Observed Recovery Rates Report Corporates

Appendix: Database & Methodology

Global Credit Data

by banks for banks

Contents

1. About Global Credit Data

At GCD, our mission since 2004 has been to empower banks and the financial industry with a deep understanding of credit risk through a unique data source. As a non-profit organization owned by 50+ member banks, we collect valuable data directly from banks' books. GCD's data pools support the key parameters of banks' credit risk modelling: Probability of Default (PD), Loss Given Default (LGD), 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 results and can therefore confirm results and test them on customised sub-sets of data. The LGD database now totals over 300,000 non-retail defaulted loan facilities from around the world to more than 150,000 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.

2. The GCD LGD/EAD Database

Global Credit Data has built up the world's largest non-retail bank loan loss database, which includes over 300,000 defaulted facility observations. This extensive database covers all non-retail Basel asset classes and represents a total value of over €700 billion. Among GCD's databases, the LGD/EAD platform is the largest and was established in 2005. Since its inception, it has experienced rapid growth and currently provides detailed information on tens of thousands of defaulted counterparties. This includes valuable data on cash flows and collateral. The driving principles and objectives of the Global Credit Data databases are:

- **Confidentiality**
  GCD ensures data is fully anonymised
- **Comparability**
  The data GCD utilises is only from banks with common definitions
- **Data Quality**
  GCD ensures the highest standard of data quality through a rigorous process of validations, auditing and scoring
- **Granularity**
  A full database is always returned to member banks, not just the aggregate
- **Reciprocity**
  Member banks must contribute to the databases before they can benefit from them by asset class and year
- **Best practice sharing**
  GCD utilises method workshops, surveys and HPE
- **Research standards**
  There is always a common basis for shared analysis and research

The data collected by GCD comes from over 60+ banks across the span of 20 years. Most defaults occurred from the year 2000 forwards. Global Credit Data returns the complete database of detailed transactions to member banks, excluding borrower names and Lender ID's. Members also receive some pre-calculated metrics (EAD,

LGD, etc.) but are encouraged to calculate their own, using their own methods. Members are then free to sample the database, clean the data, calculate their own metrics and extract drivers.

Each member chooses the data pools (asset classes) to which it contributes. The most popular is the large corporate asset class, comprised of loans to corporates where the borrower group has a turnover of €50m or less. Over 60 banks in total have contributed to this asset class, largely driven by the fact that nearly all members have a large majority of their lending books by volume taken up by large corporate lending. Next most popular are Banks and Financial Companies, SME and Real Estate Finance, again driven by the ubiquity of these assets in loan books. More specialised financing such as Aircraft or Shipping Finance comprise a more limited club of around 20 or less lenders over time.

The purpose of the database is to develop banks' understanding of both the quality of their data in comparison to other banks and the risk involved in the loans. Global Credit Data member banks receive back a database of detailed transactions, excluding borrower names and Lender ID's. Due to the "give to get" rules applying, members only receive back data for the years of default and asset classes which they submitted. Members are then free to sample the database, clean the data, calculate their own metrics and extract drivers. When the data set is returned to member banks, each member can then slice and dice it to produce matching portfolios before finally creating a representative reference data set which is the key success factor when using pooled data.

3.   Data definitions

CORPORATES

Corporates are defined according to the Basel rules as a class of corporate exposures which is not identified in one of the five specialised lending classes, as described in paragraph 218 and paragraph 219 of the Basel II Accord In this report, SME – small and medium enterprise and LC – Large Corporate are treated together.

DEFAULT

The Basel definition of default is used. According to the GCD Data Pool Regulations all Pool participants must report their resolved defaults. Reporting unresolved defaults is recommended but optional.

DEFAULT DATE

The date at which a borrower has been recorded as a default according to the Basel default definition.

INDUSTRY

Banks use a variety of industry codes (such as NAICS, NACE, SIC, etc.). GCD member banks have agreed on a set of industry groups that banks map their internal, typically very granular, industry types to. GCD provides banks with mapping tables for the most common industry types. Therefore, consistency among banks is ensured. In this report the following grouping of detailed industry codes has been applied:

Group in Dashboard	Detailed Group in Database
Agriculture	Agriculture, Hunting and Forestry
	Agriculture
	Hunting
	Forestry
	Fishing and Fishing Products
Communications	Communications

Group in Dashboard	Detailed Group in Database
Construction	Construction
Hotels and Restaurants	Hotels and Restaurants
	Hotels
	Restaurants and Food services
Manufacturing	Manufacturing
	Manufacturing - Automotive
	Manufacturing - Chemicals
	Manufacturing - Food and Beverage
	Manufacturing - Oil & Gas (Downstream)
	Manufacturing - Other
Mining	Mining
	Mining - Oil & Gas Extraction (Upstream)
	Mining - Metals and Non-metals except Oil & Gas
Real Estate	Real Estate and Rental and Leasing
Social/Health Services	Health and Social Services
Other Services	Professional, Scientific and Technical Services
	Public Administration and Defense
	Education
	Other Community, Social and Personal Services
	Private Sector Services (Household)
	Extra-Territorial Services and Organizations
Transportation	Transportation and Storage
	Transportation - Air Transport
	Transportation - Rail Transport
	Transportation - Water Transport
	Transportation - Road Transport
	Transportation - Oil & Gas (Midstream)
	Transportation - Other
	Warehousing and Storage
Utilities	Utilities
	Utilities - Oil & Gas
	Utilities - Other (excluding Oil & Gas)
Wholesale/Retail Trade	Wholesale and Retail Trade
	Wholesale and Retail Trade - Automotive
	Wholesale and Retail Trade - Durable Goods
	Wholesale and Retail Trade - Nondurable Goods
	Wholesale and Retail Trade - Other
Other	Other Industries and Unknown

OBSERVED RECOVERY RATE

Global Credit Data (GCD) members do not provide Observed Recovery Rate as an input field to the database but the underlying raw information such as outstanding amount at default or cash flows during the default. GCD calculates Recovery Rate by following different methodologies. Internal methodology as well as regional

regulatory requirements might result in different calculation methods which member banks are able to realise on the raw data they receive back. In this report the following Recovery Rate definition is used:

$$Observed\ Recovery\ Rate = \frac{Recoveries - (Costs + Interests\ Charged + Fees\ Charged)}{Default\ Amount}$$

Cash flows are nominal, undiscounted cashflows. Post default drawing are treated as part of the Default Amount. All Recovery Rates are floored at -50% and capped at 100%. The RRs are calculated on facility level.

A detailed composition of Recoveries and the default amount is given below.

Recoveries

Recoveries	The following recoveries are collected separately:
	principal payments,
	interest payments,
	recorded book value in case the bank repossesses a collateral. It is the amount with which the credit obligation of the obligor has been diminished and which has been recorded as an asset on the balance sheet of the institution,
	fees and commissions received which are recoveries on extra fees and commissions charged to the obligor post default on additional services
	outstanding amount at resolution. The oustanding amount at resolution can only be greater than 0 in case the borrower returns to a non defaulted status. In this case it can be fairly assumed that the borrower will be able to pay back its obligation and therefore the amount is treated as if it were a recovery.

Default Amount

=	Outstanding Amount at Default	The amount of the principal outstanding plus past due interest as at the default date.
+	Cash-out on Guarantee	Any cash drawing on a contingent facility. Contingent facilities have by definition an outstanding amount of 0 at default date. The cash-out converts them into a cash obligation on which the Recovery rate can be calculated.
+	Financial Claim	Financial Claims are the final adjustment of the exposure at default due by the obligor in default on a mark-to-market facility. It is the final claim, if any, of the bank against the obligor after netting all exposures and collaterals at their market value on date of liquidation.
+	Advances	Advances include additional funding extended post default with intention to help improve the borrower's financial condition as well as additional money drawn by the borrower as part of a revolving facility.

Costs

Costs	Global Credit Data captures direct costs including legal expenses, administrator or receiver fees, liquidation expenses and other external workout costs. These cost types are collected as separate items. Internal costs are not included.

RECOVERY RATE FOR UNRESOLVED DEFAULTS

Recovery Rate is most accurately calculated on closed (resolved) cases, where the outcome is anything from full repayment to complete loss, or something in between. On the other hand, this restriction leads to underestimating recent losses, since only quickly resolved, cured cases (typically with lower losses) can be included (resolution bias). Therefore, the metrics in the graph of Recovery Rate by Year of Default include projections for unresolved defaults that are displayed in separate lines. By including both observed losses and projections on incomplete recoveries, realistic and up-to-date Recovery Rates, that includes all available Information can be calculated especially for recent years.

GCD's methodology is based on an extrapolation approach that has been designed to make the best use of currently available data (blue cash flows) to estimate expected future cash flows (green cash flows). The description of the methodology in detail can be found on our website. Members have exclusive access to a more technical detailed description of the methodology as well as yearly monitoring and backtesting reports.

REFERENCE DATA SET

Reference Data set (RDS) refers to the data set after application of filters which is used for the analysis.

REGION

The GCD data set offers country information on several levels (country of residence, country of jurisdiction, collateral country of jurisdiction). The items are collected on country level and returned on country level unless there are less than three banks providing data in a certain country. In this case the countries are aggregated to regions until the minimum requirement of three banks are met. This rule was established to protect anonymity of the lender identity. In this report country information is aggregated on regional level. The regions displayed are

- Africa & Middle East
- Asia & Oceania
- Europe
- North America
- Latin America

For real estate collateral GCD furthermore collects post codes on a "give to get" basis.

RESOLUTION DATE

Generally, a default can resolve because of three reasons: first, the borrower pays back all the debt, second, the borrower returns to a non-defaulted status or third, the bank decides to stop the recovery efforts and writes off the outstanding debt (or sells it). The resolution date is an input field in the GCD database.

RESOLVED/UNRESOLVED

Defaults are considered as 'unresolved' where banks are still expecting further cash flows. All other cases where the lending bank has closed the recovery file are considered 'resolved'. This is an input field in the database.

SECURED/UNSECURED

Unsecured means the absence of collateral A loan is unsecured if there is no collateral attached to it. Otherwise, the loan is secured. It is important to note that no further requirements like a minimum Loan To Value threshold are applied. As this definition creates a relatively homogeneous dataset for the unsecured data the focus in the analytics is put on this subsample.

GCD collects a variety of information on the underlying collateral. In this report, a difference between Primary and Secondary Collateral is applied.

Primary Collateral comprise the following collateral types:

- Cash/Reserves
- Marketable Securities
- Ships
- Aircraft
- Real Estate
- Other Object for Object Finance

Secondary Collateral comprise the following collateral types:

- General Collateral
- Non-Marketable Securities
- Accounts Receivable
- Inventory
- Cars
- Commercial Vehicles
- All Assets Charge or Debenture
- Specific Fixed Assets
- Stock of Subsidiary (Capital Stock)
- Intangibles
- Commodities Under Trade
- Oil and Gas Reserves - Mines
- Telecom Networks
- Projects
- Life Policies

SENIORITY

Banks provide seniority on a slightly more granular level than displayed in this report.

- Super Senior
- Pari Passu
- Subordinated or Junior
- Equity

By definition a loan is always Pari-Passu unless the lender has made agreements with other lenders to "promote" or "demote" itself to Super Senior or Subordinated/Junior. If unknown the banks can provide an escape clause. The usage of the escape clauses is closely monitored in the audit and should be avoided if possible.

For simplicity, the first two categories have been grouped together to "Senior". Please note the grouping on obligor level: Borrowers are not always borrowing uniquely senior or subordinated. Occasionally a bank will provide facilities of differing seniority to the same borrower. The small number of bond and equity defaults as well as unknowns are also included here.

TIME TO PEAK RECOVERY

Time to peak Recovery is calculated as the centre point of recovered cashflow, and it puts a weight of the amount of the cashflow on the timing. It is defined as the cashflow weighted average period between default and cashflow. The following picture visualises the concept. The cash flow weighted time or average year of cash flow represents the weighted average of all relevant points in time between default and resolution where cash flows took place. Time to peak Recovery is by definition lower or equal to Time to Resolution.

EXHIBIT 1
CONCEPT OF TIME TO PEAK RECOVERY AND TIME TO RESOLUTION
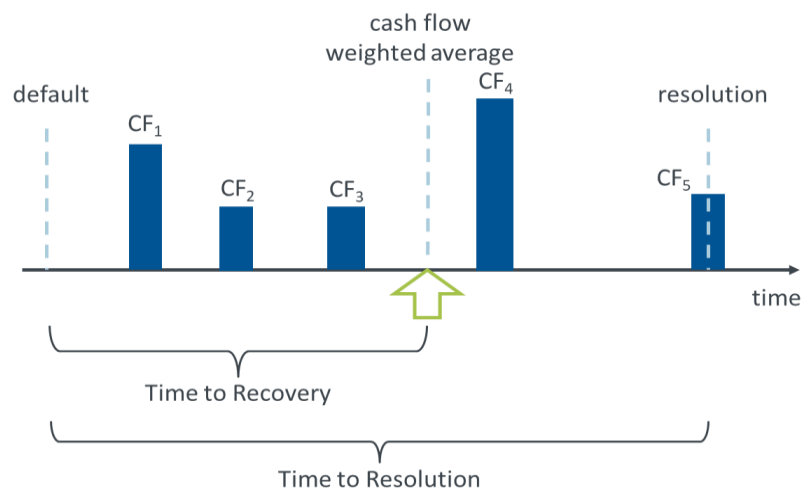
4.   Reference Data Set (RDS)

4.1.   Why and how to create an RDS?

GCD provides members with the full "raw" data set when returning data to members after the validation and auditing process. No filtering or data cleansing is done. Banks are advised to create a reference data set (RDS) from the full data set which is a subset of observations from the full data set (borrower, loans, collateral) that should resemble the referenced portfolio.

The RDS can be used for modelling the credit risk of a portfolio, benchmarking a portfolio and validation or calibration of a model for a certain target (referenced) portfolio. This RDS should have two qualities: representativeness and data quality. The creation of a RDS is the key success factor for using pooled data.

4.2.   Representativeness of data

Many regulators have set standards[1] or guidelines for data to be used in credit risk estimation models. Such guidelines cover both internal and external data. In both cases the data used needs to be representative of the target portfolio during the proposed measurement time. Even internal data from a different time period, jurisdiction or sub-portfolio should be assessed for representativeness.

The total GCD defaulted borrower data set is composed of data from the banks who have chosen to be GCD members. These banks' geographical lending footprint, loan and borrower types as well as collateral practices are merged in the database. Due to the size and long time series of the database and the contributions from banks of many countries, the data set could be seen as broadly representative of an average bank, however more accurately it represents the average of GCD member banks, weighted towards the largest member banks who provide most data. Details of GCD member banks, including geographical footprint of the data, are given on the GCD website.

NO STANDARD GCD RDS

The purpose of creating an RDS is to match as closely as possible the risk conditions of a target portfolio of a single bank. Therefore, a single standard RDS could not possibly suit all users. In this report GCD bases the analytics on a filtered data set which combines elements of representativeness and data quality.

Again, the capability of member banks to be able to narrowly define a data set aligned to the reference model significantly improves the comparability and validity of the benchmarking exercise and addresses benchmarking challenges such as differing portfolio composition, processes and policies, default definitions, weighting schemes and so forth.

4.3.   Elements of the RDS in this study

The full database of Corporate related loans contains 255,683 defaulted facilities. 168,022 facilities remain in the reference data set after applying filters. The different elements and the reasons for filtering are explained below.

---

[1] Some examples of regulatory requirements for representativeness include:
BCBS: Basel II §417, §450 and §448
European CRR: Articles 174, 179 and 185
UK PRA Internal Rating Based Approach (SS11/13) Article 10.12
EBA Guidelines on PD Estimation, LGD Estimation and Treatment of Defaulted Exposures Section 4.2.2.19
ECB's Targeted Review of Internal Models (TRIM): Section 6.2, paragraph 57 (d)(iii) and Section 2.1
US Federal Reserve: SR11-7

OBSERVED RECOVERY RATES DASHBOARD CORPORATES

UNRESOLVED CASES: CAN THE ULTIMATE RECOVERY RATE BE CALCULATED?

Recovery Rate is most accurately calculated on closed (resolved) cases, where the outcome is anything from full repayment to complete loss, or something in between. On the other hand, this restriction leads to underestimating recent losses, since only quickly resolved, cured cases (typically with lower losses) can be included (resolution bias). Therefore, the dashboard displays two different cuts of the data: First, the metrics in the tables of the dashboard are calculated restricted to resolved defaults with default date maximum three years prior to the publication date of the dashboard. Second, the metrics in the graph of Recovery Rate by Year of Default includes all default years and projections for unresolved defaults that are displayed in separate lines. By including both observed losses and projections on incomplete recoveries, realistic and up-to-date Recovery Rates, that includes all available Information can be calculated especially for recent years. Please see chapter **Error! Reference source not found.** for more details on the methodology for unresolved defaults.

TABLE 1
REFERENCE DATA SET CREATION

Filter	Stage	Unresolved	Year of Default	Small Default Amount	Incomplete Portfolio	Validation Rules	Nr of facilities	Total
Raw data set w/o filters	initial	14,515	4,681	31,210	14,719	10,036	33,418	255,682
Unresolved	before	14,515	4,681	31,210	14,719	10,036	33,418	255,682
Unresolved	after	-	4,642	29,335	13,252	9,848	30,797	241,167
Year of Default	before	-	4,642	29,335	13,252	9,848	30,797	241,167
Year of Default	after	-	-	28,822	13,252	9,535	30,434	236,525
Small Default Amount	before	-	-	28,822	13,252	9,535	30,434	236,525
Small Default Amount	after	-	-	-	11,181	6,024	22,024	207,703
Incomplete Portfolio	before	-	-	-	11,181	6,024	22,024	207,703
Incomplete Portfolio	after	-	-	-	-	4,983	21,245	196,522
Validation Rules	before	-	-	-	-	4,684	21,199	196,223
Validation Rules	after	-	-	-	-	-	20,903	191,539
Nr of facilities	before	-	-	-	-	-	20,903	191,539
Nr of facilities	after	-	-	-	-	-	-	170,636
RDS	final	-	-	-	-	-	-	170,636
RDS excl. recent years	final	-	-	-	-	-	-	168,022

YEAR OF DEFAULT: HOW TO AVOID THE RESOLUTION BIAS

An important distinction should be identified and accordingly addressed with respect to cases with a short workout period when calculating Recovery Rate. Generally, a short workout period is related to higher recoveries. In the most recent years short workout period cases are naturally overrepresented. This is known as resolution bias. Hence, including all the default years might lead to an unrealistically long-term average Recovery Rate. Therefore, when creating an RDS it is advisable to address the resolution bias by restricting the defaults to those with a reasonable window time for workout processes to conclude. The decision on this filter, like all RDS filters, should be left to the discretion of users and be aligned to what is representative of their own portfolio. For the GCD dataset the average observed workout period is two years and the latest default year available in the LGD/EAD database is 2022. To address the resolution bias caused by cured cases, it is reasonable to restrict data points to defaults up to and including 2019.

A filter is applied on the lower end of the time series in addition to the filter on the upper end. Although the earliest entry in the GCD database dates back to 1983, for some banks it is difficult to deliver all the data elements required to identify cured cases for older defaults consistently with newer defaults. Such data may still be useful for driver analysis, but the lower reported cure rate can tend to bias the resulting pre-2000 data such that the reported Recovery Rate is lower than it would have been in a full data set. In this report where the absolute level of the resultant recoveries is important because long term averages are calculated, defaults that occurred prior to 2000 are excluded.

SMALL DEFAULT AMOUNT: ARE SMALL DEFAULT AMOUNTS RELEVANT?

Default amounts in the GCD database range from zero (e.g. for uncalled contingent facilities) to several hundreds of millions of Euro. For an appropriate setup, banks are advised to compare the default amount structure to their internal portfolio. For this exercise default amounts below 100,000 EUR are excluded as they are deemed to be not representative of large corporate defaults.

INCOMPLETE PORTFOLIO: HOW TO DEAL WITH FORMER MEMBER BANK DATA

When a member bank resigns from the association and/or from a Data Pool, the most recent defaulted years that they have submitted must be incomplete as they would no longer participate to submit/update their defaults. The incomplete data contains only cases with short time to resolution which might be affected by the resolution bias. Therefore, the last three years of data of former member banks are filtered out of the RDS.

VALIDATION RULES: HOW TO DEAL WITH OLDER DATA

As described above, GCD applies a series of validation rules during the submission process which prevents inconsistent or incomplete data from being accepted automatically. This is the major data quality insurance that protects the database. The validation rules are updated and amended as required by our members for every submission. That said, some entries were integrated into the database before certain validation rules had been implemented. Those entries can still be part of the database if not updated by the member bank. GCD policy is to not remove any data as it may still contain useful information. However, for this exercise, data points with errors that affect the integrity of the database (e.g. the event date at default must be the same for all facilities of a given borrower) or the correct calculation of Recovery Rate (e.g. balancing the cash flow between the transaction and the history table) were excluded. Due to the GCD rule that every bank must update their full data submission at least every three years, there are only a small number of entries removed in this filter step.

The remaining validation rules that are triggered deal with completeness. They check e.g. if for a certain collateral type, year of construction is given. Where data was submitted before the rule was applied, the information is sometimes not provided. Since that data is not wrong it is included in the RDS. Correctness items check for example, if several connected fields are consistently filled: such as if for a syndicated loan a total syndicated amount and a currency are given. If they do not deal with crucial information used in this report the data is included in the RDS.

NUMBER OF FACILITIES PER BORROWER: HOW TO DEAL WITH FACILITY WEIGHTING EFFECTS

Borrowers with many loans which might cause overweighting are removed. The 60% of the full dataset for corporates consists of one facility per borrower. About 98% of the dataset has less than 10 facilities attached to one borrower. There are outliers with over 200 loans attached and these lower the overall average Recovery Rate due to the bimodal left skewed shape of the distribution. These are verified loan constructions which relate to specific commercial circumstances including multi-family housing with one home per apartment or equipment

finance with one loan per vehicle in a fleet. Inclusion of these would overweight the importance of a single collateral. Therefore, for more homogeneity, a filter on borrowers with 10 or more facilities is applied at obligation level.

About Global Credit Data

At GCD, our mission since 2004 has been to empower banks and the financial industry with a deep understanding of credit risk through a unique data source. As a non-profit organization owned by 50+ member banks, we collect valuable data directly from banks' books.

GCD's activities revolve around pooling credit loss data, particularly from low default portfolios. Beyond data pooling, we foster knowledge exchange, facilitate research, and provide information sharing services, creating a dynamic environment for insights and collaboration.

Join our community to access exclusive data insights, gain market understanding, and benchmark your performance against industry peers.

For additional information, please contact

Nina Brumma
Head of Analytics and Research
nina.brumma@globalcreditdata.org

www.globalcreditdata.org

Global Credit Data
by banks for banks