# PD & Default Rate Report: Large Corporates

# Appendix: Database & Methodology

**Global Credit Data**

*by banks for banks*

# Contents

# 1.    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), 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 250,000 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.

# 2.    The GCD PD & Rating Database

The Global Credit Data PD & Rating platform was created in 2009 and has grown rapidly ever since. It now gathers detailed information on estimated Probability of Default, internal ratings migrations and observed defaults over the entire banking portfolio from the platform participants. 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 returned to participants, at an aggregated risk bucket level and also at granular obligors.
- **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 30 banks across the span of over 15 years. Most of the data is collected from 2008 onwards, but some segments are available as early as 2002. Global Credit Data returns the complete database to member banks at a finely detailed risk bucket (described later). The information returned is always anonymized, excluding any identification of the lender and of the borrower. At the risk bucket level,

members receive some pre-calculated metrics (average PD, observed default rate at multiple time horizons, migration matrices, etc.). The granular data return at obligor level enables participating banks to customize the calculation of the metrics.

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 more. 29 banks in total have contributed to this asset class, largely driven by the fact that nearly all members have substantial lending to large corporates.

The purpose of the database is for participants to benchmark their credit position in comparison to other banks and the evolution of risk in their portfolios. Due to the "give to get" rules applying, members only receive back data for the reporting years and asset classes which they submitted. Members are then free to sample the database, clean the data, calculate their own metrics and perform relevant analyses. 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. Definitions

## LARGE CORPORATES

Large 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, neither as part of the SMEs – small and medium-sized enterprises – asset class, described in paragraph 273 as corporate exposures where the reported sales for the consolidated group of which the firm is a part is less than €50m.

## 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 |
| Construction | Construction |
| Hotels and Restaurants | Hotels and Restaurants |
| | Hotels |
| | Restaurants and Food services |
| Manufacturing | Manufacturing |
| | Manufacturing - Automotive |
| | Manufacturing - Chemicals |
| | Manufacturing - Food and Beverage |

| Group in Dashboard | Detailed Group in Database |
|---|---|
| | 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 |

## DEFAULTED OBLIGORS

The Basel definition of default is used. According to the GCD Data Pool Regulations, all Pool participants must report their entire portfolio per asset class, therefore also all the observed defaults in the portfolio. This requirement is verified by GCD during the audit phase during the submission cycles.

Following the calculation methodology in the PD&Rating platform, which has been agreed with the PD&Rating subcommittees and validated by the Methodology Committees; different types of defaults are observed in the database:

- Standard defaults: a borrower which is part of the portfolio at the beginning of the year defaults during the year

- Multiple defaults: a borrower defaults more than once: the defaults will only be counted as once per calendar year.
- Quick defaults: a borrower default during a year where that same borrower was not present in the portfolio at the start of the year. These quick defaults are not included in the numerator of the default rate in this report.

## COHORT SIZE

In the context of default rate calculation, the cohort size is defined as the count of performing obligors on January 1st of each year. This is equivalent in the dataset to the count of borrowers in the fourth snapshot (fourth quarter) of the previous year. Although the count of borrowers change throughout the year (with new obligors entering the portfolio, exiting obligors leaving the portfolio), only the count of performing ones at the start of the year is used for the default rate calculation, therefore it is the used and reported as cohort size.

When reported over multiple years, yearly Cohort Size (as explained here above) are averaged.

## OBSERVED DEFAULT RATES

The default rate is calculated as the ratio between the defaulting borrowers during the period of interest (in this report, always on a 1-year horizon) and the count of performing borrowers at the start of the year in the portfolio of interest. A performing borrower is an obligor for which the internal rating is not defaulted at a specific reporting date.

The default rate for a given year is calculated exclusively based on the count of performing borrowers at the beginning of that year, without considering a rolling calculation across all quarters.

Quick defaults are not considered in the numerator of the ratio because those borrowers were not part of the portfolio at the start of the period which is used as the denominator.

$$\boldsymbol{Yearly\ Observed\ Default\ Rate} = \frac{Defaults\ occuring\ within\ the\ year}{Performing\ borrowers\ at\ start\ of\ the\ year}$$

The overall default rate is the simple average of the yearly default rates across all years.

## RISK BUCKETS

To facilitate the analyses of the database, different metrics are pre-calculated in the different files of the aggregated data return. Therefore, users of the platform can rely directly on the methodology that has been agreed by members of Global Credit Data (GCD) to obtain, for example, the historical default rate in a certain country or industry. Therefore, the count of defaults, amongst other metrics, is available directly at a certain level of aggregation: the risk bucket.

The risk bucket is defined by the following fields: the asset class, industry and country of the borrower and the internal rating assigned by the lender to the borrower.

**PD AND OBSERVED DEFAULT RATE FOR LARGE DASHBOARD CORPORATES**

## RATING GRADE / INTERNAL RATINGS

The data collected in the PD&Rating platform of Global Credit Data are internal information from each of the participating members. This provides the participants with a complete perspective of the bank's lending portfolio including smaller companies not rated by external parties.

The Trough-the-cycle PDs and internal ratings are estimated by each of our Basel-compliant members. To enable comparability between the different internal rating scales, each member is requested to convert their internal ratings to the S&P equivalent, in accordance to their own regulatory external reports.

## REFERENCE DATA SET

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

While the current dashboard takes a complete perspective of the available data in the pool, it is strongly recommended for member banks to select the data in accordance to their own portfolio profile. Participating bank are welcome to contact the GCD team for such representativeness preparations.

## REGION

The GCD data set offers country information are collected on country level (country of residence) 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

**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 started collecting historical loss data in 2004, to which member banks have exclusive access. This database now totals more than 250,000 non-retail defaulted loan facilities from around the world.

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 place our databases as the global standard for credit risk data pooling.

For additional information, please contact

*Olivier Plaetevoet*
PD & Rating Executive
olivier.plaetevoet@globalcreditdata.org

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

www.globalcreditdata.org

**Global Credit Data**
*by banks for banks*