards for inputs from member banks. The effectiveness of this approach is underlined by the data's consistency over time. GCD has published updated LGD reports every year for the last three years, alongside an LGD Downturn Report in 2013, 2017 and 2020 - adding new data from each year gone by, as well as additional data from new member banks.

GCD collects default information on both resolved and unresolved defaults. Defaults are considered 'unresolved' or 'incomplete' where banks are still expecting further cash flows. All other cases where the lending bank has closed the recovery file are considered 'resolved'.

## EXECUTIVE SUMMARY

As the COVID-19 pandemic crisis continues, banks still face an uncertain impact on their credit risk assessment. Now, more than ever, it is of utmost importance to incorporate recent information into credit risk models and especially Loss Given Default (LGD) models. If these models are based on historical defaults with complete workout profiles only, most recent developments are naturally not included. This creates a resolution bias which leads to underestimating recent losses, since only quickly resolved, cured cases (typically with lower losses) can be included. Hence, average LGDs based only on resolved loans can result in unrealistic long-term average values. Knowing this, regulatory authorities explicitly ask for unresolved defaults to be incorporated into the modelling process (e.g. European EBA GL 2017/16 or US BCC Bulletin 13-5).

Establishing how to account for incomplete recovery processes or unresolved loans is therefore a key question for credit risk modellers right now. This report describes the GCD methodology for calculating LGDs for unresolved loans. The methodology benefits from GCD's detailed and granular collection of post-default cash flow data and is based on extrapolations of historical recovery cash flows refined by the usage of risk drivers.

[GCD's data](#) is detailed enough to develop and enhance internal LGD models and to be used for validation, calibration or benchmarking purposes. These models can be used to support the Advanced Internal Ratings-Based (AIRB) approach, to fulfil the credit provisioning standards IFRS9 or CECL, as well as for stress-testing, economic capital and pricing, among other uses.

The methodology has been developed on the entire GCD database. For the sake of simplicity, this report presents the results

for Large Corporates only (defined by their sales or assets being above €50m). Results for the rest of the segments are available upon request.

### **The key observations are:**

- Including unresolved default cases when assessing LGD leads to a more conservative long-run average LGD, although this effect is overall limited to 2%.
- The impact becomes more significant when looking at the most recent years – including unresolved defaults causes an increase in the long-run average LGD by 12% for 2018.
- For default years before 2013 or more than seven years in default, the completion rate of defaults is above 95% and these years can therefore be considered complete.

GCD provides unresolved LGDs to its member banks both at a loan and borrower level. This allows member banks to calculate long-run average LGDs for calibration purposes, avoiding a resolution bias when using most recent default cases.

The methodology provides a straightforward, data-driven way of incorporating incomplete workout processes in the estimation of long-run average LGDs. Extensive validation both in- and out-of-sample has shown that the method works well in predicting LGDs for unresolved defaults.

## RESULTS: FROM OBSERVED TO LONG-RUN AVERAGE LGD

The LGDs in this report do not reflect specific portfolio alignment or the addition of any statistical uncertainty add-ons or downturn adjustment. GCD members do not provide LGDs as an input field to the database, but rather the underlying raw information such as outstanding amount at default or resolution and cash flows during the default. This enables GCD to calculate LGDs in a consistent way. Long-term average calculations that only include resolved loans do not reflect the most recent developments nor complicated long in-default cases. Especially in recent years, cures and low-loss defaults are over-represented.

To measure this effect, the observed average LGD of resolved cases is compared to the average LGD including unresolved defaults for both the overall timespan as well as by year of default. The first measure relies purely on observed historical outcomes and is therefore the most objective measure possible, while the latter includes assumptions on expected future cash flows and therefore some degree of uncertainty. Both measures must therefore be used in the right context and circumstances.

Overall results are detailed in Table 1. The average values on the unresolved sample lead to more conservative values than the realised LGDs on the resolved sample only. This can be explained by the on average much longer time in default of the unresolved loans.

## REGULATORY BACKGROUND

Regulatory authorities explicitly ask for the consideration of unresolved defaults in LGD calculation. The European Banking Authority (EBA) requires unresolved defaults to be included in the calculation of the long-run average LGD [GL 2017/16]. According to the guidelines, the long-run average LGD is calculated from the observed average LGD (average LGD over all resolved defaults) plus additional adjustments for unresolved defaults and further corrections, such as representativeness or the new definition of default.

As another example, in the context of data limitations, US banking regulation requires conservative adjustments for the inclusion of unresolved defaults or missing components for LGD calculation, such as workout costs [BCC Bulletin 13-5].

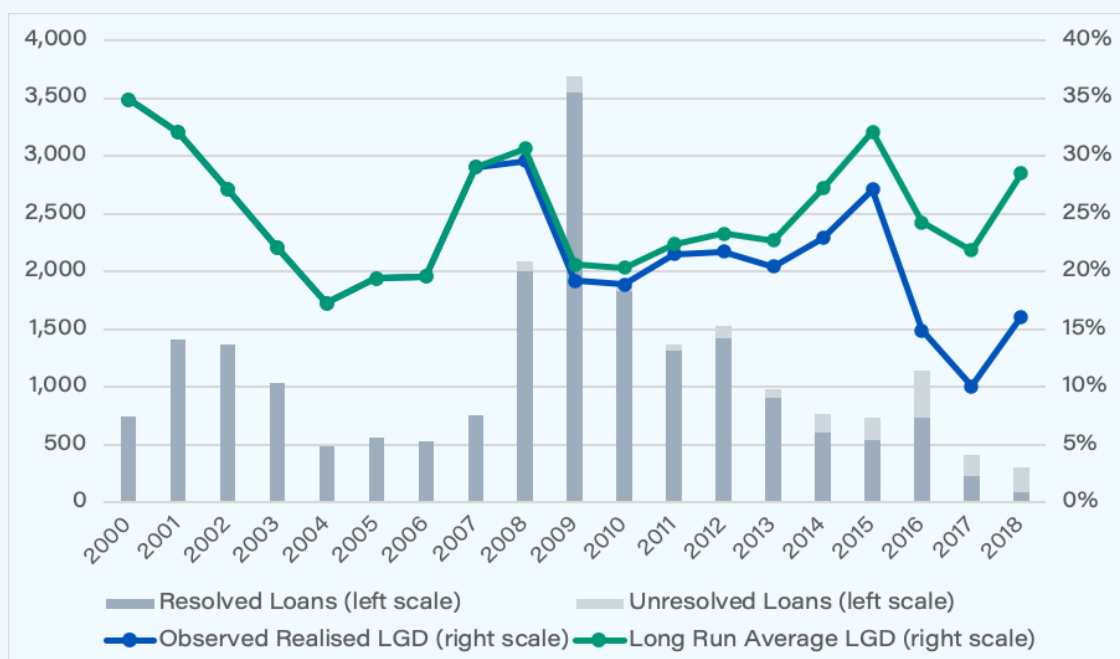
**TABLE 1**  
**OVERALL AVERAGE LGD**

	NUMBER OF FACILITIES	AVERAGE LGD
<b>RESOLVED LOANS</b>	20,139	23%
<b>UNRESOLVED LOANS</b>	1,680	44%
<b>COMBINED DATASET</b>	21,819	25%

Exhibit 1 shows the number of defaulted facilities for resolved and unresolved loans as well as the average LGD by year for the resolved cases only and for the sample including the unresolved loans. Prior to 2008, all the defaults have already been resolved and therefore the LGD for these years does not change. For the years 2008 to 2011, the completion rate is above 95%. A small number of unresolved loans exist but their inclusion does not change the average LGD significantly.

After 2012, the completion rate lowers gradually until reaching only 30% by 2018. As a result, the impact of the inclusion of unresolved loans has a gradually higher impact for more recent years.

**EXHIBIT 1**  
**LGD BY YEAR OF DEFAULT**



# THE GCD METHODOLOGY FOR UNRESOLVED LGD

The GCD methodology is based on an extrapolation approach designed to make best use of currently available knowledge.

It provides individual estimates for LGDs on facility and borrower level and can be used to calculate the long-run average LGD for a selected segment.

## EXHIBIT 2 EXTRAPOLATION SCHEME

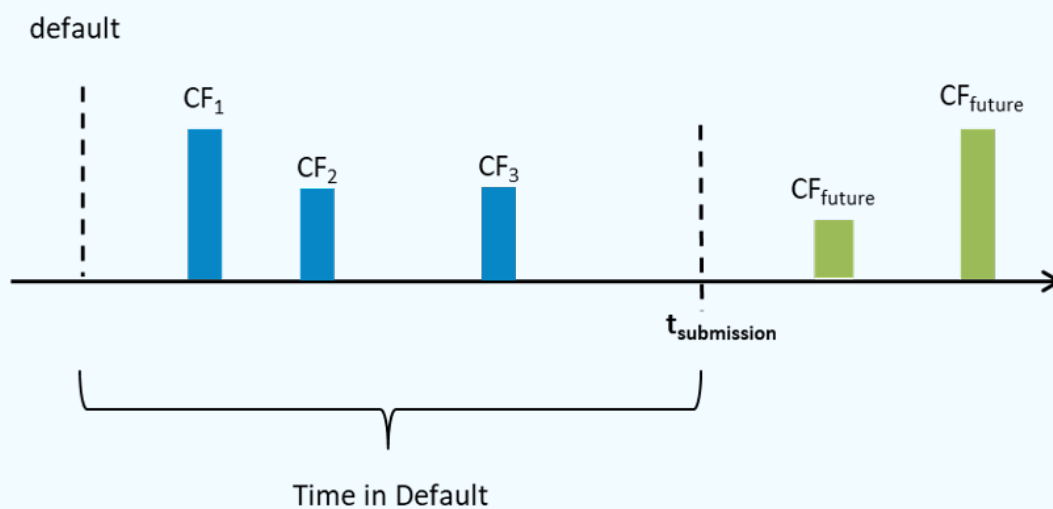


Exhibit 2 outlines the availability of cash flows for unresolved defaults. Cash flows after default up to the submission date (the blue bars) are provided by the institutions as part of the regular data collection. Including only these realised cash flows would generally lead to an underestimation of the recoveries as institutions still expect future cash flows (green bars). However, the exact date and value of future recoveries is not yet known and has to be estimated. GCD applies an extrapolation methodology which is based on the historical recovery patterns observed on those cases that are already resolved.

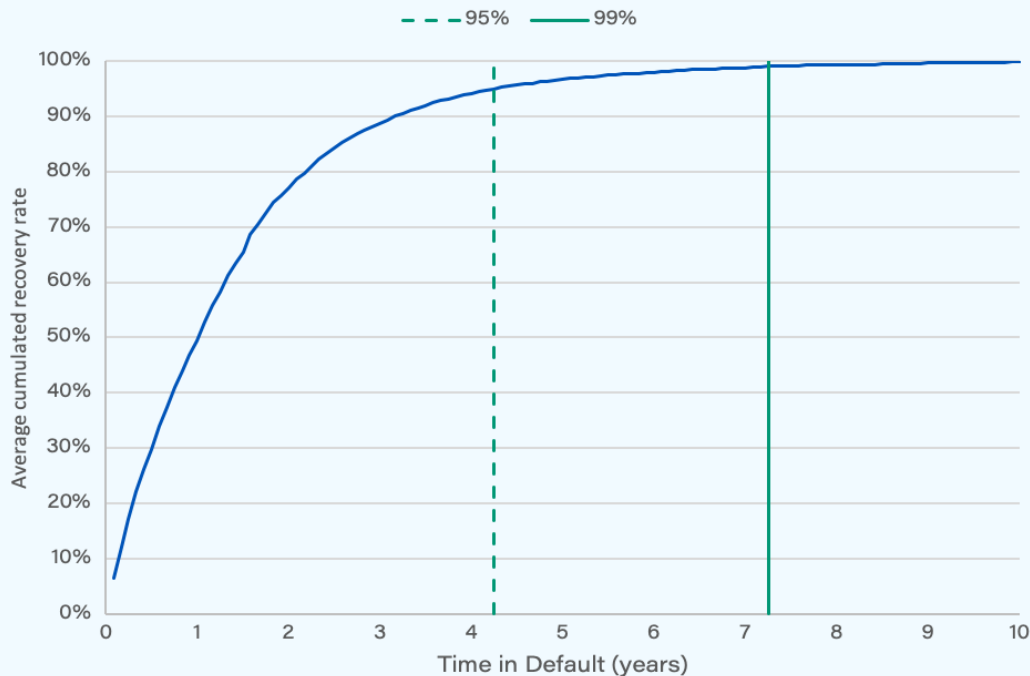
A well-known fact about LGDs is that the length of the workout period correlates with the LGD: the longer it takes to

recover the outstanding amount, the higher the LGD (see, for example, our [LGD Report for Large Corporates](#)). Therefore, a maximum recovery period (MRP) is introduced to define a maximum length of the recovery process during which additional recoveries can reasonably be expected. This threshold separates the long in-default cases from the more recent ones. For defaults exceeding the MRP, no further cash flows are estimated and LGDs are calculated based on the already received cash flows up to now.

The MRP equals seven years (which corresponds roughly to the 99% recovery level) and has been derived based on the historical recovery patterns as illustrated in Exhibit 3, where the recovery curve significantly flattened.

## EXHIBIT 3

### AVERAGE CUMULATED RECOVERY AND THE MAXIMUM RECOVERY PERIOD



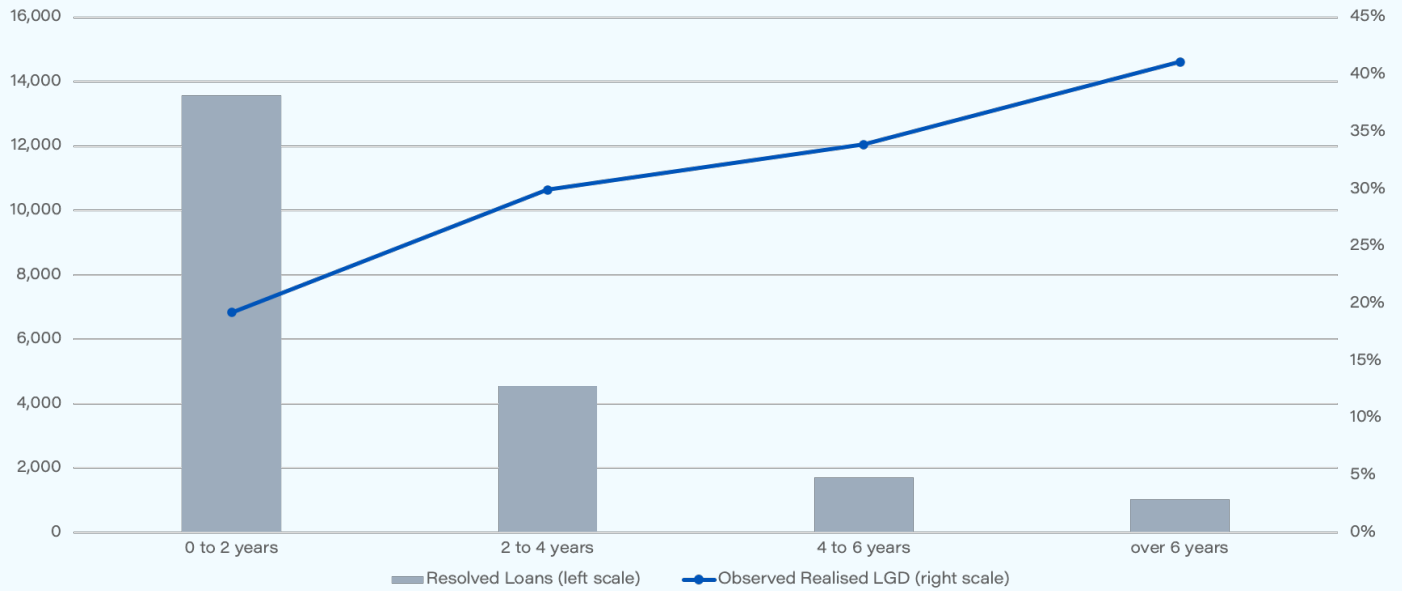
Defaults that are resolved – or that are unresolved but exceed the MRP – make up the development sample, while the unresolved loans below the MRP are the ones for which the parameters for unresolved LGD calculation are developed.

The methodology uses average LGD values calculated for multidimensional buckets with the established GCD calculation for resolved loans. Besides the facility asset class, two more bucket dimensions are used: the time to resolution and the already received recovery amount at the beginning of a time to resolution bucket. The average LGD increases with longer time to resolution periods as outlined in Exhibit 4. Furthermore, with increasing received recovery amount, the realized average LGD decreases (see Exhibit 5).

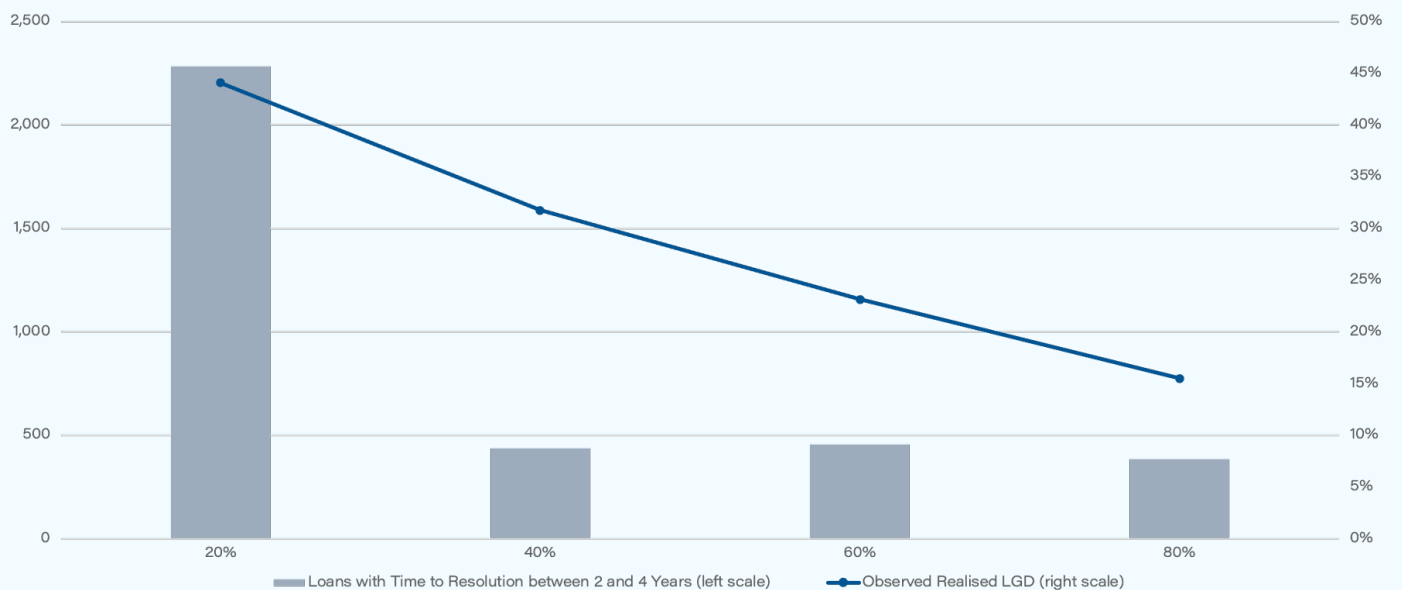
The introduction of these bucket dimensions results in a look-up table structure that is used for the extrapolation of LGD values. Each unresolved loan is assigned to a certain bucket with a certain probability.

Exhibit 6 summarises the functionality of the developed methodology (further details on the parametrisation table and the assignment procedure on individual loan level are available upon request).

## EXHIBIT 4 LGD BY TIME TO RESOLUTION

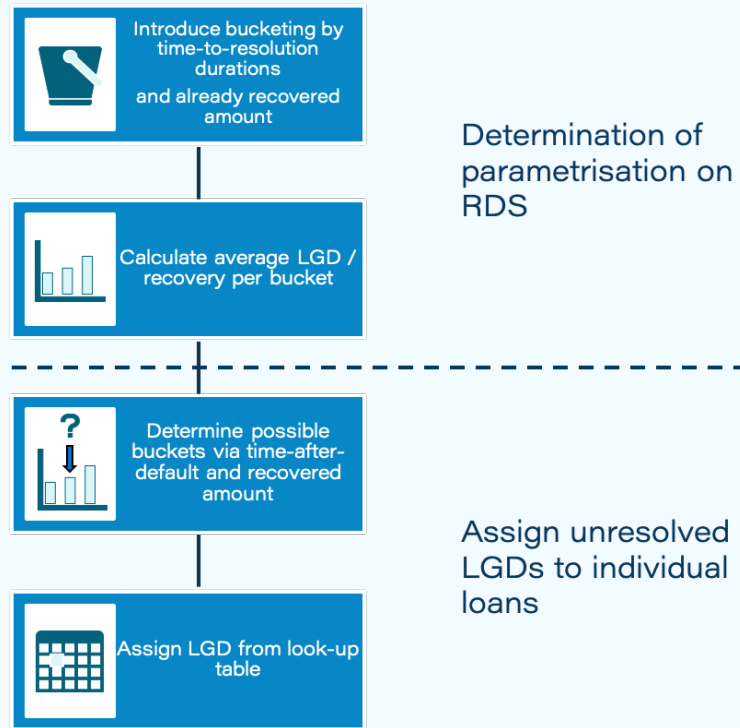


## EXHIBIT 5 LGD BY RECOVERED AMOUNT RECEIVED AFTER TWO YEARS IN DEFAULT



## EXHIBIT 6

# ILLUSTRATION OF THE FUNCTIONALITY OF THE METHODOLOGY



## BACKTESTING

In order to underpin the validity of the method, a comprehensive backtesting was carried out both at a portfolio and individual-case level. At a portfolio level, the calibration answers the question of whether or not the forecasted average LGD matches the realised average LGD. This was tested by applying a T-test and comparing the forecasted average LGD with the realised average LGD. At an individual case level, the forecasted and realised LGDs are compared qualitatively via Kernel Density plots.

To avoid overfitting, in- and out-of-sample backtesting was performed. For in-sample backtesting, the full RDS was used to develop the parametrisation. Then yearly sub-samples were created by selecting all cases that were in default in a particular year. For each yearly cohort, the unresolved

methodology has been applied using the realised cash flows up to date. Then, the outcome has been compared to the ultimately realised LGD.

For out-of-sample backtesting a sub-sample of one-year duration has been removed from the development sample. The parametrisation has been derived on this reduced development sample. Comparison of forecasted and realised LGD values took place on the omitted part of the sample. This was repeated for several sub-sample removals.

The backtesting results indicate that the method reaches a reasonably good accuracy of LGD estimates (further information regarding the backtesting methodology and detailed results is available upon request).

## CONCLUSION

The methodology and results presented here enable institutions to incorporate unresolved defaults in their LGD modelling exercises. This approach ensures that the most recent developments can be incorporated – particularly important in the light of the ongoing COVID-19 crisis, as well as for complicated long-run defaults. The insights gained from these analytics confirm the benefit of a detailed and granular collection of post-default cash flow data for both resolved and unresolved defaults.

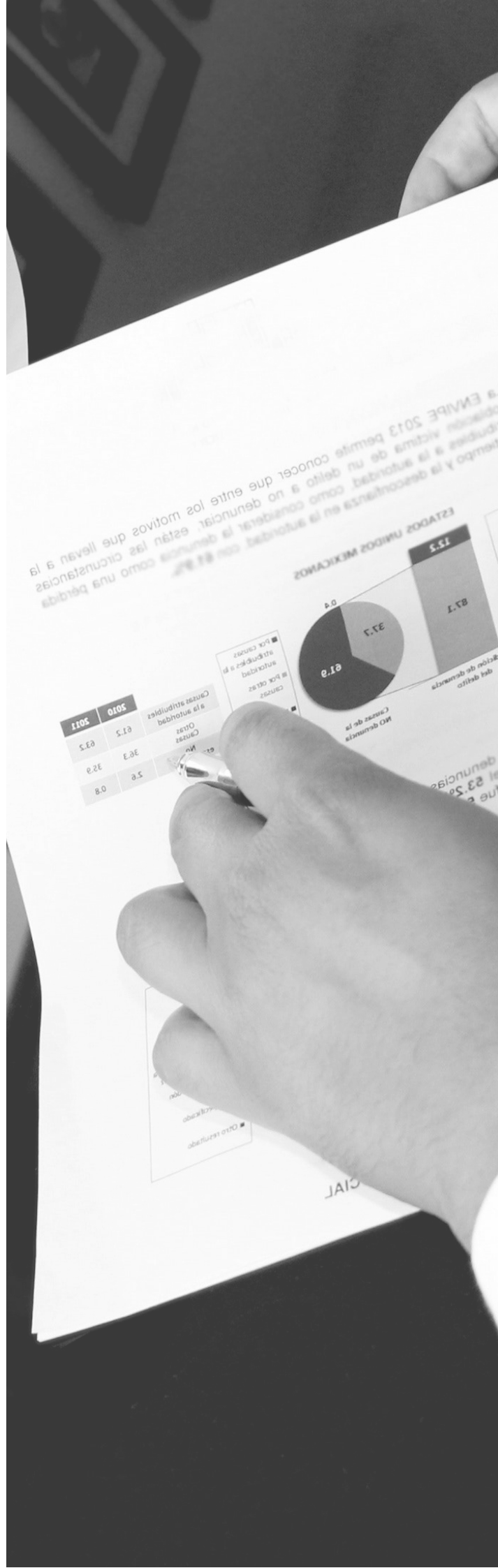
While the results presented here are for large corporate defaults only, the methodology has been developed for the overall GCD database, covering all asset classes. The detailed results can be obtained upon request.

## OUTLOOK

In addition to the selected extrapolation methodology, institutions can choose from a range of modelling approaches. Other common techniques are regression models or the provision-based approach where loan loss provisions are used as a proxy for the future loss.

More exotic approaches include now-casting or survival methods. While each approach has strengths and limitations, they both require large data volumes, which may provide an indication as to why they are less widely used in the institutions. These other options could be explored and used as challenger models versus the extrapolation method, given the large size of the GCD data.

Within the chosen framework of the extrapolation method, the data set will evolve over time as new defaults occur and resolve. Therefore, the obtained LGDs should be subject to regular validation and re-parametrisation exercises. The methodology could also be further enhanced by factoring in, for example, the degree of collateralisation and debt seniority.



## FAQS

### Which reference data set has been used for this study?

The reference data set (RDS) for this study contains resolved and unresolved defaults submitted up to December 2019. Cases with year of default from 2000 to 2018 were chosen due to completeness. Pre-2000 defaults can be biased towards long, difficult workouts, while the collection of the defaults in 2019 was not complete at the moment the data was collected.

Results are shown on loan level. 'Facility', 'loan' and 'default' are considered the same for the purpose of this report, with the term 'loan' used for consistency throughout. Most of the loans in the GCD data set are loans of some type, such as term loans or overdrafts. However, the data also includes significant numbers of contingent loans, including letters of credit or derivatives, as well as some bonds and equity.

### What is the definition of an unresolved loan?

A default is considered 'unresolved' in the instance that the lending bank is still expecting further cash flows. All other cases where the lending bank has closed the recovery process are considered 'resolved'. How is LGD and its variation over time calculated?

LGD refers to the Loss Given Default rate, which is calculated as  $1 - \text{recovery rate}$ . The recovery rate is the net of all cash flows, including external costs (using the discounted cash flows where the discount rate is equal to the risk-free rate of three months' EURIBOR as at the default date). This calculation is made at the loan level.

The LGD calculation is made using a cap of 150% and floor of 0%, where the exposure at default (EAD) is increased by the amount of

any post-default advances. The LGD levels are calculated based on raw data and do not reflect any bank-specific portfolio alignment or addition of any statistical uncertainty add-ons. Variations could include using a different discount rate based on a combination of the risk-free rate and a risk premium for systematic risk at the time of default.

#### ATTRIBUTION

Analytics were conducted by the GCD Analytics Team and supported by:  
[markus.seifert@d-fine.de](mailto:markus.seifert@d-fine.de)  
[benjamin.galow@d-fine.de](mailto:benjamin.galow@d-fine.de)

d-fine



