Financial Conduct Authority

Research Note

13/02/25

An empirical analysis of pricing differences by demographic characteristics in the UK mortgage market

Daniel Bogiatzis-Gibbons, Lawrence Charles, Luke Lawlor, Bjorn Limani, Eoghan O'Brien, Ned Staniland

FCA research notes in financial regulation

The FCA research notes

The FCA is committed to encouraging debate on all aspects of financial regulation and to creating rigorous evidence to support its decision-making. To facilitate this, we publish a series of Research Notes, extending across economics and other disciplines.

The main factor in accepting papers is that they should make substantial contributions to knowledge and understanding of financial regulation. If you want to contribute to this series or comment on these papers, please contact Kieran Keohane (Kieran.Keohane@fca.org.uk).

Disclaimer

Research notes contribute to the work of the FCA by providing rigorous research results and stimulating debate. While they may not necessarily represent the position of the FCA, they are one source of evidence that the FCA may use while discharging its functions and to inform its views. The FCA endeavours to ensure that research outputs are correct, through checks including independent referee reports, but the nature of such research and choice of research methods is a matter for the authors using their expert judgement. To the extent that research notes contain any errors or omissions, they should be attributed to the individual authors, rather than to the FCA.

Authors

Daniel Bogiatzis-Gibbons, Lawrence Charles, Luke Lawlor, Bjorn Limani, Eoghan O'Brien and Ned Staniland were all staff of the FCA at the time of writing. Daniel is also a PhD student in Politics at Birkbeck College, University of London.

All our publications are available to download from www.fca.org.uk. If you would like to receive this paper in an alternative format, please call 020 7066 9644 or email publications_graphics@fca.org.uk or write to Editorial and Digital Department, Financial Conduct Authority, 12 Endeavour Square, London E20 1JN.

Contents

1 Overview

Purpose

By matching two datasets regularly collected by the Financial Conduct Authority (FCA), this note sets out the results of a piece of research investigating differential pricing outcomes by demographic characteristics in the UK mortgage market.

Mortgage features, such as interest rates, loan-to-value (LTV) ratio and lender fees, are compared across the following demographic characteristics:

- age
- sex
- sexual orientation
- ethnic group, and
- health condition.

After analysing the differences across these groups, we quantify the influence demographic characteristics have on mortgage pricing. We built a machine learning model that predicts the interest rate of a mortgage given a range of mortgage properties, the Bank of England base rate and borrower features, including their demographic characteristics. This is one of the first analyses of some of these characteristics on a product level in a UK context and provides a novel approach for modelling interest rates.

Our focus is on gaining insights into the pricing differences observed across demographic characteristics in the mortgage market. We do not provide any opinion on what constitutes unfair treatment or discrimination from a legal standpoint.

Key findings

The results of this research are split into two components.

Firstly, statistical analysis of the matched dataset showed that there are differences in the types of mortgage products taken out by different groups, which may affect overall price paid.

Those with a health condition appear to have mortgages with higher initial gross rates of interest on average, but lower upfront lender fees, lower property values and lower household incomes. This could suggest they are more likely to take out products where payments are spread over time resulting in slightly higher overall prices paid. It was unclear if this difference was driven by consumer choice or due to the types of mortgages these consumers were able to access (though we note that, for regulated mortgage products, customers should always have the choice of paying lender fees upfront or including them in the loan amount, under the FCA's rules).

In addition, females appear to have mortgages with similar interest rates and LTV ratios as males, but lower household income and property values. These are likely indicators of wider structural inequalities, such as the gender pay gap.

People who identify as either lesbian, gay, bisexual, or asexual (LGBA) compared to people who identify as non-LGBA appear to have mortgages with higher interest rates, higher LTV ratios, lower lender fees and lower property values.

People from minority ethnic groups appear to have mortgages with marginally lower interest rates, higher loan amounts, higher household incomes and higher lender fees. They also appear to be slightly more likely to take out mortgages with higher loan-to-income (LTI) ratios. However, we were unable to explore differences across minority ethnic groups due to small sample sizes in the dataset.

Secondly, we built a machine learning model to quantify how influential demographic characteristics were in determining mortgage price. The model predicts the initial gross rate of interest of a mortgage based on a range of product features, borrower features and macro-economic variables. We included the demographic characteristics as features to quantify their influence on the predicted interest rate. The model performed well on unseen test data. We find that demographic characteristics had little to no effect on the predicted interest rates.

In conclusion, we did not identify any evidence of differences in mortgage pricing across different demographic groups from this research. Instead, we find that groups appear to have different types of mortgage products. However, we cannot conclude that there are no issues with the availability of products to different groups.

2 Research Design

Context

The potential for algorithms to give different predictions on the basis of demographic characteristics such as ethnicity and sex, is known as *algorithmic bias*. We summarise these conceptual questions in a separate Research Note.

This note will explore pricing differences by demographic characteristics in the UK mortgage market and what might explain those differences. *Mortgage pricing* here refers to the interest rate of the mortgage whilst considering the other mortgage properties that usually explain the level of interest rate applied to a mortgage, such as

- Loan-to-value (LTV) ratio
- Product fees
- Initial incentivised term – e.g., 2 and 5 years
- Type of interest rate – e.g., fixed or tracker
- Product type or special features – e.g., bridging, second charge and lifetime mortgages
- Lender

Mortgage pricing in the UK is usually determined through the range of mortgage products a firm offers. Anyone who qualifies for an offer should be getting the same price. When controlling for product type, which can be defined as some combination of the mortgage properties above that affect interest rates, any differences that remain for groups with particular demographic characteristics could be indicative evidence of bias or other unjustified pricing differences. There are multiple ways of defining fairness constraints for setting prices. The above example relates to *pricing fairness* which ensure that the prices offered to two groups, split by demographic characteristics, are nearly equal (Cohen et al. 2022). However, there could be less direct factors that affect fairness, e.g., differences in product availability or choices differing between groups. This relates to *demand fairness* which ensures access to products is as close as possible across groups (Cohen et al. 2022). This is more difficult to identify with the data utilised here as they only include prices (observed interest rates and mortgage properties) rather than offers, or more generally what products were available to that person at the time (which are unobserved).

Customers wanting a mortgage either seek out a product directly with a lender or by speaking with an advisor who will then recommend and apply on their behalf. Overall, our research will not allow us to conclude whether there is algorithmic bias in mortgage pricing as we do not observe firms' algorithms. However, it does provide indicative evidence on whether there are differential outcomes in the mortgage market which can't be explained by other factors. Where there is evidence of predictive algorithms giving different predictions or performance to different groups, this could be indicative of *group* *bias* (Verma and Rubin, 2018). Price discrimination, whereby different prices are offered to different groups, can come in many forms, such as offering different prices on the basis of a person's willingness to pay, on their age, on a person's vulnerability, or if a person is uninformed. It is considered by regulators since it is evident across different markets but can be difficult to determine whether the outcomes are fair, particularly in financial services markets (Starks et al. 2018).

Due to the lack of data on demographic characteristics for mortgages collected in the United Kingdom there are few studies of mortgage pricing differences across demographic groups. Previous work has focused on differences in the market by age because these data are already regularly collected by the FCA. For instance, young people are more likely to make mistakes getting a mortgage by picking more expensive offers (Coen, J. et. al. 2021). There is more research on mortgage outcomes by different demographic characteristics from the United States given data availability and the historical context. Whilst the findings from this research is not analogous to the UK given that difference in context, that research has helped improve understanding of the market. For example, Nami, S. et al. 2022 found that people of certain demographics have been denied mortgages in particular areas.

Differential outcomes for groups of individuals with demographic characteristics could be observed where a group may be less well served by the market compared to others. For instance, lending to older households had previously been declining and now primarily focuses on equity release products, such as lifetime mortgages that carry high interest rates (UK Finance, 2019). This may reflect changes in consumer choice, lack of competition, lack of scale, broader economic trends, or commercially driven decisions.

The question we seek to answer here is whether there are any observed price differences between groups and what might explain them.

Data

Pricing differences across demographic groups in the UK housing market are usually difficult to examine primarily due to a lack of available individual level data. The research underpinning this note utilises multiple datasets collected by the FCA to bring together loan level mortgage data and individual data on demographic characteristics. This provides a unique opportunity to analyse how mortgages differ across a range of demographic groups.

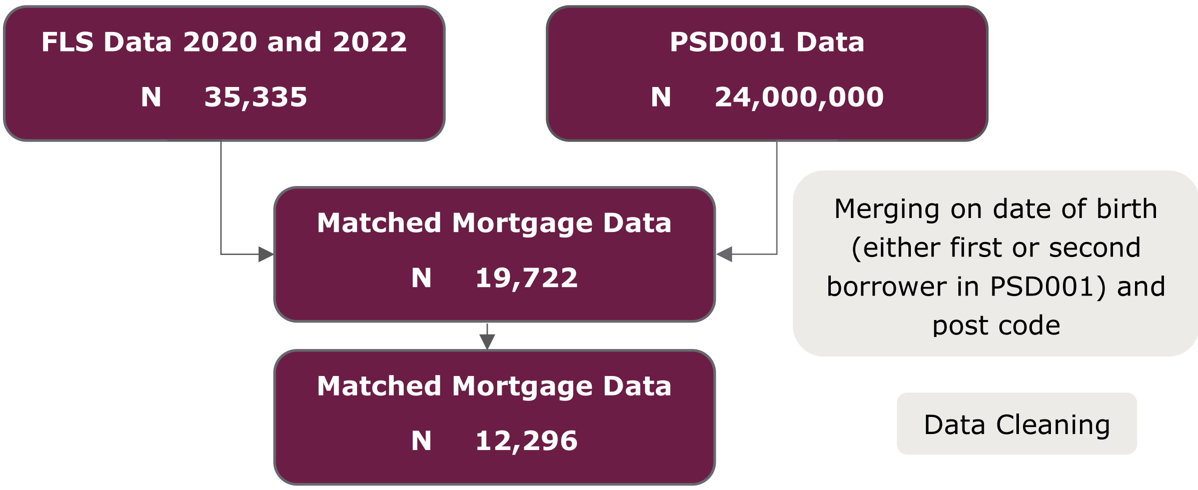
For joining the two datasets, data matching safeguards were implemented, and participants data were pseudonymised to ensure that personal identifiable information (PII) and data on demographic characteristics were not accessible together.

The Product Sales Data (PSD) comprises data collected by the FCA from firms it regulates on what products they are selling. PSD001 collects all new mortgages and their main characteristics, including mortgage and borrower characteristics. Certain variables, such as fees and income, have only been reported since the introduction of both affordability rules in 2014 and changes to the PSD reporting requirements from 2015 onwards. To account for these, we restrict our dataset to mortgages created since April 2015 where all mortgages are subject to the same affordability rules. Further work could compare these results to before the affordability checks were introduced to better understand their impact across different groups.

The Financial Lives survey (FLS) is the FCA's flagship, nationally representative survey of UK consumers. It provides information about consumers' attitudes towards managing their money, the financial products they have and their experiences of engaging with financial services firms. As a tracking survey, it provides evidence of how things are changing from the point of view of the consumer. The FLS has now run three main waves – in 2017, 2020, 2022, and two smaller recontact surveys in Jan 2023 and Jan 2024 using the respondents to the 2022 survey. For this analysis, we include the 2020 and 2022 survey respondents. We did not include the respondents to the Financial Lives 2017 survey as date of birth data were not collected.

A new dataset is created by joining PSD001 and the FLS data using date of birth and postcode as the joining key. Of the 35,335 date of birth and postcode combinations in the combined Financial Lives 2020 and 2022 dataset, 10,422 were matched in PSD001. Following data cleaning and restricting the mortgage account open date to be after April 2015, we are left with 12,296 mortgage products, including 2593 internal remortgages. This aligns with observations that approximately 30% of UK properties are owned with a mortgage (ONS 2021). We find that 2,283 mortgages holders in the FLS dataset were unmatched with records in PSD001, highlighting a potential data join issue. This could be for several reasons, such as the participant's mortgage account open date being before PSD records began, or the participant being the third or fourth name on the mortgage (who aren't required to report date of birth in PSD001).

Diagram 1: Flow diagram of dataset merging



Caveats and Limitations

The data we have used on demographic characteristics data are considered in Table 1. Here we set out some clarifications and limitations relating to the FLS data we have used, as well as how these link to demographic characteristics. Further information about the FLS variables and definition can be found in the Financial Lives 2022 publication (FLS 2023).

Table 2: Data on demographic characteristics

Characteristic	Values	Limitations
Sex	Male, female and prefer not to say.	We have taken the FLS data on sex. Gender identity is a separate question in the survey.
Age	Categorised into 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+ age groups.	Note this indicates age when taking out mortgage.
Ethnic group	Available at different levels of granularity and will be specified, e.g., minority ethnic and non-minority ethnic, or white, mixed race, Asian, black and black British, and other, etc.	
Health condition	Health conditions that affect respondents in any of the following ways: Vision, hearing, mobility, dexterity, learning, memory mental health, stamina/ breathing/ fatigue, socially or behaviourally, or in other ways. These are combined into a binary flag of health condition.	We excluded addiction from this flag. Note that we do not filter by how much the condition affects a respondent's day-to-day activities. This binary flag may miss nuances in terms of the sometimes very different levels of support a person might need or the very different impacts having a health condition has on their employment prospects.
Sexual Orientation	Heterosexual, gay or lesbian, bisexual, ace or asexual. These are combined into a binary flag of LGBA (Lesbian, Gay, Bisexual, Ace/Asexual) and non-LGBA.	This flag may not reflect the full diversity of contemporary identities that fall under sexual orientation. Due to small sample sizes and given the alternative was to not conduct analysis on this community, we accepted the limitations and grouped them for the purposes of this analysis.

There are some data missing that would likely affect mortgage pricing. For instance, credit score is not collected as part of PSD001 but likely affects a person's capability for accessing a wide range of mortgage products.

There are two caveats with the FLS/PSD001 joined dataset to highlight.

- Firstly, we join the two datasets by date of birth and postcode and ascribe the demographic characteristics (individual level) to the mortgage (loan level). In our dataset, we find that 22% of the mortgages have only one borrower assessed, 55% have 2 or more borrowers assessed, and 23% are null. However, the number of borrowers assessed does not necessarily indicate the number of borrowers on the mortgage. When creating a new variable that counts borrowers by all available borrower data, 18% have data for only one borrower and 82% have data for more than one borrower. We cannot always account for the demographic characteristics of multiple borrowers on one mortgage as all borrowers may not have been matched. Whilst we could drop these mortgages, single borrower mortgages do not represent the overall market. We keep mortgages with more than one borrower in our dataset and explicitly state when the results shown are for single or multiple borrower mortgages.

- Secondly, we are using the demographic characteristics of individuals at the time of the FLS survey (2020 or 2022), but the mortgage might have been taken out before or after (between April 2015 and May 2023).

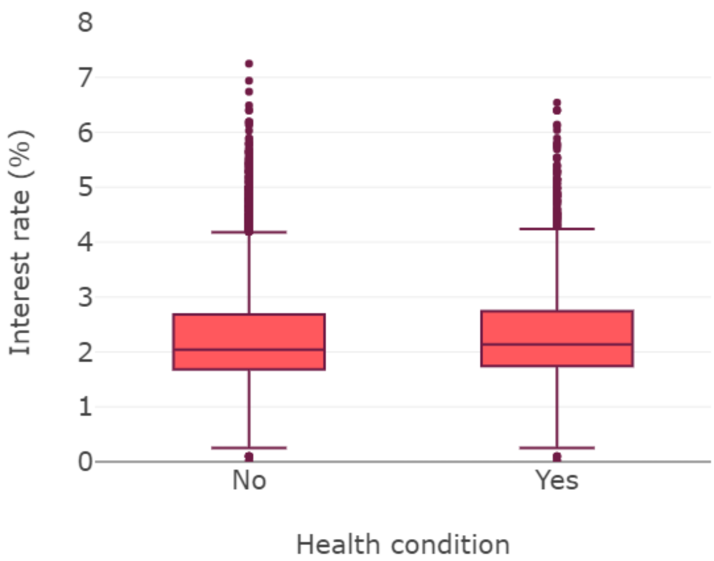
3 Results

Descriptive Analysis

Health Condition

The interest rates for people who have a health condition are on average, when comparing medians, 10 basis points higher than those who do not. Lifetime, bridging and second charge mortgages were excluded to begin for all exploratory analysis due to their higher rates of interest. Those with a health condition labelled as "unspecified", meaning either their response was "don't know" or "prefer not to say", were excluded for this part of the analysis. This difference is statistically significant at p<0.01 using a two-sided Mann Whitney U-Test. The threshold for significance we are using is at least p<0.05. When we explicitly look at single borrower mortgages, filtering by number of borrowers assessed, those who have a health condition still have higher interest rates, but the difference is no longer statistically significant.

Figure 1: Initial gross rates of interest split by health condition[1]



Relative to mortgage holders who do not have a health condition, those with a health condition tend to have lower property value, loan value and lender fees, which are all significant at the p<0.01 level, but similar LTV and LTI ratios. When just comparing single borrower mortgages, these differences remain statistically significant at the p<0.01 level except for lender fees. This pattern and statistical significance in difference in interest rates for those with a health condition is consistent across almost all regions and most age groups.

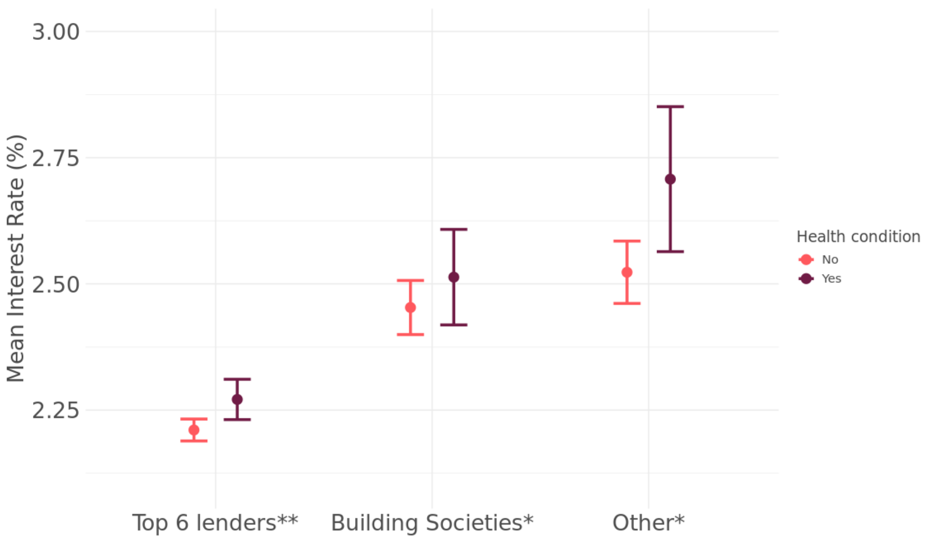
[1] , Figure 1 shows red box plots that indicate the median, interquartile range, outliers, which are data points that are more than 1.5 times the interquartile range below or above the first or third quartile respectively, and the whiskers, which are linked to the smallest or largest data point within these 1.5 times interquartile ranges.

We split lenders into three different groups:

- Top 6 mortgage lenders,
- Building societies, and
- Other.

The difference is larger for "Other" lenders compared to the top 6 mortgage lenders, as seen in Figure 1 below. The difference in the "Other" and "Building societies" groups is statistically significant at the $p<0.05$ level whilst the difference in the "Top 6 lenders" group is at the $p<0.01$ level. This difference remains largest for the "Other" group when filtering for single borrower mortgages, although the differences are no longer statistically significant across the groups.

Figure 2: Mean initial gross rates of interest and 95% confidence intervals split by health condition across different mortgage lender types[2]



These patterns suggest those who have a health condition are more likely to be taking out products with costs spread over time, resulting in higher interest rates and lower lender fees than other borrowers. We cannot say whether this is because they were the only products available to them or because they actively chose this option. Note that customers will always have a choice of paying fees upfront or including them in the loan amount for regulated mortgages as defined by the FCA's Mortgages and Home Finance: Conduct of Business Sourcebook (MCOB).

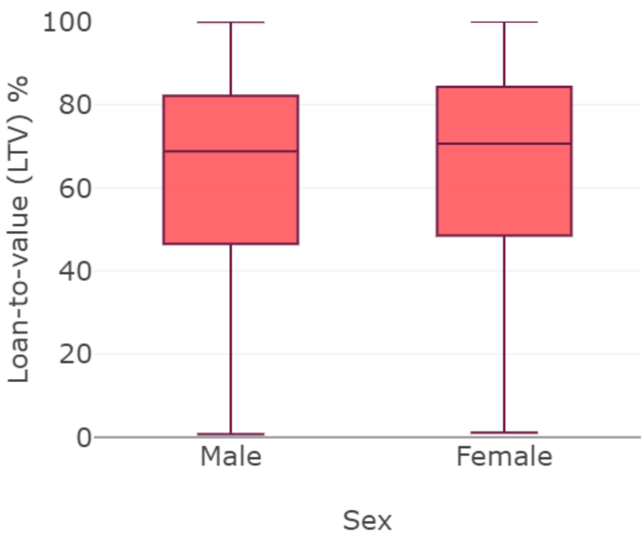
Sex

Whilst the difference in average interest rates is statistically significant between the sexes, with female rates higher and $p<0.05$, the difference is five basis points. On average females have marginally higher LTV mortgages (less than two percentage points), lower lender fees and lower property prices. These differences are statistically significant at the $p<0.01$ level. We dropped lifetime, bridging and second charges mortgages, as well as those who's sex is labelled as "prefer not to say".

[2] We use * to indicate the difference is significant at the $p<0.05$ level and ** to indicate the $p<0.01$ level.

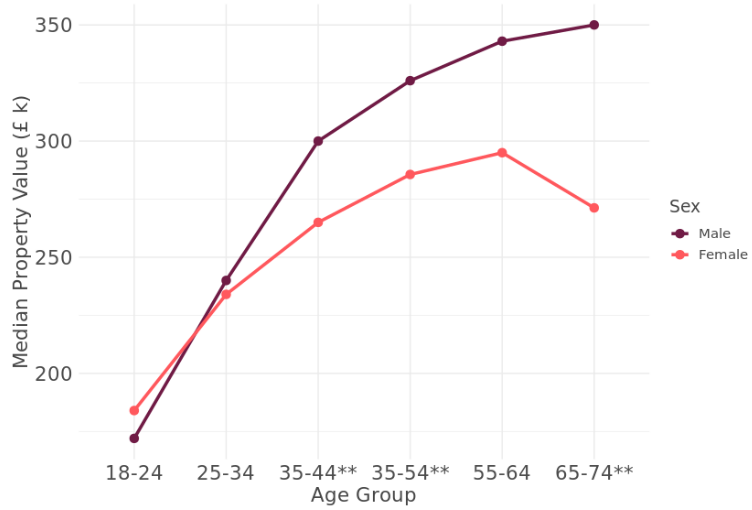
When filtering for single borrower mortgages these differences remain significant at the p<0.01 level except for lender fees, which is significant at the p<0.05 level, and interest rates, which are no longer significant. However, the difference in LTV changes direction such that it is lower for females on average for single borrower mortgages. The average LTV for women drops from 71% to 63%. It should be noted that the difference for males and females in LTV is small for both single and multiple borrower mortgages (less than five percentage points) rather than large differences that separate the groups by different LTV bands (e.g., <60% LTV compared to 90-95%).

Figure 3: Loan-to-value (LTV) ratios for males and females



Except for the 18-24 age group, females take out smaller loans than males. This difference in average loan value between sexes is statistically significant at the p<0.01 level for age groups between 25 and 64 and at the p<0.05 level for the 65-74 group. The 75+ age group was excluded due to low sample size. Females are also, on average, buying less expensive properties, except for in the 18-24 age group, with their statistical significance shown in Figure 4.

Figure 4: Comparison of median property prices for males and females across different age groups



Comparing the total household gross income, for single borrower mortgages men on average have higher income in all age groups. This difference is statistically significant at least at the p<0.05 level in all age groups except 18-24 and the 65-74 groups. When there is more than one borrower on the mortgage, those with a matched male have higher incomes except in the 18-24 age group. These differences are statistically significant for all groups except the 18-24 and 25-34 groups.

These differences likely highlight issues wider structural inequalities, such as the gender pay gap (see e.g., in a UK context ONS 2018) being taken into consideration when firms manage risk, rather than unfair lending practices. Differences in income, property value and loan amount between males and females are most prominent in the 35 to 64 age groups.

Sexual orientation

There are some differences when comparing mortgage characteristics across sexual orientation, specifically the LGBA community for this analysis, which may overlook differences between groups within that community. We dropped lifetime, bridging and second charges mortgages, as well as those who's sexual orientation is labelled as "prefer not to say".

Initial gross interest rates are nine basis points higher for people who identify as LGBA compared to non-LGBA. This difference is statistically significant at the p<0.05 level but is not found to be statistically significant when filtering for single borrowers.

The median LTV is five percentage points higher for LGBA people whilst property prices, total household income and lender fees are lower. These are statistically significant at the p<0.01 level and remain significant only for LTV when filtering for single borrower mortgages. In fact, the difference increases to eight percentage points for single borrower mortgages. However, for property prices, total household income and lender fees the difference is no longer significant when filtering for single borrower mortgages.

Ethnic group

There are some differences observed for people in minority ethnic groups compare to those not in minority ethnic groups. This may overlook differences between groups within the minority ethnic groups. We dropped lifetime, bridging and second charges mortgages, as well as those who's ethnic group is labelled as "prefer not to say".

Interest rates are one basis point lower for people from minority ethnic groups (ethnic groups other than white British), which is statistically significant at the $p<0.01$ level, even when filtering for single borrower mortgages. Their median LTV is marginally higher but is not found to be statistically significant. Loan value, total household income and lender fees are all larger and statistically significant for people from minority ethnic groups, with the latter two remaining statistically significant when filtering for single borrowers.

In addition, people from minority ethnic groups are marginally more likely to have a mortgage with LTI ratio greater or equal to 4.5 (known as the flow limit) compared to people who are not from minority ethnic groups, when analysing multiple borrower mortgages.

However, we have too few data points to draw conclusions when aggregating by more granular ethnic groups that are available as part of the FLS responses (e.g., White, Asian, Black & Black British) or to compare across other dimensions, such as region.

Age

We will not explore age here in detail since this data is not unique to our joined dataset and is collected as part of PSD001. In short, the youngest and oldest age groups have the highest average initial gross rates of interest. The 75+ age group was filtered due to the low sample size. Lifetime, bridging and second charge mortgages were filtered as they have higher associated interest rates and are more likely to be taken out by older people. Younger people on average have the highest LTV ratios which explains the high interest rates. They are also more likely to have a mortgage with LTI over the 4.5x flow limit. This likely reflect differences in product choices across age groups. For instance, those who are older are more likely to have expensive variable rate mortgage products due to their flexibility. These trends do not change when filtering by number of borrowers assessed.

Pricing Model

Aims

Considering the differences in pricing we have observed we will now explore whether these differences remain when we control for a range of factors related to the mortgage and the borrower.

We will develop our own machine learning model that predicts the interest rate of a mortgage using information about the mortgage product type, risk factors and borrower information including their demographic characteristics. This allows us to examine whether demographic characteristics contribute to differences in predicted interest rates, or if the differences are driven by these other factors.

Note that this model is not a replica or reproduction of any individual firms pricing model but seeks to predict and explain prices across the whole market for individuals for whom we have data.

Strategy

A machine learning model is developed to predict the initial gross rate of interest based on several features. We use a subset of key mortgage properties as features, as well as data on demographic characteristics matched on the borrower of the mortgage, and the Bank of England base rate at the time of purchase. We also reintroduce all mortgages that were excluded from the exploratory analysis, such as lifetime and second charge mortgages.

Methodology Detail

Following light touch pre-processing, the dataset was randomly split 90:10 into a training and testing set. The sets were stratified by health condition as this is the demographic characteristic we are most interested in based on our exploratory analysis.

A Cat Boost encoder (Prokhorenkova et al. 2019) was used to numerically encode the categorical data. This is a target encoder, meaning that it encodes a categorical feature using information about the relationship between the feature and the target. To avoid information about the target variable being leaked to the feature set, it uses ordered-target encoding, whereby data is fed in sequentially to determine target statistics in relation to the feature. These are learnt from the training data and used to transform both the training and test datasets.

A gradient-boosting decision tree model, LightGBM (Ke et al. 2017), was used. Boosting is an ensemble approach whereby decision trees are grown sequentially, such that the next tree is fit on the last tree's residuals. LightGBM benefits over other boosted tree-based algorithms due to its faster model training without a reduction in performance. It also natively handles missing values by ignoring them when splitting the data. LightGBM can result in overfitting due to its leaf-wise growth compared to level-wise growth that is commonly used by boosted tree-based models. This can be controlled by tuning the hyperparameters such as tree depth, using cross-validation and comparing results with a simpler model such as a random forest or linear regression.

A random grid search was performed over a parameter space to find the hyperparameters which maximise R^2 score over 5-fold 2-repeat cross-validated folds of the training set. The hyperparameters tuned for were the number of leaves, the learning rate, the maximum tree depth, the L1 regularization parameter and the number of trees.

The best fitting model with parameters as selected by the grid search constituted the final model and was tested on the unseen test set.

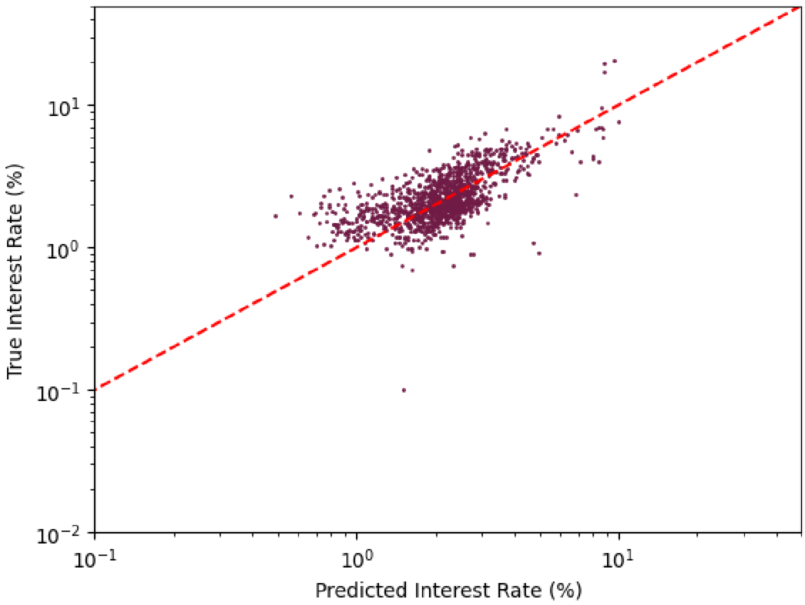
Results

Model Evaluation

The model performs well on a holdout dataset, with an R^2 of 0.84, a mean absolute error (MAE) of 0.30, and a root mean squared error (RMSE) of 0.53. This suggests a high level of predictive power. We would expect the model to perform particularly well since almost all the variables we expect to price a mortgage were included in the dataset as they are a reporting requirement of PSD001. In literature, Benneton et al. 2021 used similar datasets, including the mortgages PSD and the Bank of England's Housing survey that is akin to the FLS survey except that it does not collect data on demographic characteristics. They modelled mortgage supply and demand and achieved good predictive performance.

The model tends to make accurate predictions on more standard mortgages but can struggle making predictions for less typical mortgages. For one of the test set mortgages with an interest rate of 0.02%, our model predicted an interest rate of 1.83%, notably the worst prediction the model made on unseen data. Structural issues with the mortgage reporting, such as data quality issues, could be a source of this low performance. Exclusion of atypical interest rate mortgages (less than 0.5% and greater than 8%) produces similar R^2 score but reduced MSE.

Figure 13: True vs predicted interest rates



If we assume all relevant data used to price a mortgage was captured by PSD001 we would expect the R^2 score to be closer to 1. The UK inflation rate and SONIA interest rate benchmark were tested as features alongside our feature set. There were minor improvements in some but not all the performance metrics. The lack of credit scoring data is likely a source of the reduced performance. Whilst it is a risk considered when pricing a mortgage, it is not collected as part of PSD001. However, it was deemed unnecessary to quantify this effect here as it is already known, and firms likely use their own in-house credit scores to determine mortgage prices.

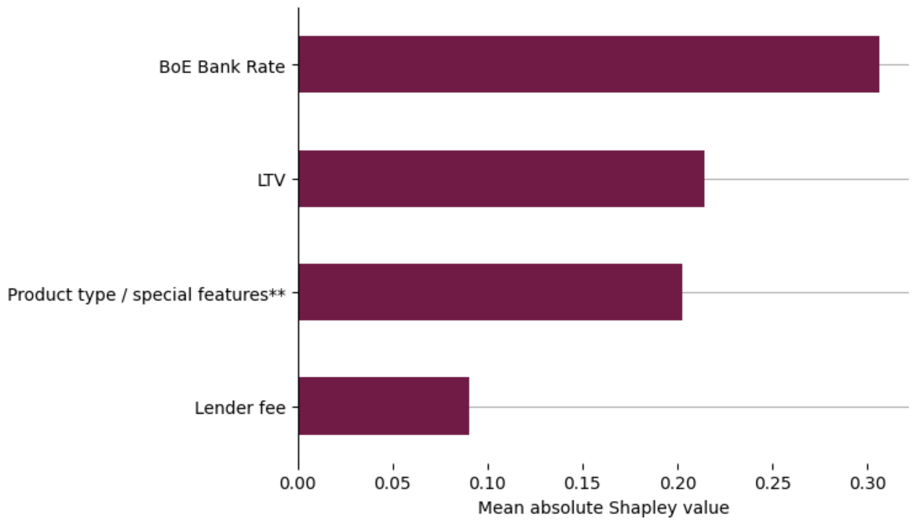
A simple baseline linear regression model was built for comparison. It performed worse across all performance metrics than the LightGBM model. The distribution of the model residuals was not normally distributed with heavy tails. Transforming the features and target did not improve performance, hence the final choice of a non-linear tree-based machine learning model.

Feature Importance

To determine the influence of demographic characteristics on predicted interest rates we calculate and aggregate the Shapley value for each feature (Lundberg et al. 2017). Shapley values originate from cooperative game theory (Shapley 1953) and provide a measure of the average marginal contribution of each feature to an instance level prediction compared to the average prediction for the dataset. To compute a global measure of feature importance we find the mean absolute Shapley value per feature for all test set predictions. This will show which of our features have the biggest influence on predicted interest rates. Ranking our feature set by this measure will indicate how much demographic characteristics contribute to the predicted interest rate compared to the mortgage features.

The top four features ranked by Shapley value are the Bank of England base rate, LTV ratio, product type/special features (e.g., lifetime, bridging or second charge mortgages) and lender fees. The demographic characteristics all rank low, with health condition ranked highest in terms of mean absolute Shapley value. Health condition has a mean absolute Shapley value of 0.0047, contrasted with 0.084 for lender fees. Also note that this measure of feature importance tells us to what extent the given feature influences the predicted interest rates. It is not necessarily indicative of statistical significance. In addition, whilst controlling for such a broad range of factors, there is a risk that a variable (such as region) that is highly correlated with a demographic characteristic we tested for (and therefore may link to potential unfairness in relation to that characteristic, despite not being the same characteristic), is significant.

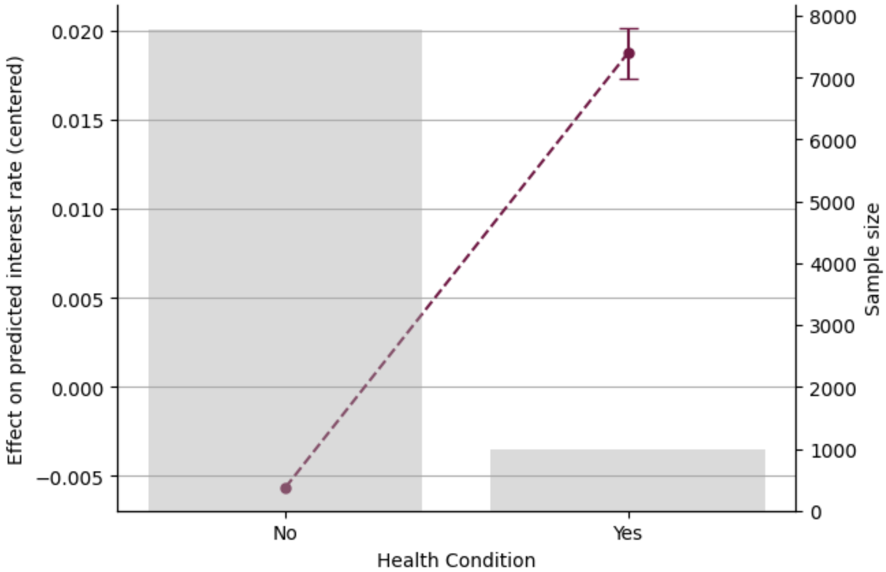
Figure 14: The top 4 features ranked by mean absolute Shapley value



Quantifying the influence of health conditions

To quantify the difference in predicted interest rates for those who have a health condition, we use accumulated local effect plots. These are like partial dependence plots but perform better when some of the features are correlated (Apley, W. and Zhu J. 2016), as is likely the case in our dataset. Analysing the accumulated local effect of health condition in Figure 5, where we have removed cases where health condition was reported as “not specified”, the difference in predicted interest rate across health conditions is approximately 0.014 percentage points, where the y-axis is in the same units as the target. This shows that health condition alone has almost no effect on interest rates, both statistically and in real monetary terms. This difference is also likely below the error bounds of the model’s predictive capabilities. However, it must be made clear that this model isn’t sufficient to determine causality. There are likely other factors affecting rates that we are not accounting for, such as credit score, which could change our findings if included.

Figure 15: Accumulated Local Effect of health condition, comparing the difference in predicted interest rates



4 Conclusion

Using data collected by the FCA a new dataset was developed that contains individual demographic characteristics and loan level UK mortgage data for a subset of mortgages taken out after the introduction of the 2014 mortgage affordability rules. This allowed for an analysis of differences in mortgage pricing and products across groups sharing demographic characteristics to examine whether there is evidence of differential outcomes.

The results of this research are split into two components.

Firstly, statistical analysis of the matched dataset was undertaken to explore the differences in mortgage properties when split by different demographic characteristics.

Differences in mortgage product features were identified amongst some of the groups sharing particular demographic characteristics.

For instance, those who have a health condition appear to have on average higher interest rates, lower lender fees, similar LTV ratios and lower household incomes. This indicates that those who have a health condition are more likely to have mortgages on less expensive properties where costs are spread over time. Though we note that, for regulated mortgage products, customers should always have the choice of paying lender fees upfront or including them in the loan amount, under the FCA's rules.

In addition, females appear to have similar interest rates and LTV ratios as males, but lower household income and property values. These are likely indicators of wider structural inequalities, such as the gender pay gap, rather than unjustified pricing differences.

Secondly, we built a machine learning model to quantify how influential demographic characteristics were in determining mortgage price once we have controlled for the mortgage product and borrower features. A machine learning model was built that takes a range of mortgage properties, the Bank of England base rate and the demographic characteristics of the matched individuals and predicted their mortgage interest rate.

When comparing the influence of these features on the predictions, demographic characteristics had little to no effect. Whilst a difference in interest rates was observed as part of the exploratory analysis, the difference can be explained once controlling for the features of the mortgage, including both product and borrower features. For instance, it may be that individuals with a health condition select different products on average, such as those with lower upfront fees, leading to different outcomes regarding interest rate when viewed alone.

From this research, whilst we observe that there is no evidence of a lack of direct pricing fairness (through differences in pricing by demographic characteristics alone), we cannot conclude that there are not issues with "demand fairness" via the availability of products to different groups. There are other potential drivers of differential outcomes, which have been discussed in this research note, such as the impact of society on the financial conditions of individuals at the point of taking out a mortgage. Future attempts to build upon this research could include further analysis as to what does drive the differences in outcomes and further work to determine causality.

Annex 1: Data

Dataset construction

In Section **Error! Reference source not found.** we discussed the data cleaning process a nd listed the demographic characteristics used for the analysis. Alongside these features we used a number of variables gathered as part of the PSD001 regulatory return. Below is the full list of mortgage features from PSD001 we used alongside demographic characteristics to predict initial gross rates of interest.

Mortgage properties used for prediction

Variable	Description	Coverage/Limitations
Advised sale flag	Advice at point of sale, either Yes or No	Complete coverage
Age when opening account	In years	Complete coverage
Bank group	Firms split by Bank type, including Big Five, Challenger, and Other	Complete coverage
Bank of England base rate	Base rate at time of account open date	Complete coverage
Borrower type	Type of borrower, such as first-time buyer and remortgagors	Complete coverage
Credit history summary	Including secondary arrears and Individual Voluntary Arrangement	>75% missing values
Dwelling type	Including semi-detached house, terraced house, etc.	Complete coverage
Fees added to loan	Fees or charges added to the loan	Complete coverage
First borrower CCJ	County Court Judgements	10-25% missing values
FRN	Firm reference number	Complete coverage
Government supported initiative	Was the mortgage advanced under a government supported initiative – Yes, No or Unspecified	Complete coverage
Impaired credit indicator	Either Yes or No	Complete coverage
Income verification	Income evident by lender or third party	10-25% missing values

Interest rate type	Including fixed, discounted rate, LIBOR tracker, etc.	Complete coverage
Intermediary or third-party fees	Fees and charges added by mortgage intermediary or third party included in the annual percentage rate (APR) of charge in relation to the mortgage	Complete coverage
Internal product transfer	Flag showing if mortgage an internal product transfer	Internal product transfers only reported if completed on or after 1st April 2021
Lender fees	Fees and charges added by lender included in the annual percentage rate (APR) of charge in relation to the mortgage	Complete coverage
Length of early repayment charge	In years	5-10% missing values
Length of incentivised rate	In years	0-5% missing values
Loan value	Total loan amount	Complete coverage
LTI	Loan-to-income ratio	10-25% missing values
LTV	Loan-to-value ratio	10-25% missing values due to missing property value for certain products, such as remortgages
Main borrower age at maturity	In years	0-5% missing values
Mortgage lender type	Firms split by mortgage lender type, including Big 6, building societies, and other lenders	Complete coverage
Mortgage term	In years	0-5% missing values
New dwelling flag	Either Yes, No, or Unspecified	Complete coverage
Number of bedrooms	Integer	10-25% missing values
Number of borrowers assessed	Either 1, 2 or 3 (with 3 meaning 3 or more)	10-25% missing values
Property value	Property value at time of purchase	10-25% missing values, not required for remortgages or further advances
PTI	Payment-to-income ratio	25-50% missing values
Region	Region of property	Complete coverage
Repayment strategy	Repayment strategy, such as buy-to-let, lifetime mortgage, bridging and second charge mortgages	25-50% missing values

Repayment type	Type of repayment, including capital and interest, interest only and mixed	Complete coverage
Reverse rate type	Including tracker, standard variable rate, etc.	Complete coverage
Sales channel	How the sale was made, including face to face, telephone, intermediary etc.	Complete coverage
Second borrower age at maturity	In years	25-50% missing values
Second borrower CCJ	County Court Judgements	25-50% missing values
Total credit commitments	In sterling pounds	10-25% missing values
Total gross income	Total gross income of all borrowers on mortgage	10-25% missing values
Town	Town of property	Complete coverage

Dataset validity

Comparing the distribution of demographics within the joined dataset and the FLS 2020 and 2022 dataset, after data cleaning we find that the most remain unchanged, increasing or decreasing within three percentage points.

For minimum sample size calculations needed to construct a 95% confidence, we used the standard binomial formula for required sample size given some error rate (Watts 2022). We used the subsample means as estimated proportions for each category (i.e., white vs. not white, Asian vs. not Asian etc.) to have a binomial distribution. We then apply the formula using the estimated proportions to determine the minimum sample size required to create a 95% interval for 3% error rate (typically 1%, 3% and 5% are standard acceptable ranges). We found that we only just have enough data in all the ethnic group subsamples (the smallest category requirement was n = 80 for black and black British, and we have 85 data points). We advise caution when using these small subsamples, especially when aggregating by more than just ethnicity, such as region, as these values are a minimum requirement. We were satisfied that all other demographic characteristics we included were populated enough for this analysis using this test.

Statistical analysis

When comparing distributions and medians in Section 3 we used the Mann Whitney U-Test. This is a test of stochasticity due to the ranking procedure. It tests whether it is equally likely that a randomly selected value from the first sample will be less than or greater than a randomly selected value from the second sample.

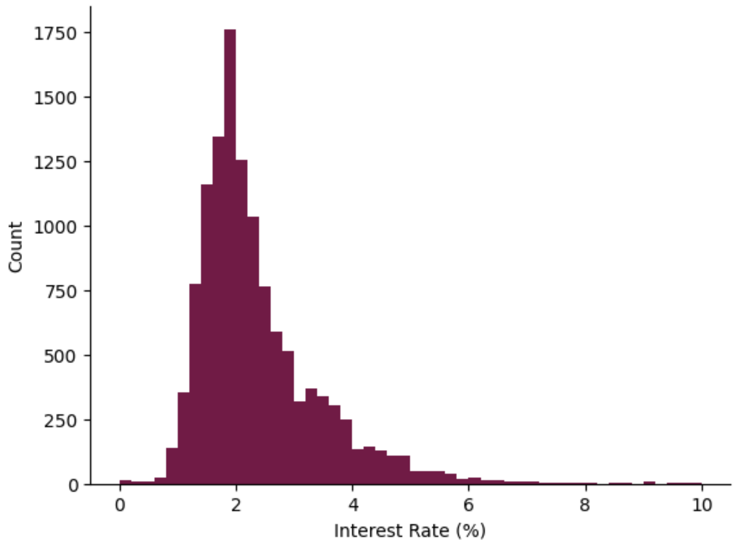
This test is used because the assumption of normality in the unpaired two-sample t-test cannot be satisfied. This was identified when analysing the distribution of the key mortgage properties, such as lender fees that has a bimodal distribution. Thus, the Mann

Whitney U-Test is a robust non-parametric alternative to the t-test. Note this test only acts as a valid comparison between medians of different groups under the assumption of similar distribution of data between independent groups. The base implementation in R includes a continuity correction as standard.

Exploratory analysis

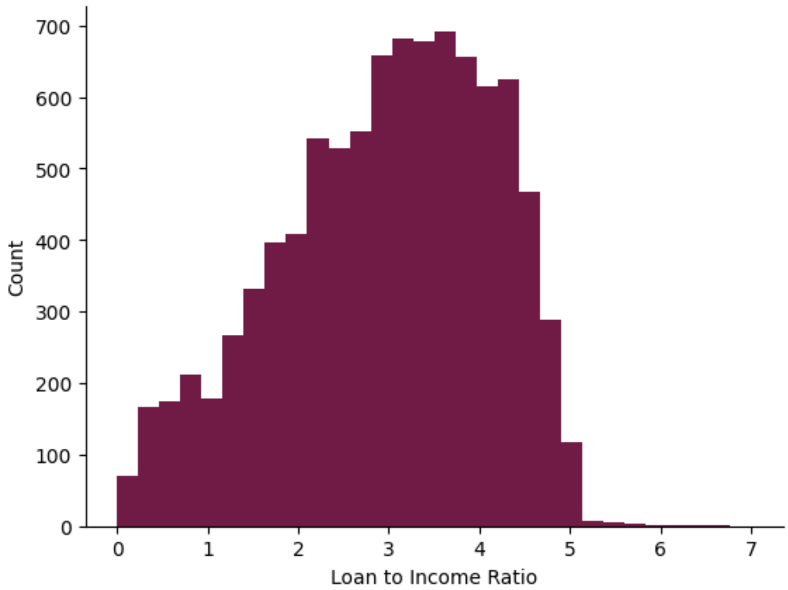
Before exploring the differences in mortgage properties across demographic characteristics, we analysed the distribution of the key mortgage properties across the whole dataset.

Figure 1: Distribution of initial gross rate of interest of mortgages with interest rates below 10% (99.8% of all mortgages in dataset)



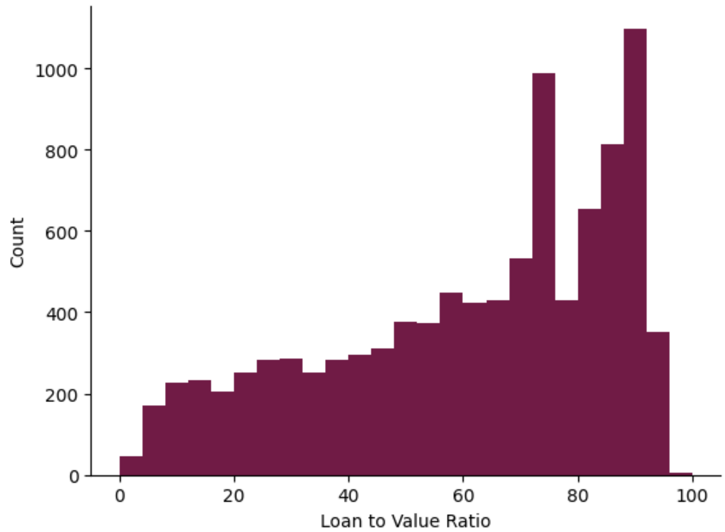
The initial gross rate of interest distribution shows a positive skew with a relatively long tail, and the highest interest rate observed being 26.82%. We considered applying a log transformation to the initial gross rate of interest prior to modelling. This was ultimately not taken forward as it did not lead to improved model performance, nor did it allow for further insight in exploratory analysis.

Figure 2: Distribution of loan to income ratio of mortgages with ratios below 7 (99.8% of all mortgages in dataset)



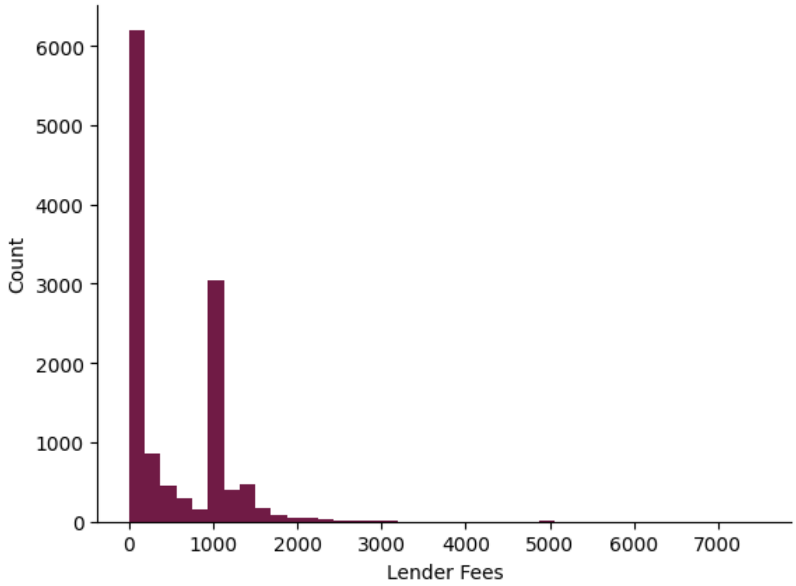
The distribution of loan-to-income ratios (LTIs) in the mortgages in our dataset provides expected results – we can see the impact of the flow limit regulatory requirement stating that mortgage providers must not provide more than 15% of their total mortgage lending on mortgages with LTI greater than or equal to 4.5 times.

Figure 3: Distribution of loan-to-value ratios of mortgages



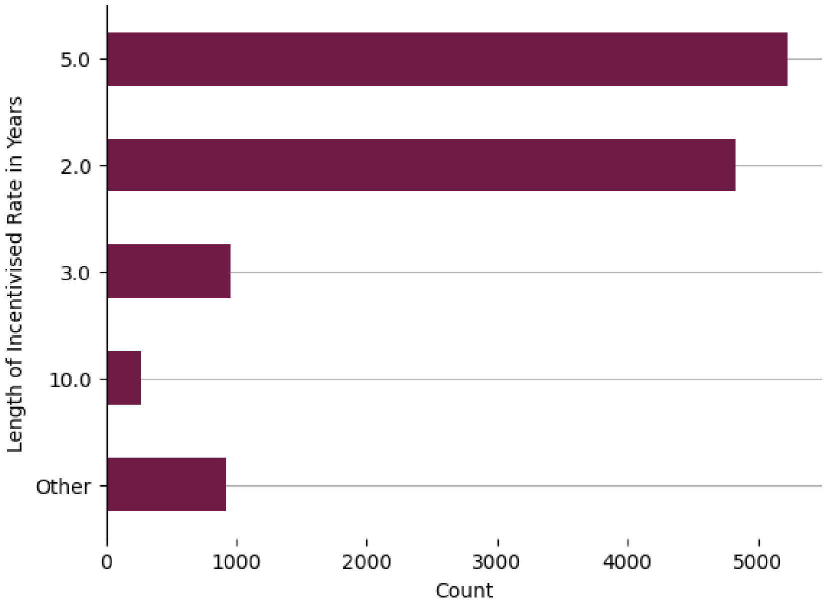
Similarly to LTIs, the distribution of the LTVs of the mortgages in the dataset displays expected characteristics –the most common LTVs occur in the 60-95% range, and negative skewness reflecting that many consumers purchase properties with higher deposits enabling lower LTVs.

Figure 4: Distribution of lender fees



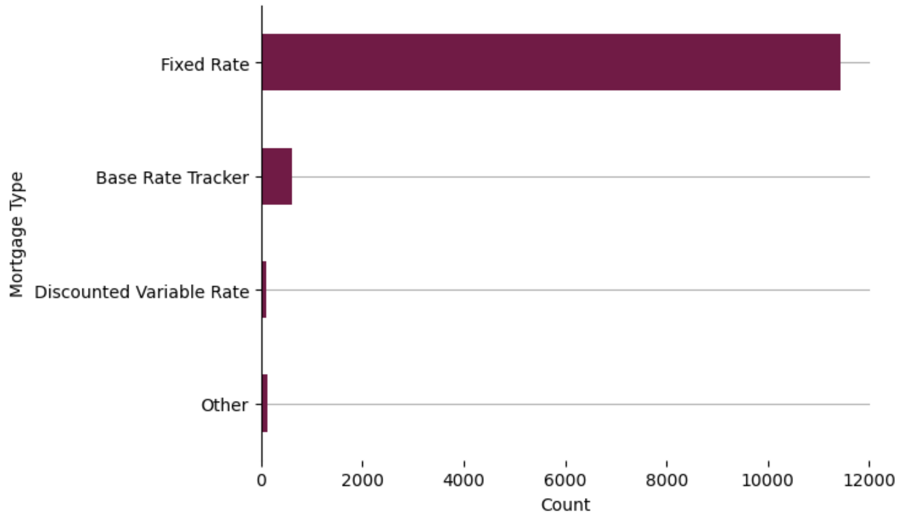
The distribution of lender fees across mortgages is bimodal, as mortgages generally fall into one of two groups – having substantial lender fees or not. Furthermore, there is a long right tail. Of mortgages with lender fees, 98% have lender fees under £2,000.

Figure 5: Most common lengths (years) of incentivised rate



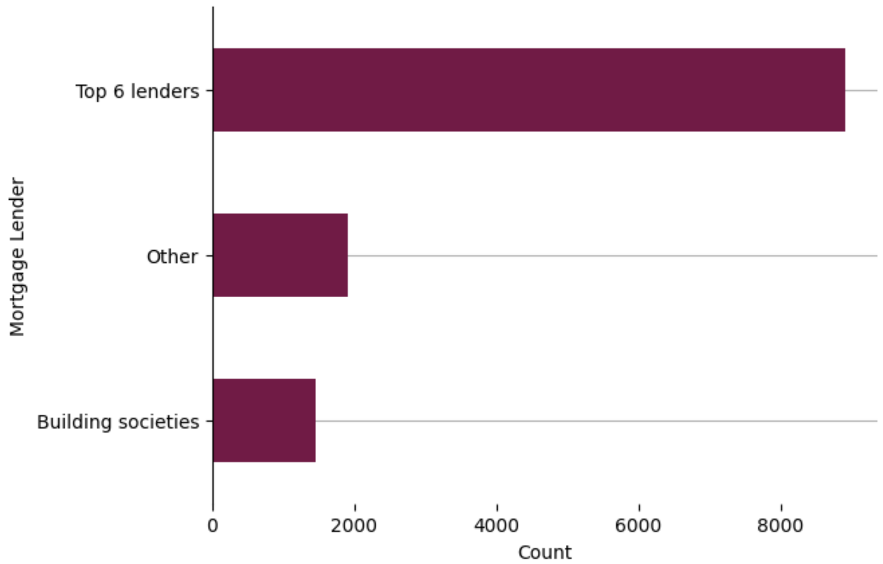
Most mortgages have an initial incentivised term of 2 or 5 years, with terms outside of these two options proving the exception.

Figure 6: Mortgages by interest rate type



Most mortgages in the dataset are fixed rate mortgages, with most tracker mortgages being base rate trackers.

Figure 7: Mortgages by lender type

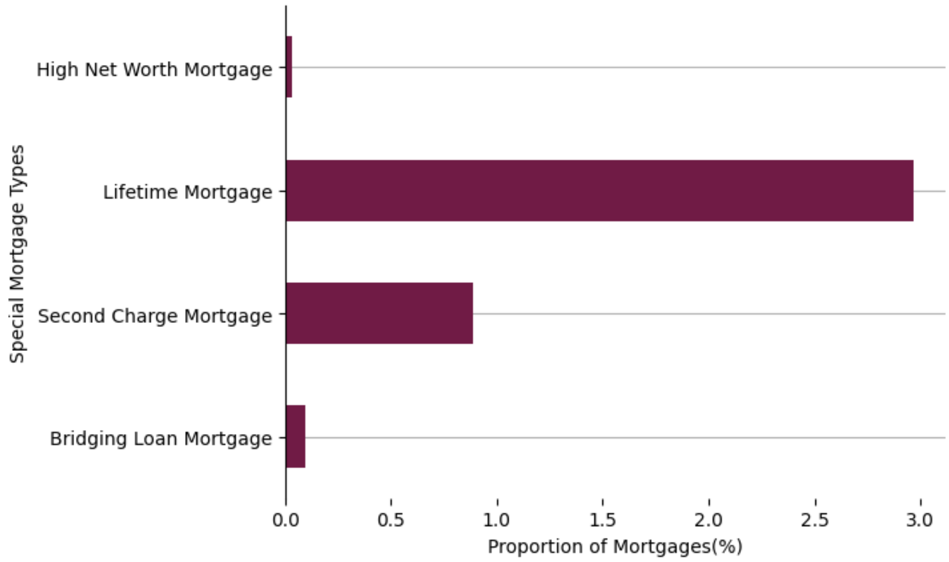


Most mortgage lenders in our dataset are one of the top 6 lenders, with more lenders falling in the "Other" category compared to the building societies.

Figure 8: Mortgages by borrower type

Our dataset contains an even split of borrower types, with the largest group being the external remortgagors. This distribution slightly differs to that of the entire PSD001 during our period of interest in that there are slightly more 2nd or subsequent buyers and first-time buyers than internal remortgagors in PSD001.

Figure 9: Prevalence of rarer mortgages



The prevalence of rarer mortgage types was generally low, with the largest group being those who have a lifetime mortgages (note that this was only 3% of our total dataset).

In general, the distributions of mortgage features as outlined in this section holds for the wider PSD001 data, satisfying our assumption that our dataset is representative of the UK mortgage market.

Annex 2: Glossary

Accumulated local effects: describe how features influence the prediction of a machine learning model on average and can handle features that are correlated.

Algorithmic bias: the potential for algorithms to give different predictions on the basis of demographic characteristics such as race and sex.

Bias: differences by group membership in predictions of risk or other future outcomes like willingness to pay that typically feed into decisions about the prices to charge and decisions to provide products to individual consumers.

Catboost encoder: encodes a feature using information about the relationship between the feature and the target.

Cross validation: resampling procedure used to evaluate machine learning models whereby the training dataset is split into different chunks to calculate the distribution and average performance for some given metric.

FLS: the Financial Lives survey (FLS) is the FCA's flagship, nationally representative survey of UK consumers. It provides information about consumers' attitudes towards managing their money, the financial products they have and their experiences of engaging with financial services firms.

Hyperparameter tuning: finding the optimal set of hyperparameters for a machine learning model that maximizes the model's performance, minimizing a predetermined loss function.

Hyperparameter: features of a machine learning model that need to be define before fitting a model.

LightGBM: a boosted tree-based machine learning model. Boosting is an ensemble approach whereby decision trees are grown sequentially, such that the next tree is fit on the last tree's residuals. LightGBM benefits over other boosted tree-based algorithms due to its faster model training without a reduction in performance.

MAE: Mean Absolute Error, defined as the sum of the absolute residual (non-negative predicted minus true value) divided by the number of predictions.

Mann Whitney U-Test: a statistical test that compares distributions and medians. This test is suitable when the assumption of normality in the unpaired two-sample t-test cannot be satisfied.

Mortgage pricing: in this context refers to the interest rate of the mortgage whilst considering other mortgage properties that affect the interest rate, such as fees and loan-to-value ratio.

Predictive algorithm: models used to predict future events or outcomes by analysing patterns in a given set of input data.

Price discrimination: charging different prices to consumers, for instance based on their willingness to pay.

Pricing fairness: ensuring that the prices offered to two groups, split by demographic characteristics, are nearly equal, as defined by Cohen et al. 2022.

PSD001: the Product Sales Data (PSD) comprises data collected by the FCA from firms they regulate on what products they are selling. PSD001 collects all new mortgages and their main characteristics, including mortgage and borrower characteristics.

R^2 score: a measure of the goodness of fit of a regression model that indicates how much of the variance in the predicted variable is captured by the model.

Random grid search: selecting and testing a random combination of hyperparameters, where the number of tests is pre-defined and the best set of hyperparameters is defined by selecting the best model performance using some predetermined metric.

Regression: a statistical method to determine the strength and character of the relationship between one dependent variable and several other variables.

RMSE: Root Mean Squared Error, defined as the sum of the squared residual (predicted minus true value) divided by the number of predictions, sometimes corrected by subtracting one.

Shapley values: provide a measure of the average marginal contribution of each feature to an instance level prediction compared to the average prediction for the dataset.

Supervised machine learning: a subcategory of machine learning and artificial intelligence that uses labelled datasets to train algorithms to classify data or predict outcomes accurately.

Annex 3: References

Apley, D.W., & Zhu, J. (2016). Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. arXiv:1612.08468. doi:10.48550/arXiv.1612.08468

Benetton, M., Gavazza, A., & Surico, P. (2021). Mortgage pricing and monetary policy. Bank of England Staff Working Paper No. 936. Available at: https://www.bankofengland.co.uk/working-paper/2021/mortgage-pricing-and-monetary-policy

Coen, J., Kashyap, A., & Rostom, M. (2021). Price discrimination and mortgage choice. Bank of England Staff Working Paper No. 926. Available at: https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2021/price-discrimination-and-mortgage-choice.pdf

Cohen, C.M., Elmachtoub, A.N., & Lei, X. (2022). Price Discrimination with Fairness Constraints. Management Science 68(12):8536-8552. doi: 10.1287/mnsc.2022.4317

Financial Lives (2023). Key findings from the FCA's Financial Lives May 2022 survey. Available at: https://www.fca.org.uk/publication/financial-lives/financial-lives-survey-2022-key-findings.pdf

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W.. Ye, Q., & Liu, T. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., & Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.

Lundberg, S., & Lee, S.I. (2017). A Unified Approach to Interpreting Model Predictions. arXiv:1705.07874. doi:10.48550/arXiv.1705.07874

Namin, S., Zhou, Y., Xu, W., McGinley, E., Jankowski, C., Laud, P., & Beyer, K. (2022). Persistence of mortgage lending bias in the United States: 80 years after the Home Owners' Loan Corporation security maps. Journal of race, ethnicity and the city, 3(1), 70–94. https://doi.org/10.1080/26884674.2021.2019568

Office for National Statistics. (2018). Understanding the gender pay gap in the UK. [online]. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/articles/understandingthegenderpaygapintheuk/2018-01-17#modelling-the-factors-that-affect-pay.

Office for National Statistics. (2021). Subnational estimates of dwellings and households by tenure, England: 2021. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/housing/articles/researchoutputssubnationaldwellingstockbytenureestimatesengland2012to2015/2021#:~:text=This%20model%20estimates%20that%20there,be%20owner%2Doccupied%20in%202021.

Prokhorenkova, L., Gusev, G., Vorobev, A., Veronika Dorogush, A., & Gulin, A. (2017). CatBoost: unbiased boosting with categorical features. arXiv:1706.09516. doi:10.48550/arXiv.1706.09516

Shapley, L. (1953). 17. A Value for n-Person Games. In H. Kuhn & A. Tucker (Ed.), Contributions to the Theory of Games (AM-28), Volume II (pp. 307-318). Princeton: Princeton University Press. https://doi.org/10.1515/9781400881970-018

Starks, M., Reynolds, G., Gee., C., Burnik, G., & Vass, L. (2018). Price discrimination in financial services: How should we deal with questions of fairness? Financial Conduct Authority Research Note.

UK Finance. (2019). The Changing Shape of the UK Mortgage Market. Available at: https://www.ukfinance.org.uk/policy-and-guidance/reports-publications/changing-shape-uk-mortgage-market

Verma, S. and Rubin, J., 2018, May. Fairness definitions explained. In Proceedings of the international workshop on software fairness (pp. 1-7). doi: https://doi.org/10.1145/3194770.3194776

Watts, V. (2017). Introduction to Statistics: An Excel-Based Approach. Fanshawe College Pressbooks.

12 Endeavour Square, London E20 1JN
Telephone: +44 (0)20 7066 1000 Website: www.fca.org.uk