EUROPEAN CENTRAL BANK

EUROSYSTEM

Occasional Paper Series

João Gaudêncio, Agnieszka Mazany, Claudia Schwarz

The impact of lending standards on default rates of residential real estate loans

No 220 / March 2019

Contents

Abstract

This paper analyses the impact of lending standards for residential real estate (RRE) loans on default rates, using a novel loan-level dataset from the European DataWarehouse (EDW) that covers eight euro area countries. To the best of the authors' knowledge, this paper is the first to use, for this purpose, a consistent set of loan-level data on loans originated in multiple euro area countries. Previous literature has used either national loan-level data, which does not allow for cross-country comparisons, or aggregate cross-country data. The dataset is first explored through an extensive descriptive analysis and this is followed by static probit regressions. The findings confirm the key influence of lending standards – in particular, loan-to-value and loan-to-income ratios at origination, original loan maturity and borrower employment status – on loan default rates. The impact of other variables, such as interest rate fixation and payment type, varies depending on the country of loan origination. These results are particularly relevant for microprudential supervisors in their ongoing assessment of banks' credit policies. The highlighted country specificities should be taken into account in macroprudential policymaking.

Keywords: loan defaults, lending standards, residential real estate, loan-level data, default probability

JEL codes: C25, G21

Executive summary

Following the financial crisis, non-performing loan (NPL) ratios in many euro area countries increased considerably. However, while this is to be expected in a recession context, NPL ratios are still high ten years later.

From a microprudential perspective, bad credit can have adverse impacts on banks' capital, profitability and funding costs. First, NPLs require risk-weight penalties, so banks have to put aside capital, which is then unavailable for other purposes. Second, high NPL levels imply a large set of non-income-generating assets, thus adversely affecting profitability. Lastly, these conditions can increase funding costs, as investors require a larger risk premium, on account of the information asymmetry on the quality of the loan book.

Therefore, identifying what is driving NPLs is paramount from a microprudential supervision perspective. While macroeconomic trends certainly feature among the drivers, supervisors typically treat the general economic climate as an exogenous factor when assessing banking risks and supervisory measures. However, microeconomic factors such as banks' lending policies are within a supervisor's grasp. These represent an important supervisory target, as they influence a bank's loan portfolio performance in a given economic environment.

Lending for residential real estate (RRE) is a major activity for euro area banks, so that market can be considered of great importance from both a microprudential and a systemic risk perspective, given its ties with the real economy and the financial sector.

With this in mind, the present paper seeks to assess the impact of certain lending standards on the default rates/probability of RRE loans. To do this, a novel dataset from the European DataWarehouse (EDW) was used, consisting of loan-level data on RRE loans from eight euro area countries.

For each unique loan, EDW reports several lending standard measures, including both borrower and loan-specific characteristics. Using these, this paper addresses the main research question in two steps: first, through a rich descriptive analysis at euro area aggregate and individual country levels; and second, by assessing, in a probit regression setting, the magnitude and significance of the patterns discovered in the first step. While the sample consists of safer than average loans, meaningful results can still be drawn from this extensive and granular dataset. The findings are not a description of the general lending habits of banks in the euro area, but drivers of default conditional on specific loan characteristics as described in the paper.

It was found that, even when controlling for a range of other factors, for an average borrower, a 10 percentage points increase in loan-to-value (LTV) ratio at origination (OLTV) raises the default probability by 0.2 percentage points. Increasing the borrower's loan-to-income (LTI) ratio by 1 elevates the risk of defaulting by 0.1 percentage point. The maturity of the loan also appears to act as a default driver: each additional year to maturity raises the default probability by 0.1 percentage point.

Loans declared for renovation or remortgaging purposes seem to default more than those for house purchase. The paper confirms that default is more likely when the borrower is unemployed or self-employed, rather than when she is employed by a third party. Loans to legal entities seem subject to particularly high default rates.

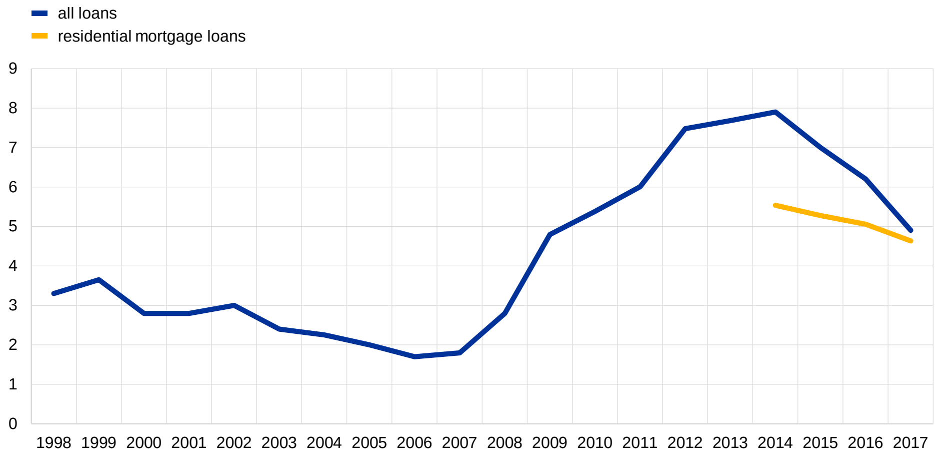
On the other hand, when controlling for LTI, the borrower's income was not found to be significant in the probit. This indicates that the borrower's wealth in relation to the size of the loan is a better indicator of default probability than the level of income itself. Lastly, floating-rate or hybrid loans were deemed to default more frequently than fixed-rate ones.

1 Introduction and literature review

The global financial crisis and the euro area sovereign crisis left euro area banks with large volumes of non-performing assets. The aggregate NPL ratio of "significant institutions" in the euro area peaked in 2014 at well above pre-crisis levels (Chart 1), as the slowdown in economic growth and slump in property prices depleted the value of collateral and weighed heavily on the quality of banks' balance sheets. While a reduction in the euro area aggregate NPL ratio and volumes has been observed in recent years, a significant number of institutions still display high NPL ratios and tackling NPLs remains a key priority for European Banking Supervision.

Chart 1
Non-performing loan ratio in the euro area

(x-axis: years; y-axis: NPL ratio in the euro area, in percentage, per year)



Sources: Up to 2013 World Bank, bank non-performing loans to gross loans for euro area, retrieved from FRED, Federal Reserve Bank of St. Louis; from 2014 onwards ECB supervisory banking statistics (based on "significant institutions" only).

Large stocks of NPLs have significant negative implications, from a systemic, macroeconomic and (most relevant for the purposes of this paper) microprudential point of view. As regards the latter, a high NPL ratio often puts considerable constraints on a bank's ability to operate optimally, as mentioned in BIS (2017), through:

- capital restrictions – NPLs require a higher risk-weight than performing loans, leading to increased capital requirements. The extra capital becomes tied up, whereas it could otherwise be used for other profitable opportunities including, for instance, further lending;
- decreased profitability – banks need to set provisions aside when a loan becomes non-performing. This adversely impacts net income, thus reducing profitability. Furthermore, NPLs generate less revenue than a normal, performing loan portfolio; and

- increased funding costs – potential investors tend to require a larger risk premium on account of lower expected profits and constrained capital.

The prolonged period of very low interest rates and subdued bank profitability, coupled with the ensuing search for yield, might have resulted in excessive risk-taking by banks, increasing the potential for a build-up of future NPLs. As microprudential supervision aims to ensure the safety and soundness of individual institutions, assessing the drivers of NPLs in order to gauge the riskiness of supervised banks' loan portfolios has long been a major supervisory concern. Certainly, the overall macroeconomic environment, real GDP growth and house price developments play a pivotal role in driving NPLs. For instance, Beck et al. (2013), Glen and Mondragón-Vélez (2011) and Jordà et al. (2013) analyse in depth how macroeconomic variables such as GDP growth, interest rates, unemployment rates and private sector leverage relate to a build-up of NPLs. However, a microprudential authority or a single bank operating in competitive markets has limited influence over such variables. Taking a microprudential perspective, this paper thus assesses the impact of lending standards as drivers of NPLs.

A number of studies have helped to determine the relevance of other microeconomic variables as default drivers, by assessing the impact of a plethora of bank-level and loan-level variables on NPL ratios.

From a bank perspective, Berger and DeYoung (1997) find that cost efficiency may negatively impact default rates. Salas and Saurina (2002) confirm that finding and show evidence of a relationship between banks' problem loan ratios and their size, growth policies and managerial incentives. Klein (2013) posits that higher capitalisation and profitability tend to lead to lower NPL ratios in the future. Anastasiou et al. (2016) agree in what regards profitability, while suggesting that the size effect is relevant only in peripheral euro area countries. They also indicate that elevated moral hazard (measured by a loans-to-deposits ratio) drives up default rates significantly.

As regards individual borrower and loan characteristics ("lending standards"), Dell'Ariccia et al. (2012) find that delinquency rates during the financial crisis were higher in areas with larger credit booms, which are associated with lower lending standards. Demyanyk and van Hemert (2011) find evidence of the same lending standard deterioration and point out that high-OLTV borrowers became increasingly risky in the run-up to the crisis. This correlation between high OLTV ratios and default rates is often confirmed by the literature (e.g. Vandell and Thibodeau (1985); Deng et al. (2000); Epley et al. (1996); Kelly and O'Toole (2016); Dietsch and Welter-Nicol (2014)). OLTV ratios tend to drive defaults, not only because they serve as a proxy for the borrowers' wealth, but also because they are inherently correlated with their future equity positions (see Campbell and Dietrich (1983)).

There is also empirical evidence linking default rates to borrowers' indebtedness levels, as measured by LTI or debt-service-to-income (DSTI) ratios. Campbell and Dietrich (1983), and Kelly and O'Toole (2016) find that higher DSTI ratios are linked to a higher probability of default. Quercia et al. (2012) find that LTI ratios do not impact default rates directly, but rather affect constrained borrowers' reactions to changes in their equity position or in other default triggers. In contrast, Foote et al. (2010) find that elevated LTI ratios explain few foreclosures.

Kelly and O'Toole (2016) find that longer-maturity loans are less likely to enter into default, due to lower instalments, whereas Epley et al. (1996) find evidence of the opposite, due to an adverse selection process in which riskier borrowers select a longer maturity precisely because of the smaller payments.

The borrower's income and employment status are also assessed in this context. Quercia et al. (2012) and Vandell and Thibodeau (1985) posit, respectively, that borrowers with lower income and lower wealth are more likely to default. The former go on to stress that the correlation is stronger where the borrower is also unemployed, while the latter find that self-employed borrowers are more likely to default.

As regards the loan's interest rate type, Campbell and Cocco (2015) discover that fixed-rate loans default more when interest rates and inflation go down, whereas floating-rate loans default more when they go up. However, they suggest that, during the housing crisis, floaters defaulted more, even in a climate of falling interest rates, since they are preferred by borrowers with riskier income, particularly if their income is tied to interest rates.

Von Furstenberg (1970) and Kelly and O'Toole (2016) find that defaults become less likely as loans get older, with the former finding that defaults peak around the third or fourth year of a mortgage.

When granting loans, a bank decides what borrower and loan characteristics are acceptable, and specifies these in its loan origination policies. As such, they have the potential to be targeted by the supervisor's actions, particularly as they have the potential to influence the credit quality of the bank's portfolio in any given macroeconomic environment and its performance compared to competitors' in a given economic area.

As a result, supervisors assess banks' lending policies on an ongoing basis. Their task is to understand the banks' risk position and appetite, while requiring them to hold adequate levels of capital. Specific requirements can be applied in relation to certain lending standard indicators and capital add-ons can be imposed, in order to prevent excessive risk-taking. The country dimension of this analysis is important from a microprudential perspective, as different legal frameworks, cultures, housing market developments, etc. also determine the influence of certain lending standards on the probability of default.

In view of the importance and relevance of lending standards, in particular from a microprudential supervisory perspective, the need to formalise their relationship with default rates in the euro area is self-evident.

This paper attempts to do precisely that, focusing in particular on the RRE sector.[1] RRE lending is a key activity for the euro area banks and in 2017 represented

[1] Although NPL ratios tend to be higher in the non-financial corporations sector than in the RRE sector, the former were not in the scope of this paper.

about 40% of euro area GDP and 14% of those banks' total assets.[2] Moreover, adverse developments in the RRE market can have severe repercussions on the real economy and the financial sector, given the close links between them (see ESRB 2015), as evidenced for example in the recent financial crisis.

This paper explores the relationship between lending standards and default rates in the euro area RRE market using a novel dataset extracted from the European DataWarehouse (EDW). The dataset consists of loan-level data on RRE loans from eight euro area countries. To the best of the authors' knowledge, this paper is the first to use, in this context, a consistent set of loan-level data on loans originated in multiple euro area countries. The literature investigating the drivers of defaults tends to use either aggregate cross-country data, which hampers a detailed analysis on the influence of certain loan and borrower characteristics on loan default rates, or national loan-level data, e.g. from central credit registers, which rules out cross-country comparison.

The EDW is an independent institution, created at the ECB's initiative to increase transparency on asset-backed securities (ABSs) used as collateral in ECB operations. Eurosystem banks that wish to make an ABS eligible for such purposes, even if the collateral is not posted, are obliged to report a series of fields on the underlying loans. The resulting sample is therefore entirely made up of loans underlying residential mortgage-backed securities (RMBSs) that are eligible for such operations.

This selection procedure generates two important potential biases. First, jurisdictions where loan securitisation is more prevalent are over-represented in the sample so that a country such as Germany, with a large RRE market but a small proportion of securitisations (given its preference for *Pfandbriefe*), is under-represented. Second, the decision as to whether or not to securitise may be dependent on the quality of the loan. There is some debate on whether securitised loans tend to be of higher or lower quality. For example, Albertazzi et al. (2015) find, using Italian loan-level data, that securitised prime mortgages have a lower probability of default than non-securitised loans. Similarly, Ambrose et al. (2005) conclude that securitised loans tend to be less risky as, due to regulatory capital arbitrage, banks sell safer loans, keeping riskier assets in the balance sheet. Conversely, investigating US subprime mortgages, Keys et al. (2010) found that harder-to-securitise loans were more heavily scrutinised than non-securitised loans, since the bank recognises that they are more likely to stay on the balance sheet, thus leading to lower default rates. Elul (2009) and Dell'Ariccia et al. (2012) come to similar conclusions. It does not fall within the scope of this paper to compare the default rates for securitised or non-securitised loans.

Given the above caveats, the dataset might provide a biased picture of domestic lending markets in some countries, as illustrated in Section 2, where the cross-country incidence of fixed vs. flexible interest rate loans in our sample is compared with information from other sources. Accordingly, this paper does not seek to describe euro area mortgage lending markets, but only to analyse the behaviour of loans, given the observed borrower and loan characteristics.

[2] Source: MFI balance sheet items statistics, ECB.

For each loan in the RMBS deals, the EDW data providers report, inter alia:

- the loan's original balance and vintage year;
- the country in which the loan was originated;
- OLTV and LTI ratios at origination;
- the maturity at origination;
- the loan's repayment type and interest rate type (i.e. fixed vs. floating);
- the loan's guaranteed portion, if any;
- the loan's intended purpose (i.e. house purchase, remortgaging, etc.); and
- the borrower's income and employment status.

These constitute the lending standards in this analysis. In addition, "guarantee" and "non-performing" flags were constructed using a composite of other variables. With this information, this paper first sets out to establish the relationship between lending standards and default rates by means of a rich descriptive analysis. For each variable, "buckets" were created and default rates in each bucket were calculated to assess how increases in each lending standard affect defaulting. This was done at both aggregate and country level.

Following this initial investigation, an econometric analysis in the form of a probit model was conducted, in order to control for some unavoidable country-specific characteristics and to help formalise any significant relationships discovered in the descriptive analysis.

The main takeaways, resulting from both methodologies, seem to be that the following indicators/types of loan are associated with a higher probability of default:

- higher OLTV ratios and lower LTI ratios;
- floating interest rate loans (as compared with fixed-rate loans);
- loans to unemployed and self-employed borrowers or legal entities (as compared with those employed by a third party); and
- loans being declared for renovation or remortgaging purposes (as compared with those for house purchase).

The rest of the paper is organised as follows:

Section 2 gives a detailed description of the dataset and its particularities, offering an overview of how the various lending standards have developed in recent years and clarifying how the project fits into the general supervisory framework. Sections 3 and 4 present the main results from the descriptive analysis and from the probit regression of the impact of lending standards on loan default rates, respectively. Section 5 draws conclusions.

2 Lending standards in the euro area

Lending standards in the euro area have fluctuated significantly over the last two decades. The years prior to the global financial crisis saw a substantial easing of lending standards amid booming house prices. Everything changed with the collapse of the housing market in the United States and the euro area, and the sovereign debt crisis, and banks have tightened their lending standards considerably. As well as changing over time, lending standards also differ significantly across euro area countries. This must be taken into account when examining their role in driving loan default rates. In particular, in a supervisory context, it is important to consider lending standards in the light of national customs and macroprudential regulations. A high average OLTV ratio does not necessarily imply a loose lending policy on the part of the originating bank if, for instance, national regulation dictates that high-OLTV loans must be tied to a savings account. Alongside general euro area trends, such particularities are also relevant to an accurate assessment of the riskiness of banks' credit portfolios. As highlighted in this section, there are many other examples in the dataset which are explicitly controlled for in the methodology.

2.1 EDW dataset

The novel dataset used for this paper constitutes one of its main contributions to the literature on the topic. EDW is a statistical data warehouse in Frankfurt, set up as part of the ECB's loan-level data initiative to increase transparency on ABSs used as collateral in ECB operations. It contains loan-level data on loans underlying such ABSs, including car loans, corporate and consumer credit, leasing agreements, credit cards and, in the case of RMBSs, RRE loans. As this project focuses on the latter, loan-level data on all available RMBS deals were downloaded.

The analysed dataset for euro area RMBS deals comprises over 11.6 million loans from eight euro area countries (Belgium, Germany, Spain, France, Ireland, Italy, the Netherlands and Portugal), with more than 150 variables each, including various identifiers for loans and borrowers, borrower information (e.g. primary and secondary income, employment status, etc.), loan characteristics (e.g. OLTV ratio, original maturity, interest rate type, etc.), a flag for the presence of a guarantee and information on the loan's performance (e.g. number of months in arrears, account status, etc.).[3] Reporting is mandatory for only 55 of these variables, while the others may be reported optionally. Data for a subset of 21 of the variables were downloaded for all available submission dates from 2013Q4 to 2017Q4. As the interest lies in static variables (loan and borrower characteristics at the origination date), downloading a snapshot from the last available submission date would suffice for the purpose of this

---

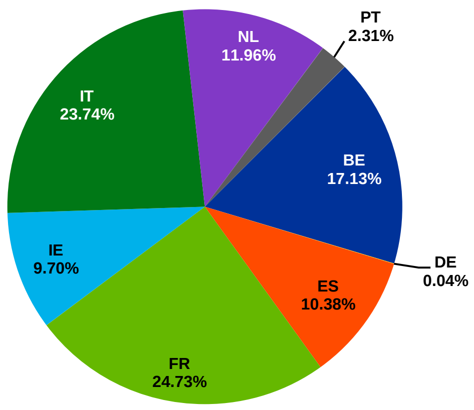
3 Flags for guarantee and performance are created from a composite of other variables. A loan is flagged as guaranteed if any of six guarantee-related fields is not empty or zero. A loan is flagged as defaulted if: i) the data provider describes it as such; ii) it is in arrears for three months; iii) it reports a date of default; or iv) it reports a default or foreclosure amount.

analysis. Further submissions were downloaded so as to extend the dataset with deals that have been closed in the meantime. Also, as banks do not have to report certain variables upon default of a loan, older submissions were used to retrieve this information. As stated above, variables recovered in that way refer to figures at origination and are therefore independent of the time of submission.[4]

The reader should understand the specificities of the dataset in order to interpret the findings correctly. As mentioned above, since the dataset consists entirely of RRE loans underlying RMBS deals that are eligible for collateral in ECB operations, it cannot be used to describe the total euro area RRE market. For instance, in Germany (arguably the largest of the eight RRE markets analysed), RRE loans are not commonly packaged into RMBSs, but rather into *Pfandbriefe* (a type of covered bond). As a result, Germany's sample contribution (0.04%) is much lower than would be expected from such a large RRE market (Chart 2). Because of this very low data coverage,[5] Germany is excluded from further descriptive analysis; however, it is taken into account in the model-based analysis, where country effects are controlled for.

Chart 2
Country composition of dataset

(share of total balance in the sample originated in a given country)



Sources: EDW and ECB staff calculations.
Note: Figures obtained from a final dataset with 3.4 million observations to which the cleaning procedures in Box 1 were applied.

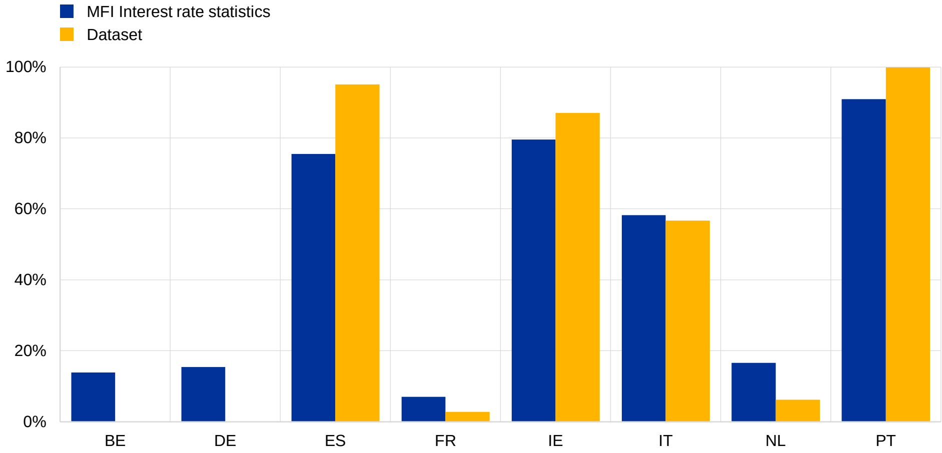
One way to see how well the dataset describes domestic lending markets is by comparing in-sample ratios of fixed vs. flexible interest rate loans with information from other sources for all countries. This exercise was performed comparing percentages of new flexible interest rate originations in the sample with MFI interest rate statistics. As shown in Chart 3, the EDW dataset reflects correctly whether a given national RRE loan market is dominated by fixed or flexible interest rate lending. However, differences can be found in the average shares of flexible interest rate loans in total new lending in certain countries, likely motivated by the potential biases described in Section 1.

[4] Despite this being the expected scenario, there are often issues related to these static fields changing with new data submissions. A more detailed description of this can be found in Box 1.

[5] The total balance at origination of loans originated in Germany in the dataset is around EUR 170 million.

Chart 3
Dataset's interest rate type representativeness

(y-axis: share of new loan balances, in a given year, with a floating interest rate (average: 2007-2016), per sample and country of origination)



Sources: MFI interest rate statistics, EDW and ECB staff calculations.

Naturally, it is possible that this inherent selection procedure also influences sample default rates. However, the NPL ratio as such is not the variable of interest in this analysis and is not calculated for deals or countries. In EDW, all loans in a given deal have to be reported until the deal is closed (even after the loans are defaulted, foreclosed, have matured or been sold), so such a ratio is not comparable to an NPL ratio reported by a bank. Banks report defaulted loans on their balance sheet (not foreclosed, not written off, not sold and not matured) at a specific point in time.[6] Table 1 shows the percentage of observed defaulted loan balance in the sample.

Table 1
Observed defaults in the sample

Observed defaults divided by all reported loans

(EDW: Share of total balance in the sample being reported as defaulted)

Source	BE	ES	FR	IE	IT	NL	PT	Total sample
EDW	1.6%	14.7%	1.5%	19.9%	8.1%	0.4%	3.4%	6.16%

Source: EDW and ECB staff calculations.

Box 1
Data treatment processes

Loans with missing values and outliers were removed from the initial dataset. That dataset contains a significant number of mortgages that are divided into parts. Segmenting mortgages into parts upon securitisation is a relatively common practice in some of the euro area countries included in the sample (e.g. the Netherlands). The loan is typically, but not exclusively, separated into the interest-bearing part and the principal. Regardless of the number of resulting parts, all of them are

6 Chart 3 remains unaffected by this issue, since both sources used concern only new originations in comparable periods.

reported in EDW as distinct observations; as it is impossible to "reconstruct" the original loan, all loan parts were dropped from the sample. This cleaning procedure meant that certain deals were left with higher or lower observations of defaulted loans to total loans, as compared to what was observed in the untreated dataset. In order to avoid potential misrepresentation of these deals, some RMBS deals were deleted entirely. These are deals for which over 95% of the observations were deleted in the data treatment process due to data quality issues, or deals for which the observed default rate roughly halved or doubled when compared to the rate in the untreated dataset.

There were also a considerable number of deals for which no default was reported. Such deals are dropped from the dataset as this seems to indicate misreporting.

The final sample resulting from the application of these procedures contains over 3.4 million loans.

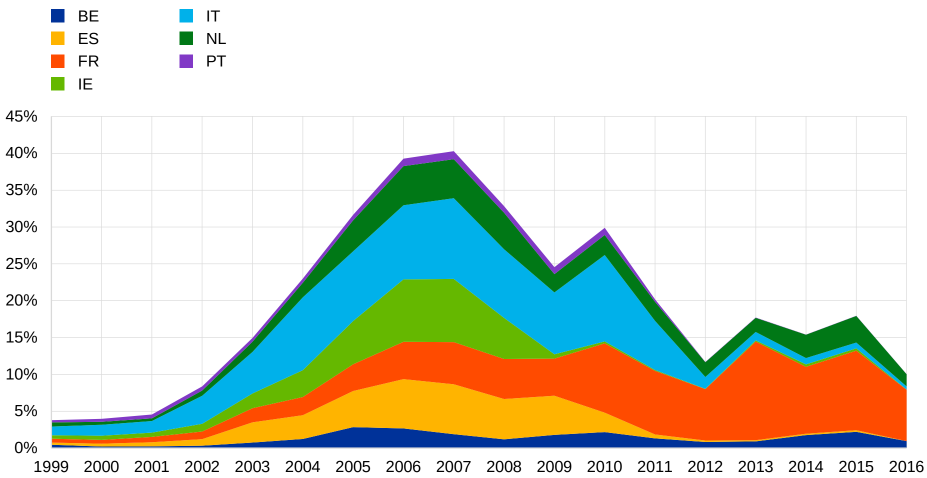
2.2 Developments in RRE lending standards over time

The dataset makes it possible to study the development of certain lending standards over time. Although the sample includes loans originated as far back as the 1970s, only around 1% of the loans in the dataset (in terms of original balance) were originated before 2000. Therefore, the analysis focuses on the period from 2000 to 2016.[7]

Chart 4

Distribution of loans by year of origination and country

(x-axis: year of origination; y-axis: share of total original balance in the sample (EUR billion) originated in a given country, per year of origination)



Sources: EDW and ECB staff calculations.

When looking at aggregate euro area time series, it is worth noting that the country composition of the dataset varies over time. For example, in the period between 2005 and 2008, when the volume of loans originated peaks in the dataset, the sample was

[7] Loans from 2017 were excluded from the descriptive statistics as not many of them have been securitised and fed into the EDW database.

distributed mainly across Italy, Ireland, Spain, France and the Netherlands. Following the global financial crisis, reported RRE loan origination activity decreased significantly to reach a trough in 2012. From then onwards, the sample is dominated by loans originated in France, followed by the Netherlands and Belgium, while loan volumes in Ireland, Spain and Italy decreased (see Chart 4). This might be related to the decreases in lending activity and/or RMBS issuance in those countries following the euro area sovereign debt crisis. Therefore, when looking at changes in aggregate lending standards over time, the reader should bear in mind that they might be driven in part by changes in country contributions.

Notwithstanding this caveat, the dataset provides a number of insights into developments of lending standards in euro area RRE lending markets over the last 20 years. The data point to a substantial easing of lending standards before the global financial crisis and to a tightening in the years thereafter, when depressed macroeconomic and financing conditions resulted in a deterioration of banks' balance sheets and higher funding costs. Banks became more risk-averse and naturally more selective towards their customers, tightening their lending standards particularly over the crisis episodes of 2008-2009 and 2011-2012.[8]

The pre-crisis easing of lending standards in euro area RRE markets and the subsequent tightening are reflected in the maturity of loans at origination. Loans with longer maturity are often perceived to be more risky, since the time span over which a debtor can default is longer. The aggregate maturity of loans at origination shows an increasing trend in the years leading up to the global financial crisis (Chart 5). Average loan size at origination (Annex A, Chart A.1) shows a similar pattern – increasing until 2007 and decreasing thereafter. This may be due to house prices rising faster than borrowers' income,[9] forcing borrowers to take on larger loans, combined with banks relaxing their policies with respect to loan affordability in this period (as seen from the increase in LTI ratios in the years leading up to the financial crisis – see below). In both cases, a longer maturity would be needed to render the repayment of a loan affordable. In particular, loans with over 30 years' maturity constituted over a quarter of total loans originated in 2007. The 2008-2009 recession reversed this trend and loan maturity at origination started to fall, stabilising from 2014 at around 21 years, with a maturity of 15-25 years for around 50% of loans originated.

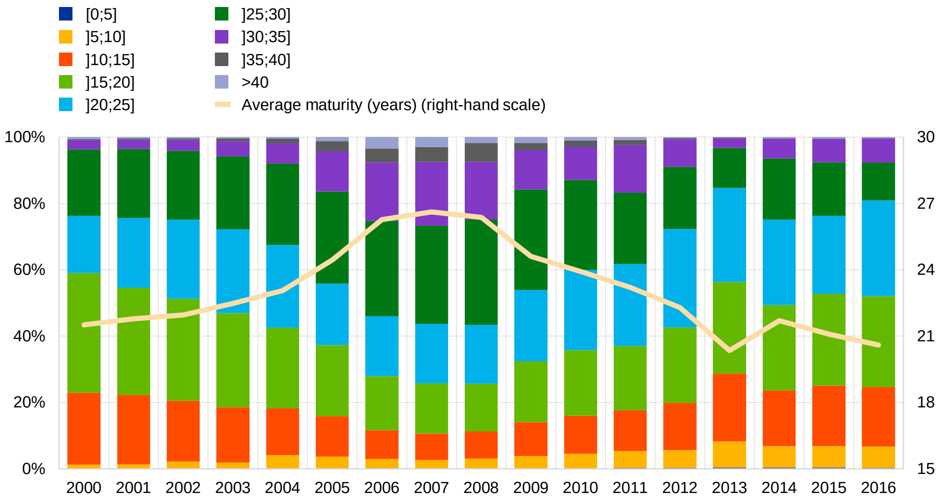
8 See the ECB Bank Lending Survey for developments and drivers of bank lending standards in the euro area.

9 For developments of price-to-income ratios in the euro area, see for example ECB (2017).

Chart 5

Distribution of loans for real estate lending by original maturity at aggregate level

(x-axis: year of origination; left-hand scale: share of total original balance in the sample, per original maturity bucket and year of origination; average original maturity, per year of origination)

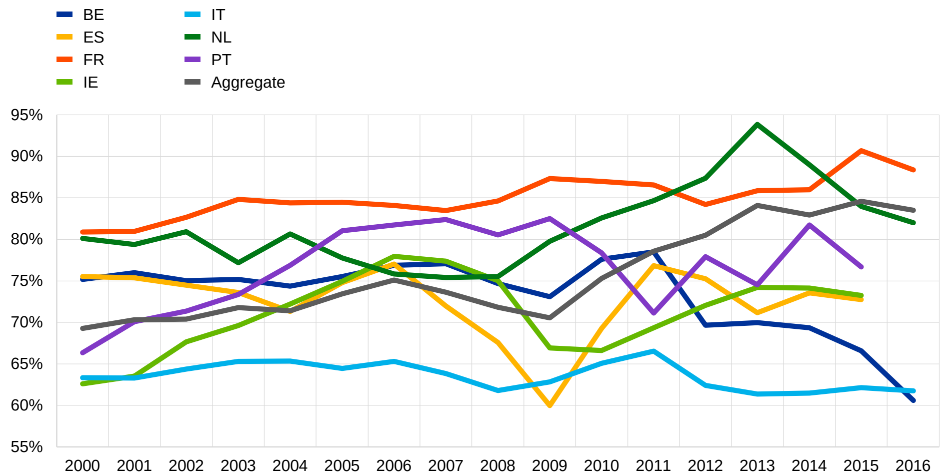


Sources: EDW and ECB staff calculations.

Chart 6

Developments in OLTV by country

(x-axis: year of origination; y-axis: average OLTV ratio, in percentage, per country and year of origination)



Sources: EDW and ECB staff calculations.
Note: Insufficient number of observations available for Spain, Ireland and Portugal in 2016.

OLTV is a key indicator for lending standards, as it relates the volume of a loan to the value of the underlying property. Its diverse developments (Chart 6) may reflect different house price movements over time and varying macroprudential regulations across euro area countries. Some jurisdictions impose caps on LTV ratios or bring them down via other policy measures (as was often the case in the post-crisis years).[10]

[10] For instance, in Ireland caps on OLTV are set at 80% OLTV for non-first-time buyers (FTBs) and at 90% for FTBs. In 2018, Ireland introduced stricter criteria for the preferential weighting of residential mortgage loans. In Portugal, the limit on OLTV varies between 80% and 100% depending on the purpose of the credit.

Private or public guarantee systems providing borrowers with insurance schemes might support higher LTV ratios, as part of the credit risk is transferred away from the lender (see ECB (2009)).

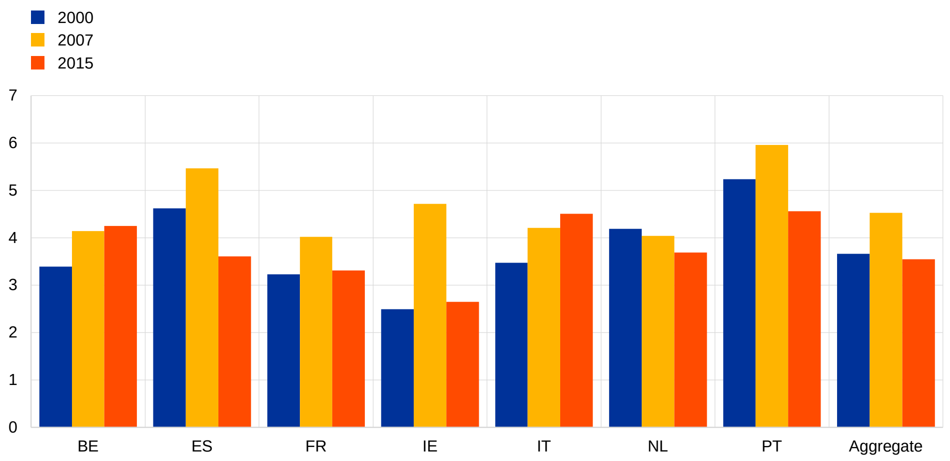
The fact that private guarantee schemes are common in France may explain its relatively high OLTVs compared to other countries. Similarly, mortgages in the Netherlands are subject to a strict recourse law, whereby the collateral can be liquidated without any court interference and the lender generally retains full non-expiring recourse rights to any remaining debt.[11] In addition, tax deductibility of interest paid increased the incentive to borrow large amounts while putting savings into savings vehicles. Aggregate OLTV ratios in the Netherlands rose after the financial crisis and reached a peak in 2013. The subsequent decrease may be due to the introduction of an OLTV cap in 2012 (starting at 106%, with annual reductions to 100% OLTV in 2018) and the reduction of the tax incentives.

The crisis episodes led to sharp declines in aggregate OLTV ratios in some countries. This could have been driven by banks becoming more risk-averse and tightening lending standards in that period, especially in countries where the real estate bubble burst.[12] The largest reduction in OLTV ratios could be observed during the 2008-2009 crisis in Spain and Ireland, which were also the countries where house prices collapsed most in that period. The euro area sovereign crisis was followed by a narrowing in OLTV ratios, mostly in Belgium, Portugal and, to a lesser extent, Italy, Spain and France.

Chart 7

Loan-to-income ratio across euro area countries

(x-axis: country of origination; y-axis: LTI ratios per country and year of origination)



Sources: EDW and ECB staff calculations.

Borrowers' indebtedness relative to their income (the LTI ratio) is another important determinant of loan quality. LTI ratios increased in most of the countries in the

11 Dutch Banking Association (2014).

12 See the ECB Bank Lending Survey for developments and drivers of bank lending standards in the euro area.

pre-crisis period, peaked in 2007 and decreased afterwards (Chart 7). Again, this illustrates how lending standards eased in the years leading up to the global financial crisis and tightened subsequently. The sharp increases in house prices in the years preceding the crisis are most likely the major driver of the sharp increases in the original balance of loans relative to borrowers' income – this is especially visible in the peaks in LTI ratios in Spain, Ireland and Portugal.

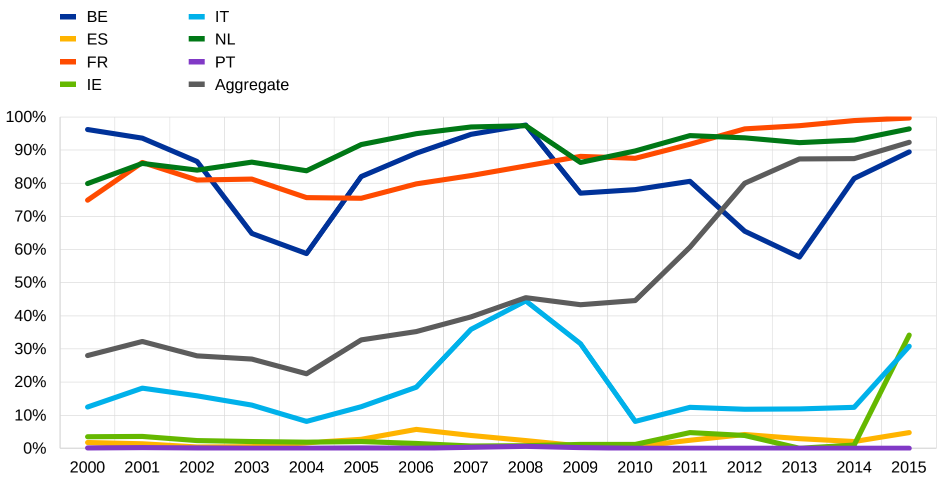
Interest rate type seems to be a loan characteristic that is driven largely by country specificities (see also Section 2.3) and is more stable over time. As shown in Chart 8, there is a clear distinction between countries that originated mainly new loans with a fixed interest rate from 2000 (Belgium, France and the Netherlands) and those where other types of interest rate (mainly floating) were more common for new loans (Spain, Ireland, Italy and Portugal). The significant increase in the proportion of fixed interest rate loans in the aggregate series stems from the changes in country composition, as countries such as Belgium and France dominate the sample after 2012. However, France has seen a growing shift towards fixed interest rates over the years, as has the Netherlands in the post-crisis period. This might have been supported by the post-crisis low interest rate environment.

As regards borrowers' employment status, as can be expected, the majority of loans were granted to employed individuals (Chart 9). The self-employed make up the second largest category – such loans are most prominent in Italy, where they constitute almost 20% of the sample, followed by Spain, Ireland, Belgium, France and the Netherlands (9-14% of the sample). In Italy and the Netherlands, a significant percentage of loans were granted to pensioners (6% of the sample). The proportion of loans granted to employed individuals has grown somewhat over time across countries, with the largest increases in Ireland and Portugal, mostly due to decreases in the proportion of loans to self-employed borrowers.

Chart 8

Proportion of loans with fixed interest rate, by country

(x-axis: year of origination; y-axis: share of total original balance in the sample originated in a given country being reported as having a fixed interest rate, per year of origination)

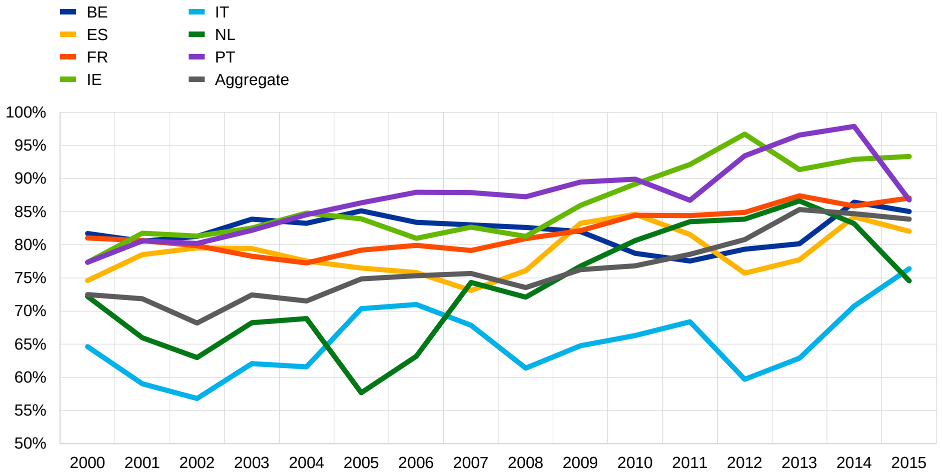


Sources: EDW and ECB staff calculations.
Note: Other categories for interest rate type include 'floating' and 'other'.

Chart 9

Proportion of loans granted to employed borrowers, by country

(x-axis: year of origination; y-axis: share of total original balance in the sample originated in a given country being granted to employed borrowers, per year of origination)



Sources: EDW and ECB staff calculations.
Note: Other categories for employment status include 'self-employed', 'student', 'legal entity', 'pensioner', 'unemployed' and 'other'.

2.3 Country-specific characteristics in lending

As mentioned above, this paper assesses the EDW dataset first through an extensive descriptive analysis and then applies a more formal econometric framework, in a probit setting. Before delving into the results of the former, it is relevant to highlight one of the main reasons for the latter: the need to control for certain country-specific characteristics. The examples in this section serve to illustrate the main challenges these characteristics posed to the interpretation of the results of the descriptive analysis that are to follow in Section 3.

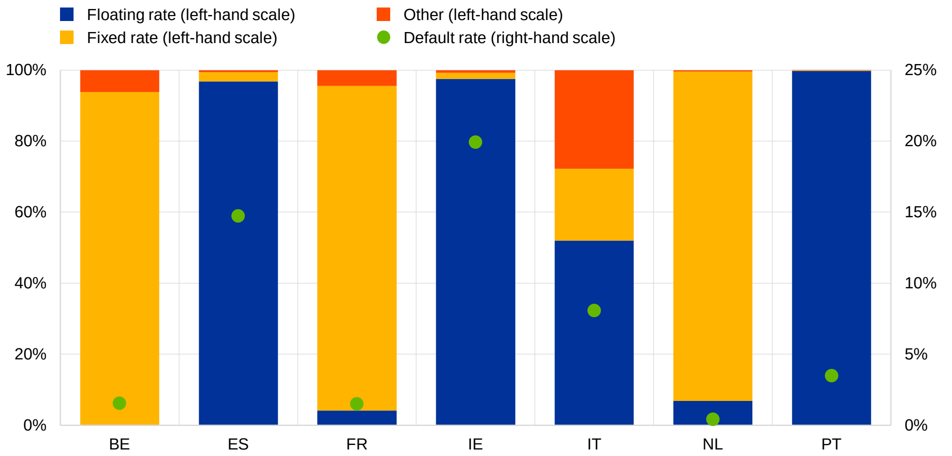
A common theme surrounding these challenges is a general problem affecting the analysis of default rates according to loan characteristics. Ideally, a simple breakdown of default rates by borrower employment status, for instance, should suffice to discern which employment categories can be considered more risky, at least in the context of the sample. However, the probability of a loan being originated with a certain characteristic is often not entirely exogenous to other relevant aspects of the loan. Namely, three instances can be cited where the country of origination seems to play a role in determining a set of loan characteristics.

First, when looking at default rates by type of interest rate (i.e. floating or fixed), the sample shows that there is a clear dichotomy between two distinct sets of countries. RRE loans in Belgium, Germany, France and the Netherlands predominantly exhibit fixed rates, but the reverse is true for Spain, Ireland, Italy and Portugal (Chart 10). This gives rise to a possible data misinterpretation, whereby it would appear that floating-rate loans report higher default rates, but in reality this would merely be a reflection of the fact that those loans are more common in countries where default rates have been higher, due to other factors such as current macroeconomic conditions.

Chart 10

Types of interest rate, by country

(x-axis: country of origination; left-hand scale: share of total original balance in the sample, per interest rate type and country of origination; right-hand scale: share of the sample's total original balance being reported as defaulted, per interest rate type and country of origination)



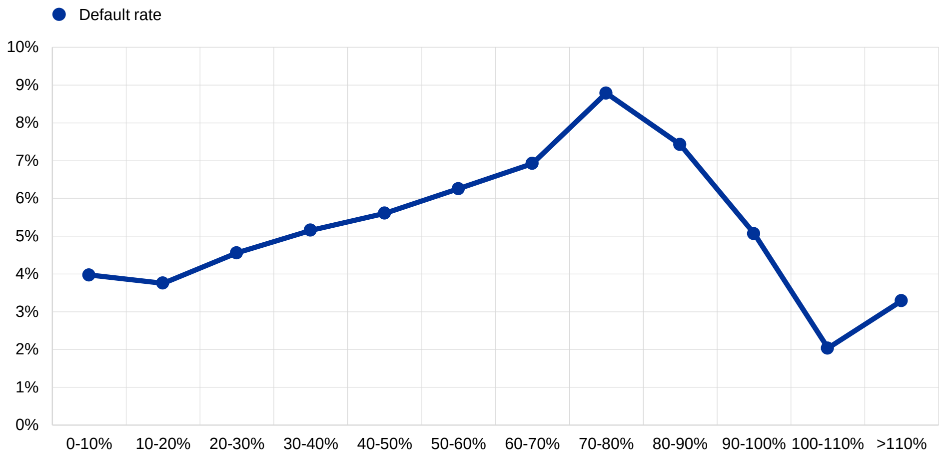
Sources: EDW and ECB staff calculations.
Note: Fixed-rate loans with future periodic resets are included as fixed-rate loans; the 'other' category includes, but is not limited to, hybrid and capped loans.

Second, when segmenting the sample into buckets according to their OLTV ratios, it appears that loans with OLTV ratios above 80% exhibit lower default rates than their less leveraged counterparts (Chart 11). However, while Italian loans make up almost 50% of the overall sample, the proportion falls to less than 7% in the 80-90% OLTV bucket, while the proportions of French and Dutch loans increase for the higher OLTV buckets (Chart 12). As the default rates in Italy were significantly higher than in France or the Netherlands, the proportions per country give rise to the counter-intuitive result (see above) that euro area aggregate default rates are lower for higher OLTV buckets.

Chart 11

Aggregate default rates by OLTV bucket

(x-axis: OLTV bucket; y-axis: share of total original balance in the sample reported as default, per OLTV bucket)

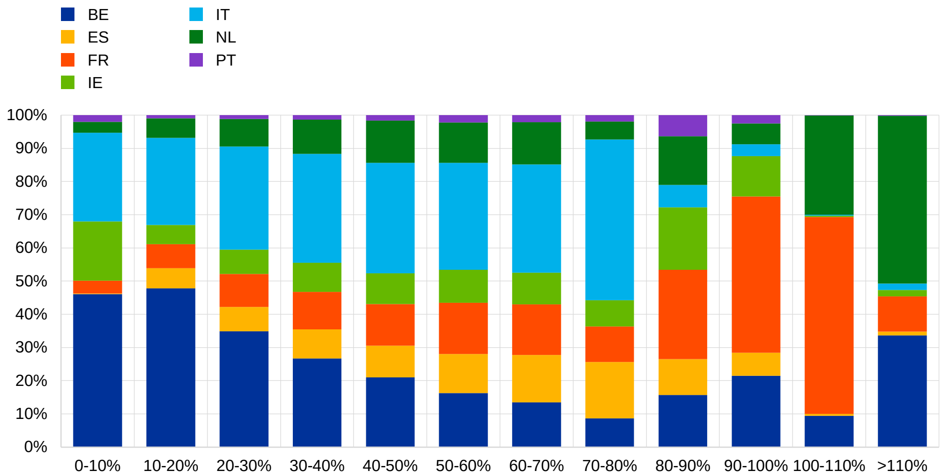


Sources: EDW and ECB staff calculations.

Chart 12

Country distribution by OLTV bucket

(x-axis: OLTV bucket; y-axis: share of total original balance in the sample, per country of origination and OLTV bucket)



Sources: EDW and ECB staff calculations.

Third, for all euro area data, bullet loans seem to be safer than credits repaid as an annuity, or with linear payment schemes (Chart 13). Dutch loans are the main driver of this apparent relationship, as they make up for almost 98% of the bullet loan subsample (Chart 14). This high proportion is driven by the full tax deductibility of interest payments up until 2013, which made it optimal for borrowers to take out interest-only mortgages[13] with a single payment at the end of the lifetime of the loan.[14]

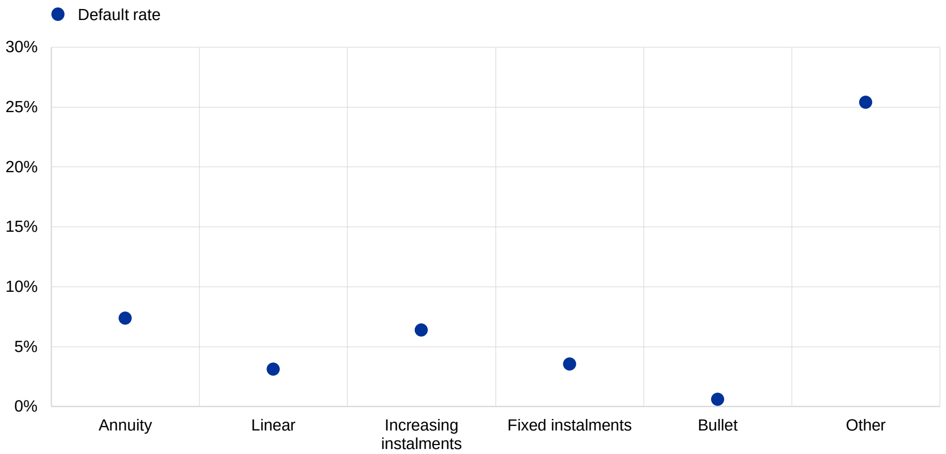
13 Dutch Banking Association (2014).

14 These "bullet" loans are often associated with a life insurance policy or a savings account which ensures repayment at the end of the lifetime of the loan and effectively acts as a periodical instalment, although not as an amortisation *per se*.

Chart 13

Aggregate default rates by payment type

(x-axis: payment type; y-axis: share of total original balance in the sample reported as default, per payment type)

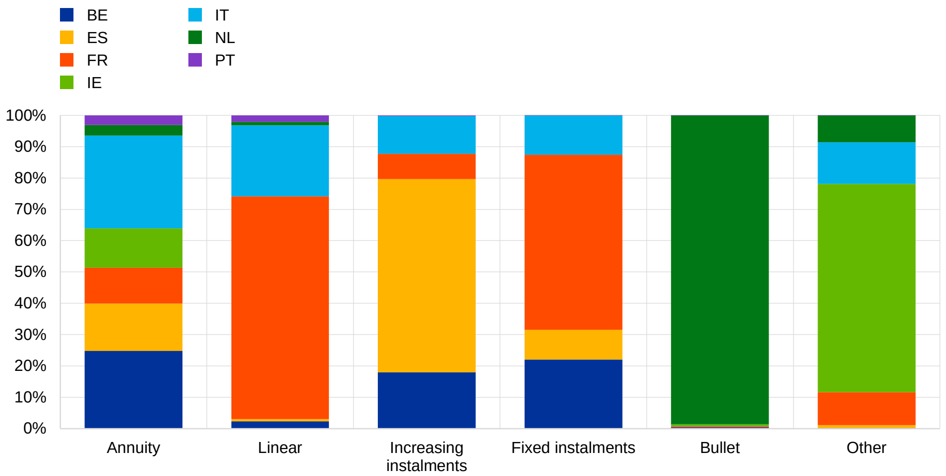


Sources: EDW and ECB staff calculations.

Chart 14

Country distribution by payment type

(x-axis: payment type; y-axis: share of total original balance in the sample, per country of origination and payment type)



Sources: EDW and ECB staff calculations.

These are just a few examples to illustrate the need to interpret results in the light of the sample's characteristics, particularly as regards country specificities. This has to be borne in mind in the next section, which describes the main findings of the descriptive analysis.

3 Impact of lending standards on default rates: a descriptive analysis

This section provides a descriptive analysis of the distribution of the key variables in the sample. Its main aim is to provide preliminary indications as to which lending standards seem to be linked to higher default rates. The statistical significance and magnitude of the effects are assessed in the econometric analysis in the following section.

For this purpose, the sample was segmented into buckets for each lending standard – for continuous variables by creating equal-sized intervals and for categorical variables by creating a breakdown per category.

Each loan is flagged as “performing” or “defaulted”;[15] the sum of the defaulted loan balances constitutes the defaulting portion in each bucket, which is expressed as a proportion of the total loan amount in that interval (i.e. the bucket’s default rate). This default rate is not comparable to an NPL ratio as reported by a bank, as loans are not removed from the dataset when they mature, are sold, foreclosed or written off. In this way, one can easily assess how changes in each individual lending standard affect defaulting, at both aggregate and country level.

The lending standards of a loan can be assessed by looking at the borrower characteristics (e.g. employment status and income) and the contract terms and conditions (e.g. OLTV and maturity at origination, purpose, payment and interest rate type, and presence of a guarantee).

Box 2

Definition of “default”

Data providers reporting to the EDW address loan performance by filling a field on the loans’ account status, which includes several categories: “performing”, “in arrears”, “defaulted”, “redeemed” and “repurchased”. However, the account status is not filled in consistently by all data providers, so that being reported in EDW as “defaulted” is a sufficient but not necessary condition for a loan to be flagged as “non-performing” in the dataset. Other variables are used to construct the “defaulted” flag, even in the absence of such a clear indication by the data provider. For instance, a default is also considered to have occurred[16] where the date of default is not empty or the foreclosure amount is larger than zero for an observation, regardless of the reported account status.

The same is true where the number of reported months in arrears exceeds three. However, data providers also report this field with some notable inconsistencies: some interpret the number of months in arrears as the number of months after an instance of default. In such cases, they report the

[15] See Box 2 for more information on the definition of default used in this paper.

[16] Even if a loan is cured after having been reported as defaulted, it is still counted as a single default in the dataset, as the scope of the paper relates merely to the occurrence or absence of a default, disregarding multiple instances or any reasons for potential recovery or curation.

number of months in arrears as one, although four months have elapsed since the last scheduled payment. Also, for instance, certain RMBS contracts establish that a loan in the pool is not considered to be defaulted until after a full year in arrears; here, one month in arrears would be the equivalent of 13 months in a normal scenario.

In such instances, the EDW team performed an analysis of the complete submission time series and applied expert judgement to determine the correct loan status and date of default.

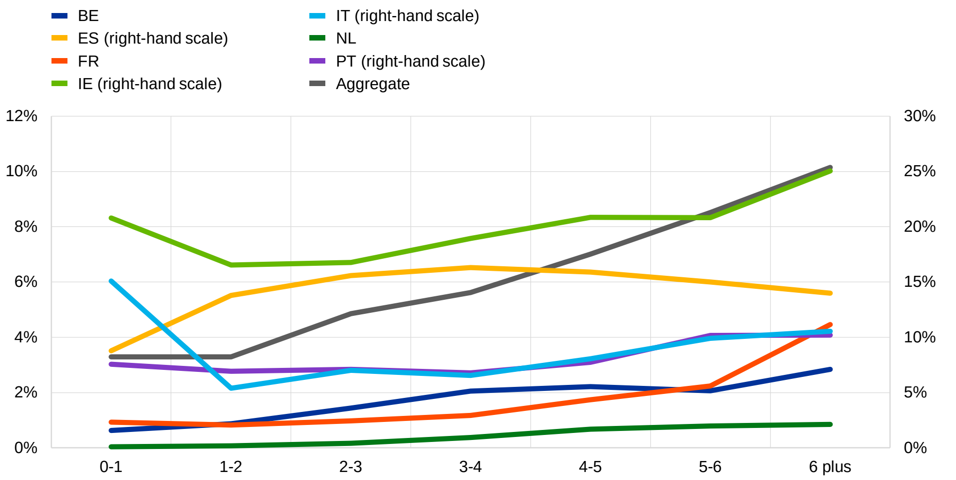
3.1 Borrower-specific characteristics

The descriptive analysis points to a strong correlation between the size of a loan relative to the borrower's annual gross income (LTI ratio) and the default rate (Chart 15). This correlation holds in all analysed countries. However, some irregularities are observed in Italy, where loans in the lowest LTI ratio bucket (0-1) have the highest rate of observed defaults. In Spain, default rates decrease somewhat for loans with an LTI ratio above 4. On aggregate, loans in the highest (>6) LTI ratio bucket have the highest default rate, whereas the default rate increases most between the 5-6 and the >6 LTI ratio buckets. At country level, this holds for Belgium, France and Ireland. In Italy and Portugal, the default rate increases most between the 4-5 and the 5-6 LTI ratio bucket. Overall, this indicates that loans with LTI ratios above 5 are more risky.

Chart 15

Loan default rate, by LTI ratio bucket

(x-axis: LTI buckets; y-axis: share of total original balance in the sample originated in a given country being reported as defaulted, per LTI bucket)

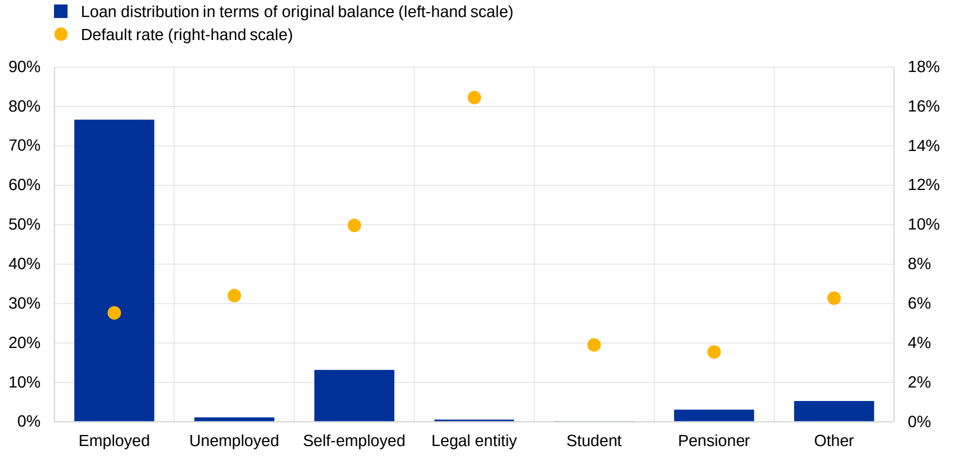


Sources: EDW and ECB staff calculations.

Chart 16

Loan default rate, by employment status

(x-axis: employment status; left-hand scale: share of total original balance in the sample, per employment status; right-hand scale: share of total original balance in the sample being reported as defaulted, per employment status)



Source: EDW and ECB staff calculations.

No clear pattern emerges from the data in the sample as regards the role of income level in driving probability of default (Annex A, Chart A.2). The expected inverse relationship is observed in Belgium, France and Portugal, but not in other countries nor on aggregate, suggesting that borrowers' income is more crucial for assessing loan riskiness when related to the size of the loan.

Employment status also influences loan default rates (Chart 16). On aggregate, the default rate is highest for loans to legal entities. However, such loans constitute only a marginal part of our sample (0.6%). More interesting are differences between default rates of employed and self-employed borrowers, who together constitute 90% of the sample. As could be expected, on aggregate, default rates among the self-employed are considerably higher than among employed borrowers. This relationship holds in all countries in our sample except Spain.

3.2 Loan-specific characteristics

The literature finds a correlation between default rates and the OLTV ratio, which is often considered to be one of the most important lending standards. For this reason, OLTV limits are used as a macroprudential tool in some countries. Three countries in our sample have introduced OLTV caps in recent years (the Netherlands in 2012, Ireland in 2014 and Portugal in 2018).[17] Analysis of data in the sample confirms the strong positive relationship between OLTV and default rates across all countries apart from Spain (Chart 17). Across OLTV buckets, default rates increase particularly for loans with an OLTV ratio above 100% in France, Italy and Portugal. For Belgium, the highest increase in default rate across buckets is observed for loans in the 80-100% OLTV category, while in Ireland loans with OLTV above 60% seem to have much

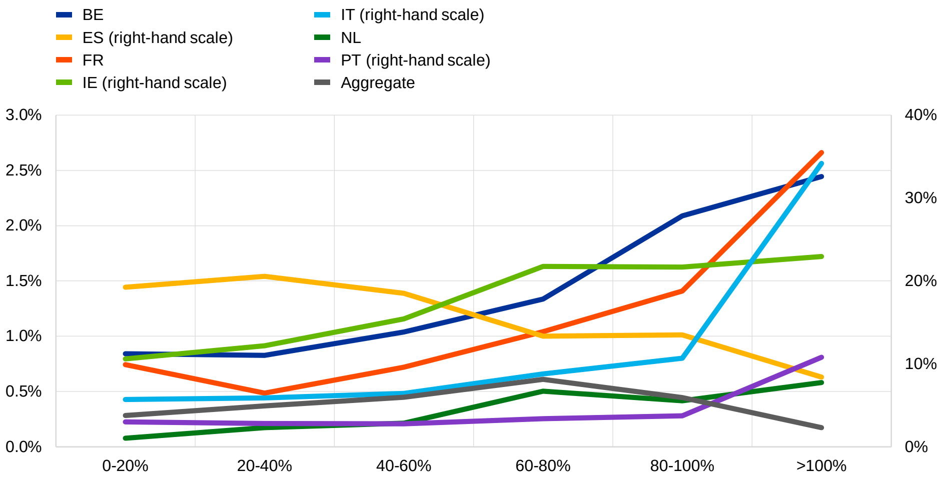
17 See the ESRB overview of national measures of macroprudential interest in the EU/EEA.

higher default rates than those below this threshold. On aggregate, an unambiguous correlation between OLTV and default rates could not be found. This is probably due to country specificities in lending practices and differences in country contributions to the aggregate across OLTV buckets.

Chart 17

Loan default, by OLTV bucket

(x-axis: OLTV buckets; y-axis: share of the sample's total original balance originated in a given country being reported as defaulted, per OLTV bucket)

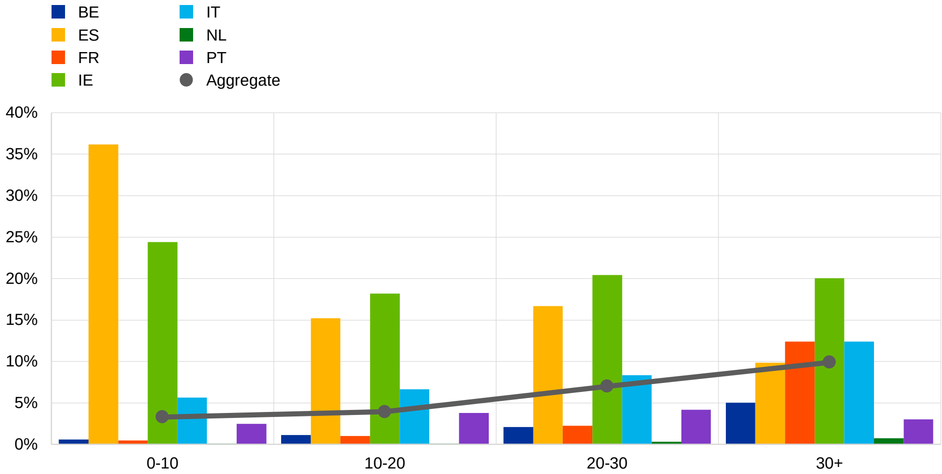


Sources: EDW and ECB staff calculations.

Chart 18

Loan default rate, by original maturity bucket

(x-axis: maturity buckets in years; y-axis: share of the sample's total original balance originated in a given country being reported as defaulted, per maturity bucket)



Sources: EDW and ECB staff calculations.

Default rates also tend to increase in line with loans' original maturity (Chart 18). This relationship holds on aggregate and also at country level for all countries apart from Ireland and Spain. In those two countries, the highest default rate can be observed in the lowest-maturity bucket. This might be a result of higher instalments in low-maturity buckets, so that borrowers have more difficulty repaying in the event of a shock.

Moreover, short-maturity loans (in particular, loans with five years' maturity or less) may be taken with an aim to roll-over at expiration; this is corroborated by the fact that most of these short-maturity loans in the sample default after their initially predicted expiration date.

The relationship between default rates and loan purpose differs across countries (Chart 19). On aggregate, loans taken as a remortgage seem to default most; this is driven by high default rates in this category in Ireland, Spain and Portugal. The loans subject to the highest default rate in France are those taken for construction purposes, in Italy those for renovation, while in Belgium and the Netherlands are those for house purchase and "other purposes". Loans for house purchase constitute the biggest proportion of loans in the sample, ranging from 51% in Belgium to 90% in Spain (Annex A, Chart A.3). The second largest category is remortgaging, which accounts for a significant part of the sample in Belgium, France and Ireland.

Given the country specificities as regards payment type and interest rate type (see previous section), it is difficult to draw clear conclusions as to their role in determining the probability of default without controlling for country effects. On aggregate, the data indicate that loans with floating interest rates and the payment type "other" are subject to the highest default rates, followed by bullet loans (Chart 13 and Annex A, Chart A.4). These relationships are investigated more thoroughly in the model-based analysis section below.

The presence of a guarantee seems to reduce default rate probability (Chart 20). This holds on aggregate and in all countries apart from Spain.[18] The relationship is not straightforward. While the existence of a guarantee should transfer credit risk away from a bank, it may also indicate a riskier borrower, as suggested for example in Calcagnini et al. (2009).[19] On the other hand, as suggested by Jimenez and Saurina (2003), if a loan is guaranteed by a bank or a government, it might be subject to double evaluation, which should lower the probability of default.

---

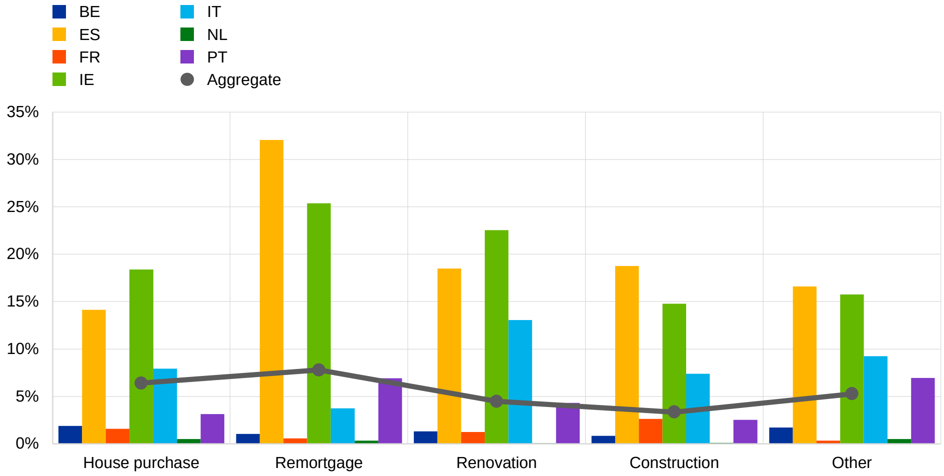
18 The data are not available for Belgium.

19 The authors formulate a hypothesis that "banks behave 'lazily' by replacing screening and monitoring activities with personal guarantees".

Chart 19

Loan default rate, by purpose

(x-axis: loan purpose; y-axis: share of total original balance in the sample originated in a given country being reported as defaulted, per loan purpose)

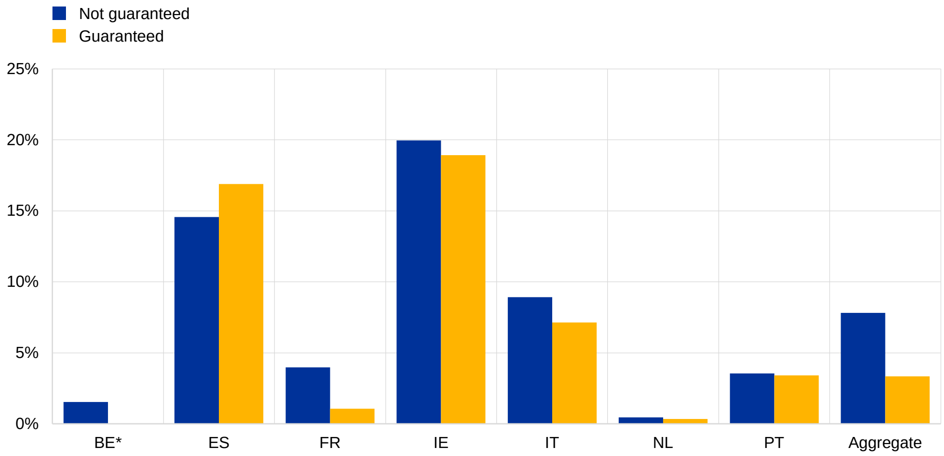


Sources: EDW and ECB staff calculations.

Chart 20

Loan default rate, by presence of guarantee

(x-axis: country of origination; y-axis: share of total original balance in the sample being reported as defaulted, per guarantee status and country of origination)



Sources: EDW and ECB staff calculations.
Note: * No data on guarantees available for Belgium.

4 Model-based analysis: taking country heterogeneity into account

As mentioned in the previous sections, many variables in the dataset exhibit substantial heterogeneity across countries. The descriptive analysis makes it clear that the country specificities can significantly distort euro area aggregates. As such, in order to control for the main sources of heterogeneity, a more formal econometric analysis is required.

Much of the literature on the usage of loan-level data concerns itself with the determinants not only of default rates, but also of prepayment. Hence, competing risk models are quite often used (e.g. Calhoun and Deng (2002), Pennington-Cross and Ho (2010), Deng et al. (2000), Campbell and Dietrich (1983)), as they allow for a non-binary, but still categorical outcome, since multiple equations exist – one for each potential outcome. Such models are outside the scope of this paper, which concerns itself only with default drivers, but parts of them (in particular, the equation pertaining to the default category) can still be of use.

A simple (default equation only) probit approach (as used in Knapp and Seaks (1992), Keys et al. (2010) and Jiang et al. (2014), for instance) was chosen for this analysis. Detailed econometric specifications can be found in Box 3. The variables covered are those described in Section 3, although controls have been added for the vintage year and the country of origination.

Box 3
Model specifications

The full model is as follows:

$$P(y = 1 \mid X) = F(y^*) = F(X'\beta)$$

$$y^* = \sum_{n=1}^{m} \beta_n x_n + \sum_{j=m+1}^{q} \beta_j COUNTRY_j + \sum_{k=q+1}^{p} \beta_k VINTAGE_k$$

where F is the standard normal cumulative density function (CDF), x_n are the variables representing the lending standards, $COUNTRY_j$ is a set of country dummies and $VINTAGE_k$ is a set of dummies for the loans' vintage years.

The model is estimated by maximum log likelihood. From the density of each observation y_i given x_i:

$$f(y_i \mid x_i'\beta) = G(x_i'\beta)^{y_i} * [1 - G(x_i'\beta)]^{1-y_i},\ for\ y = 0{,}1$$

Taking logs:

$$\ell_i(\beta) = y_i \log[G(x_i'\beta)] + (1 - y_i)\log[1 - G(x_i'\beta)],\ for\ y = 0{,}1$$

So that the log likelihood – to be maximised – for the sample is:

$$\mathcal{L}(\beta) = \sum_{i=1}^{n} \ell_i(\beta)$$

Clustered standard errors were estimated at deal level, as it is possible that residuals are correlated within a given deal.

Throughout this paper, results are presented through the usage of marginal effects rather than the simple regression coefficients. Traditionally, a coefficient could be interpreted as the elasticity, or response, of the predicted value to a change in a given regressor. This applies in a linear setting, such as:

$$y = X'\beta$$

Since the derivative of y in respect to a certain variable x_n, for each individual observation i, is simply given by said usual regression coefficient:

$$\frac{\partial y_i}{\partial x_{i,n}} = \beta_n$$

In a probit setting, however, for each observation i, given that observation's set of covariates, X_i', the analogous derivative is given by:

$$\frac{\partial y_i}{\partial x_{i,n}} = f(X_i'\beta)\beta_n$$

where $f(.)$ is the normal probability density function. As can be seen, the derivative, or marginal effect, for a given regressor effectively depends on the values of all other covariates. Therefore, no unique marginal effect for a variable exists (unlike in a linear model). A common solution for this is to take marginal effects by fixing the set of covariates at certain representative values, such as their means, in order to obtain a meaningful measure of the dependent variable's response to a change in a certain dependent variable. These marginal effects at means (MEM) are given by:

$$MEM = f(\bar{X}'\beta)\beta_k$$

The main advantage of the MEM is its straightforward interpretation: it shows the impact of variable x_k on y, when looking at a hypothetical average observation.

4.1 Main regression results

Table 2 shows results obtained following the approach outlined in Box 3 (country and vintage year controls[20] were excluded from the output):

Table 2

Probit regression results

Dependent variable: NPL flag	Marginal effects (at means)[1)]	Standard error	P-value
Employment status (binary)			
Unemployed	0.014	0.004	0.001***
Self-employed	0.019	0.004	0.000***
Legal entity	0.081	0.038	0.033**
Student	-0.006	0.003	0.060*
Pensioner	-0.005	0.006	0.426
Other	0.005	0.007	0.477
Payment type (binary)			
Linear	0.000	0.005	0.964
Increasing inst.	-0.012	0.007	0.059**
Fixed inst.	0.004	0.006	0.461
Bullet	-0.003	0.006	0.619
Other	0.045	0.020	0.021**
Interest rate type (binary)			
Floating	0.014	0.005	0.006***
Hybrid	0.025	0.013	0.066*
Other	-0.000	0.005	0.989
Purpose			
Remortgage	0.010	0.003	0.001*
Renovation	0.015	0.006	0.009**
Construction	0.001	0.002	0.614
Other	0.004	0.005	0.386
Guarantee presence (binary)	-0.008	0.005	0.070*
OLTV (%)	0.0002	0.000	0.003***
Maturity at origination (years)	0.001	0.000	0.004***
Borrower income (log)	-0.001	0.002	0.657
LTI at origination (ratio)	0.001	0.000	0.004***

Notes: *** Significant at 1% level; ** Significant at 5% level; * Significant at 10% level.
Coefficients and marginal effects are deemed to be significant at least at a 5% level in the context of this paper.
Number of observations: 2,934,462.
Pseudo R-square: 0.1594.
Percentage of defaulted observations: 5.47%.
Percentage of correctly classified observations (cut-off: 5.47%): 66.07%.
Percentage of correctly classified defaulted observations (cut-off: 5.47%): 82.78%.
1) See Box 3 for an explanation of the computation of these marginal effects.

As regards borrowers' employment status, it should come as no surprise that unemployed and self-employed borrowers exhibit significantly higher probabilities of

20 Loans without a reported vintage date were automatically excluded from this and subsequent regressions.

default (1.4 percentage points and 1.9 percentage points higher,[21] respectively) than their employed counterparts. These results are in line with those in Section 3 and with the findings of Quercia et al. (2012) and Vandell and Thibodeau (1985). The coefficient associated with loans to legal entities is also significantly positive, indicating an 8.1 percentage points higher probability of default than among employed borrowers. A contributing (and, in a way, incentivising) factor could be the limited personal liability of the legal entity in the event of default. Furthermore, the repayment incentives may also be greater for the primary home of a private person.

As regards payment type, the baseline is an annuity-based repayment profile. The "other" category has a significant effect on default probability, which is 4.5 percentage points higher. This is supported by Mayer, Pence and Sherlund (2009), who find higher default rates for loans with "atypical" payment schedules.

Floating-rate loans show a significantly positive coefficient indicating that they are more likely to default, by 1.4 percentage points, than the fixed-rate baseline. Chart 10 hints at higher default rates for floating-rate loans, but strong country-level preferences for either fixed- or floating-rate loans mean that the effect could not be distinguished from other effects stemming from the country of loan origination. However, even when controlling for country-level effects in the regression, the relationship holds. Floating-rate loans shift interest rate risks to borrowers, so that the amount of future instalments is unknown and they are more likely to have difficulties servicing their instalments in case the interest rate rises. Moreover, there might be an adverse selection problem, as floating-rate loans could be seen as more affordable or easier to obtain than fixed-rate loans (Pennington-Cross and Ho (2010)), which would encourage riskier borrowers to self-select themselves for this type of loan (Posey and Yavas (2001)).

The intended purpose of a loan can also be an indicator of the likelihood of default. For example, loans for renovating a property show a significantly positive coefficient equivalent to a 1.5 percentage points higher default likelihood than traditional loans for house purchase. This appears to contradict the evidence in Chart 19, where the aggregate default rate is lower for the "renovation" category. However, that could be motivated by the fact that Belgian loans dominate the "renovation" subsample. At country level, four countries show a trend in line with the results described here. When compared to that same baseline, remortgages also display a higher (1 percentage point) default likelihood, which is line with the results in Chart 19.

Moving on to the continuous variables in the model, the coefficients associated with the loan's maturity at origination and OLTV ratio are significantly positive, indicating that default probabilities tend to go up both with longer repayment schedules (0.1 percentage point higher for each extra year) and with higher levels of leverage in the loan (0.02 percentage point higher for each 1 percentage point increase in OLTV ratio). Both results are in line with the evidence presented in Section 3 and confirm the

[21] These values are representative for the average loan (with all covariates set at their means, in line with the approach set out in Box 3). This applies to all figures in this section.

findings of Epley et al. (1996) regarding maturity and of Demyanyk and van Hemert (2011) and Dietsch and Welter-Nicol (2014), for instance, regarding the OLTV ratio.

As regards borrowers' income, both relative and absolute measures were included in the model (in the form of an LTI ratio and the log of income, respectively). As posited in Section 3, the LTI ratio is shown to be the only significant measure of income, as a higher level of relative indebtedness appears to be associated with a higher likelihood of default (0.1 percentage point higher per extra ratio unit), as seen in Campbell and Dietrich (1983) and Kelly and O'Toole (2016). Only when LTI is not included in the regression does log income become significant (Annex B, Table B.1).

Given the inclusion of income in the regression both as a standalone variable and as part of the LTI ratio, a correlation coefficient between the two was computed, in order to assess any potential multicollinearity issues. This coefficient was found to be -0.54, indicating that there is no significant multicollinearity problem.[22]

4.2 Regression with country interaction terms

A recurring theme throughout this paper has been the predominance of country-specific effects in the dataset. Therefore, a potential improvement to the model described in Section 4.1 could be including interaction terms that used the loan's origination country as one of the variables. As mentioned in Section 3.2, the interest rate type and payment type fields are heavily affected by country specificities, so there is a good case for including them in the model under such terms. Interacting both those variables with the origination country yielded the results in Table 3.

[22] Correlation coefficients between all other regressors have also yielded low figures.

Table 3

Probit regression results – with interaction terms[1)]

Dependent variable: NPL flag	Coefficient[2)]	Standard error	P-value
Employment status (binary)			
Unemployed	0.198	0.059	0.001***
Self-employed	0.254	0.055	0.000***
Legal entity	0.699	0.214	0.001***
Student	-0.118	0.064	0.065*
Pensioner	-0.086	0.122	0.482
Other	0.099	0.097	0.307
Payment type (binary)			
Linear	-0.048	0.105	0.645
Bullet	-0.274	0.142	0.053
Other	0.114	0.171	0.506
Payment type * country			
Belgium * linear	-0.436	0.129	0.001***
Belgium * bullet	0.671	0.219	0.002**
Germany * bullet	0.468	0.149	0.002***
Ireland * bullet	0.386	0.177	0.029**
Spain * linear	-0.573	0.256	0.025**
Portugal * linear	0.396	0.174	0.023**
Interest rate type (binary)			
Floating	-0.000	0.095	0.999
Other	-0.556	0.176	0.002***
Interest rate type * country			
France * floating	0.568	0.186	0.002***
Ireland * floating	-0.671	0.210	0.001***
Portugal * floating	-0.557	0.165	0.001***
Purpose			
Remortgage	0.137	0.039	0.000***
Renovation	0.204	0.070	0.003***
Construction	-0.011	0.038	0.760
Other	0.067	0.072	0.355
Guarantee presence (binary)	-0.129	0.072	0.075*
OLTV	0.004	0.001	0.003***
Maturity at origination	0.010	0.004	0.008***
Borrower income (log)	-0.008	0.029	0.774
LTI at origination	0.012	0.004	0.005***

Notes: *** Significant at 1% level; ** Significant at 5% level; * Significant at 10% level.
Coefficients and marginal effects are deemed to be significant at least at a 5% level in the context of this paper.
Number of observations: 2,934,462.
Pseudo R-square: 0.1637.
Percentage of defaulted observations: 5.47%.
Percentage of correctly classified observations (cut-off: 5.47%): 65.77%.
Percentage of correctly classified defaulted observations (cut-off: 5.47%): 83.38%.
1) For the sake of brevity, only significant interaction terms (at 5% level) are reported in Table 3.
2) Since the marginal effect computation would distribute the effect of the interactions among its components (i.e. it would not be possible to compute a standalone marginal effect for interactions), coefficients are displayed here (in contrast to Table 2, where marginal effects are shown).

In order to ensure that the model remains parsimonious, the categories for the variables included in the interaction terms were redefined, with certain types of interest rate (hybrid) and payment (increasing and fixed instalments) being merged into the "other" category.

Nearly all the results in Table 2 still hold, with the only exceptions being precisely the payment type and interest rate type dummies, since those standalone variables now refer only to the baseline country (the Netherlands) and are not significant.

Analysing the results for the remaining countries, it can be seen that floating-rate loans appear to have a lower default rate in Ireland and Portugal as compared to the baseline, whereas the opposite is true in France. In all three countries, loans with the least common type of interest rate in that jurisdiction are perceived as most likely to default. In Portugal, for instance, only 0.2% of the total loan amount in the sample[23] is subject to fixed interest. It could thus be posited that these loans are "special" cases, where this uncommon characteristic is symptomatic of an *a priori* higher obligor riskiness.

As regards payment type, the results show that Belgian, German and Irish bullet loans tend to default more than annuities. This is also true for Portuguese loans with a linear payment structure, whereas the reverse applies for Belgian and Spanish loans. A principle similar to that for the interest rate type variable may also apply to bullet loans, since these are the least common payment types in the countries in question (although this is also true of every other jurisdiction in the sample). This reasoning does not apply, however, for loans with a linear payment structure, since they do not represent the most common payment type in Belgium and Spain, where they tend to default less often.

4.3 Assessing the predictive power of the model

In general, assessing the predictive ability of any econometric model involves comparing the "fitted" values obtained from the model (i.e. using the estimated coefficients to obtain a predicted value for each observation) with the actual status of each observation (i.e. in this case, whether or not the observation is, in reality, defaulted). For binary choice models, such as the probit used here, the "fitted" values can be interpreted as predicted probabilities. As such, one can state that in case the predicted probability of default for a given loan exceeds a predetermined threshold, e.g. 50%, then the loan will be predicted to default.

With this in mind, one can construct a classification table comparing the predicted defaults and non-defaults with the observed ones:

[23] In France, 4.1% of the total loan amount in the sample refers to loans with floating interest rates, whereas in Ireland the figure (for fixed-rate loans) is 1.7%.

Table 4
Classification table

	Observed default	Observed non-default
Predicted default	Number of true positives	Number of false positives
Predicted non-default	Number of false negatives	Number of true negatives

When assigning a default prediction to all loans with a probability of default above 50%, as in the previous example, one finds that 94.53% of all observations in the current model are correctly predicted (in sample).

However, looking at just this one figure can be misleading. In fact, on closer inspection, with this 50% threshold, only 0.03% of the actual defaulted loans were correctly predicted. In other words, the sensitivity of the model (or its true positive ratio) was rather low (0.03%) with this threshold.

This tends to happen in samples with a low number of positive observations (defaults, in this case). In such instances, since the default ratio of the sample is very small, it will be rare to find high predicted probabilities of default. As such, when a high threshold is used, such as 50% in this example, very few cases will be predicted to default. However, as is the case here, nearly all non-defaulted loans will be correctly predicted (in other words, there is a high specificity or true negative ratio). Since non-defaulted loans make up the bulk of the sample, then the total percentage of correctly predicted observations (94.53%) ends up being very high.

This shows that there is a trade-off between a model's sensitivity and specificity, i.e. between its true positive and true negative ratios. However, from a microprudential perspective, it is rather more important to correctly identify the defaulted observations. This indicates that a threshold of less than 50% should be used to meaningfully assess the predictive ability of the model.

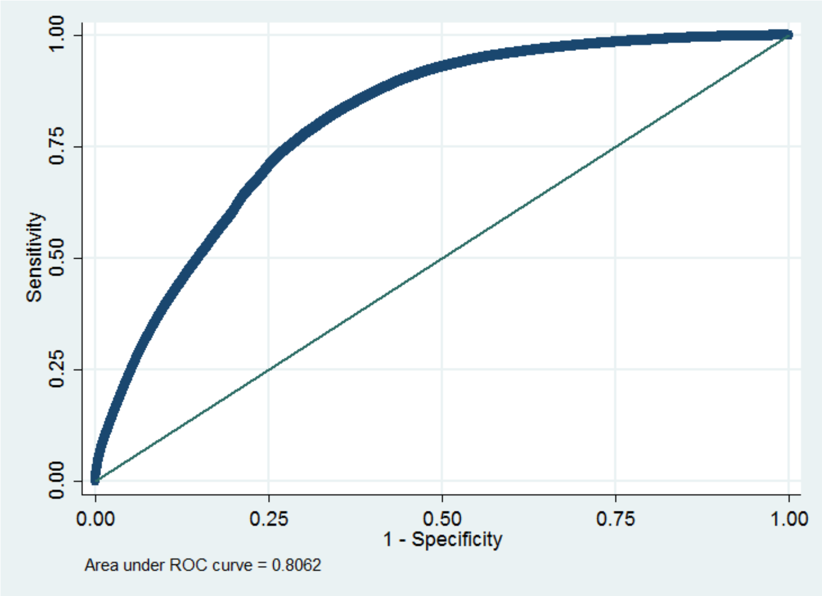
By lowering the cut-off value to, for instance, the sample default ratio (5.47%), the model correctly predicts 66.07% of the observations. However, the percentage of actual defaults that are correctly predicted, i.e. the model's sensitivity, sharply increases to 82.78%.

As shown, these figures are clearly impacted by the chosen threshold/cut-off value. In order to overcome this issue and obtain a single figure that is representative of the model's predictive ability, one can plot a receiver operating characteristic (ROC) curve which shows the performance of a model across all possible cut-off values. The curve plots the sensitivity and specificity values for each possible cut-off value. Chart 21 shows (in dark blue) the ROC curve for the in-sample estimation of the probit model without country interaction terms in Table 2.

Chart 21

Receiver operating characteristic (ROC) curve

(y-axis: sensitivity, or the true positive rate; x-axis: 1-specificity, or the false negative rate; decreasing cut-off rates from left to right)



Sources: EDW and ECB calculations

A random model (such as a coin toss, for instance) that flagged as defaulting 50% of loans would be accurate precisely 50% of those times. In other words, its true positive rate would always be equal to the cut-off value in use. Such a model is represented in Chart 21 by the 45° line. It represents the baseline with which the ROC curve of any other model can be compared.

The predictive ability of the main probit model used in this paper (see Table 2) can be assessed by the area under the ROC curve (AUROC). The computed AUROC in this case was 80.62% (significantly higher than the 50% AUROC for the random model, which indicates that the model in use has a good predictive ability).[24]

[24] The AUROC for the probit model with interactions estimated in Table 3 was 80.92%.

5 Conclusion

This paper investigates and, most importantly, quantifies the preponderance of lending standards as default rate drivers. Such efforts are particularly significant in the current environment, as several euro area national credit markets are still burdened by elevated NPL ratios (legacies from last decade's financial crisis).

Derived from a consistent dataset, the results presented here provide a unique basis for formulating relationships between lending standards and default rates at euro area level. They are intended to serve as a useful tool for both macro- and microprudential supervisors to compare their analysis and formulate or confirm their decisions regarding bank lending policies across euro area countries. Such supervisory activities are of paramount importance to the functioning of euro area credit markets, as they have the power to influence, both directly and indirectly, the lending policies of supervised financial institutions and hence the future development of NPL ratios in the euro area.

Regarding borrower-specific characteristics, loans to unemployed and self-employed individuals appear to be markedly riskier than those to employed clients. The same can be said for loans granted to legal entities. Furthermore, elevated LTI ratios were found to be correlated with higher probabilities of default, showing that a borrower's relative indebtedness is a good indicator of creditworthiness. While one can interpret the LTI ratio as a measure of relative income, this paper also proves that wealthier clients should not necessarily be thought of as safer obligors, particularly if granted comparatively large loans.

In what concerns loan-specific variables, it can be stated that floating- and hybrid-rate loans carry higher probabilities of default. However, this result seems to depend on the national market in which the loans are inserted, and whether or not fixed- or floating-rate loans are favoured there. In Portugal and Ireland, for instance, the reverse correlation applies, with floating-rate loans defaulting less than their fixed-rate counterparts.

Default rates also appear to increase with higher-leverage loans (as measured by loans' OLTV ratios) and with longer maturities at origination. Lastly, remortgages and loans granted for renovation purposes appear to exhibit higher riskiness than other categories, as is the case for loans with more "exotic" payment structures.

These conclusions are confirmed by both the descriptive and the regression analyses, which produce broadly similar results. While the regression analysis makes it possible to correct for country specificities or sample composition effects, the descriptive charts can be used to investigate the role of certain lending standards at country level.

Apart from technical/data-related improvements – such as including loan parts in the analysed dataset and resolving general data quality issues that often stem from the data provider (something which is being increasingly addressed with each new submission date) –, this paper could benefit from an extension of the study in it contained to other lending markets and/or non-securitised market segments, in order to avoid potential sample selection biases. As regards extending the work to other lending markets, the upcoming Anacredit loan-level dataset is expected to provide useful opportunities to investigate lending to non-financial corporations. Another possible extension would involve constructing a dynamic model, in order to take certain relevant time-varying macroeconomic factors into account, which was not possible with the available dataset.

References

Albertazzi, U., Eramo, G., Gambacorta, L. and Salleo, C. (2015), “Asymmetric Information in Securitization: an Empirical Assessment”, *Journal of Monetary Economics*, Volume 71, pp. 33-49.

Ambrose, B., LaCour-Little, M., Sanders, A. (2005), “Does Regulatory Capital Arbitrage, Reputation, or Asymmetric Information Drive Securitization?”, *Journal of Financial Services Research*, Volume 28, Issue 1-3, pp. 113-133.

Anastasiou, D., Louri, H. and Tsionas, M. (2016), “Non-performing Loans in the Euro Area: are Core-periphery Banking Markets Fragmented?”, *Working Paper*, No. 219, Bank of Greece, Athens, December.

Beck, R., Jakubik, P. and Piloiu, A. (2013), “Non-performing Loans: What Matters in Addition to the Economic Cycle?”, *Working Paper Series*, No. 1515, European Central Bank, Frankfurt am Main, February.

Berger, A. and DeYoung, R. (1997), “Problem Loans and Cost Efficiency in Commercial Bank”, *Journal of Banking & Finance*, Volume 21, Issue 6, pp. 849-870.

BIS (2017), “Resolution of Non-performing Loans – Policy Options”, *FSI Insights on Policy Implementation*, No. 3.

Calcagnini, G., Farabullini, F. and Giombini, G. (2009), "Loans, Interest Rates and Guarantees: Is There a Link?," *Working Paper Series in Economics, Mathematics and Statistics*, University of Urbino Carlo Bo, No. 2009/04, Urbino, April.

Calhoun, C. and Deng, Y. (2002), “A Dynamic Analysis of Fixed- and Adjustable-Rate Mortgage Terminations”, Journal of Real Estate Finance and Economics, Volume 24, Issue 1-2, pp. 9-33.

Campbell, J. and Cocco, J. (2015), “A Model of Mortgage Default”, *Journal of Finance*, Volume 70, Issue 4, pp. 1,495-1,554.

Campbell, T. and Dietrich, J. (1983), “The Determinants of Default on Insured Conventional Residential Mortgage Loans”, *Journal of Finance*, Volume 38, Issue 5, pp. 1,569-1,581.

Cocco, J. F., 2013. “Evidence on the Benefits of Alternative Mortgage Products,” Journal of Finance 68, pp. 1,663-1,690.

Dell’Ariccia, G., Igan, D. and Laeven, L. (2012), “Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market”, *Journal of Money, Credit and Banking*, Volume 44, Issue 2-3, pp. 367-384.

Demyanyk, Y. and van Hemert, O. (2011), “Understanding the Subprime Mortgage Crisis”, *Review of Financial Studies*, Volume 24, Issue 6, pp. 1,848-1,880.

Deng, Y., Quigley, J. and van Order, R. (2000), “Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options”, *Econometrica*, Volume 68, Issue 2, pp. 275-307.

Dietsch, M. and Welter-Nicol, C. (2014), “Do LTV and DSTI Caps Make Banks More Resilient?”, *Débats Économiques et Financiers*, No.13, Banque de France, Paris, August.

Dutch Banking Association (2014), *The Dutch Mortgage Market*, Amsterdam.

ECB (2009), "Housing Finance in the Euro Area", *Occasional Paper Series*, No. 101, European Central Bank, Frankfurt am Main, March.

ECB (2017), *Recent House Price Increases and Housing Affordability*, ECB Economic Bulletin, Issue 1/2018, Box 3, European Central Bank, Frankfurt am Main.

Elul, R. (2009), “Securitization and Mortgage Default: Reputation vs. Adverse Selection”, *Working Paper No. 09-21*, Federal Reserve Bank of Philadelphia, Philadelphia, September.

Epley, D., Liano, D. and Haney, R. (1996), “Borrower Risk Signaling Using Loan-to-Value Ratios”, *Journal of Real Estate Research*, Volume 11, Issue 1, pp. 71-86.

ESRB (2015) “Report on residential real estate and financial stability in the EU”, *ESRB Reports*, European Systemic Risk Board, Frankfurt am Main, December 2015.

Foote, C., Gerardi, K., Goette, L. and Willen, P. (2010), “Reducing Foreclosures: No Easy Answers”, *NBER Macroeconomics Annual 2009*, Volume 24, pp. 89-139.

Glen, J. and Mondragón-Vélez, C. (2011), “Business Cycle Effects on Commercial Bank Portfolio Performance in Developing Economies”, *Review of Development Finance*, Volume 1, Issue 2, pp. 150-165.

Jiang, W., Nelson, A. and Vytlacil, E. (2014), “Liar’s Loan? Effects of Origination Channel and Information Falsification on Mortgage Delinquency”, *The Review of Economics and Statistics*, Volume 96, Issue 1, pp. 1-18.

Jimenez, G. and Saurina, J., (2003), “Loan characteristics and credit risk”, *Proceedings*, No. 857, Federal Reserve Bank of Chicago, Chicago.

Jordà, Ó., Schularick, M. and Taylor, A. (2013), “When Credit Bites Back”, *Journal of Money, Credit and Banking*, Volume 45, Issue s2, pp. 3-28.

Kelly, R. and O’Toole, C. (2016), “Lending Conditions and Loan Default: What Can We Learn From UK Buy-to-Let Loans?”, *Research Technical Paper*, No. 04/RT/15, Central Bank of Ireland, Dublin, June.

Keys, B., Mukherjee, T., Seru, A., Vig, V. (2010), “Did Securitization Lead to Lax screening? Evidence from Subprime Loans”, *The Quarterly Journal of Economics*, Volume 125, Issue 1, pp. 307-362.

Klein, N. (2013), “Non-Performing Loans in CESEE: Determinants and Macroeconomic Performance”, *Working Paper*, International Monetary Fund, March.

Knapp, L. and Seaks, T. (1992), “An Analysis of the Probability of Default on Federally Guaranteed Student Loans”, *The Review of Statistics and Economics*, Volume 74, No. 3, pp. 404-411.

Mayer, Ch., Pence, K. and Sherlund, S. (2009), “The Rise in Mortgage Defaults,” Journal of Economic Perspectives, 23, No. 1, pp. 27-50.

Quercia, R., Pennington-Cross, A. and Tian, C. Y. (2012), “Mortgage Default and Prepayment Risks Among Moderate- and Low-Income Households”, *Real Estate Economics*, Volume 40, Issue 1, pp. 159-198.

Pennington-Cross, A. and Ho, G. (2010), “The Termination of Subprime Hybrid and Fixed-Rate Mortgages”, *Real Estate Economics*, Volume 38, Issue 3, pp. 399-426.

Posey, L. and Yavas, A. (2001), “Adjustable and Fixed Rate Mortgages as a Screening Mechanism for Default Risk”, *Journal of Urban Economics*, Volume 49, Issue 1, pp. 54-79.

Salas, V. and Saurina, J. (2002), “Credit Risk in Two Institutional Regimes: Spanish Commercial and Savings Banks”, *Journal of Financial Services Research*, Volume 22, Issue 3, pp. 203-224.

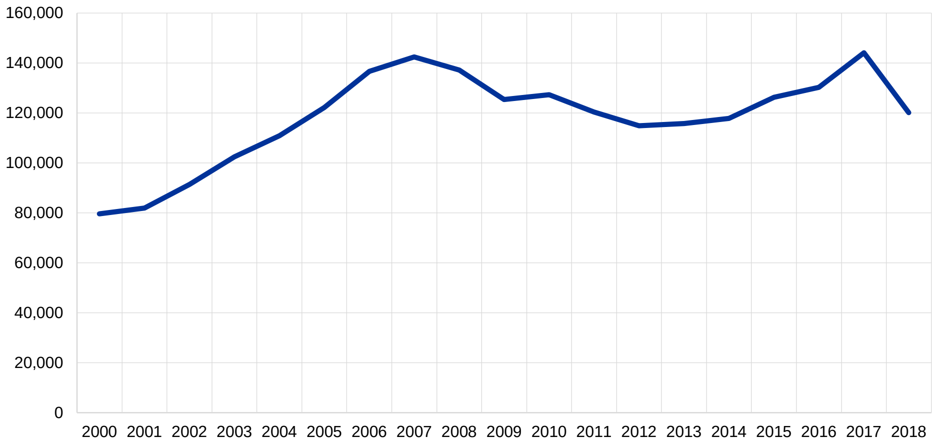
Vandell, K. and Thibodeau, T. (1985), “Estimation of Mortgage Defaults Using Disaggregate Loan History Data”, *Real Estate Economics*, Volume 13, Issue 3, pp. 292-316.

Von Furstenberg, G. (1970), “The Investment Quality of Home Mortgages”, *The Journal of Risk and Insurance*, Volume 37, Issue 3, pp. 437-445.

Annex A

Chart A.1
Average loan size by vintage

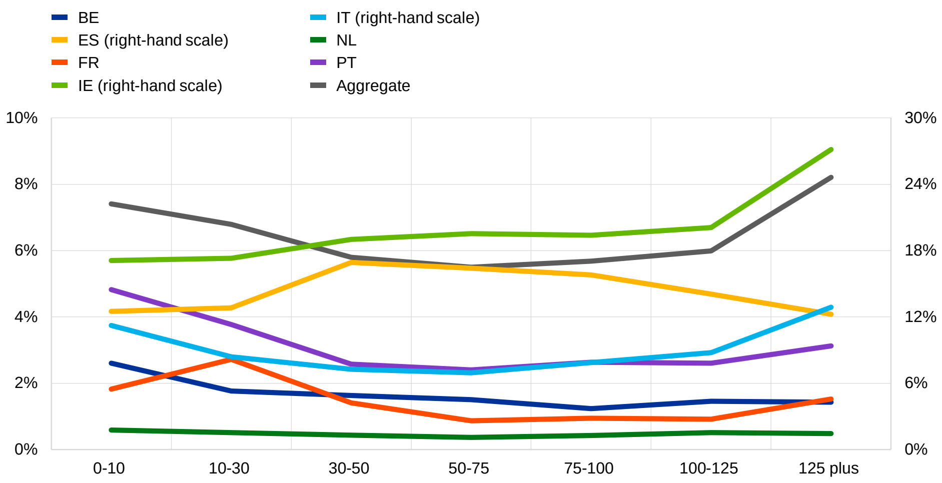
(x-axis: year of origination; y-axis: average loan size, measured by the loan's original balance, in EUR, per year of origination)



Sources: EDW and ECB staff calculations.

Chart A.2
Loan default rate by borrower's income bucket

(x-axis: gross annual borrower's income in EUR thousands; y-axis: loan default rate in percentages)

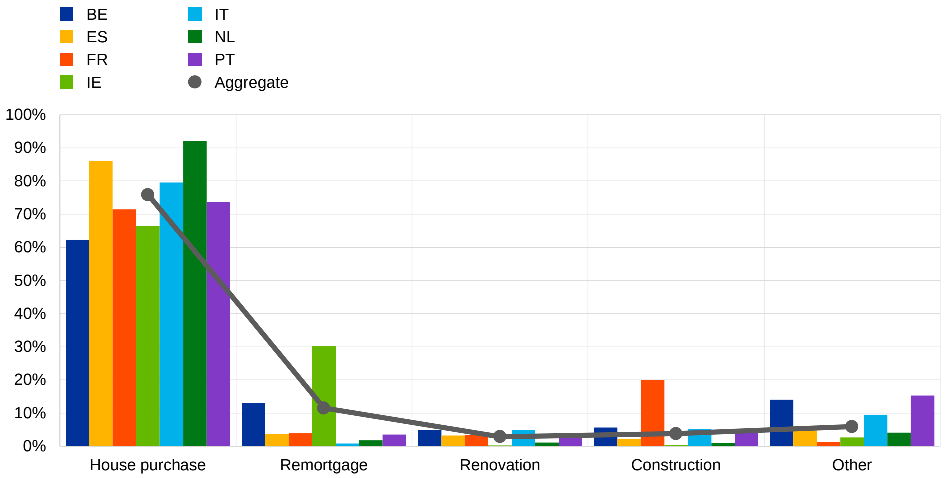


Sources: EDW and ECB staff calculations.

Chart A.3

Distribution of loans by purpose

(x-axis: loan purpose; y-axis: share of total original balance in the sample originated in a given country per loan purpose; percentages)

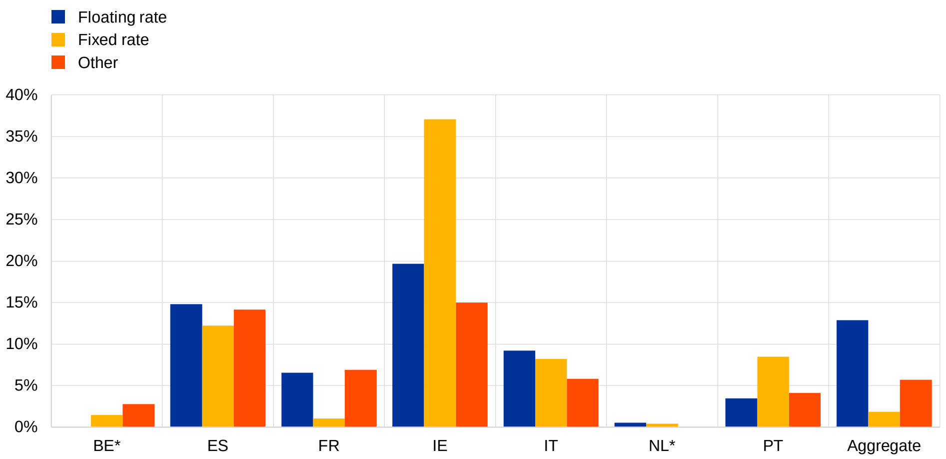


Sources: EDW and ECB staff calculations

Chart A.4

Loan default rate by interest rate type

(x-axis: country of origination; y-axis: loan default rate in percentages per interest rate type)



Sources: EDW and ECB staff calculations.
Note: * No floating-rate loans reported for Belgium and “other” interest rate type loans reported for the Netherlands.

Annex B

Table B.1
Probit regression results – without LTI as explanatory variable

Dependent variable: NPL flag	Marginal effects (at means)[1]	Standard error	P-value
Employment Status (binary)			
Unemployed	0.014	0.004	0.001***
Self-employed	0.019	0.004	0.000***
Legal entity	0.085	0.042	0.041**
Student	-0.006	0.003	0.070*
Pensioner	-0.005	0.006	0.419
Other	0.005	0.007	0.477
Payment type (binary)			
Linear	0.000	0.005	0.964
Increasing inst.	-0.012	0.007	0.061*
Fixed inst.	0.005	0.006	0.435
Bullet	-0.003	0.006	0.624
Other	0.045	0.019	0.021**
Interest rate type (Binary)			
Floating	0.014	0.005	0.006***
Hybrid	0.024	0.013	0.068*
Other	-0.000	0.005	0.973
Purpose			
Remortgage	0.010	0.003	0.001***
Renovation	0.017	0.005	0.002***
Construction	0.001	0.002	0.563
Other	0.005	0.004	0.239
Guarantee presence (binary)	-0.009	0.005	0.063*
OLTV	0.0002	0.000	0.007***
Maturity at origination	0.001	0.000	0.023**
Borrower income (log)	-0.004	0.002	0.010***
Loan size at origination (log)	0.005	0.002	0.005***

Notes: *** Significant at 1% level; ** Significant at 5% level; * Significant at 10% level.
Coefficients and marginal effects are deemed to be significant at least at a 5% level in the context of this paper.
1) See Box 3 for an explanation of the computation of these marginal effects.

Acknowledgements
The authors would like give their thanks to Prof. Thorsten Beck for his valuable insights and guidance, to Klaus Düllmann, the RIA NCA Network and the RRE Expert Group members for their helpful comments at different stages of the project, and to Paul Mayer, Adrien Amzallag and the members of European DataWarehouse team for their helpful support with the database and data related issues. We are also grateful to Calogero Brancatelli for his work on the analytical and data infrastructure at the early stages of the project.

João Gaudêncio
European Central Bank, Frankfurt am Main, Germany; email: joao.gaudencio@ecb.europa.eu

Agnieszka Mazany
European Central Bank, Frankfurt am Main, Germany; email: agnieszka.mazany@ecb.europa.eu

Claudia Schwarz
European Central Bank, Frankfurt am Main, Germany; email: claudia.schwarz@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

PDF ISBN 978-92-899-3685-9, ISSN 1725-6534, doi:10.2866/04696, QB-AQ-19-003-EN-N