NPLMARKETS

17 January 2020[1]

Burkhard Heppe, PhD
Chief Technology Officer
NPL Markets Ltd.
burkhard.heppe@
nplmarkets.com
+44 (0) 20 3984 6288

Collecting historical recovery data for transactions of non-performing loans

Overview

In this article, we explain the ongoing work on the provision and collection of standardised recovery cash flow data for non-performing loans (NPL). We will first compare a specific part of the EBA NPL and ESMA NPL securitisation data template, the table concerning historical collections, with other available data sets collecting individual cash flows from the workout of defaulted loans. Second, we point out the importance of such cash flow data for NPL ABS investor reporting in addition to bank model development and NPL valuation. The new ABS reporting requirements from ESMA impose a major challenge on NPL ABS data providers as such transactions often securitise a mix of different asset classes requiring detailed loan-level reports including historical recovery cash flows. The NPL add-on template also requires many additional fields on the legal workout process and the collateral. Finally, we propose a simple statistical method to capture the information from collection data for loss modelling and NPL valuation. The recovery curves derived from collection data can be used in the valuation analytics of the NPL Markets platform, which values loans based on standardized input data sets following the EBA NPL template.

Loan-level data templates in Europe

We compare different loan-level data templates in Europe. Other data templates exist for financial and regulatory reporting, but those data sets do not include individual recovery cash flows and are not intended to be shared with investors or other financial institutions.

Table 1 compares five different loan-level data standards, the two data templates from the European securitization market defined by the ECB and ESMA, respectively, the EBA NPL template, the collection

---

[1] With revisions on February 7, 2020

of loan loss data from Global Credit Data (GCD), and the supervisory collection of loan-level data for the Eurosystem (AnaCredit)[2]. GCD is a not-for-profit association of international banks, which started 15 years ago in Europe, but now collects loan loss data globally. Whereas the ECB ABS and AnaCredit templates are designed to report consecutive snapshots of loans including some loss and recovery data, the EBA NPL and GCD templates follow a more complex relational structure with separate data tables dedicated to historical cash flows. The ESMA ABS template requires portfolio snapshots for performing loans securitisations whereas NPL ABS also require the provision of historical collection cash flows. As of year end 2019, only the ECB ABS and GCD bank loan loss data collections offer meaningful data volumes for banks or investors. The AnaCredit data are confidential and accessible only to central banks. The reporting under the ESMA template started in 2019 while the collection of data in central repositories is expected for 2020. There is currently no public repository for EBA NPL data. In the following, we summarize the five templates and the two existing data sets of loan loss data.

Template	Published	Main user base	Asset classes	Data volume	Collected by
ECB ABS loan-level initiative	ECB 2013	ABS investors	RMBS, SME, Auto, Consumer, Cards, CMBS	>35m loans from >1,200 ABS transactions	European DataWarehouse eurodw.eu
ESMA ABS	ESMA 2018	ABS investors	15 templates including NPL ABS	In progress	ESMA and Securitization Repositories
EBA NPL	EBA 2017	Buyers and sellers of NPL	All non-performing loan types	In progress	Under discussion
Global Credit Data (GCD)	Since 2004, data shared with GCD members only	Global AIRB banks to calibrate LGD models	10 wholesale asset classes, including large corporate, SME and Real Estate Finance	>100,000 defaulted loans	globalcreditdata.org, non-public data
AnaCredit	AnaCredit Regulation EU 2016/867	Most banks and branches located in the Euro zone	Loans to entities that are not natural persons, > Euro 25,000	Not available yet	National central banks and ECB, non-public data

Table 1: Selected loan-level data collection initiatives

---

[2] ECB: European Central Bank, EBA: European Banking Authority, ESMA: European Securities & Markets Authority

ECB ABS loan-level initiative

The ECB ABS loan-level initiative started in 2013 and established specific loan-by-loan information requirements for ABS accepted as collateral in Eurosystem credit operations. Its main objective is to increase transparency by making more timely information on the underlying loans and their performance available to market participants in a standard format. In addition, the loan-level data initiative facilitates the risk assessment of ABS used as collateral by Eurosystem counterparties in monetary policy operations.

Since its inception in 2013, the data collected by the main repository, the European DataWarehouse, has increased to 35.5 million active loans of which 71% relate to residential mortgages (EDW Data Availability Report Q1 2019). Dutch, French, Spanish, UK and Italian mortgages are the largest mortgage portfolios followed by Spanish small and medium enterprise loans (SME), German auto loans and Italian SME and leases.

Some loss and recovery information is available especially for residential mortgages, but the data require significant cleansing and preprocessing. The data quality in EDW overall has improved since the inception of the ECB initiative, but loss data is relatively rare due to the recent benign credit conditions and the reporting rules for inactive loans. Some of the reported loss and recovery data lack consistency. Loss data from securitization transactions may be biased vis-à-vis all bank loans as ABS originators go through a loan selection process meeting certain eligibility criteria and sometimes securitized loans are repurchased by the originator prior to repayment. While EDW provides rich data for many retail assets classes and countries, there is practically no data for large corporate loans, specialized finance or commercial real estate.

The ESMA securitisation reporting templates

ESMA published loan-level data reporting templates for ABS in August 2018 as an annex to the technical standards on disclosure requirements under the Securitisation Regulation (EU) 2017/2402. The ESMA templates are expected to replace the ECB ABS templates later in 2020. The XML reporting schema was initially published by ESMA in July 2019 and updated on 20 December 2019 to reflect feedback from the European Commission. The standard reporting requirements are expected to become effective towards the end of the first quarter 2020. The ESMA templates are based on the ECB ABS templates with several changes. ESMA published 15 templates including a template for NPL securitisations. ESMA dropped the concept of optional data fields with all ESMA fields being mandatory. Some of the optional fields in the ECB templates were converted to mandatory and other optional fields have been removed from ESMA ABS. For example, the Amortization Profile section in the ECB template for SME has been removed. The

additional NPL-related fields that need to be reported for NPL ABS have a direct equivalent in the EBA NPL data template discussed in the next section. Other fields relating to performing loans have equivalent fields in the EBA NPL and ESMA ABS templates, but use different field names and definitions.

The new reporting requirements impose a major challenge on NPL ABS data providers for several reasons. NPL ABS did not report under the ECB ABS standard. Whereas most performing loan securitizations reference a single asset class like residential mortgages and hence will use a single ESMA template, securitized NPL pools are often a mix of secured and unsecured loans to corporate or retail clients which each have different ESMA templates and the reporting requirement applies loan by loan. The NPL add-on template requires many additional fields on the legal workout process and the collateral in addition to the historical collection information. The combination of performing loan data fields plus detailed NPL add-on fields will require NPL ABS data provider to change and increase the level of detail in their quarterly investor reports for those NPL ABS issued after January 1, 2019. The historical collection data fields required by ESMA for NPL ABS are shown in Table 2. The table shows the name and description of the historical fields and the use of no data options (ND1-ND5). The no data option is allowed in principle for many NPL fields including collections, however, ESMA made it clear in their published Q&A that this does not mean that these fields are optional. All ESMA fields are mandatory and the use of no data options must be carefully justified. Table 2 shows the requirement to report historical collections and unpaid balances on a monthly basis for the last 36 months.

Historical collections information section				
FIELD CODE	FIELD NAME	CONTENT TO REPORT	ND1-ND4 allowed?	ND5 allowed?
NPEH1	Unique Identifier	Report the same unique identifier here as the one entered into field NPEL1.	NO	NO
NPEH2	Underlying Exposure Identifier	Unique underlying exposure identifier. This must match the identifier in field NPEL3. The reporting entity must not amend this unique identifier.	NO	NO
NPEH[3-38]	Legal Unpaid Balance at month n	History of total legal unpaid balance in the thirty-six months previous to the data cut-off date, each monthly amount reported in a separate field. Start with the most recent month in field NPEH3 and end with the oldest month in NPEH38.	YES	YES
NPEH[39-74]	History of Past-Due Balances at month n	History of total past-due balance in the thirty-six months previous to the data cut-off date, each monthly amount reported in a separate field. Start with the most recent month in field NPEH39 and end with the oldest month in NPEH74.	YES	YES
NPEH[75-110]	History of Repayments - Not from collateral sales at month n	Repayment made by the obligor in the thirty-six months previous to the data cut-off date, excluding collateral sales, including collections by external collection agencies, each monthly amount reported in a separate field. Start with the most recent month in field NPEH75 and end with the oldest month in NPEH110.	YES	YES
NPEH[111-146]	History of Repayments - From collateral sales at month n	Repayment made by the collateral disposal in the thirty-six months previous to the data cut-off date, each monthly amount reported in a separate field. Start with the most recent month in field NPEH111 and end with the oldest month in NPEH146.	YES	YES

Table 2: Historical collection data defined by ESMA NPL ABS reporting standards in Annex 10 (Underlying exposures - add-on non-performing exposures).

The EBA NPL template

The EBA NPL template was published in late 2017 and a slightly revised version in September 2018. The EBA NPL template was published to create a common data standard for the screening, due diligence and valuation of NPL transactions with the aim to make the secondary market for European NPL more efficient by overcoming information asymmetries between buyers and sellers. The EBA NPL template aims to cover most bank loan asset classes including residential and commercial mortgages, corporate loans and SME, unsecured consumer loans, leasing and asset-based finance as well as specialized finance. Across all asset classes EBA has defined 460 fields of which 155 fields are deemed critical. The use of the EBA NPL template is currently not mandatory. Figure 1 shows the main tables of the EBA NPL template. In addition to many data fields relating to the borrower, the loan and any property collateral, which are also found in the performing loan securitisation reporting templates of ESMA, there are tables dedicated to non-property collateral which includes financial guarantees, a table for forbearance measures, for external collections and an enforcement table with details on the legal procedure and the foreclosure of collateral.
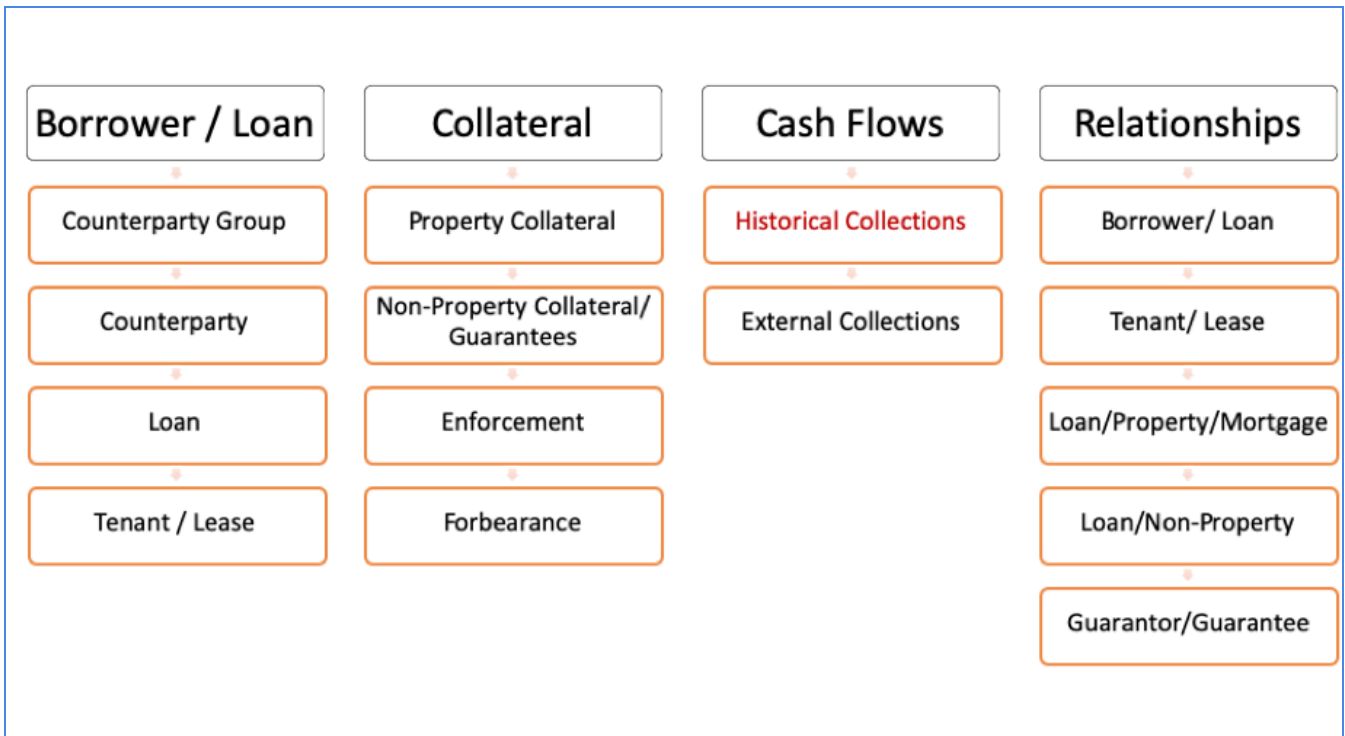

Figure 1: Main data tables in the EBA NPL template.

Table 3 shows the data fields in table 8 of the EBA NPL template called Historical Collections and Repayment Schedule. The fields 8.010 to 8.012 are only applicable to unsecured retail loans. While the

fields related to forward-looking repayment schedules (8.002-8.004) are of interest to performing and re-performing loans or where the borrower operates under a payment schedule, such fields are generally not relevant where the loan is non-performing and has been terminated. In contrast, the history of repayments (8.007-8.009) is relevant for non-performing loans, but less relevant for performing loans. The histories of balances and repayments are the same fields used in the ESMA NPL add-on for historical collections (Table 2). Like ESMA, EBA requests all histories and schedules in the Historical Collection table to be provided monthly covering a period of at least three years. In practice, we see NPL sellers providing a table with total historical collections (8.007), but we rarely see the break out by asset sales (8.008, 8.009) or a history of unpaid or past-due balances (8.005, 8.006). The table External Collections does not appear to play an important role in the current practice of non-performing loan sales.

Index	Field	Importance
8.000	Contract Identifier	Identifier
8.001	Instrument Identifier	Identifier
8.002	Total Repayment Schedule	Important
8.003	Principal Repayment Schedule	Important
8.004	Interest Repayment Schedule	Important
8.005	History of Legal Unpaid Balances	Critical
8.006	History of Past-Due Balances	Critical
8.007	History of Total Repayments	Critical
8.008	History of Repayments - Not From Asset Sales	Critical
8.009	History of Repayments - From Asset Sales	Critical
8.010	History of Outstanding Balances Since Charge-off	Important
8.011	History of Charges/Interest Since Charge-off	Important
8.012	History of Payments Since Charge-off	Important

Table 3: Historical collections and repayment schedule from EBA NPL template v1.1 (EBA 2018).

Global Credit Data's LGD database

GCD is a non-profit association owned by 55 member banks with the aim to help banks better understand and measure their credit risks through data pooling and benchmarking activities. GCD started collecting historical loss data in 2004 and member banks only have access to the data on a "give to get" basis. The database now totals over 185,000 non-retail defaulted loan facilities from around the

world. While the data is private to members, GCD regularly publishes reports about its activities and selectively makes data available for academic research.

GCD member banks wishing to sell wholesale NPL should consider using GCD data in their communication with investors. Figure 2 shows the different event dates for which GCD collects data prior and post default. Post default until resolution, GCD records each cash flow when it occurs. Figure 3 shows the distribution of ultimate LGD for large corporate loans and highlights the importance of the cure rate (GCD 2019).
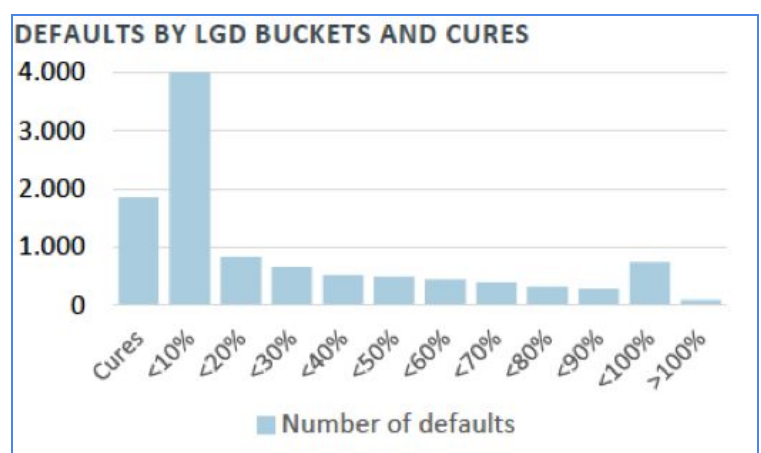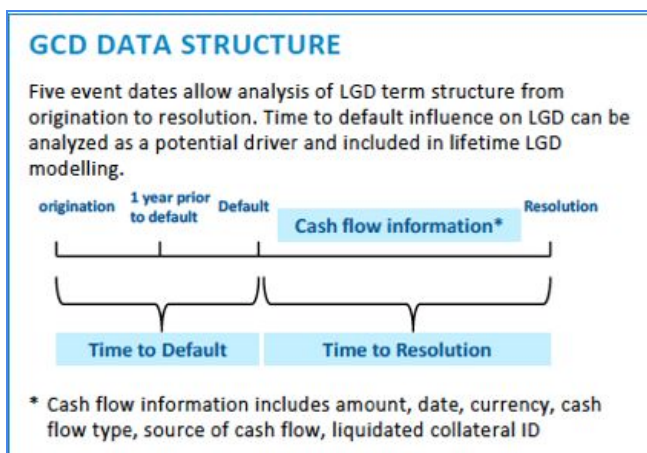
Figure 2: GCD recovery cash flow data structure.   Figure 3: LGD distribution for large corporate loans
(copyright GCD 2019, reproduced with permission).[3]

The source and nature of the recovery cash flows is often a difficult item to collect accurately from the source systems of banks especially for legacy defaults that were resolved in the past. GCD continues to improve the comparability of their data and good data consistency not only requires a well-defined standard but a continued effort from all data providers to improve data quality. For instance, data from different banks could differ in the definition of what constitutes a real estate finance loan or which loans are classified as unlikely to pay. The GCD template is more detailed than the EBA NPL template regarding the collection of historical cash flows. The transaction types collected by GCD after default include legal and other expenses, charge-offs and provisions and any further advances on undrawn commitments.

The EBA NPL template provides data for the sale of performing and non-performing loans. GCD in contrast collects historical data for modelling and benchmarking purposes and most of the data collected by GCD relate to resolved defaults where the workout process has been completed. Resolved defaults would not normally be considered for sale by banks as the loan is repaid, re-performing or no further recovery proceeds are expected with residual claims written off. Other important differences

---

[3] The views expressed in this article are those of the author and do not necessarily represent those of Global Credit Data.

between EBA NPL and the GCD template include additional enforcement and forbearance data fields in EBA, with EBA also collecting more detail on property collateral. GCD with its focus on wholesale asset classes in turn requests more detailed data on asset finance and specialized finance (e.g. aircraft, shipping, commodity and project finance).

AnaCredit data collection by the Eurosystem

AnaCredit (analytical credit data sets) is a data set containing detailed information on individual bank loans in the eurozone, harmonised across all EU member states. The project was initiated in 2011 and came into law as Regulation EU 2016/867 on the collection of granular credit and credit risk data. The data collection started in September 2018. The objective is to improve the statistical information basis for the Eurosystem. Credit institutions and branches of non-resident banks located in the euro area must report loans to legal entities exceeding EUR 25,000 on a monthly or quarterly basis. Loans to private households are currently not covered.

AnaCredit requires around 100 data fields for each exposure, including information about the counterparty (covering some financial income and balance sheet information), the instrument, collateral or other protection and risk mitigation measures. The table on accounting data includes the status of forbearance measures, cumulative impairments and the cumulative recoveries after default. If banks report the cumulative recoveries accurately then period-by-period recovery cash flow and unpaid balances since default can be calculated at instrument level similar to those required in the EBA or ESMA NPL table for historical collection data. Hence, banks that have aligned their IT infrastructure for regular and accurate reporting under the AnaCredit requirements should manage to supply historical collection data (albeit without further detail on the source of the recoveries) to NPL investors without much additional effort. Table 4 shows the data fields of the accounting table in the AnaCredit template with the information on recoveries and forbearance status highlighted.

Use of historical collection data for NPL valuation and modelling

Historical collection information is relevant for many use cases. The collection data can be used by banks or specialist servicers to monitor the performance of the loans and compare against projected recovery cash flows estimated during the acquisition of the portfolio or upon transfer of the loans to the workout department. Rating agency of NPL securitizations require updated collection histories to compare the current performance of the portfolio with their expected recovery curves when rating the portfolio at issuance. Regular cash flows from non-performing borrowers indicate that the borrower is willing to engage with the lender and/or is following a payment plan.

GCD collects historical recovery cash flows for benchmarking and modeling of bank regulatory risk parameters such as LGD and EAD and IFRS 9 loan loss provisions. Investors in NPL portfolios can use

historical collections to better understand the borrower behavior and status of the borrower workout process. If borrowers adhere to a payment plan, then such payments can be projected into the future (possibly with some risk adjustments). If the seller has already collected cash flows from certain assets sales or guarantees, then those cash flows cannot reoccur in the future and must be excluded from future projections.

AnaCredit data set	Data attribute	Level of Granularity	Frequency
06. Accounting data	Reporting agent identifier	Instrument	Quarterly
	Observed agent identifier		
	Contract identifier		
	Instrument identifier		
	Accounting classification of instruments		
	Balance sheet recognition		
	Accumulated write-offs		
	Accumulated impairment amount		
	Type of impairment		
	Impairment assessment method		
	Sources of encumbrance		
	Accumulated changes in fair value due to credit risk		
	Performing status of the instrument		
	Date of performing status of the instrument		
	Provisions associated with off-balance sheet exposures		
	Status of forbearance and renegotiation		
	Date of the forbearance and renegotiation status		
	Cumulative recoveries since default		
	Prudential portfolio		
	Carrying amount		

Table 4: Accounting data table of the AnaCredit data template from the European Central Bank.

We previously analyzed the timing and likelihood of recovery cash flows. In contrast to static models of ultimate recovery, we pursued a dynamic approach in modeling the timing of recovery cash flows after default. Our approach is based on a survival analysis that helps banks to include recovery and loss data of loans for which the workout process is continuing, so called unresolved defaults. Other applications of our analysis include models for risk provisions and the optimization of workout processes. We also show how investors can include our findings in a dynamic analysis of ABS tranches or for the valuation of NPL portfolios.

The key contribution is to develop a consistent statistical framework for loan recoveries based on the concept of censoring and considering recovery and loss allocation as competing risks. We believe the methodology is the first to capture further advances (negative recoveries through left truncation), unresolved defaults (through right censoring) and consistent estimation of cumulative recoveries in a competing risk framework where repayment and loss allocation are treated as mutually exclusive events.

We noted that the cumulative recovery as a percentage of EAD can be approximated well by a simple exponential function. Such an exponential recovery curve can be determined by two parameters, the weighted average life (WAL) of the recovery cash flows and the ultimate undiscounted recovery rate (REC, the asymptotic value for long workout times). The competing risk framework models the recovery process as a Markov chain with constant hazard rates for loss and recovery resulting in such exponential recovery curves. The constant hazard rates can be regressed against loan, borrower and collateral characteristics or determined by expert judgment if adequate data is missing. The results are closed-form recovery curves in discrete or continuous time with just two parameters, WAL and REC. The formula can be adjusted to generate future recovery cash flows from a reference date after default.

An example of parametric recovery curves are shown in Figure 4 which shows the observed and modelled recoveries for three different Italian retail NPL portfolios. The graph shows the cumulative recovery for each quarter after default as a percentage of exposure at default.[4] The two parameter model fit is reasonable and captures the main recovery dynamic well. Pool 1 (dark blue) has a much lower ultimate recovery rate than Pool 2 (light blue) and Pool 3 (red) and a shorter weighted average life of less than 1 year, whereas Pool 2 and 3 have an average life of more than two years. If the three pools were sold 10 quarters after default, then Pool 1 would attract a lower price than the other pools given the flatness of the extrapolated curve. Pool 3 should attract a higher price than Pool 2 based on the extrapolated curve despite Pool 2 having shown higher recoveries until the point of sale at quarter 10.

In summary, historical recovery cash flow data are important for the development of statistical LGD models and help refine the valuation of NPL. NPL ABS are required to report such cash flows for the last three years according to the ESMA reporting standard at loan level. First, the historical data is clustered and the assumption of a constant loss hazard resulting in the simple exponential shape of the recovery curve is tested visually and statistically for each cluster. If the hazard is not constant other more flexible parametric functions like splines or non-parametric estimates can be used. Non-parametric hazards are well known in survival analysis like the Kaplan-Meier estimator or the non-parametric baseline hazard in the Cox regression model (please contact us for details).

---

[4] NPL sale transactions typically refer to gross book value rather than exposure at default.
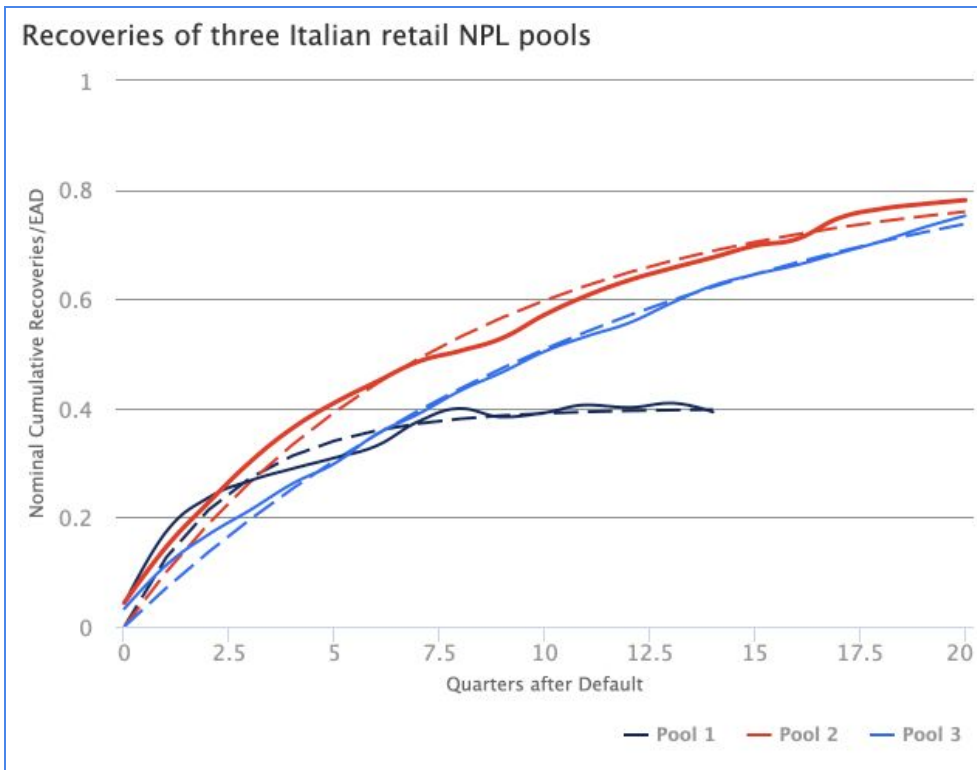
Figure 4: Observed recovery curves for three Italian NPL and model fit.

Second, the model parameters, the transition hazards for loss and recovery, are calibrated to data or determined by expert judgement. Expert judgment is easy to elicit for the intuitive quantities weighted average workout time WAL and ultimate undiscounted recovery REC, respectively. Third, the model is used to project future recovery cash flows which can then be discounted at an appropriate discount rate. For NPL investors, the discount rate will be the expected internal rate of return. For bank LGD models, the appropriate discount rate is a matter of debate and some bank regulators prescribe the discount rate which banks must use.

Often loss recovery data is scarce and the investor may wish to use a combination of expert judgment and available collection data. A sound statistical way to combine a prior assumption on the key parameters with available data is given by Bayesian statistics where the investor assigns a prior mean and uncertainty (or sample size) to the parameter and then uses the Bayes formula to update the prior with the likelihood of the observed data.

In this article we have summarized the different data templates of loan recovery data currently in use across Europe. We explain why this data is important for the valuation of NPL and outline an intuitive yet powerful way to model recovery curves of granular loan portfolios. Providing accurate and detailed data to investors to overcome the asymmetry of information is a great challenge that a detailed data

standard like the EBA NPL template aims to overcome. We highlight that banks who wish to sell their distressed loans can benefit from their existing data gathering efforts for LGD models, for GCD data sharing or the recent AnaCredit reports created for central banks. The recently adopted templates from ESMA for NPL securitisation will greatly increase the level of detail reported to investors. Bank originators, sponsors or special purpose vehicles who have issued NPL ABS in 2019 or are planning to do so going forward must comply with securitisation regulation and the technical reporting standards of ESMA which are expected to come into force by March 2020.

References

EBA 2018: Revised standardized NPL data templates
https://eba.europa.eu/-/eba-revises-standardised-npl-data-templates

European Central Bank. AnaCredit
https://www.ecb.europa.eu/stats/money_credit_banking/anacredit/html/index.en.html

European DataWarehouse 2019, Data Availability Report Q1 2019,
https://eurodw.eu/wp-content/uploads/EDW_Data_Availability_Report_20190917_web.pdf

European DataWarehouse 2018, Gap analysis between the ECB ABS loan-level data templates and the draft templates proposed by ESMA on 22 August 2018. Version 2.0
https://www.eurodw.eu/wp-content/uploads/EDW_ED_GAP_Analysis_SAMPLE-VERSION-ONLY_9.2018.pdf

Global Credit Data 2019: LGD Report 2019: Large corporate borrowers. Global Credit Data.
https://www.globalcreditdata.org/system/files/documents/gcd_lgd_report_large_corporates_2019.pdf

Disclaimer