BANK OF ENGLAND

Staff Working Paper No. 760

Determinants of distress in the UK owner-occupier and buy-to-let mortgage markets

Vladimir Lazarov and Marc Hinterschweiger

October 2018

BANK OF ENGLAND

Staff Working Paper No. 760

Determinants of distress in the UK owner-occupier and buy-to-let mortgage markets

Vladimir Lazarov[(1)] and Marc Hinterschweiger[(2)]

Abstract

The mortgage market has played a central role in the global financial crisis. One particularly pressing question surrounds the conditions under which mortgage borrowers enter distress, ie get into arrears or default. This paper develops a novel micro dataset from residential mortgage loans which UK banks and building societies have pre-positioned with the Bank of England for use as collateral in exchange for central bank funding. The dataset is used to investigate the determinants of borrower distress as a function of borrower and loan-level stock/flow characteristics over the loans' lifetime in the buy-to-let (BTL) and owner-occupier (OO) mortgage markets. We find systematic differences between these two markets, controlling for a range of loan and borrower characteristics as well as macro variables. Our main result shows that, adjusting for affordability, the loan-to-value ratio is reliably more important for borrower distress in the OO market than for distress in the BTL market, contradicting McCann's (2014) results.

Key words: Distress, default, arrears, mortgage lending, loan level, micro data, United Kingdom.

JEL classification: C23, C55, G21.

(1) Bank of England. Email: vladimir.lazarov@bankofengland.co.uk
(2) Bank of England. Email: marc.hinterschweiger@bankofengland.co.uk

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Publications and Design Team, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 3461 4030 email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1. Introduction

The central role of the mortgage market in the global financial crisis has led to a surge in academic research on the topic, especially in a US context where micro level data is readily available. One particularly pressing question surrounds the conditions under which mortgages become distressed or enter default. Understanding the determinants of distress of residential mortgages is important for a number of reasons. First, as the single largest liability on the UK household sector's balance sheet, mortgages can have important effects on the economy, for example when borrowers reduce their spending in an economic downturn to keep up with their mortgage payments. Second, mortgage lending can pose risks to financial stability as it is the largest asset class on UK banks' balance sheets. Indeed, these risks were flagged by the Financial Policy Committee of the Bank of England (the "Bank") in its June 2017 Financial Stability Report (FSR) [1]. Together with the December 2015 FSR[2], it focused in particular on the increase in the UK buy-to-let sector in recent years, which had almost doubled in size since the period before the financial crisis. These developments are the main drivers for this research paper.

We are able to shed light on these questions in the context of the UK by virtue of a novel dataset that tracks the performance of residential mortgage loans which have been pre-positioned with the Bank of England[3] by banks and building societies[4] for use as collateral in exchange for central bank funding. Our paper thereby increases the availability of mortgage micro data in the UK, alleviating Aron and Muellbauer's (2016) concerns around the limited number of empirical studies on distress in the UK mortgage market. Our sample has monthly data, collected from January 2013 to June 2016. It comprises 3.5 million loans, worth £469bn (as at end June 2016), that are originated between 1972 and 2016. They account for 41% of all UK mortgage market loans. Our variables cover borrower characteristics, loan characteristics, and property/collateral information; importantly, the dataset tracks the performance (e.g. arrears and default) of residential mortgage loans over their lifetime.

We exploit the richness of this micro dataset to understand the determinants of the probability of distress (arrears and default) in the owner-occupier (OO) and buy-to-let (BTL) mortgage markets as a function of borrower and loan-level stock/flow characteristics both at origination and over the loans' lifetime, as well as macro variables. In so doing, we investigate the differences in the performance of loans between the BTL and OO markets. To the best of our knowledge, previous academic research has either only focused on the OO market or has been based on loan-level stock/flow data at origination. In this context, the paper tests an important finding by McCann (2014) that the loan-to-value (LTV) ratio is more important for explaining default in the BTL market than in the OO market. We find that the relationship is robustly significant in the opposite direction in our borrower-level specifications, i.e. the LTV ratio is a

[1] See https://www.bankofengland.co.uk/-/media/boe/files/financial-stability-report/2017/june-2017.pdf
[2] See https://www.bankofengland.co.uk/-/media/boe/files/financial-stability-report/2015/december-2015.pdf
[3] See https://www.bankofengland.co.uk/-/media/boe/files/quarterly-bulletin/2014/quarterly-bulletin-2014-q2
[4] Banks and building societies are jointly referred to as lenders, banks, or firms throughout this paper.

less important driver of distress for the BTL sector than for OO sector (noting that our sample covers a period of overall low distress, which may influence the results). This is because the data allows us to reduce a source of bias that may be present in loan or property-level data that cannot be aggregated at the borrower level; such data may miss important linkages between mortgages.

The rest of the paper is organised as follows: Section 2 briefly reviews the literature most relevant for this paper. Section 3 discusses the data. Section 4 details the strategy and methodology of our modelling. Section 5 presents our main results as well as several robustness checks and extensions. Section 6 concludes.

2. Related Literature

The empirical literature on the determinants of borrower distress in the mortgage market is extensive. This is especially the case in a US context, where micro data at the loan and/or borrower level are relatively readily available publically. Quercia and Stegman (1992) review the US literature from the 1960s until the early 1990s, and Jones and Sirmans (2015) expand and extend their review until 2014. In the UK, less micro data is publically available, and therefore the evidence in the UK is scarcer, or relies on aggregate data. Aron and Muellbauer (2016) give an overview of disaggregated and macro-based studies on mortgage arrears and repossessions in the UK.

These studies usually take into account a range of loan and borrower characteristics as well as market and economic conditions. Loan characteristics that are often found to be very significant drivers of distress or default are a loan's initial LTV ratio, current LTV ratio, or negative equity. For example, Aron and Muellbauer (2016) find a significant effect of negative equity on repossessions and arrears in their estimation of a system of three equations based on UK data. Combining loan-level mortgage data with credit bureau information in the US, Elul et al. (2010) highlight both negative equity and illiquidity as main drivers of default. They also discover that these two factors interact: the effect of illiquidity increases with the current LTV ratio. Using data from the UK Help to Buy Equity Loans scheme, Benetton, Bracke and Garbarino (2018) find that smaller down-payments are associated with a higher ex-post default rate. Authors often identify additional variables that are relevant in their data. For example, Lanot and Leece (2016) use data from a single UK originator to find that unobserved heterogeneity and self-certification are significant predictors of default.

Studies that focus on borrower characteristics often consider changes in individual circumstances, such as divorce or illness, but in particular unemployment. Due to the lack of micro data, aggregate unemployment at the regional level is often used as a substitute. However, Gyourko and Tracy (2014) find (in a US context) that this is often a poor proxy for individual unemployment and can lead to default risk being underestimated by a factor of more than 100. Nevertheless, Aron and Muellbauer (2016) find that the (log of the) aggregate unemployment rate is positive and significant with regards to repossessions and arrears. Similarly, Aron & Muellbauer (2011) use regional county court claims and orders for mortgage possession in a quarterly panel data model, identifying the debt service ratio, negative equity, and the unemployment rate (all measured at the regional level) as important drivers. Using data from the British Household Panel Survey (BHPS) from 1991 to 1997, Böheim & Taylor (2000) find that both individual and regional unemployment are significantly positively related with housing finance problems (though the coefficient for the duration of unemployment is negative). Health problems and divorce have a similar effect. Including subsequent waves of the BHPS until 2006, Gathergood (2009) reports comparable results. Authors also discover relevant variables that have not been previously considered. For example, Chauvet, Gabriel and Lutz (2016) create a real-time mortgage default risk index based on Google search query data that can predict housing returns, mortgage delinquency indicators, and subprime credit default swaps.

The so-called double trigger hypothesis often combines loan and borrower characteristics in describing mortgage default risk (set out, for example, by Aron and Muellbauer, 2016; Financial Conduct Authority, 2015; and Foote, Gerardi, and Willen, 2008). According to this hypothesis, negative equity, potentially a consequence of falling house prices, is not a sufficient condition for default. In addition, payment difficulties, maybe as a result of unemployment or illness, also need to be present for borrowers to stop making their mortgage payments. While this hypothesis can shed some light on the low incidence of default among highly leveraged borrowers, Gerardi, Herkenhoff, Ohanian, and Willen (2018) conclude that the evidence regarding this hypothesis is mixed.

On market characteristics, Lambrecht, Perraudin and Satchell (2003) provide evidence, based on a small sample of UK mortgages provided by a mortgage insurer, that the economic cycle has an important influence on the resolution of delinquent mortgages. This affects the amount and timing of recoveries and therefore the riskiness of mortgages.

To the best of our knowledge, the majority of papers that model the probability of mortgage default have focused on the owner-occupier housing market and have mainly used either aggregate data or point-of-sale loan-level data. Some research, such as Financial Services Authority (2009), has focused on estimating probabilities of default for a large sample of lenders covering 80% of the owner-occupier market and about 45% of the BTL market. But it focuses only on a single snapshot of lenders' own books. In contrast to our dataset, the UK Finance Buy-to-Let Mortgage Survey ("UK Finance") data does not track the performance of mortgages during their lifetime.

In terms of the data used, the research paper closest to ours is McCann (2014). It is the first academic research paper to use loan-level data from multiple UK monetary and financial institutions to estimate UK mortgage defaults for both the BTL and OO markets. The paper uses a Markov multi-state model to predict the probabilities of default for the BTL and OO sectors and models mortgage default as a function of housing equity (proxied by LTV), regional unemployment, loan interest rates and a range of other variables. In an extension, the paper finds that OO defaults are less sensitive to housing equity than are

BTL defaults. This is of central interest to us as we wish to study how systemically different the BTL and OO sectors are and how they are different in the sensitivities of loan distress to their respective relevant characteristics. However, McCann (2014) mainly uses variables which are updated annually and semi-annually, with only 'months-in-arrears' data updated monthly in the period from 2010 to 2013.

To summarise, our paper extends these analyses in several ways:

- Wider market sample: our panel has a larger sample of firms than comparable studies
- Wider range of controls: our panel includes more variables varying over the loan lifetime
- More recent data (2013 to mid-2016) (which covers a period of relative calm in the market and hence low levels of distress)
- Higher frequency of available data
- Abundance of borrower and borrower-level loan characteristics which allows us to carry out our analysis at the borrower level as well as at the loan level

3. Data

a) Description of the dataset

This paper uses proprietary and confidential loan-level data owned by the Bank of England's Financial Risk Management Division (FRMD) comprising mortgage loans that UK banks and building societies use as pre-positioned collateral portfolio of loans ("loan pools"). Pre-positioned collateral may be used by banks and building societies to participate in the Bank of England's operations in the sterling money markets[5], such as the Sterling Monetary Framework (SMF) (i.e. Discount Window Facility (DWF), Indexed Long-Term Repo (ILTR), Contingent Term Repo Facility (CTRF)), Funding for Lending Scheme (FLS), and Term Funding Scheme (TFS) operations[6].

The FRMD database contains loan-level information on mortgages that 41 UK banks and building societies have been submitting since 2013, updated on a monthly basis[7]. This database has 202 fields, containing mandatory and optional variables, covering borrower characteristics, interest rate, loan characteristics, the lender and servicer of the loan, performance (e.g. arrears), and property / collateral information. 70 of these variables are dynamic, i.e. are updated by the reporting institutions (i.e. banks and building societies) and vary over the lifetime of the loans. The remaining 132 variables are static, i.e. are collected at origination and, unless the banks voluntarily submit an update (which rarely happens), are not updated after submission. This database consists of 95% residential and 5% other loans. Of the residential loans, about 84% are OO loans and about 16% are BTL loans (with very few second/holiday

[5] See https://www.bankofengland.co.uk/-/media/boe/files/markets/sterling-monetary-framework/loan-collateral-pooling-explanatory-note.pdf
[6] See https://www.bankofengland.co.uk/-/media/boe/files/markets/sterling-monetary-framework/operating-procedures.pdf
[7] See https://www.bankofengland.co.uk/-/media/boe/files/markets/eligible-collateral/residential-mortgages/residential-mortgages-loan-level-data-template.xls

homes). These numbers are representative of the UK mortgage market as of 2016, with home-owners and BTL buyers comprising 83.3% and 16.7% of the market, respectively[8]. Table 1 in the appendix presents the definition of the variables we make use of in this analysis. What is of particular significance in this database is that we have valuable (anonymised) borrower information such as a unique borrower identifier and a unique collateral identifier for each loan for a given lender that allows us to perform our analysis at the borrower level in addition to the loan level.

The total database consists of about 3.5 million loans originated between 1972 and 2016 (all loans originated before 1999 represent 0.6% of all loans). Chart 1 in the appendix shows current and original balances in this database since 1999. Chart 2 draws the coverage of our database relative to the stock of all Monetary and Financial Institution's (MFI's) mortgage lending over our study window. As of end June 2016, our database covers about 41% of all MFI's mortgage lending (worth about £469bn), in both the BTL and owner-occupier markets, and (in most recent years) about a third of mortgage originations. The database follows these mortgages until they are repaid by the borrowers or withdrawn from their pool by firms (which is very rare). This is reflected in Chart 3 by the average current and original loan-to-value (LTV) balances by origination year. As we are interested in the impact of LTV on distress probabilities for the OO and BTL sector, Chart 4 shows changes over time of the median LTV in each market. Another loan feature we are interested in is the origination channel, shown in Chart 5 for both the OO and BTL markets.

Overall, the main advantages of this database over existing datasets are as follows:

- Relative to the Financial Conduct Authority's (FCA) Product Sales Data (PSD)[9] (which is a loan-level point-of-sale flow dataset of regulated mortgages), our database tracks mortgages through time (from an earlier period and monthly), has a wider range of variables (arrears, etc.) and includes BTL loans.
- The FCA's new dataset with performance data (PSD 007) has only a few quarters of data points.
- Similar to the PSD, the UK Finance Buy-to-Let Mortgage Survey ("UK Finance") loan-level BTL dataset[10] for 2014-2016 does not follow mortgages during their lifetime as the FRMD database does.

We are limited to a random sample from this database as running regressions on the whole database is infeasible for us due to operational reasons. But, as described in Section 5, we are able to extend our preferred model to encompass all impaired loans in the database by virtue of the particular regression design that we choose.

[8] See https://www.cml.org.uk/industry-data/key-uk-mortgage-facts/
[9] For more information, see https://www.fca.org.uk/firms/product-sales-data
[10] Provided to the Bank of England by UK Finance (the Council of Mortgage Lenders, CML, which originally provided the data, has recently become part of UK Finance)

Despite its various advantages, this data may suffer from the following biases, discussed in detail below: a) there may be some adverse selection in the loans banks choose to preposition with the Bank; b) there may exist some selection bias towards larger banks due to their participation being more likely in the Bank's operations[11]; and c) the incidence of mortgage defaults/distress is relatively low in the UK market over the time horizon under study, which is reflected in our data sample.

We address these in turn. On point a), banks have become less selective over time in the mortgages they choose to pre-position with the Bank of England (as part of the loan collateral pooling under the Sterling Monetary Framework and FLS within a set of eligibility criteria[12]). And any initial bias will diminish as more recent mortgages enter the database and old ones disappear (e.g. because they get paid off). Moreover, there are fewer niche loans ("pockets") and fewer smaller loans in recent submissions, as the average size of loans is increasing. Historically, most banks had pre-positioned enough of collateral originated by 2012 and did not need to keep posting collateral (e.g. as a fixed percentage of new originations) although a few banks kept bringing new loans[13]. Hence, while banks have the option to add and remove mortgages, we do not think there is a material difference compared with other similar datasets (see below).

To address b), we also compare this database of loans with characteristics of the other most relevant datasets available – PSD and UK Finance data, which have a more balanced representation of firms by size of their balance sheet. We examine the representativeness of our sample of pre-positioned loans by comparing them to stock/flow estimates from PSD and UK Finance. This comparison is necessarily limited to the variables available in both datasets. Table 2 summarises the findings of our comparison. It needs to be kept in mind, firstly, that the PSD and UK Finance datasets are different from our sample in a number of respects. The UK Finance dataset available to the Bank of England is a point-of-sale (flow) dataset and cannot be directly compared. The PSD dataset was also generated on a flow basis until 2015, when, additionally, performance data on stocks have become available; however, for this piece of research, we were only able to obtain a single snapshot for June 2016, which permits us to make a point-in-time comparison with the final date of our dataset. Secondly, the PSD dataset has many more (smaller) banks than are in the FRMD database. Thirdly, the PSD dataset has no information on the BTL market. Lastly, the UK Finance data and PSD contain less frequent data points (quarterly) than the FRMD database (monthly). After taking all of these aspects into consideration, we find that according to Table 2, our owner-occupier data sample is similar to the PSD dataset in arrears incidence, repayment structure, and market size (in terms of the value of mortgages).

To address c), we use arrears as well as default in a measure we call distress. We define distress as all loans in arrears for two or more months (including default), which addresses the relatively low incidence

[11] This kind of bias is not relevant from the point of view of collateral valuation, where the aim is to understand the characteristics of those specific mortgages likely to be used in the Bank's operations.
[12] See https://www.bankofengland.co.uk/-/media/boe/files/markets/sterling-monetary-framework/level-c-loan-collateral.pdf
[13] Bank of England's FRMD policy is to not take loans less than 2 months into origination.

of default in the collateral database[14]. While past studies have often used default[15], we believe that distress as defined by us is an appropriate modelling choice for two reasons. Firstly, arrears are generally a very good predictor of default beyond the first month (which could reflect temporary factors such as standing orders failing or lack of funds in mortgage accounts). Secondly, this definition helps us attain more variability in our modelling.

b) Descriptive statistics

Table 3 in the appendix provides descriptive statistics which show that our randomly selected sample of loans is representative of the entire database. We then clean our data sample to remove various inconsistencies and data errors. Table 4 shows descriptive statistics of our dependent variable and independent variables, including interaction variables, after cleaning the sample.

To shed light on the structure of our dataset over time, Chart 6 shows the total number of mortgages in our sample in every year. In order to track when mortgages enter our sample, the bars are broken down by the year when a given mortgage appears in our sample for the first time. This should not be confused with the vintage of a mortgage, i.e. it is possible that a mortgage first enters our sample in, say, 2014, but was originated in 2005, though the years of origination and inclusion may also coincide. Therefore, this chart shows the stock of mortgages in each year, not the flow of new mortgages. In each year, the stock of mortgages in our sample is broadly representative of the wider market, as described above.

The LTV and affordability variables are of particular interest in our analysis. Naturally, properties cannot get valued every month. Banks report the valuation at the origination of the mortgage and the Bank of England requests an audit of the lender's data submission to be performed by an external audit firm. Typically, valuation gets updated every few years (at the beginning of the year) and at key events, such as further advances, second liens, or remortgaging. But this practice differs by lender. In a rising market, it is possible that this may cause a certain downward bias in the current LTV ratios, but we think this effect should be limited, due to the re-valuations as described above. Charts 7 and 8 provide loan level and borrower-level LTV distributions for our random sample. While the LTV distribution at the loan level (Chart 7) shows the usual distribution for LTVs[16], Chart 8 incorporates a borrower's other loans (with the same lender) and has a smoother, bimodal distribution. Chart 9 shows the properties of our (personal income) affordability proxy in our random sample. On average, borrowers in our sample tend to spend about 17% of their income on mortgage loans.

---

[14] The low incidence of defaults in our database is a reflection of the low incidence of defaults in the UK market.

[15] For example, the Bank for International Settlements (BIS) defines default for exposures secured by real estate of as 90 days past due (see www.bis.org/bcbs/publ/d403.pdf, p.6), or the PRA allows default for IRB firms' exposures secured by residential real estate in the retail exposure class to be defined as 180 days past due (see SS11/13, p.14, www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/supervisory-statement/2017/ss1113update.pdf).

[16] The shape of the loan-level LTV distribution has multiple mass points due to the way business tends to be underwritten by banks and building societies in many cases, i.e. with LTV ratios in 5% increments – e.g. 60%, 65%, 70%, 75%, etc.

As the dependent variable is of particular importance to our analysis, we provide additional descriptive statistics for distress. First, we show distress rates at both the loan and borrower level for each year of our sample (Table 5). The rates tend to fall somewhat over time in our sample, for both the owner-occupied and the buy-to-let markets.

Second, we calculate the distress rates by the year of origination of the mortgages where sufficient observations are available (Chart 10). It is important to note that these rates are conditional on the mortgages being observed in our sample, so they suffer from survivorship bias. For example, a mortgage originated in 2005 can only be in our sample if it has not been defaulted on. Therefore, the numbers shown in 2005 in the chart do not reflect the probability of distress of a mortgage in 2005, but the probability of distress of a mortgage originated in 2005 at any point in time between 2013 and 2016, conditional on the mortgage being observed in our sample. The most obvious pattern is that mortgages originated roughly before the global financial crisis have a substantially higher probability of distress than mortgages originated after the crisis (though this pattern occurs a bit later in the BTL sector). One explanation is that underwriting standards for post-crisis mortgages may be higher than before the financial crisis (because of closer regulatory/supervisory scrutiny and/or because of changes in banks' own risk appetite), with the looser pre-crisis standards still echoing today in the distress risk of mortgages that have survived through the global financial crisis. One caveat to interpreting the chart is that loans originated earlier would have by definition more scope to go into distress.

Finally, Table 6 shows the transition probabilities between being in distress and not being in distress, and vice versa, for both the loan and the borrower level. While the chance of getting into distress is relatively low in every given month, mortgages/borrowers that are in distress have an about 10% chance of becoming current again in the next month.

4. Modelling

a) Strategy

As discussed in the section below on the level of modelling, we prefer to perform our main analysis at the borrower level. This is why we first construct a cleaned (unbalanced) panel dataset of a random sample of loans from the FRMD database amounting to 10% (150,000 borrowers) of all borrowers (1.5 million) in the entire FRMD collateral database and track all of these borrowers' loans with all of their borrower, loan, and collateral characteristics from January 2013 till the end of our study period – June 2016[17]. Next, we model our data sample econometrically and determine which model works best on this 10% random sample. Lastly, we are able to extend our sample, due to our preferred modelling

[17] We follow this approach as extracting and modelling the entire database of loans (which is 1.2 TB of data) from an Oracle database has proven too time-intensive for the current state of technology.

technique's feature where only distress transitions have an effect on optimisation, to all impaired loans in the whole database (i.e. all borrowers that ever had a loan at least two months in arrears).

b) Empirical Methodology

We draw a series of random selections of loans by repeated sampling from the database amounting to 10% of all borrowers that ever appeared in the database, and track all theirs loans over time as they appear in the database. We have chosen this approach since, as explained in more detail below, we need to make sure that, when modelling performance of loans, we do not omit variables that can influence the probability of distress which would depend on certain borrower characteristics in cases where a borrower has more than one loan or remortgages / refinances their loan. This, in fact, raises the more general question as to the appropriate level of modelling borrower or loan distress – this we address in the 'Level of modelling' section below.

From those loans, we have removed from our data sample pre-positioned loans to high net worth individuals and loan portfolios explicitly flagged as residential mortgage-backed securities (RMBSs) because they are a small percentage of the number of loans in the entire database and are special cases that might bias our results. What we are left with are secured residential mortgage loans to individuals.

i. Dependent variable

Our dependent variable, called distress, is a dummy equal to 1 if a loan is in default or in arrears for two or more months, and 0 otherwise.[18]

ii. Independent variables

In line with the literature set out above, we expect the following categories of variables to have the most explanatory power in relation to our dependent variable:

- Loan characteristics
- Borrower characteristics
- Property/collateral information
- Interest rate information
- Performance data
- Macroeconomic indicators

[18] This definition was also used in the Bank of England's Financial Stability Report 40, Nov 2016, see https://www.bankofengland.co.uk/-/media/boe/files/financial-stability-report/2016/november-2016.pdf, p.35, Chart A

Based on these categories, previous studies, theoretical research, market intelligence, and Table 1, we construct the following variables that we think would influence the probability of distress (we explain in the 'Key hypotheses' section below how we would expect these variables to be associated with our dependent variable).

Most importantly, we calculate the Loan-to-Value ("LTV") ratio as the ratio of the outstanding balance to the current valuation of the loan – obtained from two variables: 'Total (Current Balance)' and 'Total (Current Valuation Amount)'. Similarly, we construct a debt service coverage ratio (which we call "Affordability"), obtained as 'Payment Due' divided by 'Joint Monthly Income', which takes into account both the primary and secondary borrower's incomes. We also use the variable 'Current Interest Rate' of the loan.

We also construct several dummy variables. Our Buy-to-Let dummy is equal to 1 if 'Occupancy Type' is 'Non-owner-occupier/buy-to-let', and 0 otherwise. Partially owner-occupier and Holiday homes are a very small proportion of the FRMD database and have been removed so that our reference group is the OO market. The Self-employed dummy equals 1 if 'Employment status of the primary applicant' is 'Self-employed'. Similarly, the Remortgaging dummy equals 1 if the 'Loan purpose' is 'Remortgage', 'Remortgage with Equity Release', or 'Remortgage on Different Terms'. We also include dummy variables for Increasing instalment, Bullet payment, and Broker-intermediated loans. They equal 1 if a loan's 'Payment Type' is 'Increasing instalments', 'Bullet'[19], or if 'Origination channel' is 'Broker', respectively (see Table 1 as well as Chart 5 for the latter). In line with the literature, we also calculate a dummy for negative equity, equal to 1 if the LTV ratio is greater than 100%. In certain of our regressions, we include a First time buyer dummy equal to 1 if the loan's borrower is a first time buyer, and a Right-to-buy dummy equal to 1 if the loan's borrower is a right-to-buy buyer.

In order to assess any differential effects of the relevant dummy variables above between the BTL and OO sectors, we calculate a series of interaction variables with the BTL dummy, respectively.

Additionally, we compute the number of 'Further Loan Advances' for a given borrower at any point in time. We also use the age of the primary borrower (in years).

In terms of the macroeconomic variables, our regional unemployment rate variable is the seasonally adjusted regional unemployment rate according to the UK NUTS classification (integrated with the collateral database's loan-level region) from the ONS. Similarly, the 'regional house index' variable is the regional house index of UK regions according to the NUTS classification. Last, regional house sales volume represents the regional number of house sales transactions from the ONS.

[19] Bullet type loans combined with savings deposits, life insurance, or investment portfolios account for an extremely small proportion of our results.

iii. Level of modelling

Many studies have looked at modelling loan default at the loan level, usually because of a lack of data on borrowers or inconsistent data. However, we would argue that when looking at ratios like loan-to-value (LTV), loan-to-income (LTI), debt service coverage ratio (DSCR), etc. as significant predictors of distress, we would make a mistake if we did not incorporate into those ratios information pertaining to other loans of the same borrower. Otherwise, we would only have a partial picture of the driving factors of distress with respect to the borrower serving these mortgage loans. In our sample, for instance, we have found that, on average, a borrower has 1.25 loans (at a given point in time), which should be taken into account in the regression design.

We have valuable borrower information such as a unique collateral identifier and unique borrower identifier for each uniquely defined loan that we incorporated in our modelling. This allows us to form a total LTV ratio for each borrower based on all of his/her available loans in the database. We do not have perfect information when, for example, a borrower would have different advances on top of their loans with other banks not in our collateral database. Even if all of a borrower's loans are with institutions in our data sample, differences in lenders' reporting systems[20] may prevent us from allocating all relevant loans to the respective borrower. However, it is relatively rare that a borrower would have advances with the same collateral at a different institution, or sometimes even other mortgage loans with other banks. Nevertheless, within our database and sample, we have ensured that, if we selected any given loan, we identified and extracted all other loans to whomever this loan belongs to and incorporated this information in our respective borrower-level variable. This is how we drew our random sample – we took repeated random samples of borrowers in the database totalling 150,000 borrowers with all of their loans, and tracked each of these borrowers' loans over their lifetime in our study window.

Therefore, we consider that borrower-level modelling would be more justified, other than in the cases where we would be interested in distress driver features which are not meaningful at the borrower level. For example features such as first-time buyer and right-to-buy buyer would be better suited to be explored at the loan level. Therefore, to complement and compare our analysis and explore the effects on distress of first-time buyers and right-to-buy buyers, we also run regressions at the loan level.

Our panel's cross-sectional variable is borrower (or loan), and our panel's time variable is months. In constructing our borrower-level variables with the list above, we have made the following assumptions:

1. For each borrower-month combination, we have formed a total current LTV ratio that incorporates all loans of a given borrower at that time and divided it by the total current valuation of his/her properties.
2. If a borrower has at least one BTL loan, then that borrower is flagged as BTL in our BTL dummy.

---

[20] Banks are not required at this time to follow a coordinated and unique reporting of their loans, borrowers, and collateral across their systems, and naturally reporting systems at different institutions in use differ.

3. A borrower is self-employed if he/she is either self-employed and the only borrower on her/his loans; or, in case there is a secondary borrower, if both are self-employed (this is a static variable).
4. The remortgaging, further advances, bullet-payment loans, increasing instalment, and broker-intermediated loans dummy variables are a sum of the respective flags for all of a borrower's loans.
5. The negative equity dummy for a borrower is based on the total current LTV across their loans.

iv. Estimating the probability of distress

Econometric approach

Unobserved borrower-level specific factors may bias our results. In our baseline model, we introduce fixed effects at the borrower level to account for this. We also control for macroeconomic effects (which affect all borrowers at a point in time) but are not captured by the variables we included to account for this.

We have experimented with different modelling approaches, such as panel Tobit and panel Poisson models (see discussion in Section 5). Our preferred approach is a fixed-effects logistic model at the borrower level. We perform F-tests and Hausman tests and find that it is best to go with a fixed-effects, rather than a pooled or random-effects, model. Results of these tests are shown in Table 7.[21]

Another consideration in picking the appropriate model concerns what the focus of our analysis is. If we were interested in the probability of distress of a borrower taking a loan at time, say, t_2, with high LTV, then we would be doing an analysis from the viewpoint of a bank or building society and, in this case, the appropriate model specification might have been a pooled model. However, we are more concerned with removing unobserved heterogeneity which may confound the effects of our variables of interest on borrower distress. In this case, a fixed effects framework which exploits both time- and cross-sectional variation in covariates is more appropriate.

Empirical specification

We first estimate a panel logistic regression on our random sample of the general specification:

[21] We have checked with F-tests and Hausman-type (1978, 1981) post-regression testing that non-pooled OLS estimators and random-effects estimators are not optimal (and their nulls are rejected in favour of the fixed-effects estimator). To decide between the fixed effects and simple pooled OLS estimator we use Hausman's test. This is because a significant F-test would mean that the fixed effects are non-zero, but an F-test does not tell us if pooled OLS estimates would be biased if Cov (Xit, ui) ≠ 0. Whether this latter condition is true is addressed by the Hausman test because it might be that the fixed effects are non-zero, but that they are uncorrelated with our time-varying explanatory variables.

$Y_{it} = \alpha_i + \beta X_{it} + \gamma_t + \varepsilon_{it}$, where the α_i's are at the borrower level.

In particular, we model Equation (1):

$$Distress_{it} = \alpha_i + \beta_1 LTV_{it} + \beta_2 Affordability_{it} + \beta_3 BTL_{it} + \beta_4 (BTL * LTV)_{it} + \beta_5 (BTL * Affordability)_{it}$$

$$+\theta(other\_characteristics)_{it} + \rho(other\_BTL\_interactions)_{it} + \varphi(macro\_variabless)_t + \varepsilon_{it}, \quad (1)$$

where *i* denotes borrower, *t* denotes month (t = 1, …, 41, where 1 = 2013m1), the dependent variable is 1 if months in arrears is greater than or equal to 2 and otherwise zero. The α_i's are the model's fixed effects, ($\beta_1 \dots \beta_5$) are the coefficients on the main variables of interest, θ, ρ, and ϕ are vectors of coefficients on other control variables, and ε is the error term.

A theoretical issue discussed in estimating non-linear panel models is with finite-dimensional panel models where we have a large cross-sectional dimension and relatively large (or borderline large) time dimension to be able to estimate the fixed effects parameters on which the coefficients of interest depend (the so called incidental parameters problem, see Neyman and Scott (1948)), discussed most recently in Fernandez-Val and Weidner (2016). For these non-linear panel models there is no 100% analytical form of bias correction but there exists a type of split panel jackknife correction estimator that Fernandez-Val and Weidner develop (although p-values tend to go larger to an extent). However, in our logistic specification we use the conditional MLE estimator which for our particular logit (as opposed to probit) form of the optimisation function properly drops all cross-sectional units for which there is no change in the dependent variable across the whole period (i.e. where the distress dummy is either zero or one for all time periods). We are thus able to integrate out the fixed effects and do not run into the incidental parameters problem, as discussed in Wooldridge (2002, pp.490-491). We therefore follow Wooldridge's (2002), and in particular Chamberlain's (1980), approach to use maximum likelihood estimation, and we get consistent estimates of the β's. It should be noted that with this model, we cannot make prediction since the α_i's cannot be recovered.

We do not include time fixed effects as we believe that they would not contribute to our analysis for two reasons – first, the period of study is marked by low incidence of distress and we do not see any particular sub-periods of interest; and second, we already have included some key macroeconomic parameters that are likely to influence all or part of our cross-sectional unit variation at given points in time. The absence of time effects allows us to avoid having to generalize the conditional maximum likelihood (CMLE) approach to include two fixed effects for the logit model as developed in Charbonneau (2014), and thus we can estimate our parameters of interest consistently under CMLE with fixed effects (assuming no serial correlation and heteroscedasticity).

We also prefer not to use a dynamic model under this specification for two reasons – first, our dependent variable, distress, moves slowly; and second, we found that such a model does little to the explanatory value of our model.

v. Key hypotheses

Our main conjecture, contrary to McCann's (2014) finding, is that the effect of the LTV ratio on distress would be lower for the BTL sector than for the OO mortgage market when assessed at the borrower level. This may seem counterintuitive as there is theoretical and empirical evidence that high LTV ratios, and in particular negative equity, may induce moral hazard (as an example, see the reasons set out in McCann, 2014). And this may be more so for investment properties than for an owner-occupied home, the latter of which a home-owner may want to avoid losing under all circumstances.

Much of this evidence is based on an analysis at the loan- or property-level. In fact, we can confirm this result in our loan-level robustness checks, as set out further below. But there may be arguments why this relationship may not hold at the borrower level, that is, by taking into account the portfolio (i.e. the number and types) of properties that a particular borrower owns. Everything else equal, a BTL borrower is more likely to be able to withdraw equity from at least one of their properties, and use these funds to cross-subsidize other properties, if necessary. This could take the form not only of keeping up with mortgage payments, but also of upgrading such properties, thereby increasing their value, instead of walking away from an under-water mortgage. Similarly, owning several properties may have diversification benefits. To illustrate this, take the extreme case in which a BTL investor owns dozens of properties. If one or a few of these properties have negative equity, the investor may be willing to ride out such instances, especially if house prices in the different areas where the properties are located are not highly correlated. Instead, an owner-occupier having negative equity on their single property may be less willing to take the risk of the property still being under water in the future when she may want to sell it (for example to move to a bigger house), and instead default on the mortgage strategically. For these two reasons, we might observe a higher ability and willingness of BTL investors to avoid falling into distress at a given LTV ratio by virtue of the additional flexibility that owning several properties may give them. These effects would not be directly observable in loan- or property-level regressions.

Secondly, we construct our affordability proxy based on personal (labour) income reported by borrowers at the time of taking a loan as opposed to rental income (which is in the database but is of relatively poor quality in that it is inconsistently reported and there is no rental income at time of mortgage origination – we hope to refine this in further research). While we believe that our thus defined affordability proxy should be positively significant in the OO market (as greater indebtedness would lead to higher probability of distress in serving a loan), we believe this (personal) income to be of lower significance and impact for the BTL market – and in fact to be insignificant if the assumption of non-substitutability of rental for personal income holds (including in times of distress).

Thirdly, we construct our models to be able to differentiate the probability of distress between the BTL and OO markets, providing for affordability and a range of other loan characteristics. This does not mean that, all other factors being equal, BTL borrowers (and loans) should have a lower probability of distress. In fact we suspect the opposite, i.e. that BTL loans would have a higher probability of going into distress, since BTL properties are not the main residence of a borrower and are therefore more dispensable.

We have also identified other variables which may impact distress probability. We think that higher interest rates should increase the probability of distress in the OO market, and even more so for the BTL market. Remortgaging, further advances, and especially negative equity should increase the distress probability for both markets. Further advances and remortgaging in the BTL sector can be even more pronounced due to the mainly interest-only nature of this sector. Bullet-payment loans or alternatively increasing instalment loans could be viewed as a disciplining factor and decrease distress probabilities.

5. Results

a) Random sample

We first discuss results from our random sample and then extend it to a full sample of all impaired loans in the next section.

The results from running Equation 1 on our estimation sample can be found in Table 8. The key finding from this table is that β_4, the coefficient of the interaction term of BTL with LTV, is reliably negative. This contradicts McCann's (2014) result and confirms our hypothesis. As expected, a higher LTV, per se, leads to a higher probability of going into distress, as indicated by the positive β_1. As this model is highly non-linear, we can only say that the probability of distress for a given change in LTV in a borrowers' loans position in the OO market is about three times higher than the one in the BTL market, but in order to assess the overall effects of the LTV ratio on distress, we need to look at the marginal effect (also reported in Table 8). By comparison with the LTV, the probability of distress is much less sensitive to changes in the affordability situation of a borrower as measured by the payment due with respect to the joint income of the borrower(s). The coefficient on affordability (β_2) is as expected positive and is not significantly different for the BTL sector (β_5). Borrowers in the BTL sector seem to have a higher probability of going into distress, all other things being equal, than the ones in the OO market (β_3's marginal effect).

Positive interest rate changes on the borrower's loans are positively associated with distress probability, as expected, but there seems to be no difference in this proclivity for borrowers with BTL loans. Self-employed borrowers seem to have a smaller probability per se to go into distress situations, however, marginally, this is not so for the BTL borrowers. Borrowers who remortgage are more likely to face higher distress, but (in this specification only) that seems to be much more so for those in the BTL market. Each further advance acts in the same way with an even more pronounced difference for the BTL sector. The sign for negative equity is as expected.

Increasing instalment loans seem to act as discipline compared to other types of loans and, as expected, lower the distress probability the same way for both markets. Bullet-payment loans act the same way but their effect is opposite for the BTL market[22]. Broker-intermediated loans have a positive effect on borrowers' distress on average in the same way for both BTL and OO markets. Regional unemployment (in this specification) seems not as relevant for a borrower's servicing of her debt, however housing index movements appear to lower borrowers' chance of being in distress. As mentioned in the literature review above, the insignificant coefficient that we find on unemployment may be due to the use of regional rather than individual unemployment data.

The (pseudo) R^2 of this regression is relatively low at 2.0%; however, it is not unusual to find such a low goodness of fit measure in borrower and loan level regressions due to the variety of unobserved individual drivers of distress, for example employment status over time, divorce, illness, etc.

It should be noted that care should be taken when extrapolating these results to other time periods outside of our sample (or to other countries). In the UK, the period from 2013 to 2016 can be seen as being characterised by relatively low volatility (e.g. no financial crisis or recession in the UK, relatively low and stable interest rates, no general downturn in the housing market). We would therefore caution against generalising these results to periods of higher volatility. As more data become available over time, further analysis may be able to shed more light on borrower distress under different economic circumstances.

b) Sample of impaired borrowers

We can improve upon these results by exploiting a feature of the particular regression design of our baseline model: only transitions in our dependent variable (i.e. borrowers moving into and out of distress) are relevant for the optimisation function of the estimation of a logistic CMLE model. We therefore extend our sample by extracting all loans from the FRMD database that experience at least one transition from being not in distress to being in distress, or vice versa (and aggregate them by borrower). Thus, we can effectively make estimations based on the full FRMD database, thanks to the choice of our regression design. The results from Equation 1 on all impaired loans in the database are presented in Table 9. We find that a 1% increase in the average LTV of borrowers leads to a marginal increase of 0.26% in the probability of borrowers going into 2 or more months in arrears. More importantly, for the BTL market this sensitivity to LTV changes is about a third weaker (by comparing the magnitude of the coefficients). Additionally, a one-percent rise in indebtedness in terms of the borrowers' debt service coverage ratios

[22] This finding for this 10% random sample specification may be partially due to collinearity issues from the BTL dummy as most BTL loans are bullet-paying.

(based on non-rental income) for borrowers in the OO market influences the distress probability by 0.03%, and this is the case for the BTL sector as well which confirms the results from our random sample.

Borrowers with BTL sector loans on average tend to have a greater probability of going into distress by 2.6% than ones solely in the OO market. Unit interest rate changes on a borrower's loan in the OO sector lead to a 1.63% increase in their probability to go into arrears for two or more months, while for borrowers in the BTL market this is about a third of this number.

Self-employed borrowers for both sectors tend to have a smaller probability of going into distress by 2.6%. This finding may be due to the fact that the self-employed who ultimately end up getting a mortgage are the more robust payers. On average, remortgaging in both sectors leads to a higher probability of distress by 2.8% (for the BTL sector this might be marginally lower if looking at the coefficients). Similarly, each further advance increase this probability additionally by 0.5% (for the BTL sector this may be marginally higher). On the other hand, increasing instalments and bullet-paying loans impose a discipline on borrowers and decrease distress probability, for both sectors by 1.6% and 0.9%, respectively. Broker-intermediated loans increase the probability of a borrower going into distress by 4.5%. Borrower age seems to play a role in raising distress probability for the average OO market buyer of 0.04% (for every year with respect to average age).

Macroeconomic variables relevant for borrower distress that we have identified are regional unemployment, housing indices and sales volumes. Unemployment in the region increases borrower distress probability by 0.6%, while positive changes in the housing index decrease this probability by 0.1%. Regional house sales volumes also decrease distress probability but to a very small extent.

c) Robustness checks and extensions

We explore several robustness checks and extensions of our baseline model, both for the random sample as well as the extended sample (where possible)[23]. These are briefly described in turn below.

i) Specifications of the dependent variable

Banks and building societies are encouraged to use their own definition of arrears. This has led to some banks submitting arrears equal to the fraction of actual payments due in arrears while most banks have stuck to integer numbers of months in arrears.

[23] Where not shown, all results are available from the authors upon request.

The nature of our arrears data would therefore lend itself to a range of alternative models. For example, rounding all the arrears data to the closest integer would enable the use of a count data model (such as the Poisson model). This is usually considered to represent stronger assumptions on the underlying distribution of the data. We experimented by rounding arrears and employing a panel zero-inflated Poisson model but this approach ran into convergence issues (which does not apply to the logistic model we used).

Alternatively, interpreting the distribution of arrears to be censored at zero, one could use a panel Tobit model. But this would unlikely be a good fit because of the data's near-count nature. For these reasons, we decided to use a logistic panel model for our baseline regressions.

ii) Pooled regressions

We run pooled regressions for both samples, with standard errors clustered at the borrower level. The results are qualitatively similar to the fixed effects regressions, with some notable exceptions. Most importantly, the signs of the BTL coefficient as well as some interactions terms have flipped (though the BTL coefficient is not significant in the random sample regression). Similarly, the coefficient on the broker-intermediated dummy is negative. We interpret these differences to result from the misspecification of the pooled regression since unobserved heterogeneity at the borrower level is not controlled for. Having said that, many of our key variables (such as LTV, affordability, and the interest rate) maintain the same sign and level of significance.

iii) Loan-level regressions

While we believe that the most suitable regression model is based on borrowers rather than loans as the cross-sectional units, certain features of the mortgage market and their impact on borrower distress can only be explored and have meaning at the loan level. This allows us to include several variables, such as first-time buyer effects (FTB) and right-to-buy (RTB)[24] loans, which would be difficult to interpret in a borrower setting as they are mainly a characteristic of the underlying mortgage rather than of the borrower.

Results from a loan-level run of Equation 1 are shown in Table 10. For reasons discussed before, here we are not interested in the value of the coefficients $\beta_1 \dots \beta_5$ (as we think the proper level of modelling is at borrower level – e.g. LTV's coefficient (β_1) is reliably negative but fails to incorporate important information on other loans of the same borrower) but rather in the direct impact on the coefficients on FTB and RTB loans. On the basis of the table we can say FTB loans are less likely by go into distress by

[24] For more information on RTB, see https://www.gov.uk/right-to-buy-buying-your-council-home/overview

about 19% than non-FTB loans. This is supported by evidence from Ireland[25]. Everything else equal, first time buyers may be particularly unwilling or unable to go into arrears, compared to other buyers, since this may imply losing their home. This is particularly true compared to borrowers who own more than one house that they could possibly move to. There may also be a larger degree of emotional attachment to one's first home. It is also worth noting, as set out above, that the coefficient on the interaction term of BTL with the LTV ratio is positive in this loan-level regression, contrary to our borrower-level results. We interpret this finding to indicate that some important information may potentially not be taken into account when relying on loan-level regressions.

iv) Bootstrapped errors

We use bootstrapping techniques to cross-check our results, following Brownstone and Valletta (2001). We rerun the above regression but bootstrap the standard errors to provide for robust errors. As Table 11 shows, our results are mainly confirmed but notably β_2 and β_3 turn insignificant in the random sample dataset. However, this is not the case in the larger sample based on the impaired borrowers, confirming the reliability of all of our main coefficients of interest ($\beta_1 \ldots \beta_5$) as well as the rest of the coefficients (Table 12).

v) Sub-samples

As discussed above, we caution against generalising our results beyond the time period for which our data was collected, i.e. 2013 – 2016, due to the prevalence of low volatility. Albeit an imperfect way to proxy the sensitivity of our results to different volatility regimes due to survivorship bias, we split our sample, based on the origination date of the first loan of a given borrower, into borrowers with loans that were originated before the global financial crisis (before 2008) and loans originated in 2008 and after. The results from the two sub-samples are very similar qualitatively to each other and to the full random sample. The most notable exception is that the coefficient on the affordability variable turns negative but insignificant in the sub-sample with loans originated in 2008 and after. It is possible that low interest rates after 2008 have acted as less of a constraint on borrowers' ability to service their debt than before the financial crisis when interest rates were considerably higher.

We also split the sample into two sub-samples based on whether a borrower gets into distress or whether they get out of distress. The results are also very similar qualitatively. In the impaired sample, the results are virtually the same, except that further advances turn significantly negative in the "getting into distress" sub-sample. In the random sample, BTL turns negative but insignificant in the "getting into

[25] See Review of Residential Mortgage Lending Requirements, pp. 28-29, based on CBI Research Technical Papers, Vol. 2015(2) (http://www.centralbank.ie/docs/default-source/publications/research-technical-papers/research-technical-paper-02rt15.pdf?sfvrsn=8)

distress” sub-sample, and affordability turns negative at the 5% confidence level in the “getting out of distress” sub-sample.

vi) Modelling the loan-to-value (LTV) ratio

An argument can be made that the LTV ratio should enter the regression model in a quadratic manner as well. This is the case if borrower distress is considerably higher for high LTV ratios, say, above 80%, but relatively flat below this point. This could be due to moral hazard or other unobserved factors.

We include a quadratic LTV ratio term in our regressions and also interact it with the BTL variable. In both samples, both the original and the quadratic term are positive and highly significant. The interaction term of BTL with LTV turns positive and significant, while the interaction term of BTL with the squared LTV ratio is significantly negative. However, the inclusion of the quadratic LTV ratio term did not improve the fit of the model substantially.

vii) Logs

It is possible that, for some variables, what matters are proportional changes rather than absolute changes. For example, one may argue that an increase in interest rates from 1 to 2% may have a higher impact on borrower distress (since interest rates double) than if interest rates increase from 4 to 5% (since interest rates increase by ‘just’ 25%). We therefore take the log of the relevant variables, i.e. the interest rate as well as our macro variables (house price index, sales volume index, and regional unemployment rate).

The sign and significance level of the affected variables remain essentially unchanged.

viii) Lagged independent variables

In addition to taking the logarithm of the variables mentioned in the previous section, we create first-differenced variables for them at the 1-month, 3-month, and 12-month horizon. The intuition is that changes in these variables over some time in the past might matter more than their contemporaneous level. For example, there may be a lag until an increase in regional unemployment rates affects arrears. We run separate regressions for each of the three lags in both of our samples.

The results are inconclusive. The signs of some of the affected variables flip, in particular at shorter horizons. It is difficult to reconcile these results with the established theoretical and empirical evidence. For example, the log of the interest rate is significantly negative in all regressions. But it is difficult to argue from an economic point of view why an increase in the (log of the) interest rate over the last one, three, or twelve months would be associated with a lower borrower distress rate today. We defer a more detailed analysis of these findings to future research.

6. Conclusion

This research paper contributes to improving our understanding of the determinants of borrower distress in both the buy-to-let and owner-occupier sectors. In particular, it can inform an assessment of the drivers of risk that may differ between both markets.

Many studies using micro data have so far been based on loan-level data at origination. Our research extends the literature with a novel mortgage dataset, collected by the Bank of England for its sterling monetary framework operations and the Funding for Lending Scheme in the form of pre-positioned collateral. This is the first paper to use this dataset for academic research purposes. Compared to some of the other datasets available in the UK, it includes information on the stock of mortgages in both the OO and BTL markets, contains a large amount of borrower, lender, mortgage, and collateral characteristics, and is available at monthly frequency. It allows us, within certain limits, to identify different loans belonging to the same borrower.

Our results show that, in periods of low distress and adjusting for non-rental income affordability, McCann's (2014) hypothesis that the loan-to-value ratio is more important for defaults in the BTL market than in the OO market does not hold at the borrower level, and that this relationship is reliably in the opposite direction. Furthermore, we find that BTL borrowers are by about a third less sensitive to changes in the average LTV ratio. On the other hand, we find that BTL borrowers are by about 2.6% more likely to go into distress than OO borrowers. Shocks to affordability (defined as mortgage payments as a share of joint income) contribute another 3.3% probability of borrowers in both mortgage markets experiencing distress.

We find that borrowers without BTL loans experience a higher distress probability to increases in their loan's interest rates than BTL borrowers. Other loan- and borrower-characteristics play a role as well. Self-employment, increasing instalment, and bullet payment loans discipline borrowers and decrease distress probability by 2.6%, 1.6%, and 0.9%, respectively. Remortgaging, further advances, and broker-intermediated loans act in the opposite direction – by 2.8%, 0.5%, and 4.5% respectively.

Based on our sample, first-time borrower loans are less likely to go into distress by about 16% than non-FTB ones. Macroeconomic variables such as regional unemployment and house price indices affect borrowers' distress probability by 0.6% and 0.1%, respectively.

Our main results are robust to alternative specifications, modelling approaches, and error assumptions. We find that our core hypotheses and results for the behaviour of our main variables of interest hold and are resilient to robustness checks when modelling all impaired borrower loans for about 41% of the UK owner-occupier and buy-to-let markets.

One extension would be to model the entire database (rather than a sample) with a different model that would be able to account not only for transitions but also for those loans that never go into arrears. This extension was not possible in our study due to computational limitations and the way the database has been compiled historically in the Bank of England's systems.

While it may be challenging to obtain detailed data on each BTL borrower's rental income over time for each property as this variable is reported very inconsistently, another possible extension is to obtain more detailed postcode-level data on rental income for the BTL sector. One way to do this is to get information at the first-level postcode (the first three or four letters in the postcode) and integrate this time-varying BTL rental income information into our data sample (which contains at least this first-level postcode information on a relatively consistent basis). This would refine our BTL affordability proxy.

Similarly, some other FRMD database characteristics may be included in the model which may shed light on questions of interest, such as originator and servicer categorical variables to see effects of different lending institutions, repayment method and other loan characteristics, borrower's credit score and additional borrower characteristics, among others.

References

Aron, J. & Muellbauer, J., (2011). Modelling and Forecasting with County Court Data: Regional Mortgage Possession Claims and Orders in England and Wales. *SERC Discussion Papers, Spatial Economics Research Centre, LSE 0070*, Spatial Economics Research Centre, LSE February 2011.

Aron, J., & Muellbauer, J. (2016). Modelling and forecasting mortgage delinquency and foreclosure in the UK. *Journal of Urban Economics, 94,* 32-53.

Benetton, M., Bracke, P., & Garbarino, N. (2018). Down payment and mortgage rates: evidence from equity loans. *Bank of England Staff Working Paper No. 713*.

Böheim, R. and Taylor, M. (2000). My Home Was My Castle: Evictions and Repossessions in Britain. *Journal of Housing Economics, 9 (4),* 287-319.

Brownstone, D. & Valletta, R. (2001). The Bootstrap and Multiple Imputations: Harnessing Increased Computing Power for Improved Statistical Tests. *Journal of Economic Perspectives, vol. 15, no. 4,* 129-141.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *Review of Economic Studies 47,* 225-238.

Charbonneau, K. (2014). Multiple Fixed Effects in Binary Response Panel Data Models. *Bank of Canada Working Paper 2014-2017*.

Chauvet, M., Gabriel, S., & Lutz, C. (2016). Mortgage default risk: New evidence from internet search queries. *Journal of Urban Economics, 96,* 91-111.

Elul, R., Souleles, N. S., Chomsisengphet, S., Glennon, D., & Hunt, R. (2010). What "triggers" mortgage default? *American Economic Review, 100(2),* 490-94.

Fernandez-Val, I., & Weidner, M. (2016). Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics, 192,* 291-312.

Financial Conduct Authority (2015). Methods for analysing mortgage markets. *Occasional Paper 11,* 20-21.

Financial Services Authority (2009). Financial Risk Outlook.

Foote, C. L., Gerardi, K., & Willen, P. S. (2008). Negative equity and foreclosure: Theory and evidence. *Journal of Urban Economics, 64(2),* 234-245.

Gathergood, J. (2009). Income Shocks, Mortgage Repayment Risk and Financial Distress Among UK Households. *Discussion Papers 09/03,* University of Nottingham, Centre for Finance, Credit and Macroeconomics (CFCM).

Gerardi, K., Herkenhoff, K. F., Ohanian, L. E., & Willen, P. S. (2017). Can’t pay or won’t pay? Unemployment, negative equity, and strategic default. *The Review of Financial Studies, 31(3),* 1098-1131.

Gyourko, J., & Tracy, J. (2014). Reconciling theory and empirics on the role of unemployment in mortgage default. *Journal of Urban Economics, 80,* 87-96.

Hausman, J. (1978). Specification Tests in Econometrics. *Econometrica, 46,* 1251-1271.

Hausman, J., & Taylor, W. (1981). Panel Data and Unobservable Individual Effects. *Econometrica, 49,* 1377-1398.

Jones, T., & Sirmans, G. S. (2015). The underlying determinants of residential mortgage default. *Journal of Real Estate Literature, 23(2),* 167-205.

Lambrecht, B., Perraudin, W. R., & Satchell, S. (2003). Mortgage default and possession under recourse: A competing hazards approach. *Journal of Money, Credit, and Banking, 35(3),* 425-442.

Lanot, G., & Leece, D. (2016). Mortgage Loan Characteristics, Unobserved Heterogeneity and the Performance of United Kingdom Securitized Subprime Loans. *Real Estate Economics, 44(4),* 771-813.

McCann, F. (2014). Modelling default transitions in the UK mortgage market. *Research Technical Paper 18/RT/14,* Financial Stability Division, Central Bank of Ireland.

Neyman, J., & Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica, 16 (1),* 1-32.

Quercia, R. G., & Stegman, M. A. (1992). Residential mortgage default: a review of the literature. *Journal of Housing Research,* 341-379.

Wooldridge, J. (2002). Econometric analysis of cross section and panel data. MIT Press, Cambridge, MA.

Appendix – Charts and Tables

Chart 1. Original and current balances in the Bank of England's FRMD collateral database

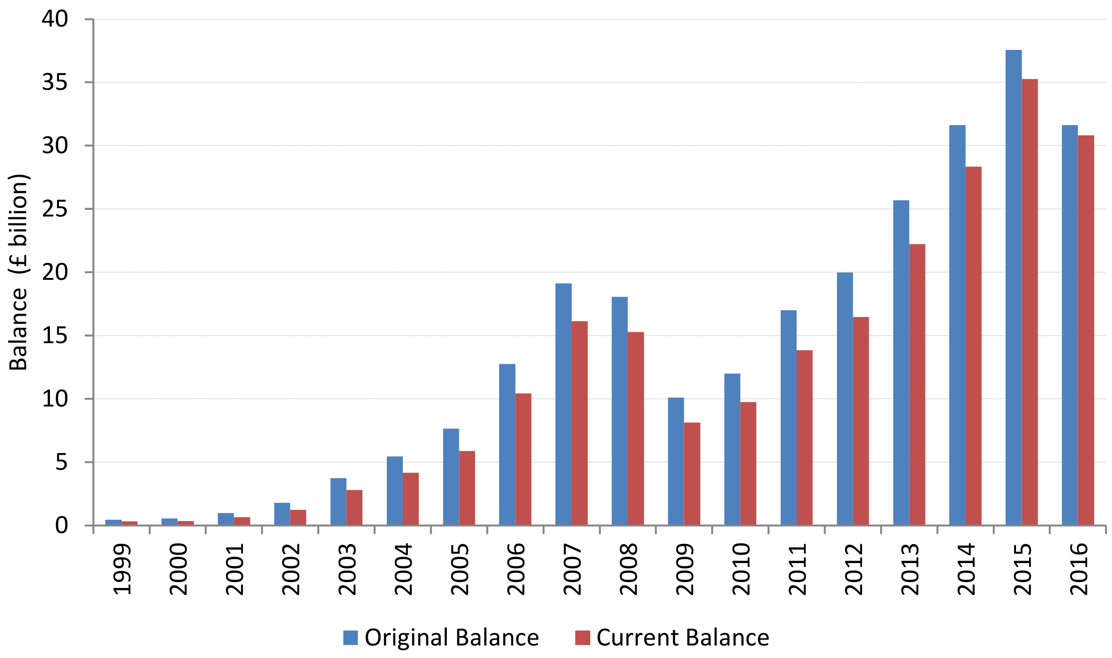

Chart 2. Coverage of collateral database as % of all MFI mortgage balances outstanding

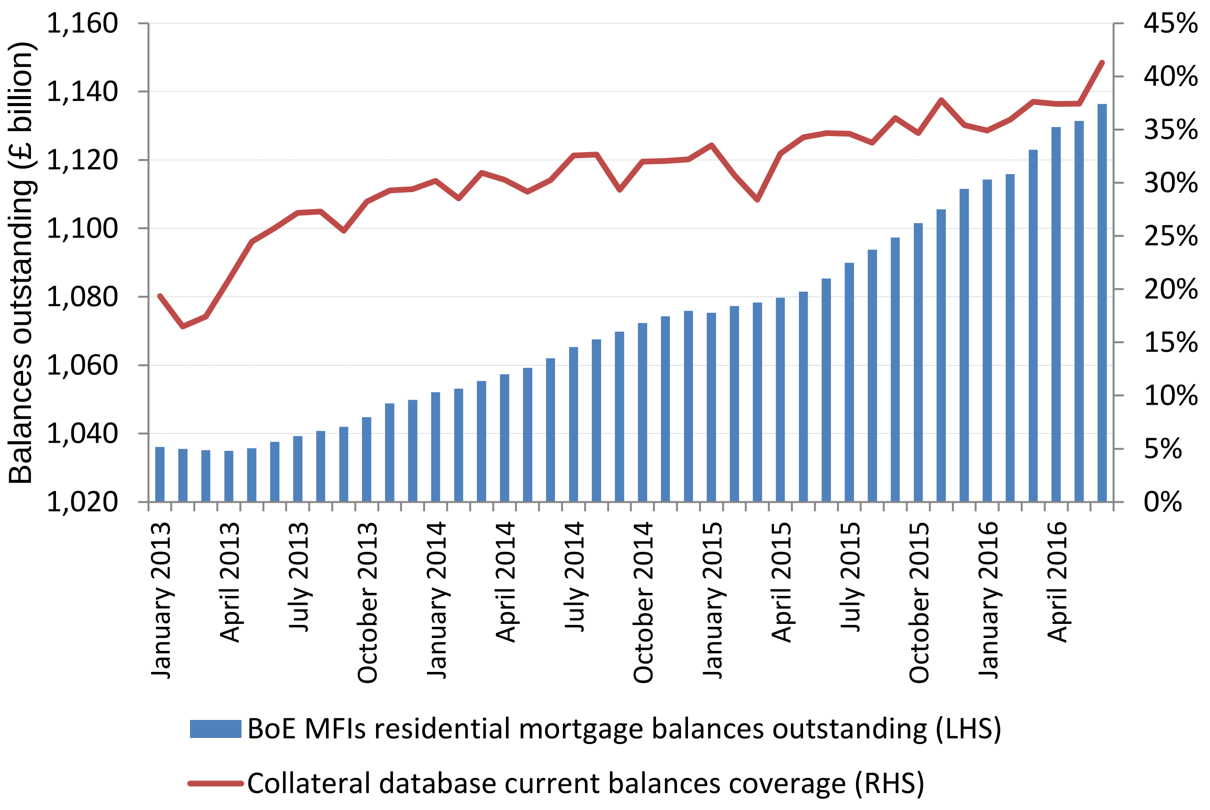

Chart 3. Current and original average LTV by origination year in the full database

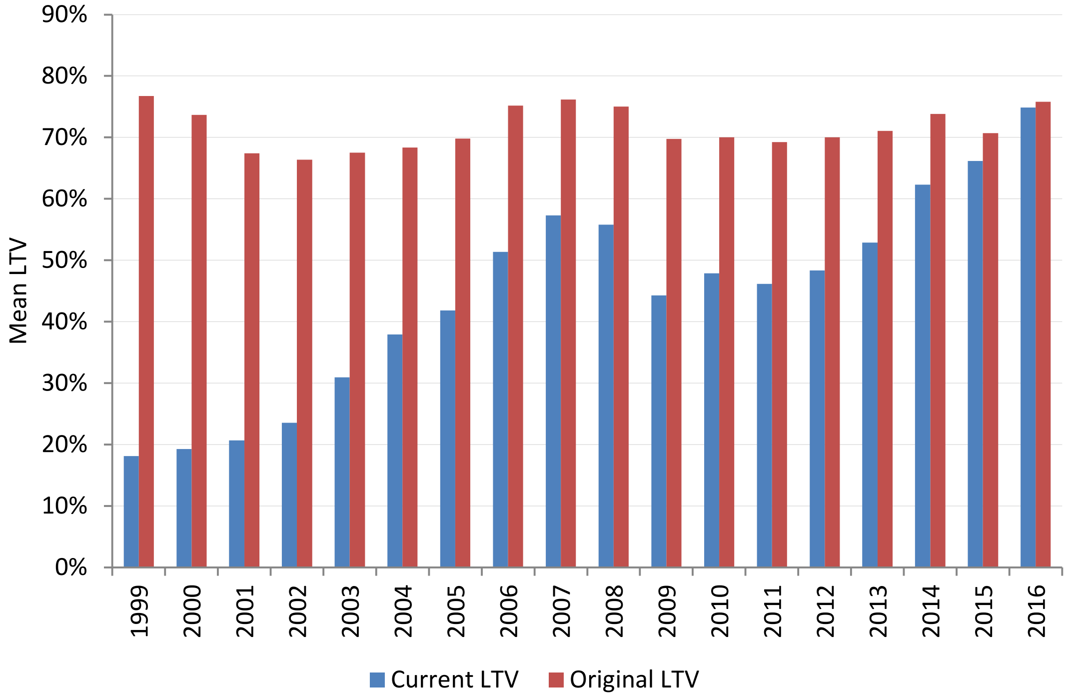

Chart 4. Median LTV over time for the OO (left) and BTL (right) markets in the full database

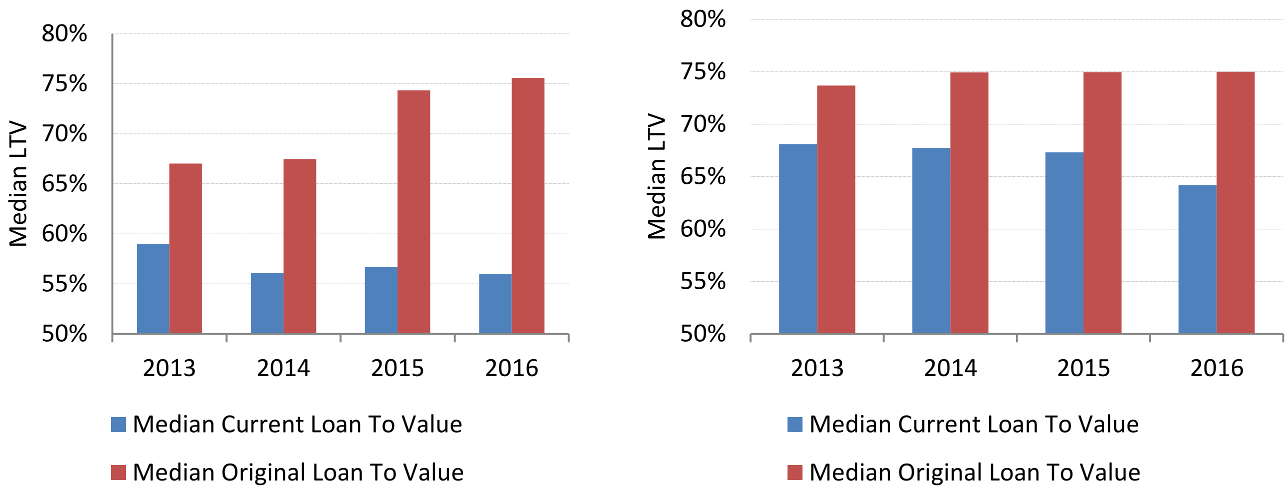

Chart 5. Loans by origination channel in the OO (left) and BTL (right) markets in the full database

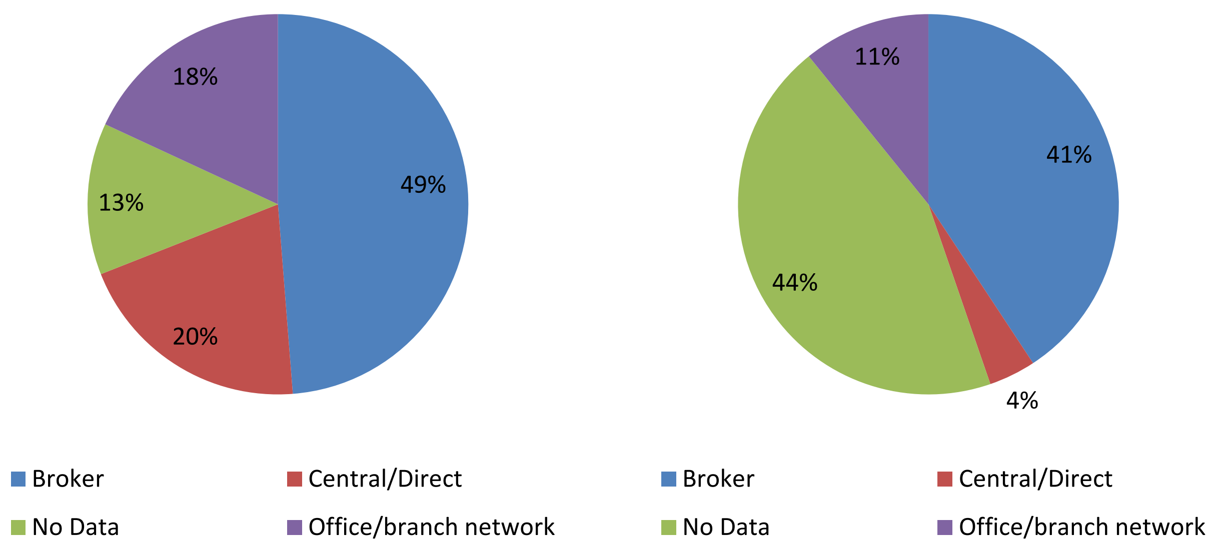

Chart 6. Number of loans by first year of inclusion in random sample

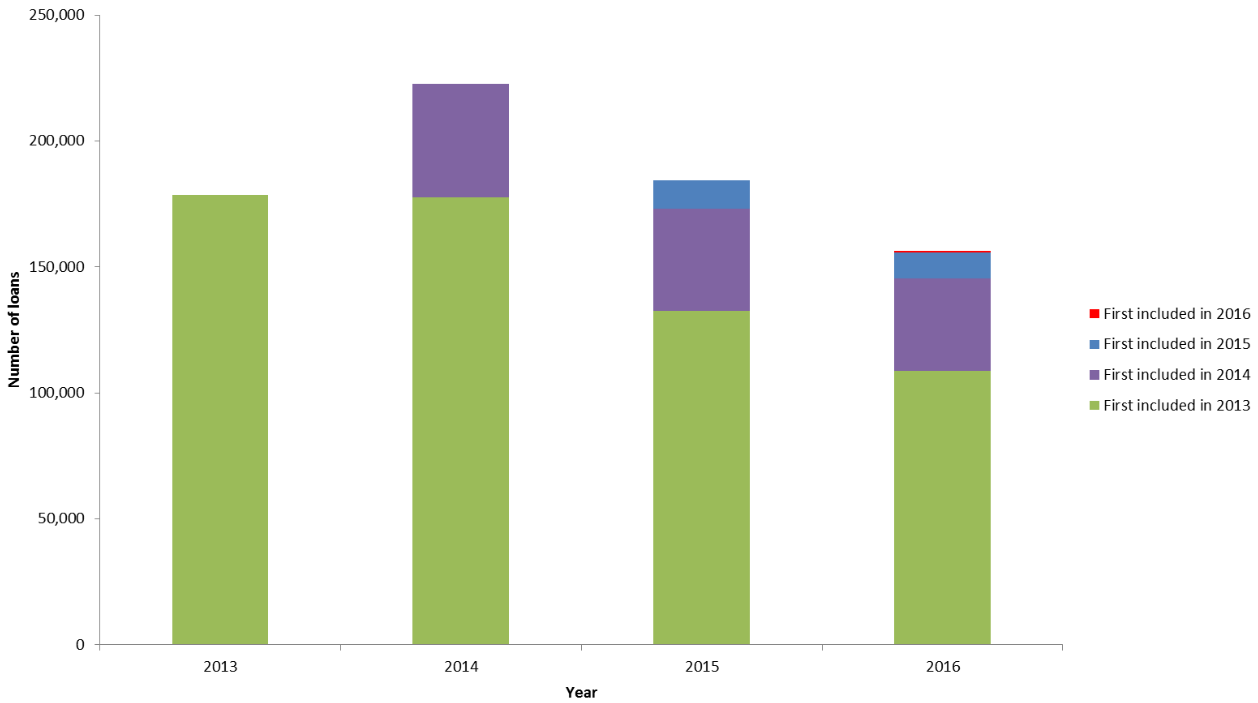

Chart 7. Loan-level current LTV ('CLTV') distribution for random sample of 150,000 borrowers

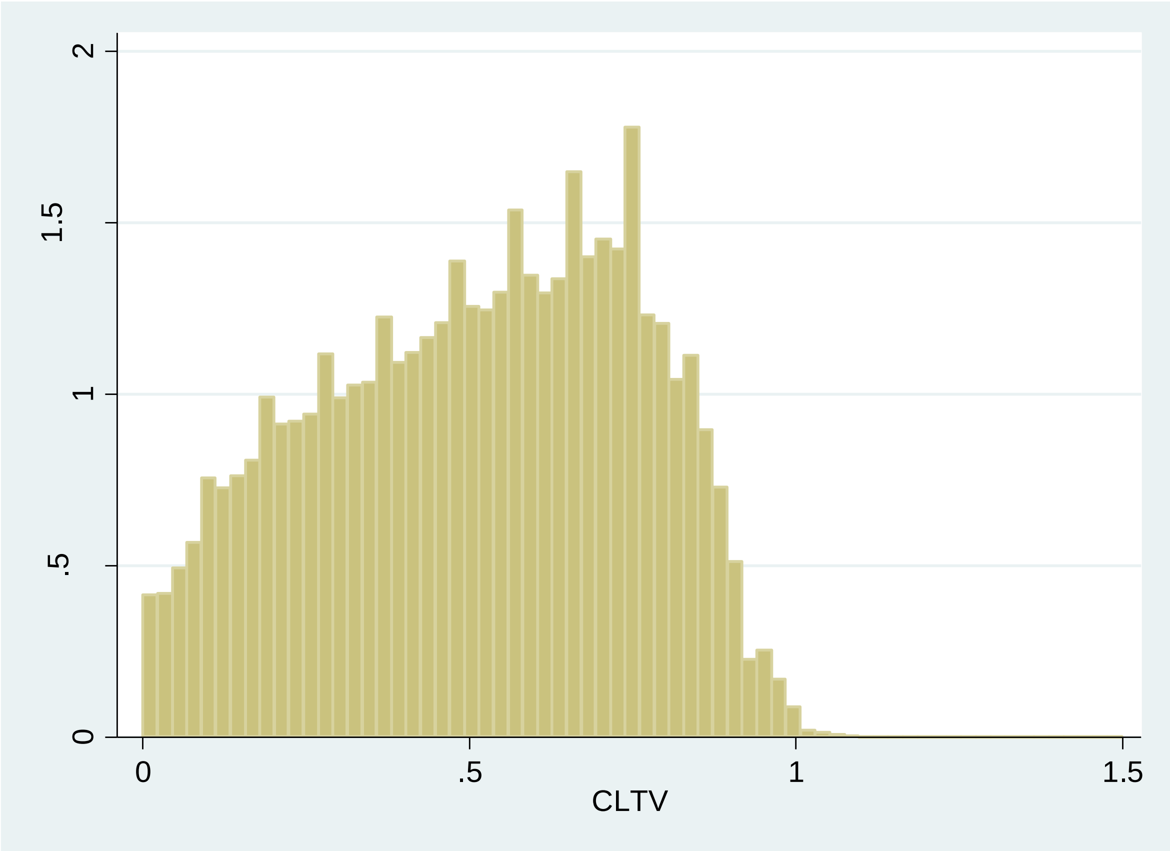

Chart 8. Borrower-level current LTV ('cltv') distribution for random sample of 150,000 borrowers

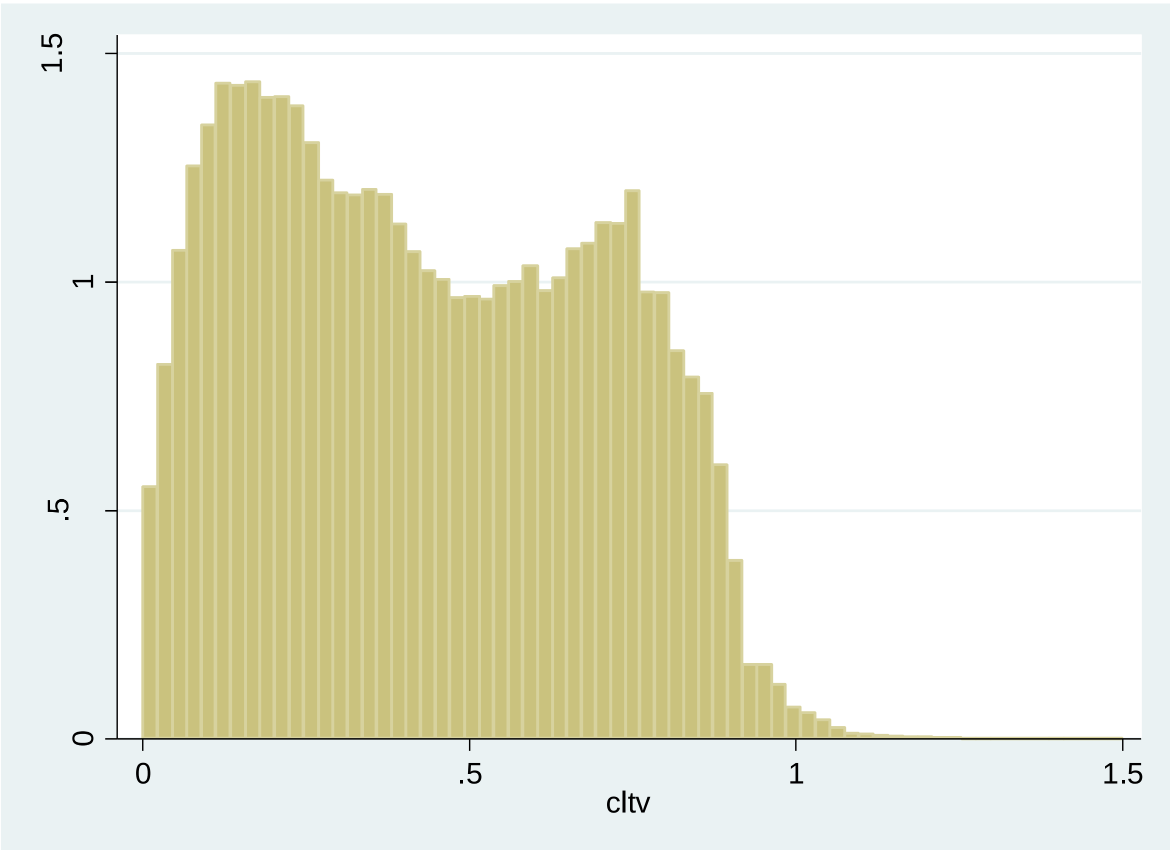

Chart 9. Distribution at borrower-level of Affordability (debt service coverage ratio, or indebtedness) constructed as "Monthly payment due / Monthly joint non-rental income" (designated 'dscr2inv') for the random sample (including both OO and BTL borrowers)

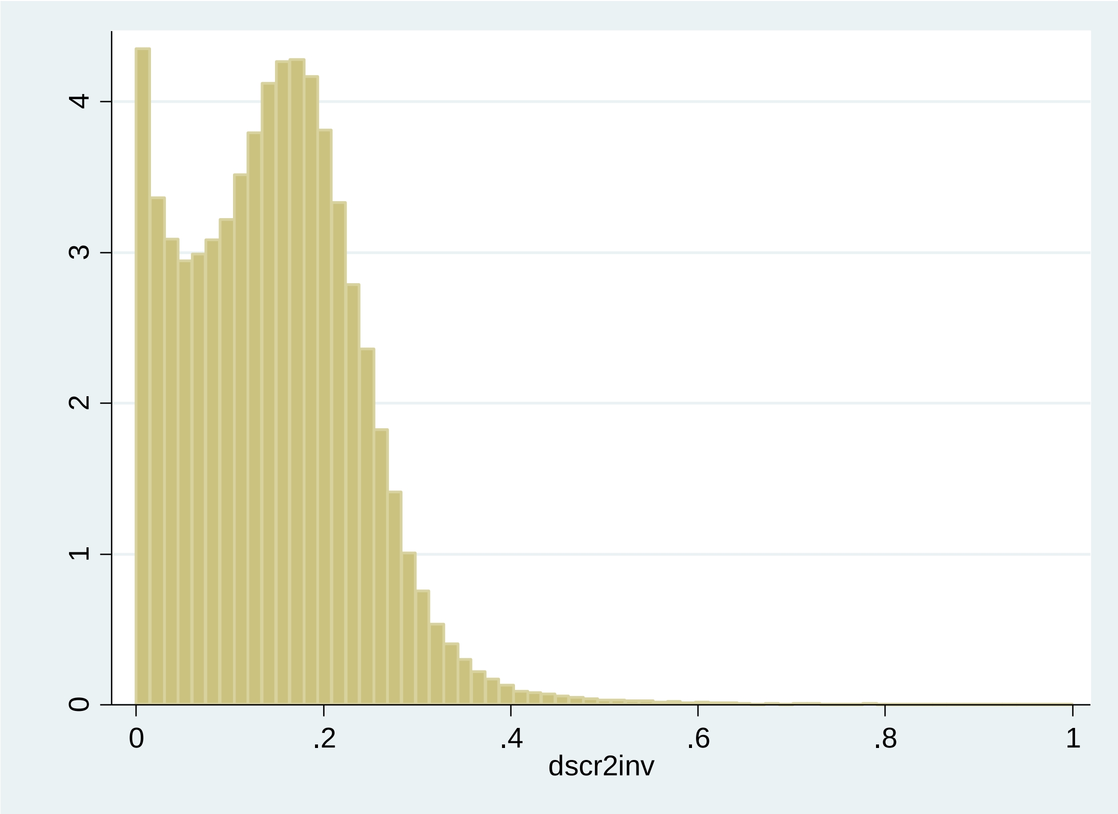

Chart 10. Loan distress rate in 2013-2016 by year of origination

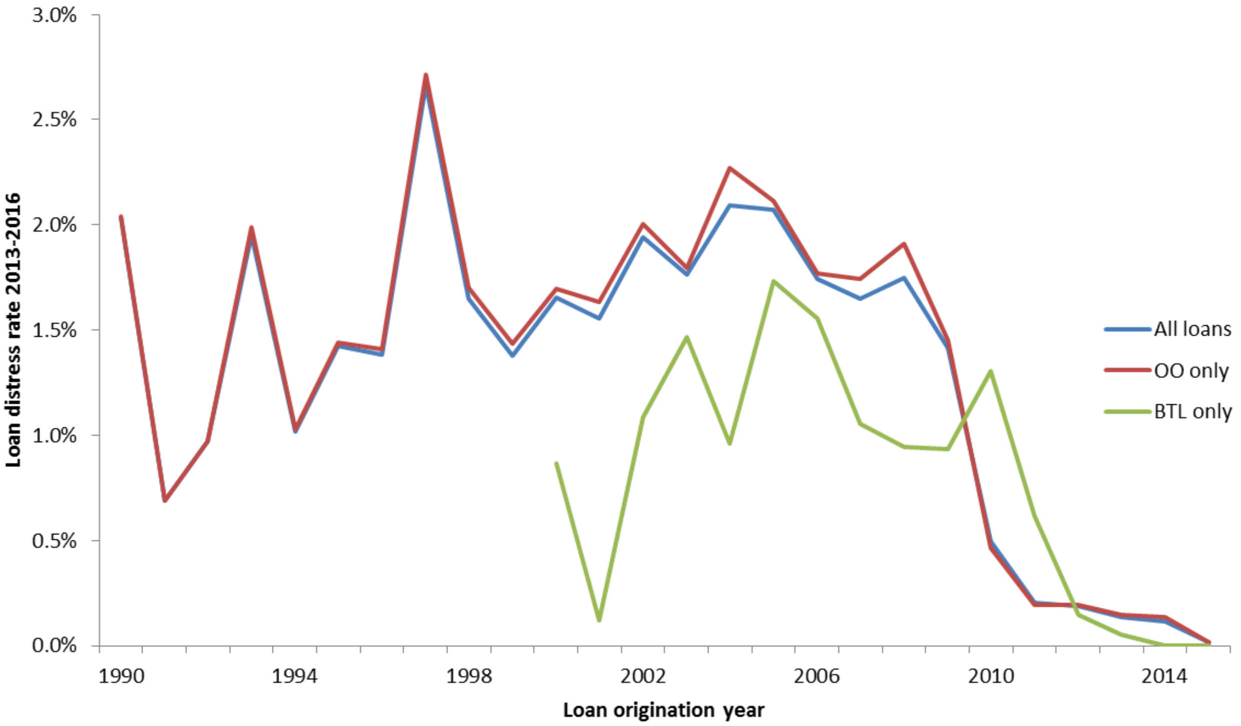

Table 1. Variables in the Bank of England's FRMD collateral database used in the construction of covariates

Variable Name	Variable Type	Description	Data Type	Dynamic	Mandatory
Pool Cut-off Date	Core	Pool or Portfolio cut-off date.	Date	Y	Y
Pool Identifier	Core	Pool or Portfolio identifier / name of transaction.	Text/Numeric	N	Y
Loan Identifier	Core	Unique identifier (ID) for each loan.	Text/Numeric	N	Y
Borrower Identifier	Core	Unique identifier (ID) per borrower to enable borrowers with multiple loans in the pool to be identified	Text/Numeric	N	Y
Property Identifier	Core	Unique identifier per property to enable properties with multiple loans in the pool to be identified	Text/Numeric	N	Y
Borrower Year of Birth	Borrower	Borrower year of birth. YYYY format.	Date	N	N
Number of Debtors	Borrower	Number of borrowers to the loan	Numeric	N	Y
Borrower's Employment Status	Borrower	Employment status of the primary applicant: Employed or full loan is guaranteed (1) Employed with partial support (company subsidy) (2) Protected life-time employment (Civil/government servant) (3) Unemployed (4) Self-employed (5) No employment, borrower is legal entity (6) Student (7) Pensioner (8) Other (9) No Data (ND)	List	N	Y
First-time Buyer	Borrower	First time buyer flag	Y / N / ND	N	N
Right to Buy	Borrower	Right to Buy (RTB) flag	Y / N / ND	N	N
Primary Income	Borrower	Primary borrower underwritten gross annual income (not rent)	Numeric	N	Y
Secondary Income	Borrower	Secondary borrower underwritten gross annual income	Numeric	N	Y
Origination Channel / Arranging Bank or Division	Loan	Origination channel, arranging bank or division for the loan: Office / branch network (1) Central / Direct (2) Broker (3) Internet (4) Packager (5) No Data (ND)	Text	N	Y
Purpose	Loan	Loan purpose. Permissible answers: Purchase (1) Re-mortgage (2) Renovation (3) Equity release (4) Construction (5) Debt consolidation (6) Other (7) Re-mortgage with Equity Release (8) Re-mortgage on Different Terms (9) Combination Mortgage (10) Investment Mortgage (11) Right to Buy (12) Government Sponsored Loan (13) SCPI (14) Besson (15) Perissol (16) DOM (Défiscalisation Métrople) (17) Other (18) No Data (ND)	List	N	Y

Original Balance	Loan	Original loan balance (inclusive of fees)	Numeric	N	Y
Current Balance	Loan	Amount of loan outstanding as of pool cut off date. This includes any amounts that are secured by the mortgage and is classed as principal in the transaction	Numeric	Y	Y
Repayment Method	Loan	Type of principal repayment: Interest Only (1) Repayment (2) Endowment (3) Pension (4) ISA/PEP (5) Index-Linked (6) Part & Part (7) Savings Mortgage (8) Other (9) No Data (ND)	List	N	Y
Payment Due	Loan	Periodic contractual payment due (the payment due if there are no other payment arrangements in force)	Numeric	Y	Y
Payment Type	Loan	Principal payment type: Annuity (1) Linear (2) Increasing instalments (3) Fixed instalments (changing maturity) with structural protection (4) Fixed instalments (changing maturity) without structural protection (5) Bullet (6) Bullet + Savings deposit (7) Bullet + Life insurance (8) Bullet + Investment portfolio (9) Bi-annual (10) Tri-annual (11) Offset mortgage (12) Other (13) No Data (ND)	List	N	Y
Further Loan Advance	Loan	Total value of further advances made on loan	Numeric	Y	Y
Interest Rate Type	Interest Rate	Interest rate type: Floating rate loan (for life) (1) Floating rate loan linked to Libor, Euribor, BoE reverting to the Bank's standard variable rate (SVR), ECB reverting to Bank's SVR (2) Fixed rate loan (for life) (3) Fixed with future periodic resets (4) Fixed rate loan with compulsory future switch to floating (5) Capped (6) Discount (7) Other (8) No Data (ND)	List	N	Y
Current Interest Rate	Interest Rate	Current interest rate (%).	Numeric	Y	Y
Property Postcode	Collateral	First 2 or 3 characters must be provided at a minimum. Do not supply the full postcode.	Text	N	N
Occupancy Type	Collateral	Type of property occupancy: Owner-occupied (1) Partially owner-occupied (A property which is partly rented) (2) Non-owner-occupied/buy-to-let (3) Holiday/second home (4) No Data (ND)	List	N	Y

Property Type	Collateral	Property type: Residential (House, detached or semi-detached) (1) Residential (Flat/Apartment) (2) Residential (Bungalow) (3) Residential (Terraced House) (4) Multifamily house (properties with more than four units securing one loan) with recourse to the borrower (5) Multifamily house without recourse to the borrower (6) Partially commercial use (property is used as a residence as well as for commercial use where less than 50% of its value derived from commercial use, e.g. doctor's surgery and house) (7) Commercial/business use with recourse to the borrower (8) Commercial/business use without recourse to the borrower (9) Land Only (10) Other (11) No Data (ND)	List	N	Y
Original Loan to Value	Collateral	Originator's original underwritten Loan To Value ratio (LTV). For 2nd lien loans this is the combined or total LTV.	Numeric	N	Y
Valuation Amount	Collateral	Property value as of date of latest loan advance prior to a securitisation.	Numeric	N	Y
Current Loan to Value	Collateral	Originator's current Loan to Value ratio (LTV). For 2nd lien loans this is the combined or total LTV	Numeric	Y	Y
Current Valuation Amount	Collateral	Most recent valuation amount (if e.g. at repossession there were multiple valuations, this should reflect the lowest).	Numeric	Y	Y
Gross Annual Rental Income	Collateral	Gross Annual Rental income for Buy To Let (BTL) properties	Numeric	N	Y
Number of Buy to Let Properties	Collateral	Total number of properties in portfolio, including those mortgaged with other lenders (BTL loans only)	Numeric	N	Y
Debt Service Coverage Ratio	Collateral	For Buy to Lets the Debt Service Coverage Ratio (DSCR) - Monthly Gross Rental Income divided by the Mortgage Payment For borrowers the DSCR is the Monthly Income divided by the Mortgage Payment.	Text/Numeric	N	Y
Account Status	Performance	Current status of account: Performing (1) Arrears (2) Default or Foreclosure (3) Redeemed (4) Repurchased by Seller (5) Other (6) No Data (ND)	List	Y	Y
Number Months in Arrears	Performance	Number of months this loan is in arrears (at pool cut off date) according to the definition of the issuer	Numeric	Y	Y

Table 2. Comparison of our data sample to PSD and UK Finance datasets at loan level

	Our data sample (2013-2016)[26]			PSD 007 (OO) (stock as of 2015 H1[27])	UK Finance (flow of new BTL lending from 14Q1 to 16Q4)
	All	OO	BTL		
Distress (payment shortfall)	n/a	n/a	n/a	4.1%[28]	n/a
Distress (in arrears)	2.20%	2.24%	1.78%	2.9%[29]	n/a
Distress (months in arrears ≥2)	1.14%	1.16%	0.99%	2.2%[30]	n/a
Repayment – capital & interest	73.83%	77.84%	34.55%	78.57%	25.36%
Repayment – interest only	19.79%	15.71%	59.75%	15.27%	73.04%
Interest rate type[31] – floating	58.43%	56.59%	76.45%		
Interest rate type – tracker	2.68%	2.36%	5.80%	18.74%	11.02%
Interest rate type – fixed	34.05%	35.90%	15.92%	46.82%	83.88%
Interest rate type – SVR				28.52%	
Number of lenders	41			155	17
Market size (volume)	132,356 (2015m6)	117,788 (2015m6)	14,467 (2015m6)	6,102,265	n/a
Market size (value)	£27.9b	£25.6b[32]	£2.27b	£725b	n/a
Market size (volume) - not cleaned	167,871 (2015m6)	152,426 (2015m6)	15,252 (2015m6)	n/a	n/a
Market size (value) - not cleaned	£37.8b	£35.4b	£2.36b	n/a	n/a
Current LTV (weighted)	47.95%	47.28%	57.70%	59.15%	n/a

Table 3. Proportions of performance categories across loans in database and random sample

Performance metric	Number of loan observations (full database)	Proportion in full database	Proportions in 10% random sample (before cleaning)
Current	132,263,388	95.95%	95.58%
In arrears	3,029,522	2.20%	2.20%
Default	64,954	0.05%	0.05%
Redeemed	365,987	0.27%	0.27%
Repurchased	46,929	0.03%	0.03%
Other	102	0.00%	0.00%
No Data	2,080,236	1.51%	1.87%
Total	137,851,118	100.00%	100.00%

[26] Data remaining after cleaning random data sample, unless stated otherwise. Lost loans due to cleaning amount to 28% of the total sample taken at random.
[27] Cleaned data only available for this point in time at the time of writing
[28] Deduced from payment shortfall: months in arrears = payment shortfall / monthly payment
[29] Based on arrears date
[30] Deduced from payment shortfall: months in arrears = payment shortfall / monthly payment
[31] Different definitions are used between different datasets (e.g. floating – SVR)
[32] In order to make this number comparable with the one of PSD 007, we need to make several adjustments. We need to account for the effects of a) cleaning the data which removes 5% of our sample for June 2015, b) taking a random sample of 10%, and c) the whole FRMD database only comprising 41% of the overall OO market. Adjusting for these factors, the total market size would be about £657b (£25.6b * 100 / 95 * 100 / 10 * 100 / 41).

Table 4. Descriptive statistics of random data sample of 150,000 borrowers' loans

Variable	Observations	Mean	Std. Dev.	Min	Max
Distress (dep. variable)	4,658,201	0.011	0.104	0	1
LTV* (ratio)	4,658,201	0.480	0.256	0	3.00
Affordability* (ratio)	4,658,201	0.168	0.158	0	3.00
Buy-to-Let (dummy)	4,658,201	0.090	0.286	0	1
InterestRate (%)	4,653,025	0.034	0.012	0	0.10
Self_Employed (dummy)	4,658,201	0.125	0.331	0	1
Remortgages (number)	4,658,201	0.456	0.649	0	20
Further_Advances (number)	4,658,201	0.184	0.515	0	20
Negative_Equity (dummy)	4,658,201	0.006	0.076	0	1
Increasing_Instalments (number)	4,658,201	0.210	0.667	0	20
Bullet_Payment (number)	4,658,201	0.228	0.577	0	28
Broker_Intermediated (number)	4,658,201	0.450	0.598	0	14
Age (years)	4,657,017	48.72	10.24	18	100
Regional_Unemployment (%)	4,497,098	6.108	1.492	3.2	10.2
Regional_HouseIndex (index)	4,497,098	99.17	7.11	77	119
Regional_SalesVolume (number)	4,486,050	8765	3411	1622	20152
Interaction variables					
BTLxLTV	4,658,201	0.051	0.175	0	2.81
BTLxAffordability	4,658,201	0.010	0.051	0	2.94
BTLxInterestRate	4,653,025	0.003	0.010	0	0.10
BTLxSelf_Employed	4,658,201	0.024	0.154	0	1
BTLxRemortgages	4,658,201	0.038	0.233	0	10
BTLxFurther_Advances	4,658,201	0.010	0.122	0	15
BTLxNegative_Equity	4,658,201	0.001	0.026	0	1
BTLxIncreasing_Instalments	4,658,201	0.001	0.048	0	11
BTLxBullet_Payment	4,658,201	0.069	0.339	0	15
BTLxBroker_Intermediated	4,658,201	0.062	0.301	0	10
BTLxAge	4,657,017	50.72	10.21	23	93

***** To reduce the impact of outliers on our results, we have restricted LTV and Affordability to less than 3.

Table 5. Distress rates in baseline sample

	Loan level			Borrower level		
	All	OO	BTL	All	OO	BTL
2013	1.23%	1.23%	1.20%	1.40%	1.38%	1.52%
2014	1.15%	1.17%	1.00%	1.29%	1.29%	1.28%
2015	1.09%	1.11%	0.93%	1.19%	1.20%	1.07%
2016	1.04%	1.06%	0.84%	1.10%	1.11%	0.96%
2013-2016	1.14%	1.15%	1.00%	1.26%	1.26%	1.23%

Table 6. Transition probabilities of being in and getting out of distress in baseline sample

	Loan level			Borrower level		
in %	not in distress at t+1	in distress at t+1	Total	not in distress at t+1	in distress at t+1	Total
not in distress at t	99.88	0.12	100	99.89	0.11	100
in distress at t	10.16	89.84	100	10.29	89.71	100
Total	98.84	1.16	100	98.91	1.09	100

Table 7. Sample model specification tests – FE logistic regression versus OLS and RE estimates

Test 1	(β)	(B)	(β-B)	sqrt(diag(V_β-V_B))
FE vs OLS	FE	OLS	Difference	S.E.
LTV	4.321	1.378	2.943	0.165
Affordability	0.307	0.202	0.105	0.131
…				

Ho: difference in coefficients not systematic

$\chi^2(23)$ = (β-B)'[(V_β-V_B)^(-1)](β-B) = 1089.20 Prob > χ^2 = 0.0000

Test 2	(β)	(B)	(β-B)	sqrt(diag(V_β-V_B))
FE vs RE	FE	RE	Difference	S.E.
LTV	4.321	3.148	1.173	0.146
Affordability	0.307	0.342	-0.035	0.080
…				

Ho: difference in coefficients not systematic

$\chi^2(22)$ = (β-B)'[(V_β-V_B)^(-1)](β-B) = 220.80 Prob > χ^2 = 0.0000

Table 8. Results from Equation (1) on random sample of borrowers' loans

Variable	Coefficient	Marginal effect	BTL-sector interactions	Coefficient
LTV	4.321***	0.332***	**btl*LTV**	-2.918***
	(0.000)	(0.000)		(0.000)
Affordability	0.307**	0.0268**	**btl*Affordability**	0.715
	(0.022)	(0.016)		(0.241)
Buy-to-Let	1.361***	0.067***		
	(0.000)	(0.000)		
InterestRate	13.458***	1.068***	btl*InterestRate	0.475
	(0.000)	(0.000)		(0.922)
Self_Employed	-0.453***	-0.034***	btl*Self_Employed	0.530*
	(0.000)	(0.000)		(0.055)
Remortgages	0.169***	0.015***	btl*Remortgages	0.596***
	(0.000)	(0.000)		(0.000)
Further_Advances	0.0665**	0.007***	btl*Further_Advances	0.437***
	(0.027)	(0.009)		(0.016)
Negative_Equity	0.872***	0.070***	btl*Negative_Equity	0.390
	(0.000)	(0.000)		(0.283)
Increasing_Instalments	-0.249***	-0.019***	btl*Increasing_Instalments	0.265
	(0.000)	(0.000)		(0.479)
Bullet_Payment	-0.165***	-0.012***	btl*Bullet_Payment	0.327***
	(0.002)	(0.006)		(0.002)
Broker_Intermediated	0.923***	0.074***	btl*Broker_Intermediated	0.276
	(0.000)	(0.000)		(0.123)
			Pseudo R-squared	0.020
Regional_Unemployment	0.014	0.001	$\chi 2(23)$	1902***
	(0.221)	(0.166)	Prob > χ^2	(0.000)
Regional_HouseIndex	-0.007***	-0.001**	Number of observations	127,330
	(0.000)	(0.016)	Number of groups	3617
Regional_SalesVolume	0.000004	0.000	Average observations per group	35.2
	(0.332)	(0.321)		

*Note: Dependent variable is probability of borrower distress (months in arrears ≥ 2). Main coefficients of interest are bolded. For definitions of variables, see Table 1. P-values are in parentheses with *** p<0.01, **p<0.05, *p<0.1.

Table 9. Results from Equation (1) on all borrowers with impaired loans in the database

Variable	Coefficient	Marginal effect	BTL-sector interactions	Coefficient
LTV	2.698***	0.263***	**btl*LTV**	-0.948***
	(0.000)	(0.000)		(0.000)
Affordability	0.333***	0.033***	**btl*Affordability**	0.021
	(0.000)	(0.000)		(0.752)
Buy-to-Let	1.309***	0.026***		
	(0.000)	(0.000)		
InterestRate	17.067***	1.630***	btl*InterestRate	-11.128***
	(0.000)	(0.000)		(0.000)
Self_Employed	-0.26***	-0.026***	btl*Self_Employed	0.074
	(0.000)	(0.000)		(0.317)
Remortgages	0.222***	0.028***	btl*Remortgages	-0.067*
	(0.000)	(0.000)		(0.064)
Further_Advances	0.047***	0.005***	btl*Further_Advances	0.032*
	(0.000)	(0.000)		(0.061)
Negative_Equity	0.063**	0.070**	btl*Negative_Equity	-0.258***
	(0.062)	(0.175)		(0.009)
Increasing_Instalments	-0.162***	-0.016***	btl*Increasing_Instalments	0.056
	(0.000)	(0.000)		(0.349)
Bullet_Payment	-0.094***	-0.009***	btl*Bullet_Payment	0.026
	(0.000)	(0.000)		(0.444)
Broker_Intermediated	0.923***	0.045***	btl*Broker_Intermediated	-0.102**
	(0.000)	(0.000)		(0.019)
Age	0.0039***	0.00037***	btl*Age	-0.003**
	(0.000)	(0.000)		(0.017)
			Pseudo R-squared	0.019
Regional_Unemployment	0.060***	0.006***	$\chi2(26)$	15,188***
	(0.000)	(0.000)	Prob > χ^2	(0.000)
Regional_HouseIndex	-0.010***	-0.001***	Number of observations	1,029,729
	(0.000)	(0.000)	Number of groups	33,528
Regional_SalesVolume	-0.000008***	-0.0000008***	Average observations per group	35.2
	(0.000)	(0.000)		

*Note: Dependent variable is probability of borrower distress (months in arrears ≥ 2). Main coefficients of interest are bolded. For definitions of variables, see Table 1. P-values are in parentheses with *** p<0.01, **p<0.05, *p<0.1.

Table 10. Results from Equation (1) on random sample of loans at loan level

Variable	Coefficient	Marginal effect	BTL-sector interactions	Coefficient
LTV	2.962***	0.279***	btl*LTV	1.621***
	(0.000)	(0.000)		(0.006)
Affordability	0.840***	0.072***	btl*Affordability	-0.481
	(0.000)	(0.002)		(0.730)
Buy-to-Let	-4.917***	[not estimable][33]		
	(0.000)	[not estimable]		
FTB	-1.828***	-0.163**		
	(0.009)	(0.032)		
RTB	2.469*	0.222*		
	(0.056)	(0.065)		
InterestRate	17.807***	1.810***	btl*InterestRate	26.825***
	(0.000)	(0.000)		(0.001)
Self_Employed	0.106	0.029***	btl*Self_Employed	2.410***
	(0.322)	(0.004)		(0.000)
Remortgages	-0.746***	-0.058***	btl*Remortgages	1.194**
	(0.000)	(0.002)		(0.033)
Further_Advances	0.254***	0.026***	btl*Further_Advances	0.421
	(0.000)	(0.000)		(0.263)
Negative_Equity	-1.555***	-0.308	btl*Negative_Equity	-20.919
	(0.000)	(0.943)		(0.969)
Increasing_Instalments	17.278	1.555	btl*Increasing_Instalments	[omitted]
	(0.983)	(0.983)		
Bullet_Payment	-0.476**	-0.169	btl*Bullet_Payment	-15.727
	(0.031)	(0.974)		(0.980)
Broker_Intermediated	2.171***	0.219***	btl*Broker_Intermediated	2.972***
	(0.000)	(0.000)		(0.000)
			Pseudo R-squared	0.016
Regional_Unemployment	0.079***	0.007***	$\chi 2(24)$	1880***
	(0.000)	(0.000)	Prob > χ^2	(0.000)
Regional_HouseIndex	-0.005***	-0.00045**	Number of observations	143,171
	(0.000)	(0.047)	Number of groups	4991
			Average observations per group	29

*Note: Dependent variable is probability of loan distress (months in arrears ≥ 2). Coefficients of interest are bolded. For definitions of variables, see Table 1. P-values are in parentheses with *** p<0.01, **p<0.05, *p<0.1.

[33] This BTL marginal effect was not estimable with our preferred delta method.

Table 11. Error bootstrapping results with 300 repetitions of Equation (1) on 10% random sample

Variable	Coefficient	BTL-sector interactions	Coefficient
LTV	4.321***	**btl*LTV**	-2.918**
	(0.000)		(0.039)
Affordability	0.307	**btl*Affordability**	0.715
	(0.327)		(0.621)
Buy-to-Let	1.361		
	(0.253)		
InterestRate	13.458***	btl*InterestRate	0.475
	(0.001)		(0.970)
Self_Employed	-0.453*	btl*Self_Employed	0.530
	(0.084)		(0.710)
Remortgages	0.169	btl*Remortgages	0.596
	(0.134)		(0.233)
Further_Advances	0.067	btl*Further_Advances	0.437
	(0.359)		(0.463)
Negative_Equity	0.872**	btl*Negative_Equity	0.390
	(0.038)		(0.722)
Increasing_Instalments	-0.249*	btl*Increasing_Instalments	0.265
	(0.065)		(0.930)
Bullet_Payment	-0.165	btl*Bullet_Payment	0.327
	(0.281)		(0.291)
Broker_Intermediated	0.923***	btl*Broker_Intermediated	0.276
	(0.000)		(0.629)
		Pseudo R-squared	0.020
Regional_Unemployment	0.014	$\chi2(23)$	143.7***
	(0.654)	Prob > χ^2	(0.000)
Regional_HouseIndex	-0.007	Number of observations	127,330
	(0.230)	Number of groups	3617
Regional_SalesVolume	0.000004	Average observations per group	35.2
	(0.564)		

*Note: Dependent variable is probability borrower of distress (months in arrears≥ 2). Main coefficients of interest are bolded. For definitions of variables, see Table 1. In parentheses are bootstrapping-implied p-values with *** p<0.01, **p<0.05, *p<0.1.

Table 12. Error bootstrapping results with 500 repetitions of Equation (1) on all borrowers with impaired loans in the database

Variable	Coefficient	BTL-sector interactions	Coefficient
LTV	2.698***	**btl*LTV**	-0.948***
	(0.000)		(0.004)
Affordability	0.333***	**btl*Affordability**	0.021
	(0.000)		(0.854)
Buy-to-Let	1.309***		
	(0.000)		
InterestRate	17.067***	btl*InterestRate	-11.128***
	(0.000)		(0.000)
Self_Employed	-0.26***	btl*Self_Employed	0.074
	(0.000)		(0.670)
Remortgages	0.222***	btl*Remortgages	-0.067
	(0.000)		(0.496)
Further_Advances	0.047**	btl*Further_Advances	0.032
	(0.046)		(0.335)
Negative_Equity	0.063	btl*Negative_Equity	-0.258
	(0.627)		(0.360)
Increasing_Instalments	-0.162***	btl*Increasing_Instalments	0.056
	(0.001)		(0.645)
Bullet_Payment	-0.094**	btl*Bullet_Payment	0.026
	(0.000)		(0.758)
Broker_Intermediated	0.462***	btl*Broker_Intermediated	-0.102
	(0.000)		(0.314)
Age	0.0039***	btl*Age	-0.003
	(0.000)		(0.149)
		Pseudo R-squared	0.019
Regional_Unemployment	0.060***	$\chi 2(26)$	15,188***
	(0.000)	Prob > χ^2	(0.000)
Regional_HouseIndex	-0.010***	Number of observations	1,029,729
	(0.000)	Number of groups	33,528
Regional_SalesVolume	-0.000008***	Average observations per group	35.2
	(0.000)		

*Note: Dependent variable is probability borrower of distress (months in arrears ≥ 2). Main coefficients of interest are bolded. For definitions of variables, see Table 1. In parentheses are bootstrapping-implied p-values with *** p<0.01, **p<0.05, *p<0.1.