BANK OF ENGLAND

Staff Working Paper No. 948

Refinancing cross-subsidies in the mortgage market

Jack Fisher, Alessandro Gavazza, Lu Liu, Tarun Ramadorai and Jagdish Tripathy

November 2021

BANK OF ENGLAND

Staff Working Paper No. 948

Refinancing cross-subsidies in the mortgage market

Jack Fisher,(1) Alessandro Gavazza,(2) Lu Liu,(3) Tarun Ramadorai(4) and Jagdish Tripathy(5)

Abstract

Evidence from a range of countries reveals that household inaction in mortgage refinancing can be pervasive despite financial incentives to take action. Inactive households may implicitly cross-subsidise active households, allowing competitive lenders to set lower average mortgage rates. To provide a money-metric assessment of cross-subsidies, we construct a model of household refinancing and structurally estimate it on rich administrative data on the stock of loans in the UK mortgage market in June 2015. We estimate sizeable cross-subsidies during this sample period, from relatively poorer households and those located in less-wealthy areas towards richer households and those located in wealthier areas. The findings over this sample period highlight how the design of household finance markets can contribute to wealth inequality. Estimated cross-subsidies may differ in more recent periods given changes in the UK mortgage market since 2015.

Key words: Mortgages, refinancing, cross-subsidies, wealth inequality, household inaction, household finance.

JEL classification: G21, N20, R21, R31, L51, D63.

(1) London School of Economics. Email: j.w.fisher@lse.ac.uk.
(2) London School of Economics. Email: a.gavazza@lse.ac.uk.
(3) Imperial College Business School. Email: l.liu16@imperial.ac.uk.
(4) Imperial College Business School and CEPR. Email: t.ramadorai@imperial.ac.uk.
(5) Bank of England. Email: jagdish.tripathy@bankofengland.co.uk.

We thank seminar participants at the Bank of England Workshop on Household Finance and Housing, SITE Workshop on Financial Regulation, Imperial College, and King's College Business School QCGBF Virtual Seminar, and Mikhail Golosov and Gregor Matvos for useful comments. Alessandro Gavazza gratefully acknowledges financial support from the European Research Council (under Consolidator Grant Agreement No. 771004). Any views expressed in this article are not meant to represent the views of the Bank of England, members of its Financial or Monetary Policy Committee, or the Financial Conduct Authority. All errors are our own.

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

Bank of England, Threadneedle Street, London, EC2R 8AH
Email enquiries@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

In markets for complex products, cross-household variation in the degree of sophistication can result in cross-subsidies that flow from less sophisticated to more sophisticated market participants (Gabaix and Laibson, 2006; Armstrong and Vickers, 2012). In financial markets, such cross-subsidies can be regressive, and contribute to the growth in inequality of financial wealth if financial sophistication is correlated with wealth and income, a common finding in the household finance literature (Calvet et al., 2009).

Carefully measuring the extent of such cross-subsidies and identifying how they are distributed across households is a challenging task, and one that we take up in this paper. We focus on analyzing and quantifying cross-subsidies in residential mortgage refinancing. Mortgages are typically the largest household financial liability (Campbell, 2006; Badarinza et al., 2016; Gomes et al., 2021), but despite their importance in household budgets, many households do not exhibit the financial sophistication to perform the often complex optimization required to manage this component of their balance sheets. A prominent example is the mortgage refinancing decision, where there is considerable heterogeneity in observed refinancing efficiency across households, with many households inactive despite strong financial incentives to take action (Keys et al., 2016; Andersen et al., 2020).

To undertake a quantitative assessment of cross-subsidies in mortgage refinancing, we build a structural model that we fit to high-quality administrative data on the stock of all outstanding mortgages in the United Kingdom in June 2015. The model facilitates counterfactual analysis, and provides quantitative magnitudes of refinancing cross-subsidies in the U.K mortgage market during this sample period. We use the model to assess how computed cross-subsidies over the sample period are distributed across households of different income levels, and those located in different regions of the U.K.. Our work contributes to the literatures on regional redistribution through the mortgage market (Hurst et al., 2016; Beraja et al., 2019), and helps to uncover how features of the financial system contribute to inequality (Campbell et al., 2019; Greenwald et al., 2021).

In the U.K., as in many other countries, the dominant mortgage form is a "discounted rate" instrument with a relatively short initial fixation period (Badarinza et al., 2018). To fully take advantage of this discounted or "teaser" rate, it is imperative to promptly refinance at the point at which the initial fixation period ends, to avoid being rolled on to a significantly more expensive "reset rate".[1] Households who fail to promptly refinance pay these higher reset rates, and this contribution to lender profits allows in turn for discounted rates to be lower in equilibrium. Households that are swift to refinance take advantage of these lower discounted rates, resulting in the cross-subsidy that we study.

The ideal data to compute cross-subsidies in this setting would contain detailed records of household-level mortgage refinancing behavior, as well as information about aggregate lender revenues. The data that we study, sourced from the Financial Conduct Authority (FCA) of the U.K., are well-suited for our purposes. They comprehensively track individual mortgages in the stock of all outstanding UK mortgage loans issued by all regulated financial institutions in the country at a semi-annual frequency.[2] These data have been used in a range of academic studies, including Benetton (2021); Robles-Garcia (2020); Cloyne et al. (2019); Best et al. (2020); Belgibayeva et al. (2020); Benetton et al. (2021), and track the stock of outstanding mortgages between 2015H1 and 2017H2.

We focus in this draft of the paper on statistics derived from 2015H1.[3] At this date, the total stock of mortgages in our sample equals £513 BN, with mortgage loans on discounted rates accounting for roughly 67% of this sum, with the remaining 33% paying the revert rate. At 2015H1, the average remaining discounted period on discounted loans is 25 months, reflecting the modal initial discounted rate fixation period of 2 years. There is also a significant spread between the (lower) average discounted rate and the (higher) average reset rate of 50bp on an equal-weighted basis, and 60bp when weighted by outstanding loan balance.

While we spend more time in the paper describing the data in detail, the simple statistics outlined above are revealing. While discounted rate mortgages comprise the major portion of outstanding mortgage loans during the sample period, a substantial fraction of mortgage loans pays the reset rate. Moreover, there is a visible and material spread between the rates paid on average between mortgages in these two categories, despite the different cohorts of

[1] This particular feature of the UK mortgage market has prompted calls for reform, prominently as a result of the implicit cross-subsidy, in addition to other undesirable features (Miles, 2004).

[2] In what follows, we denote the first and second observations in each year of our sample by H1 and H2 respectively to denote "half-years".

[3] Consequently, the cross-subsidies and costs of refinancing that we estimate in this paper are restricted to this sample period. As shown in Figure 1, the share of mortgages on reset rates has declined from 2015H1 to 2017H2. This means that drawing inferences about more recent periods in the UK will require further work, a task we intend to take up going forward.

mortgages represented in these broad aggregate statistics. These different rates, combined with the outstanding balances on both categories jointly contribute to lender revenues in what we might call the "cross-subsidy equilibrium".

Our goal is to arrive at a money-metric assessment of cross-subsidies from these and other statistics. To do so, we set up a partial equilibrium model of the UK mortgage market which is geared towards matching the main features of the data, taking observed mortgage rates as given. The model assumes that households are heterogeneous along two dimensions. The first dimension captures differences in households' preferences for owning houses (as opposed to renting them—which we model as an outside option). The second dimension is households' degree of inaction, which we model as a household-specific fixed cost of refinancing, which captures both "rational" costs à la Agarwal et al. (2013), as well as any psychological increment to these refinancing costs such as hassle factors or opportunity costs of time à la Andersen et al. (2020).

We also make several other simplifying assumptions. First, we assume that households can only refinance at pre-determined periods over their total mortgage lifetimes. This is to capture the institutional setting in the U.K. mortgage market, which features pre-determined fixation periods over which discounted rates apply, and high penalties for prepayment before these fixation periods elapse. Second, we assume away cash-out refinancing and loan maturity extensions at the point of refinancing. Third, we assume that all households have the same constant loan-to-value (LTV) ratio at the point of mortgage origination. And finally, we assume that all loans are amortizing, so outstanding balances at any period can be written as a function of the rate paid, and the maturity of the loan.

Under these assumptions, any given household's refinancing behavior can be characterized by a household-specific refinancing "cutoff date" following loan origination. Past this date, the household stops refinancing into the discounted rate on the expiration of the fixation period, and simply pays the reset rate.

The structure of the model facilitates easy aggregation of loans, permitting us to write down intuitive expressions for aggregate mortgage loan balances on the discounted rate and on the reset rate. We use this feature of the model to consider a counterfactual scenario in which all households pay a single rate for the entire duration of their mortgage contract, by assuming equality of revenues (computed as outstanding loan balances times applicable rates) across the dual-rate and single-rate pricing structures. Of course, this equality of aggregate revenues does not guarantee that individual loans will necessarily look similar across the two scenarios. Indeed, the model reveals that there are both intensive margin effects (individual loan sizes increasing or decreasing) and extensive margin effects (new loans originated, or loans not originated at all that were previously in existence), when comparing the two scenarios. The counterfactual single rate that equalizes revenues is the cornerstone of our cross-subsidy calculations, as the basis for comparison with the observed discounted and reset rates.

To quantitatively evaluate cross-subsidies, we take the model to the data for structural estimation. To facilitate this exercise, we make a few additional assumptions in this draft of the paper. First, we assume that the market is in equilibrium in 2015H1. As mentioned earlier, we therefore focus on this single snapshot of the data, mapping estimated data moments to the model-implied moments from the model's steady state.[4] Second, we fix a set of parameters at values observed in the data. Most notably, we simply read the average discounted and reset rates from the 2015H1 snapshot; we set the initial discounted rate fixation period to 2 years to correspond to the modal value seen in the data; and we set the LTV ratio at mortgage origination at 80%, a frequently observed value. Third, we assume that housing valuations are lognormally distributed across households. And finally, we assume that refinancing costs are drawn from a mixture of two lognormals, with different means and variances, to correspond to the dual-rate structure of mortgages—we assume that there are two groups, of "low cost" and "high cost" households in the data.

Applying these assumptions, we first simulate from the steady-state of the model to fit 17 different data moments, adjusting the values of 9 key unobserved parameters to maximize the match between model-implied and observed moments using minimum distance and inverse-variance moment weighting. We achieve a close fit to the moments, and estimate that borrowers' average refinancing costs are £3,144 in the population, with a standard deviation equal to £3,310. These numbers are comparable to, though slightly larger, than the average total psychological plus fixed refinancing cost estimated in Andersen et al. (2020) of roughly £1852, despite the differences in setting.[5]

[4]This assumption implies that we ignore the time-evolution of the relative fractions of mortgages in the two categories in this draft, rather than considering the richer time-dynamics that are evident from our descriptive statistics including changes in the shares of discounted and reset-rate mortgages as well as the spread in rates between them.

[5]The Andersen et al. (2020) model also considers "Calvo-style" refinancing inaction in addition to individual-specific refinancing thresholds. This difference might contribute to the greater estimated size of

We also estimate two extended versions of the model. The first estimates parameters separately for different geographical regions in the U.K., to reflect regional heterogeneity in preferences. The second estimates parameters separately for 12 income groups (bottom-eight income deciles, and the top-two deciles each additionally split into two sub-groups). When we do so, while the resulting models continue to match the aggregate moments very well, we find considerable differences in refinancing costs across regions and income groups, a harbinger of how cross-subsidies vary cross-regionally and across the income distribution.

The estimated parameters allow us to compute the counterfactual single rate through our revenue-equality assumption, and produce magnitudes for U.K. refinancing cross-subsidies by comparison with observed discounted and reset rates. We estimate that the average discounted rate in the stock of mortgages would rise by 20bp towards the counterfactual single rate, while the average reset rate would fall to the same rate by roughly 30bp. Total mortgages increase under the counterfactual, because more mortgages are issued in the single-rate world to high-refinancing-cost households who no longer pay the punitive reset rate, and fewer mortgages are issued to low-refinancing-cost households, who scale back in the face of having to pay the higher single rate rather than the discounted rate. However, the mean initial loan balance falls in the counterfactual equilibrium by 2.30 percent of the baseline average loan size, because low-refinancing-cost households scale back the size of their mortgage loans more aggressively than high-refinancing-cost households increase their mortgage loans.

How are these estimated cross-subsidies distributed across U.K. regions and income groups during the sample period under study? While all regions on average would pay slightly higher interest rates under the counterfactual, there is clear evidence that households in the richer South-West of the country would pay higher rates under the single-rate structure, and households in the relatively poorer North-East and North-West of the country would pay lower mortgage rates under the counterfactual single-rate scenario than they do in reality. These changes to rates are also accompanied by increases in the home-ownership rate, but a slight shrinking of average mortgage debt on the intensive margin, for all regions of the U.K.. A more subtle finding is that average mortgage payments in relatively richer areas of the U.K. shrink, while the reverse is true in relatively poorer areas—an endogenous response that suggests "democratization" of mortgage takeup under the counterfactual single-interest

our thresholds, which capture all heterogeneity in household refinancing behavior.

rate scenario.

Similar patterns emerge across income groups. Interest rates increase on average for all income groups under the counterfactual, as low-refinancing-cost borrowers are present in all groups and dominate the averages. However, while these increases are mild for low-income households, interest rate are substantially higher for high-income groups in the counterfactual, rising by close to 10 basis points per annum for the groups at the very top of the income distribution. This reflects the low estimated refinancing costs that we estimate for the highest income groups. These changes in interest rates translate into lower mortgage debt for all groups, but substantially lower mortgage debt for the very top income groups, who adjust their loan sizes downwards (by close to 4%) in the face of higher lifetime mortgage rates in the new single-rate world. This downward adjustment is reflected in overall lower mortgage payments for the highest income groups, an endogenous response to the more expensive single-rate they pay under the counterfactual. Conversely, mortgage payments mildly increase for lower income households, who take on larger loans to take advantage of the more beneficial rate structure in the counterfactual single-rate world—a world which eliminates costly refinancing. Overall, the picture that emerges shows that cross-subsidies in the current dual-rate structure are regressive.

As mentioned earlier, our work is connected to the literature on mortgage refinancing (Keys et al., 2016; Andersen et al., 2020), which documents and studies many of the patterns that we use in our empirical work, but does not extract quantitative magnitudes of cross-subsidies or undertake counterfactual analysis. It is also connected to the growing literature on the factors contributing to inequality of income and wealth (Alvaredo et al., 2017; Benhabib and Bisin, 2018; Fagereng et al., 2020; Hubmer et al., 2020), and more specifically on the sources of inequality in financial wealth (Campbell et al., 2019; Greenwald et al., 2021). Our structural model contributes to this literature by providing a money-metric assessment of cross-subsidies in an important household financial market. We demonstrate that high-income households are cross-subsidized by low-income households, on account of high-income households' superior ability to navigate the current mortgage market design.

Finally, we contribute to the literature on regional redistribution and its connection with the housing and mortgage markets (Hurst et al., 2016; Beraja et al., 2019)—our work helps to show that such regional redistribution can occur directly as a result of differential efficiency in using financial products, in addition to broader effects of access to finance on household outcomes.

The remainder of this paper is organized as follows. Section 2 describes the rich administrative data that we employ in our analysis. Section 3 sets up the model. Section 4 describes our structural estimation exercise and summarizes the results of the estimation. Section 5 concludes.

2 Data, Institutional Features, and Summary Statistics

Our primary data source is the Financial Conduct Authority (FCA) of the UK, which comprehensively tracks the stock of outstanding UK mortgage loans issued by all regulated financial institutions in the country at a semi-annual frequency. These data have been used in a range of academic studies, including Benetton (2021); Robles-Garcia (2020); Cloyne et al. (2019); Best et al. (2020); Benetton et al. (2021). The specific FCA database that we utilize is the Product Sales Database 007 (or PSD007, in short), which provides information about the stock of mortgage loans between June 2015 (henceforth 2015H1), and December 2017 (2017H2). The database tracks a range of loan-level characteristics for every mortgage in regulated financial institutions' portfolios, in snapshots taken at half-yearly intervals. Regulated financial institutions in the UK are legally required to report these details within 30 working days following the end of each calendar half-year.

The group of regulated financial institutions in the UK includes deposit-taking institutions (including building societies), as well as some non-bank financial institutions. Our sample focuses on the owner-occupier segment of the mortgage borrowing population, and excludes "buy-to-let" mortgages which are issued mainly to landlords on rental properties.

At each reporting date, for each mortgage, the dataset records the outstanding balance, original loan amount, original loan term, remaining term to maturity, current interest rate, currently monthly payment, and performance status, i.e., whether the loan is in arrears and if so, for how long. The database also includes information on the property location at the most granular level in the UK (6-digit postcode), and borrower characteristics such as date of birth and the opening date for the bank account associated with the mortgage. Table A.1 in the online appendix provides more detailed descriptions of the main variables from the PSD 007 data-set used in this paper. In addition to the interest rates on outstanding mortgages, lenders also report the type of interest payments contracted for each mortgage. A large proportion of UK mortgages are issued with "discounted" interest rates which are fixed for a set time period, usually between one and five years, depending on the contract chosen by the borrower. At the end of the fixation period, the mortgage rate automatically rolls over into a higher reset rate known as the "standard variable rate", unless borrowers choose to refinance the mortgage into another discounted period (for a detailed treatment of the characteristics of the UK mortgage market, please see Miles, 2004).[6] As we describe more fully in Section 3, the relative proportion of mortgages in these categories (i.e., discounted versus reset rate) in the mortgage stock, the interest rates paid on these different categories of mortgages is an important moment in the data that helps us to pin down the extent of cross-subsidies in this market.

In this version of the paper, our structural model estimates use data from a single snapshot, namely 2015H1 (we discuss the implications of this choice further below). Moreover, the PSD 007 dataset on the stock of mortgages does not include information on borrower incomes. As we seek to also assess cross-subsidies across income levels in our sample, we merge borrowers in the stock data with loan-level data on borrower characteristics shared with lenders at the time of loan origination. This results in our final sample of 4.15 million mortgages in the 2015H1 snapshot, for which we have estimates of borrower income. Appendix Section A.4 provides further details on the underlying data used to estimate borrower incomes across the stock snapshots.

The mortgage market in the U.K. has experienced a number of changes since 2015H1, the sample period for our study. In particular, there has been a notable shift in the composition of the mortgage stock, with fewer mortgages paying the reset rate, and variations in the spread between discounted and reset rates. This leads to the caveat that the cross-subsidies estimated in this draft of the paper apply to the sample period of our data. Further work is required to draw inferences about more recent periods in the UK, a task we intend to take up going forward.

[6] There is a third type of interest rate known as a tracker rate, paid on around 15% of all mortgages outstanding, which is a floating rate linked to the Bank of England base rate. We exclude such mortgages from our analysis since such mortgages share properties with both discounted rate (such as fixation periods) and reset rate (mortgages may reset to tracker rates following discounted periods) mortgages. Online appendix Figure A.1 shows the proportion of mortgages paying tracker rates over time.

Before proceeding further, we note that a number of factors could explain the drop in the number of mortgages under reversion rate since 2015H1. First, as we report in the data, there has been an increase in the spread between average revert rates and discounted rates from 2015H1. Second, as reported in Financial Conduct Authority (2019a), there has been an increase in lenders' focus on retaining existing customers through internal switching, and an increased role of intermediaries in prompting borrowers to undertake beneficial switches.[7]

We also note that the results of our study have broader applicability to countries around the world. Badarinza et al. (2018) (Table 1) provide information on mortgage interest-rate fixation periods across a broad set of countries, which shows that many large economies are similar in their average mortgage-rate fixation period to the U.K.. More specifically, the mortgage form that we study in the U.K., with discounted rates that subsequently move to revert rates following the expiration of the initial fixation periods is similar to that present in many other countries, including Ireland, Australia, India, and Spain.

Table 1 shows summary statistics for selected variables in the merged 2015H1 snapshot, which as mentioned above, tracks 4.15 million mortgages reported as of this date, comprising the (cleaned and filtered) stock of outstanding loans that are either on a discounted rate or paying the reset rate. On average, these mortgages have an outstanding balance of £123,325 pounds (amortized down from an initial average loan balance of £135,542). This amounts to a total stock of mortgages of £513 BN outstanding in the final filtered dataset on which we conduct our analysis.[8]

Taking a simple equal-weighted cross-sectional average across all mortgages, Table 1 shows that they pay an average interest rate of 3.52% at the end of 2015H1, at a spread of 2.87% over the yield on a Bank of England bond with an equivalent maturity rate, and

[7]We direct interested readers to more recent changes in the UK mortgage market aimed at facilitating switching at the time of refinancing. For instance, Financial Conduct Authority (2020a) reflects on increased use of technology and other remedies to facilitate switching; and recent policies have made it easier for for financial groups to switch customers from a group's closed book or lender to an active one (Financial Conduct Authority (2020b), with the objective to make intra-group switching easier), and modified affordability assessments while refinancing for borrowers with up-to-date payments (Financial Conduct Authority (2019b)).

[8]As discussed earlier, we do not consider the tracker mortgages in the data, which account for between 14-18% of the total outstanding balance (this shrinks over the sample period to 14% in 2017H2); or buy-to-let mortgages, which account for around 20% of the mortgage stock. The value of the stock of outstanding mortgages also grows over time: In 2017H1, the total stock of mortgages in the filtered dataset amounts to £657 BN. This means that in June 2017, our data cover 51.5% of the total stock of UK mortgages outstanding. The overall size of the UK mortgage market from 2015-2017 can be accessed here.

Table 1: Summary Statistics for the Mortgage Stock in 2015H1

	mean	sd	p10	p25	p50	p75	p90
Balance (GBP)	123,325	98,092	38,770	64,821	101,620	152,765	223,988
Interest rate (in pp)	3.52	1.00	2.39	2.58	3.49	4.14	4.78
Spread to T-bill (in pp)	2.87	1.07	1.68	2.05	2.64	3.55	4.29
Original size (GBP)	135,542	100,123	50,595	76,000	112,625	164,795	237,500
Orig. term (in months)	274	89	146	216	300	324	396
Rem. term (in months)	222	96	93	154	217	288	352
Rem. discounted period	25	18	5	12	22	37	51
Borrower age	42.78	10.43	30.00	35.00	42.00	50.00	57.00

The table above shows summary statistics of mortgages from the stock data reported in 2015H1. The sample includes mortgages in two categories, namely, those paying discounted interest rates, and those paying the Standard Variable Rate. The total sample comprises around 4.15 million mortgages, of which 60.7% are discounted rate mortgages at this point in time. Appendix Table A.1 contains a description of the underlying variables.

have a remaining term to maturity of 222 months, or around 18.5 years on average.[9] 60.7% of the 4.15 million mortgages in our final dataset pay discounted rates in this snapshot, with an average equal-weighted remaining discounted period of 25 months. In terms of the limited demographic characteristics that we have available, the average borrower age in 2015H1 is around 43 years.

Table 1 also reveals considerable cross-sectional variation in these variables. The remaining discounted rate period ranges from 5 months at the 10th percentile and 51 months at the 90th percentile, which affects the rate of transition between categories (discounted and reset rate) in any given time interval. The remaining mortgage term and the outstanding loan balance also exhibit considerable cross-sectional variation. When both remaining term and outstanding loan balance are low, this rationally reduces incentives for borrowers to refinance given the lower financial incentive from any interest rate reduction associated with doing so (Agarwal et al., 2013). The mortgages also vary considerably in terms of the overall interest rate they pay, as well as the spread over the maturity-matched government rate, a necessary condition for the presence of sizeable cross-subsidies in the market. There is also demographic variation in the mortgage stock captured in the data,

[9]Mortgage spreads are computed with respect to the yield on a nominal zero coupon UK government bond with maturity matched to the mortgage interest rate fixation period. We use the short-term interest rate for mortgages paying the reset rate. For instance, for a mortgage with t years of fixation, the spread is calculated by subtracting off the spot rate for a UK government bond maturing in t years as at the reporting date.

as seen in the age of borrowers, with both relatively young borrowers, aged 30 at the 10th percentile of the cross-sectional distribution, and older borrowers, aged 57 at the 90th percentile of the cross-sectional distribution.

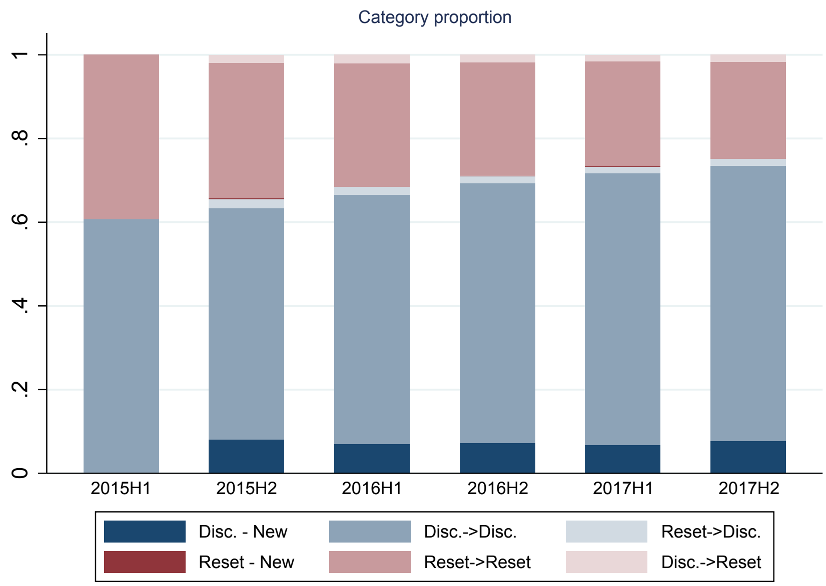
Table 1 shows summary statistics of a single snapshot of the data; we observe a total of 6 such snapshots starting in 2015H1 and up until 2017H2. While our current draft of the paper does not use this information to compute cross-subsidies, by computing differences across these snapshots, we can track how mortgages move across categories. As we describe more clearly when we present the model, our current approach is to assume that the market is in steady-state in 2015H1, and to map the steady-state equilibrium of the model to the data under this assumption. This assumption equates to assuming that the fractions of mortgages on discounted and revert rates that we observe in 2015H1 will remain constant, i.e., that flows of mortgages into, out of, and between categories preserve these observed shares. As we will see below, this assumption is somewhat at variance with the data; in future iterations of this work we will consider the evolution of the mortgage stock over the entire sample.

Across all 6 snapshots, the data track 6.00 million unique mortgages. Over the entire sample period, we can observe whether a borrower continues to pay a discounted rate or the reset rate, moves between these two categories, initiates, or fully repays a mortgage.[10] Appendix Figure A.2 shows the number of mortgages originated and discontinued between consecutive reporting periods. For instance, the figures for 2015H2 show the mortgages originated/discontinued since the mortgage cohort in 2015H1. In our sample, on average, around 345,000 mortgages are originated, and around 178,000 mortgage accounts are discontinued in the UK every 6 months. Figure 1 shows how these originations, discontinuations, and refinancings affect the proportion of mortgages under the two categories (discounted and reset rate) over time.[11] From 2015H1 to 2017H2, the fraction of all outstanding mortgages on discounted rates rises from 60.7% to over 75.2%. This trend is mainly driven by new issuance of discounted rate mortgages over the two years of our sample, and a countervailing decrease in the share of mortgages on the reset rate. As we discuss below, this trend is

[10] The data record the precise property location (6-digit postcodes, which are very granular in the UK, associated in most cases with a single street) and the borrower date of birth reported for each mortgage. This permits unique identification of a given loan, and allows us to track each loan through time, across snapshots.

[11] As our data begin in 2015H1, this decomposition is not available for this year.

Figure 1: Proportion of Mortgages under Discounted and Reset Rates



The figure above shows the proportion of mortgages under discounted rates (dark blue) and the reset rate from the mortgage stock as reported at a half-yearly period from 2015H1 to 2017H2. The proportion of mortgages that are new to a snapshot are shown using a darker shade; and the proportion of mortgages that cross categories across snapshots are shown in a lighter shade.

accompanied by an increase in the difference between the two rates over the same time period.

Figure 2 (a) shows the proportions of mortgages in different categories as a share of the total loan balance outstanding in GBP (rather than the number of loans). Discounted mortgages account for about 67% of the stock of value-weighted mortgage loans outstanding at the beginning of the sample, a fraction that increases to around 83% by the end of the sample.[12] The figure shows that there is considerable persistence in these category identities over time, with a very small fraction moving from discounted rates to reset rates and vice versa at any given point in time. The proportions of newly issued mortgages in both discounted and reset rate categories (labelled "Disc.-New" and "Reset-New", respectively)

[12]The proportion of mortgages in our database under discounted rates is somewhat overstated for the 2017H2 snapshot. As described in Appendix Section B.2, two large lenders report anomalously large loan balances in the 2017H2 snapshot for mortgages under discount rates. We restate the 2017H2 loan balance of the discount rate mortgages issued by these lenders for which the balance for 2017H1 is available. The anomalous loan balance for mortgages for which the 2017H1 loan balance is not available, i.e. discount rate mortgages issued by these lenders between 2017H1 and 2017H2, will be addressed in subsequent versions of this paper. However, the loan balances in 2017H2 have no bearing on the cross-subsidies reported in Section 4 which are estimated using data moments from the 2015H1 snapshot.

also remain roughly constant through time, though, as expected, almost all newly issued mortgages are discounted mortgages.

The persistence of mortgages in the reset rate category naturally raises questions about whether mortgage refinancing is unconstrained in the U.K.. In some markets, such as the U.S., a credit check is triggered at the point of refinancing (Keys et al., 2016), whereas in others, such as Denmark, even delinquent borrowers are able to refinance as long as there is no cash out (Andersen et al., 2020). We rely on an FCA study of the mortgage market conducted in 2018 (Financial Conduct Authority, 2019a), which studied 2 million reset rate mortgages using the same data that we employ, and concluded that roughly 30,000 of these mortgages were unable to switch despite being up to date with payments. The study goes on to report that two-thirds of these mortgages were associated with an inactive, failed lender (e.g., Northern Rock, famously subject to a run during the financial crisis); and the remainder were either interest-only mortgages that were subject to changes in lending standards following the financial crisis, or in negative home equity. Additionally, in the report, lenders claim that they do not carry out new credit or affordability checks on existing customers, suggesting that internal remortgaging is available even to those borrowers suffering a deterioration in their credit circumstances. On the face of it, this suggests that involuntary non-refinancing is not the principal reason for mortgages being on the reset rate, and indeed, therefore, for any cross-subsidies estimated in the data.[13]

Returning to Figure 2 (a), the main visible trend in the plot, as with the plot showing the shares of the number of loans, is an increase in the relative share of total loan balances on discounted rate mortgages that are refinanced within that category (or simply not discontinued) from prior snapshots, and a relative decrease in the share of reset rate mortgages.

Figure 2 (b) shows loan balance-weighted interest rates across categories, as well as for all outstanding loans in the sample. The average mortgage rate first drops and then rises slightly towards the end of the sample period. This trend is most pronounced in interest rates on newly issued discounted mortgages. The spread between the reset rate and discounted rates is an essential component of the cross-subsidy that we seek to estimate structurally in

[13]While this is reassuring, in future versions of this paper, we intend to conduct robustness checks in which we exclude borrowers with interest-only mortgages, those with negative home equity, and those experiencing payment shortfalls to ascertain the extent to which our cross-subsidy estimates change with these additional data filters.

Section 4, using the model described in Section 3, and an increase in the spread is consistent with a larger proportion of borrowers refinancing into discounted mortgages, all else equal.

Table 2 shows the averages of selected variables in the dataset across snapshots from 2015H1 to 2017H2. The average size of discounted rate loans has risen steadily over time, while the average size of loans on the reset rate has decreased. This is consistent with a change in refinancing incentives over the period as explained above, and an increase in cash-out refinancing as shown in the change in overall loan balance from discounted-to-discounted rate flows in Figure A.7. The average remaining term on discounted rate mortgages rises from around 20.1 to 20.4 years (21.6 to 22.5 value-weighted), while the average remaining term on reset rate loans decreases through the sample period, from around 16.1 years to 14.1 years (16.9 to 15.3 value-weighted). The average remaining discounted period on discounted loans is 23 to 25 months in all snapshots of the data, reflecting the modal discounted period of 2 years observed in the data.[14] Finally, we observe an increase in the average interest rate gap between loans on reset rate and discounted rates over time, from 50bp (60bp value-weighted) in 2015H1, to 98bp (107bp value-weighted) in 2017H2. In all sample periods, the loan-balance weighted rate spread is higher than the equal-weighted rate spread. This effect stems mainly from larger discounted rate mortgages having lower rates on average—consistent with wealth-based heterogeneity in mortgage refinancing efficiency detected by Andersen et al. (2020).

To reiterate, the documented time-evolution in the stock of mortgages is not a feature that we currently consider in our structural assessment of cross-subsidies, which focuses on data from 2015H1. The net effect of the current approach on the computation of cross-subsidies is somewhat ambiguous. While revert-rate mortgage loans comprise a relatively lower share of all mortgage loans towards the end of the sample, the spread between revert rates and and discounted rates also increases over this period, which means that the total effect on lender revenues is not entirely clear. We intend to extend our model-based analysis of these interesting trends in the data in future drafts of the paper.

To assess regional variation in computed cross-subsidies, we complement our administrative data with demographic variables for local authorities in the United Kingdom from the 2011 census. These variables describe socioeconomic dimensions along which policy

[14]Mortgages under the discounted period are essentially fixed-rate loans. At origination, as shown in Figure A.5, the most common discounted period is 2 years, followed by 5-year fixed-rate loans.

Table 2: Summary Statistics over Mortgage Snapshots

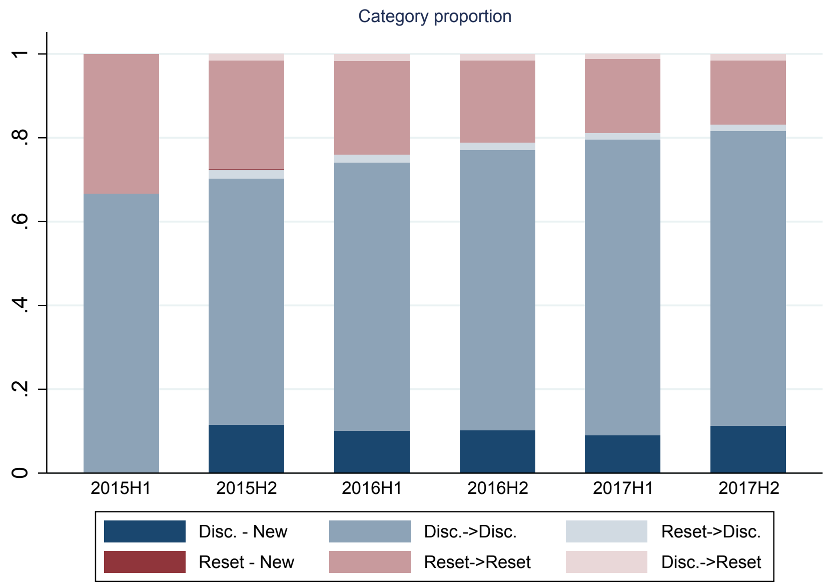
Cohorts	2015H1	2015H2	2016H1	2016H2	2017H1	2017H2
Average loan size in £						
Teaser	135620	140685	144411	148361	149805	155343
SVR	104364	101894	99199	97709	95672	95368
Average remaining term in months						
Discounted	241	242	243	244	244	245
Reset	193	187	181	178	174	169
Average remaining term (value-weighted) in months						
Discounted	259	262	264	267	268	270
Reset	203	198	193	190	186	184
Average remaining teaser period in months						
Discounted	25	25	25	24	24	24
Average remaining teaser period (value-weighted) in months						
Discounted	25	25	25	24	23	24
Average interest rate						
Discounted	3.33	3.18	3.06	2.93	2.79	2.88
Reset	3.83	3.79	3.78	3.53	3.56	3.86
Average interest rate (value-weighted)						
Discounted	3.21	3.10	3.02	2.91	2.76	2.87
Reset	3.81	3.77	3.75	3.52	3.55	3.94
Average borrower age						
Discounted	41	41	41	41	41	41
Reset	45	46	46	47	47	47

The table above share summary statistics of mortgages for the stock snapshots from 2015H1 to 2017H2. The sample includes mortgages under two categories - those under discounted rates, and under the reset rate. Please see Appendix Table A.1 for a description of the underlying variables.

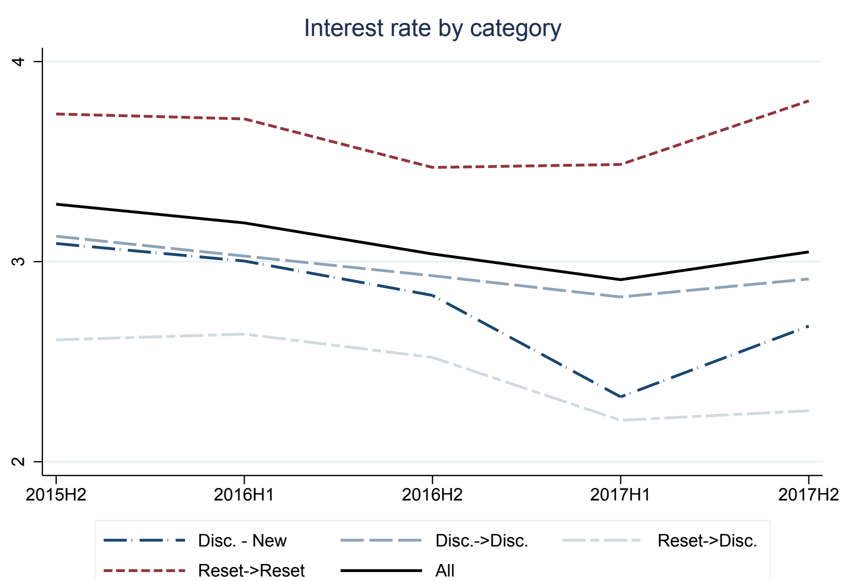
makers are concerned about equitable economic opportunities and outcomes, such as ethnic diversity, levels of deprivation, and long term economic activity. Appendix C provides further details on these data. In the next section, we turn to describing the model that we use to guide our structural estimation.

Figure 2: Proportion and Average Interest Rate of Mortgages under Discounted and Reset Rates

(a) Based on balance



(b) Value-weighted interest rates



Panel A in the figure above shows the proportion of mortgages under discounted/reset rates over mortgage stock cohorts based on the remaining balance. Panel B shows the value weighted interest rate for the substantial categories of mortgages under discounted/reset rates in the mortgage stock snapshots reported at a half-yearly period from 2015H1 to 2017H2.

3 Model

The summary statistics show that the UK mortgage market is a mix of mortgages on discounted rates and those paying the SVR. While the mortgages on discounted rates constitute the major share, there is still a considerable fraction (20-40% depending on the snapshot; 39.3% in 2015H1—the sample period that we study in this draft) of the outstanding mortgage stock paying the reset rate. Moreover, the loan-weighted reset rate in the data is on average over 100bp higher than discounted rates. To quantitatively assess the magnitude of the cross-subsidy that this dual-rate structure embeds, we develop a model in this section that we later map to the data for structural estimation.

We model a mortgage market in which a measure M of households enters in every period. Households are heterogeneous in their per-period valuation for housing v (to help us capture the diversity of initially chosen loan sizes seen in the data) and in their fixed cost k of refinancing (to capture the different types of mortgages taken by households in the data), which are distributed according to the cumulative joint distribution function $G(v, k)$ with density $g(v, k)$. All households discount the future at the common rate β.

Mortgages are long-term contracts for T periods that pay a discounted rate r for an initial period T_d, and a reset rate $R > r$ after this initial period in the event that the household does not refinance. To simplify and facilitate evaluating counterfactuals, our model is partial equilibrium, in the sense that we take both rates as given (and in our structural exercise, map them to the observed average rates). As we describe later, we estimate structural parameters under this assumption, and then evaluate cross-subsidies under a counterfactual single interest rate scenario.

We also assume that T/T_d is a (positive) integer and that households can only refinance at the point at which the discounted rate expires. Moreover, we assume that households do not change their loan balance (i.e., we currently rule out "cash-out refinancing,"), and that they do not change the final maturity of their loan at the point of refinancing. As a result of these assumptions, each household makes T payments over the life of the loan, the same as the duration of the mortgage contract.

Households choose the size of their loan at time $t = 0$. After the expiration of the discounted rate, they choose whether or not to refinance the loan. We further assume that households are disbursed the loan amount at time $t = 0$, but make the first payment at

$t = 1$. Hence, the first refinancing period is T_d, and the mortgage is fully repaid at T.

Households' flow utility equals $vh_0^\alpha/\alpha - m(l_0, r, T)$, where h_0 is the size of the house that the household picks, $0 < \alpha < 1$ is a parameter governing utility from housing, and $m(l_0, r, T)$ is the per-period mortgage payment of a household with a mortgage with initial loan balance l_0, interest rate r, and total number of payments T. We assume that l_0 is proportional to h_0, i.e., $l_0 = \frac{h_0}{\omega}$ where $1/\omega$ is the loan-to-value ratio of the mortgage, which we simply assume to be constant across households.

The periodic mortgage payment $m(l, r, T)$ follows from the amortization of the loan; given a loan of size l_0, applying the standard mortgage payment formulas, each payment equals:

$$m(l_0, r, T) = l_0 \frac{r(1+r)^T}{(1+r)^T - 1}. \tag{1}$$

We now solve the model to determine two household choices: 1) whether or not to refinance at each opportunity, i.e., in every T_d periods; and 2) the optimal size of the initial loan $l_0(v, k)$.

3.1 Optimal Refinancing

In order to determine households' refinancing behavior, we need to keep track of the evolution of the loan balance. For reasons that will become clear below, because the loan is amortizing, the incentives to refinance decline over time as the outstanding balance decreases.

In the model, each household will have a cutoff date T_{max}, which is heterogeneous across households (determined by its optimal loan size choice and its (unobserved) refinancing cost k), and determines the last period at which it pays the discounted rate r. This date fully defines a household's refinancing behavior—as it will always refinance before T_{max}, and will never refinance at opportunities arising after T_{max}, i.e., it will pay the reset rate thereafter. Some households may refinance at every opportunity that arises, for such households, $T_{max} = T$.

A household that refinances up until period T_{max} makes payments equal to $m(l_0, r, T)$ from $t = 1$ until $t = T_{max}$, and its end-of-period loan balance at each $t \in \{1, \dots, T_{max}\}$

equals:

$$\begin{aligned} l_t(r, l_0) &= l_0(1+r)^t - m(l_0, r, T)\frac{(1+r)^t - 1}{r} \\ &= l_0\frac{(1+r)^T - (1+r)^t}{(1+r)^T - 1}. \end{aligned}$$

The beginning-of-period loan balance simply equals $l_{t-1}(r, l_0)$. It is easy to verify that $m(l_0, r, T) = m(l_t(r, l_0), r, T - t)$.

In period $T_{max} + 1$, the household doesn't refinance into another discounted mortgage, and therefore pays the rate R thereafter until the maturity of the loan. At this point, the household has a beginning-of-period loan balance equal to $l_{T_{max}}(r, l_0)$. Hence, again applying the standard mortgage payment formulas, the payment and end-of-period loan balance for each $t \in \{T_{max} + 1, \ldots, T\}$ equal, respectively:

$$\begin{aligned} m(l_{T_{max}}(r, l_0), R, T - T_{max}) &= l_{T_{max}}(r, l_0)\frac{R(1+R)^{T-T_{max}}}{(1+R)^{T-T_{max}} - 1}, \\ l_t(R, l_{T_{max}}(r, l_0)) &= l_{T_{max}}(r, l_0)\frac{(1+R)^{T-T_{max}} - (1+R)^{t-T_{max}}}{(1+R)^{T-T_{max}} - 1}, \end{aligned} \tag{2}$$

and the beginning-of-period loan balance equals $l_{t-1}(R, l_{T_{max}}(r, l_0))$.

For a given loan size l_0, the household's value function at origination equals:

$$\begin{aligned} V(v, k, l_0, T_{max}) = \max_{T_{max}} &\sum_{t=0}^{t=+\infty} \beta^t v\,(\omega l_0)^\alpha/\alpha - \sum_{t=1}^{t=T_{max}} \beta^t m(l_0, r, T) - k \sum_{t=1}^{T_{max}/T_d+1} \beta^{tT_d+1} \\ &- \sum_{t=T_{max}+1}^{t=T} \beta^t m(l_{T_{max}}(r, l_0), R, T - T_{max}). \end{aligned} \tag{3}$$

The first term is the household's lifetime utility from the house; the second term is the sum of mortgage payments on the discounted rate r; the third term collects the costs k paid by the household at each refinancing event; and the last term is the sum of mortgage payments on the reset rate R. Note that T_{max} varies across households because both k and v are heterogeneous; it also follows from these assumptions that l_0 varies across households.

We solve for the optimal refinancing path (holding loan size fixed at l_0) by backward induction. Consider period $T - T_d + 1$, which is the last refinancing period, and households who have always refinanced up until this point, i.e., households for which $T_{max} \geq T - T_d$.

At this point, households with $T_{max} = T$ will refinance and households with $T_{max} = T - T_d$ will not—i.e., a household refinances if $V(v, k, l_0, T) \geq V(v, k, l_0, T - T_d)$. This allows us to define households who are indifferent between refinancing in period $T - T_d + 1$ (the last refinancing opportunity prior to period T) and not refinancing at this point. We denote such households as type $k^*(T)$, and they satisfy the condition:

$$V(v, k^*(T), l_0, T) = V(v, k^*(T), l_0, T - T_d).$$

Given the value function in equation (3), we can solve for $k^*(T)$, which equals:

$$k^*(T) = (m(l_{T-T_d}(r, l_0), R, T_d) - m(l_0, r, T)) \sum_{t=0}^{t=T_d-1} \beta^t.$$

This means that all households with $k \leq k^*(T)$ refinance at $T - T_d + 1$.

Similarly, in previous refinancing periods $T - sT_d + 1$ for integer $s \in \{2, \ldots, T/T_d - 1\}$, $k^*(T - (s-1)T_d)$ will satisfy the indifference condition:

$$V(v, k^*(T-(s-1)T_d), l_0, T_{max} = T-(s-1)T_d) = V(v, k^*(T-(s-1)T_d), l_0, T_{max} = T-sT_d), \tag{4}$$

which corresponds to:

$$\begin{aligned} k^*(T-(s-1)T_d) = {} & (m(l_r(T-sT_d), R, sT_d) - m(l_0, r, T)) \sum_{t=0}^{t=T_d-1} \beta^t + \\ & (m(l_r(T-sT_d), R, sT_d) - m(l_r(T-(s-1)T_d), R, (s-1)T_d)) \sum_{t=T_d}^{t=sT_d-1} \beta^t. \end{aligned}$$

And once again, households with $k \leq k^*(T-(s-1)T_d)$ refinance at $T - sT_d + 1$.

3.2 Optimal Loan Size

Households choose the loan size that maximizes their value function at origination, equation (3), given their v and k. From the first-order-condition, their optimal loan size choice satisfies:

$$\sum_{t=0}^{t=+\infty} \beta^t v\omega(\omega l_0)^{\alpha-1} - \sum_{t=1}^{t=T_{max}} \beta^t \frac{\partial m(l_0, r, T)}{\partial l_0} - \sum_{t=T_{max}+1}^{t=T} \beta^t \frac{\partial m(l_{T_{max}}(r, l_0), R, T - T_{max})}{\partial l_0} = 0.$$

From equations (1) and (2), we obtain:

$$\begin{aligned}
\frac{\partial m(l_0, r, T)}{\partial l_0} &= \frac{m(l_0, r, T)}{l_0} = \frac{r(1+r)^T}{(1+r)^T - 1} \equiv \lambda_r(T), \\
\frac{\partial m(l_{T_{max}}(r, l_0), R, T - T_{max})}{\partial l_0} &= \frac{m(l_{T_{max}}(r, l_0), R, T - T_{max})}{l_0} \\
&= \frac{l_{T_{max}}(r, l_0)}{l_0} \frac{R(1+R)^{T-T_{max}}}{(1+R)^{T-T_{max}} - 1} \\
&= \frac{(1+r)^T - (1+r)^{T_{max}}}{(1+r)^T - 1} \frac{R(1+R)^{T-T_{max}}}{(1+R)^{T-T_{max}} - 1} \\
&\equiv \lambda_R(T_{max}, T).
\end{aligned}$$

Hence, the optimal loan size satisfies:

$$\begin{aligned}
l_0(v, k) &= \frac{1}{\omega} \left(\frac{\sum_{t=1}^{t=T_{max}} \beta^t \lambda_r(T) + \sum_{t=T_{max}+1}^{t=T} \beta^t \lambda_R(T_{max}, T)}{\omega v \sum_{t=0}^{t=+\infty} \beta^t} \right)^{\frac{1}{\alpha-1}} \\
&= \frac{1}{\omega} \left(\frac{(1-\beta)}{\omega v} \left(\lambda_r(T) \sum_{t=1}^{t=T_{max}} \beta^t + \lambda_R(T_{max}, T) \sum_{t=T_{max}+1}^{t=T} \beta^t \right) \right)^{\frac{1}{\alpha-1}}. \qquad (5)
\end{aligned}$$

The optimal loan size equation is revealing. It shows that a household's loan size choice depends directly on its valuation for housing v, and indirectly on its refinancing cost k, through the optimal refinancing strategy defined by the household's T_{max}.

The refinancing cost determines the extent to which mortgage payments are made on the higher reset rate rather than the lower discounted rate. When households have higher k, they pay the reset rate for longer, in contrast with households that have lower refinancing costs k, who can access the discounted rate for longer. This is because obtaining the more beneficial discounted rate for a longer period of time requires incurring k across a greater number of refinancing opportunities. Anticipating this tradeoff, households with higher k will scale back the size of the loans that they initially take. We return to this issue in greater detail when evaluating counterfactuals.

Given the optimal loan size, we can define $v^*(k)$ as the valuation for housing of a household that is indifferent between getting a mortgage or not getting one: $V(v^*, k, l_0(v^*, k), T_{max}) = \frac{\bar{u}}{1-\beta}$, where $\bar{u}$ is a per-period outside (rental) option, which we assume is common to all households and fixed over time.

3.3 Aggregation: Mortgage Stocks in Steady-State

We can now calculate the total stock of mortgages paying the discounted rate and the reset rate, by assuming that the economy is in steady state.

To do so, it is useful to define:

$$\begin{aligned}
\gamma_r(t) &= \frac{l_t(r, l_0)}{l_0} = \frac{(1+r)^T - (1+r)^t}{(1+r)^T - 1}, \\
\gamma_R(t, T_{max}) &= \frac{l_t(R, l_{T_{max}}(t, l_0))}{l_0} \\
&= \frac{l_{T_{max}}(r, l_0)}{l_0} \frac{(1+R)^{T-T_{max}} - (1+R)^{t-T_{max}}}{(1+R)^{T-T_{max}} - 1} \\
&= \gamma_r(T_{max}) \frac{(1+R)^{T-T_{max}} - (1+R)^{t-T_{max}}}{(1+R)^{T-T_{max}} - 1}.
\end{aligned}$$

to be the end-of-period t share of the initial loan remaining to be repaid on the discounted interest rate r and on the reset rate R, respectively.[15] The household pays the reset rate after period T_{max}, the last period in which it refinances. As a result, T_{max} can be used to determine the remaining loan balance, the mortgage payment amount, and the cumulative share of the initial loan repaid by any period t.

We now define three groups of mortgages. The first group, group 0, comprises the mortgages of households who got a mortgage of initial size $l_0(v, k)$ and are on their initial discount period. The aggregate number $N_0(r)$ and aggregate balance $Q_0(r)$ of mortgages of this group equal:

$$N_0(r) = M T_d \int_{-\infty}^{+\infty} \int_{v^*(k)}^{+\infty} dG(v, k), \tag{6}$$

$$Q_0(r) = M \sum_{t=1}^{t=T_d} \gamma_r(t-1) \int_{-\infty}^{+\infty} \int_{v^*(k)}^{+\infty} l_0(v, k) dG(v, k), \tag{7}$$

[15] Recall that these implicitly define "paths" since households either pay the discounted rate or the reset rate and don't switch back from one to the other in this model.

where we use the beginning-of-the-period share $\gamma_r(t-1)$ to account for the loan balance before payment, and $l_0(v,k)$ follows from (5).[16]

The second group comprises the mortgages of all households who refinanced and pay the discounted rate. The aggregate number $N_1(r)$ and aggregate balance $Q_1(r)$ of mortgages of this group equal:[17]

$$N_1(r) = MT_d \sum_{s=1}^{T/T_d-1} \int_{-\infty}^{k^*((s+1)T_d)} \int_{v^*(k)}^{+\infty} dG(v,k), \tag{8}$$

$$Q_1(r) = M \sum_{s=1}^{T/T_d-1} \sum_{t=sT_d+1}^{(s+1)T_d} \gamma_r(t-1) \int_{-\infty}^{k^*((s+1)T_d)} \int_{v^*(k)}^{+\infty} l_0(v,k)dG(v,k). \tag{9}$$

The third group comprises the mortgages of all households who did not refinance, and pay the reset rate. The aggregate number $N_2(R)$ and aggregate balance $Q_2(R)$ of mortgages of this group equal:

$$N_2(R) = MT_d \sum_{s=1}^{T/T_d-1} \int_{k^*((s+1)T_d)}^{+\infty} \int_{v^*(k)}^{+\infty} dG(v,k), \tag{10}$$

$$Q_2(R) = M \sum_{s=1}^{T/T_d-1} \sum_{t=sT_d+1}^{(s+1)T_d} \gamma_R(t-1, sT_d) \int_{k^*((s+1)T_d)}^{+\infty} \int_{v^*(k)}^{+\infty} l_0(v,k)dG(v,k). \tag{11}$$

The above expressions can be directly mapped to the empirically observed stock of mortgages in each category, under the assumption that the market is in steady-state. We next turn to describing our approach to computing cross-subsidies using this model.

3.4 Cross-Subsidy

To calculate the cross-subsidy across different households, our approach is to consider a benchmark case in which all mortgages have a constant interest rate r_f for their entire

[16] To gain intuition for equation (6), recall that a mass M of households enters the market in each time period. The fraction of them getting (discounted-rate) mortgages is given by the integrals, which in this case just conditions on them satisfying the extensive margin condition, i.e., $v > v^*(k)$ (the other integral integrates across the entire k distribution, irrelevant at the point of mortgage issuance since by assumption all households are initially issued discounted rates). Such households pay the discounted rate for T_d periods, so there are T_d such cohorts represented in the steady-state stock. Equation (7) follows by weighting these mortgages by their initial loan sizes and amortization.

[17] Note that here, we are counting in steps of refinancing opportunities, and integrating over the set of households who take these opportunities at each point, which we can read from equation (4).

duration, a rate that yields the same revenue as the composite of the populations on the discounted rate and the reset rate.

Specifically, in the model, aggregate lender revenue from all mortgages (on both discounted and reset rates) equals:

$$r(Q_0(r) + Q_1(r)) + RQ_2(R). \tag{12}$$

Under a constant interest rate r_f, households do not need to refinance. Hence, their optimal loan size would equal $l_0(v,k)$ in (5) evaluated at $k = 0$, which implies that $T_{max} = T$. Therefore, the aggregate number $N(r_f)$ and aggregate balance $Q(r_f)$ of mortgages in this scenario will equal:

$$\begin{aligned}
N(r_f) &= MT \int_{-\infty}^{+\infty} \int_{v^{**}(r_f)}^{+\infty} dG(v,k),\\
Q(r_f) &= M \sum_{t=1}^{T} \gamma_{r_f}(t-1) \int_{-\infty}^{+\infty} \int_{v^{**}(r_f)}^{+\infty} l_0(v, k=0) dG(v,k),
\end{aligned}$$

where $\gamma_{r_f}(t-1)$ is the beginning-of-period-t share of the initial loan to be repaid and $v^{**}(r_f)$ is the household in this constant rate scenario that is indifferent between getting a mortgage or not, i.e.:

$$V(v^{**}, k=0, l_0(v^{**}, k=0), T_{max} = T) = \frac{\bar{u}}{1-\beta}.$$

Under the assumption of aggregate lender revenues remaining constant across the two scenarios, the interest rate r_f must satisfy:

$$r_f Q(r_f) = r(Q_0(r) + Q_1(r)) + RQ_2(R). \tag{13}$$

Based on this counterfactual r_f, the estimated parameters of the model, and the observed discounted rate r and reset rate R, we can calculate the difference in loan size between the current and counterfactual scenarios for each household (v,k), as well as the difference in mortgage payments and a measure of the lifetime cross-subsidy paid or received by the household. This household-level calculation can be aggregated up at the group level using the baseline model, or indeed, in an extended version of the model in which we estimate group-specific parameters, as we describe next.

3.5 Multiple Groups

The richness of our data allows us to calculate subsidies across different groups based on observable characteristics. While we do have access to a small set of granular demographic characteristics such as age in the mortgage dataset, we focus mainly on two different household groupings in this draft. The first groups households by income, and the second looks at households located in different regions of the UK.

Understanding variation in the extent of cross-subsidies paid or received along the income distribution helps us to understand how the design of the financial system contributes to the inequality of financial wealth, a significant theme of current economic research (Campbell et al., 2019; Greenwald et al., 2021). We also look at the extent of regional variation in mortgage cross-subsidies given the importance of regional re-distribution through the mortgage market highlighted, for example, in Hurst et al. (2016); Beraja et al. (2019).

We therefore extend the model to accommodate and interpret such rich heterogeneity. To begin with, consider different groups based on observable characteristics and indexed by $j = 1, ..., J$. Let M_j and $G_j(v, k)$ be the measure and the distribution of household preferences v and cost k in group j, respectively.

Following the analysis of previous subsections, we can define the following variables:

$$N_{0j}(r) = M_j T_d \int_{-\infty}^{+\infty} \int_{v_j^*(k,r,R)}^{+\infty} dG_j(v,k), \tag{14}$$

$$Q_{0j}(r) = M_j \sum_{t=1}^{t=T_d} \gamma_r(t-1) \int_{-\infty}^{+\infty} \int_{v_j^*(k,r,R)}^{+\infty} l_0(v,k) dG_j(v,k), \tag{15}$$

$$N_{1j}(r) = M_j T_d \sum_{s=1}^{T/T_d-1} \int_{-\infty}^{k_j^*((s+1)T_d)} \int_{v_j^*(k,r,R)}^{+\infty} dG_j(v,k), \tag{16}$$

$$Q_{1j}(r) = M_j \sum_{s=1}^{T/T_d-1} \sum_{t=sT_d+1}^{(s+1)T_d} \gamma_r(t-1) \int_{-\infty}^{k_j^*((s+1)T_d)} \int_{v_j^*(k,r,R)}^{+\infty} l_0(v,k) dG_j(v,k), \tag{17}$$

$$N_{2j}(R) = M_j T_d \sum_{s=1}^{T/T_d-1} \int_{k_j^*((s+1)T_d)}^{+\infty} \int_{v_j^*(k,r,R)}^{+\infty} dG_j(v,k), \tag{18}$$

$$Q_{2j}(R) = M_j \sum_{s=1}^{T/T_d-1} \sum_{t=sT_d+1}^{(s+1)T_d} \gamma_R(t-1, sT_d) \int_{k_j^*((s+1)T_d)}^{+\infty} \int_{v_j^*(k,r,R)}^{+\infty} l_0(v,k) dG_j(v,k). \tag{19}$$

And in this case with multiple groups, the cross-subsidy calculation can be done using:

$$r_f \sum_{j=1}^{J} Q_j(r_f) = \sum_{j=1}^{J} \left(r(Q_{0j}(r) + Q_{1j}(r)) + RQ_{2j}(R) \right),$$

where

$$Q_j(r_f) = M_j \sum_{t=1}^{T} \gamma_{r_f}(t-1) \int_{-\infty}^{+\infty} \int_{v_j^{**}(r_f)}^{+\infty} l_0(v, k=0) dG_j(v,k), \tag{20}$$

is the aggregate mortgage debt of group j when the interest rate is fixed at r_f.

We next turn to acquiring quantitative estimates of the model's parameters and an assessment of the model-implied cross-subsidy by mapping the model to the data.

4 Quantitative Analysis

The model does not admit an analytic solution for all endogenous outcomes. As a result, we choose the parameters that best match moments of the data with the corresponding moments computed from the numerical solution of the model. We then study the quantitative implications of the model evaluated at the calibrated parameters.

4.1 Calibration

In our calibration, we fix a subset of model parameters at values taken directly from the data. We estimate the remaining parameters of the model to best match key moments of the mortgage data, assuming that the model-implied moments are generated from the model's steady state.

In a set of tables below, we specify all parameters that are fixed and estimated in each of the models that we estimate. We provide a high-level summary of our estimation approach here. To begin with, we read the interest rates on discounted and reset rate mortgages directly from the underlying data, using value-weighted averages of the corresponding rates in the 2015H1 snapshot of the data. In addition, we set the unit of time in the model to be one year, the mortgage maturity at $T = 30$ years and the fixation period at $T_d = 2$ years, which is the modal initial fixation period in the UK mortgage market over the sample period. Moreover, we set the the discount rate at $\beta = 0.95$ and the parameter $\omega = 1.25$ to correspond to a loan-to-value ratio at origination of 80 percent.

Turning to the estimated parameters, we assume that borrowers' valuation v follows a lognormal distribution, i.e., $\log(v)$ follows a normal distribution with mean μ_v and standard deviation σ_v. Given that mortgage rates have a dual-rate structure (i.e., discounted and reset rates), we model k as comprising two types of households, some with low refinancing costs and others with high refinancing costs. That is, we assume that k follows a mixture distribution of two lognormal distributions: with probability η, $\log(k)$ follows a normal distribution with mean μ_{k1} and standard deviation σ_{k1}; with probability $1-\eta$, $\log(k)$ follows a normal distribution with mean μ_{k2} and standard deviation σ_{k2}. This mixture distribution allows us to separate households into low- and high- refinancing cost types, while also allowing some heterogeneity within each type.

Denoting type 1 as low-cost households and type 2 as high-cost households, we impose that the average k for the low-cost type 1 is lower than the average k for high-cost type 2, i.e., $\exp\left(\mu_{k1} + \frac{\sigma_{k1}^2}{2}\right) < \exp\left(\mu_{k2} + \frac{\sigma_{k2}^2}{2}\right)$. We also set $\eta = 0.5$, and the correlation between v and k to zero, because the empirical moments that we employ in the calibration in this version of the paper do not allow us to separately identify these parameters, as we explain below in greater detail.

Finally, our calibration recovers the parameter α of the utility function, the level of the outside option $\bar{u}$, and the size of the market M (the mass of households entering in each period).

We search for the nine parameters (α, $\bar{u}$, M, μ_{k1}, σ_{k1}, μ_{k2}, σ_{k2}, μ_v, σ_v) that minimize the distance between selected moments in the data and the corresponding moments of the model. More specifically, for each combination of these unknown parameters, we solve the model of Section 3 to find its equilibrium, characterized by the distribution of mortgage loans at origination $l_0(v,k)$ and the distribution of optimal refinancing periods T_{max}. Based on these distributions, we simulate from the model and construct the following aggregate moments:

1. the average loan balance for mortgages on the discounted rate;
2. the standard deviation of the loan balance of mortgages on the discounted rate;
3. the average loan balance for mortgages on the reset rate;
4. the standard deviation of the loan balance of mortgages on the reset rate;
5. the average remaining maturity of mortgages on the discounted rate;
6. the standard deviation of the remaining maturity of mortgages on the discounted rate;
7. the average remaining maturity of mortgages on the reset rate;
8. the standard deviation of the remaining maturity of mortgages on the reset rate;
9. the number of mortgages on the discounted rate;
10. the number of mortgages on the reset rate;
11. the fraction of mortgages on the discounted rate for the following partition of the loan balance distribution: $[0-5]$ percentile, $(5-25]$ percentile, $(25-50]$ percentile, $(50-75]$ percentile, $(75-95]$ percentile, and $(95-100]$ percentile;
12. the share of homeowners, i.e., the fraction of households that enter the housing market and choose to purchase a house taking on a mortgage loan.

The minimum-distance estimator chooses the parameters that minimize the criterion function:

$$(\mathbf{m}(\psi) - m_S)' \, \Omega \, (\mathbf{m}(\psi) - m_S),$$

where $\mathbf{m}(\psi)$ is the vector of stacked moments (the set described above) simulated from the model, evaluated at the vector ψ of parameters, and m_S is the vector of corresponding moments in the data. Ω is a symmetric, positive-definite matrix; in practice, we use a diagonal matrix whose elements are those on the main diagonal of the inverse of the matrix $E(m_S' m_S)$.

We estimate three versions of the model. In a baseline case, we pool together all mortgages in our data and assume that all households can be characterized by a single distribution $G(v,k)$, as well as common α, $\bar{u}$, and M parameters. This entails estimating 9 parameters using 17 moments.

We also pursue two richer versions of the estimation: the first one estimates the model separately for different income groups, and the second one estimates the model separately for different geographic segments of the UK. These richer versions allow us to estimate group-specific distributions $G_j(v,k)$ and group-specific parameters α_j, $\bar{u}_j$, and M_j for each group j (i.e., either an income group or a region). This gives us additional flexibility to capture heterogeneity across groups in tastes and housing opportunities.

We consider 12 income groups based on the following percentiles of the distribution of reported incomes in the PSD: 0-10, 10-20, 20-30, 30-40, 40-50, 50-60, 70-80, 80-85, 85-90, 90-95, and 95-100. We consider 12 broad geographical regions of the U.K.: North-East, North-West, Yorkshire and the Humber, East Midlands, West Midlands, East of England, Greater London, South East, South West, Wales, Scotland, and Northern Ireland.[18]

In both cases, we estimate a total of 108 parameters (9 parameters for each of the 12 groups) using a total of 204 moments (17 moments described above for each of the 12 groups).

[18] These are the 12 NUTS-1 regions of the U.K., where NUTS stands for Nomenclature of Territorial Units for Statistics.

4.2 Sources of Identification

Although the model is highly nonlinear, so that (almost) all parameters affect all outcomes, the identification of some parameters relies more heavily on certain moments in the data.

Specifically, the moments characterizing the distributions of loan sizes on the discounted and the reset rate, those characterizing the distribution of remaining maturities in each of these bins, as well as the relative numbers of mortgages in each of these categories together identify the parameters of the distribution of household preferences v and of costs k. Notably, equation (5) makes it clear that household' initial loan amounts—and, thus over time, their loan balances—depend on their housing preferences v, as well as their costs k through T_{max}.

We note that if the cost k was prohibitively high for all borrowers, almost all mortgages would be on the reset rate, and conversely, if k was extremely low for all borrowers, all mortgages would be on the discounted rate. As a result, the numbers of mortgages on the discounted rate and the reset rate contribute to the identification of the averages of k in the two types of households (high- and low-cost of refinancing).

Given a value of k, borrowers have stronger financial incentives to refinance if they have a large loan balance. Hence, the share of mortgages on the discounted rate should be increasing in loan balance.

The rate of change of the share of mortgages on each rate as loan size changes is also informative about the heterogeneity in k. The increase is fast if the heterogeneity across households is small, whereas it is slow if the heterogeneity is large. Our assumption that k follows a mixture distribution allows us to flexibly capture different rates of increase in the share of mortgages on the discounted rate at different percentiles of the loan balance distribution, thereby contributing to the identification of the refinancing cost heterogeneity parameters σ_{k1} and σ_{k2} of the mixing distribution.

Finally, the number of mortgages in total across both discounted and revert rate categories identifies the market size parameter M, and the share of owners versus renters in the U.K. data identifies the level of outside option utility $\bar{u}$.

Table 3: Calibrated Parameters, Single Group

r	3.327	R	3.829
T	30	T_d	2
β	0.950	ω	1.250
η	0.500		
μ_v	-0.806	σ_v	0.144
μ_{k_1}	4.277	σ_{k_1}	0.837
μ_{k_2}	8.683	σ_{k_2}	0.302
$\bar{u}$	1,149	M	224,025
α	0.781		

Notes: This table reports the calibrated parameters.

4.3 Parameters and Model Fit

Baseline. Table 3 reports the calibrated parameters of the model, assuming that all borrowers constitute a single group. The top of the table reports the fixed/set parameters, and the bottom of the the table reports the estimated parameters.

The set parameters take fairly standard values. We read R and r directly from the data, set T_d to 2 years, the modal discounted rate fixation period, and T to 30 years, the standard mortgage contract duration in the U.K.. ω is set to 1.25, which corresponds to an LTV of 80%, and β the rate of time discounting, is set to 0.95. Finally, we set η, the ratios of the two lognormal distributions in the mixture of distributions for k to 50% in the first instance.

Turning to the estimated parameters, they imply that households' valuation v has an average of 0.451 and a standard deviation of 0.065. There is also modest concavity estimated in household utility from housing ($\alpha = 0.781$). These parameter values mean that a household with the average v enjoys annual utility flow (i.e., consumption) of $\frac{vh^\alpha}{\alpha}$ equal to £$5,511$ from a house worth £$125,000$, for example. This corresponds to a rental yield of roughly 4.4% evaluated at these values, which is lower than the average rental yield for the whole of the U.K., but roughly in line with average industry values reported for London in this period.[19]

In the baseline model, we estimate borrowers' average refinancing cost μ_k to equal

[19]See, for example, Savill's UK Report on Rents and Returns, 2015.

£3,144 in the population, with a standard deviation equal to £3,310. These two centered moments of the population distribution combine the distribution of low-cost households, with an average cost of £102 and a standard deviation of £103, and the distribution of high-cost households, with an average cost of £6,181 and a standard deviation of £1,913. That is, the distribution of low-cost households is considerably more concentrated around its mean than the distribution of high-cost households—suggesting that there are a variety of underlying determinants for household inaction (Gabaix and Laibson, 2001; Abel et al., 2013; Matějka and McKay, 2015; Caplin, 2016), a common finding in the household finance literature across multiple components of the household balance sheet (Choi et al., 2002; Brunnermeier and Nagel, 2008).

The average estimated value of k is larger than the average total psychological plus fixed refinancing cost estimated in Andersen et al. (2020) of roughly £1,852. However, our figures are of a similar order of magnitude to those in that analysis, despite the differences in setting (UK vs. Denmark), and the fact that the Andersen et al. (2020) model also considers "Calvo-style" refinancing inaction in addition to individual-specific refinancing thresholds, while we only consider the latter in this version of the paper.

The outside option utility value is £1,149 which represents the annual net utility from renting—meaning that the model estimates that households with a net utility value (over and above all mortgage payments and refinancing costs) greater than this level from purchasing a house enter the mortgage market in each period.

Table 4 presents a comparison between the empirical moments and the moments calculated from the model at the calibrated parameters reported in Table 3. Overall, from visual inspection, the model appears to fit the data well.

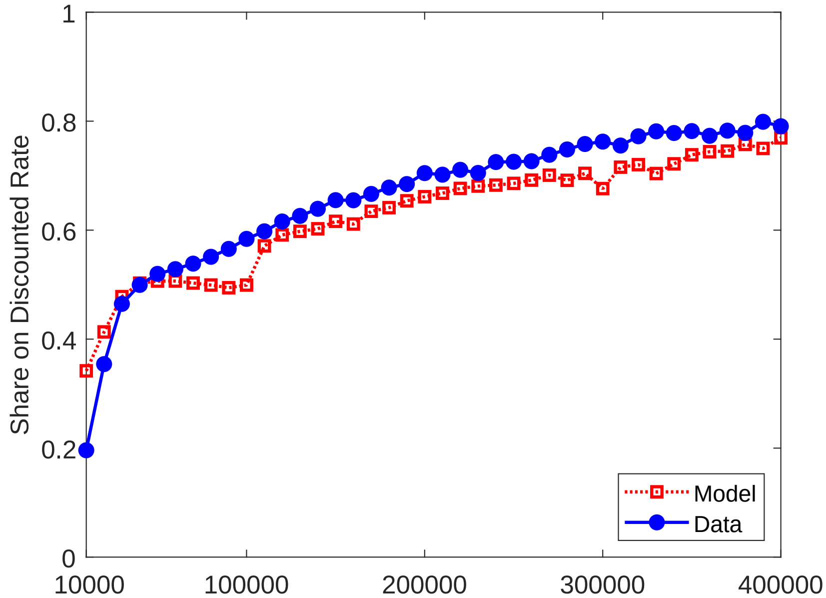
Similarly, Figure 3 displays the comparison between the model-implied shares of mortgages paying the discounted rate and its empirical analog. Notably, the model is well-able to capture the concave relationship between the two variables, with a faster rate of increase in the share of mortgages on the discounted rate at low balances and a slower rate of increase at high balances.

Table 4: Model Fit

	Data	Model
Mean Loan Balance, Discounted Rate	135,620	138,386
Standard Deviation Loan Balance, Discounted Rate	105,843	107,536
Mean Loan Balance, Reset Rate	104,364	105,658
Standard Deviation Loan Balance, Reset Rate	81,190	76,963
Mean Remaining Years, Discounted Rate	20.06	16.52
Standard Deviation Remaining Years, Discounted Rate	8.05	8.86
Mean Remaining Years, Reset Rate	16.05	14.17
Standard Deviation Remaining Years, Reset Rate	7.35	8.20
Share of Mortgages on Discounted Rate, 0-5 Percentile	36.69	37.86
Share of Mortgages on Discounted Rate, 5-25 Percentile	53.00	49.79
Share of Mortgages on Discounted Rate, 25-50 Percentile	57.21	50.57
Share of Mortgages on Discounted Rate, 50-75 Percentile	63.62	59.86
Share of Mortgages on Discounted Rate, 75-95 Percentile	70.41	66.35
Share of Mortgages on Discounted Rate, 95-100 Percentile	78.85	76.72
Number of Mortgages on Discounted Rate	2,519,789	2,290,875
Number of Mortgages on Reset Rate	1,633,813	1,759,041
Share of Owners	63.13	60.26

Notes: This table reports the values of the empirical moments and of the moments calculated at the calibrated parameters reported in Table 3.

Figure 3: Share of Loans on Discounted Rate



Notes: This figure displays the share of loans paying the discounted rate as a function of its loan balance in the data (solid line) and in the model evaluated at the calibrated parameters (dashed line).

Table 5: Interest Rates and Loan Sizes, Single Group

	Model	Counterfactual
Discounted Rate	3.33	3.54
Reset Rate	3.83	3.54
Mean Initial Loan Amount	206,590	201,839
Standard Deviation Initial Loan Amount	117,273	110,539
Mean Loan Balance	124,171	121,509
Standard Deviation Loan Balance	96,835	92,783
Number of Mortgages	4,049,916	4,106,370

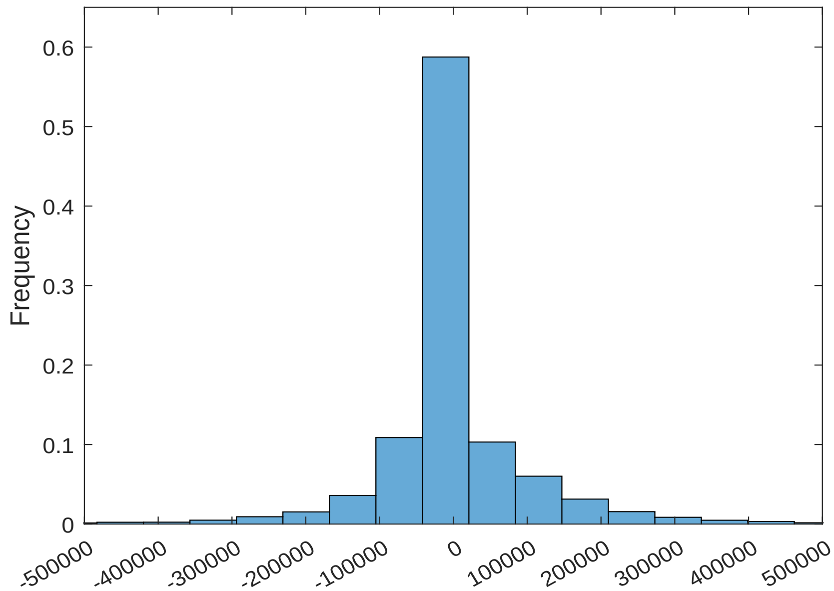
Notes: This table reports the statistics on the mortgage market in the baseline model and in a counterfactual market with constant interest rates.

4.4 Model Implications: Constant Interest Rate

Baseline Model. Table 5 reports the results of the calculation of the constant interest rate that satisfies the constant-revenue equation (13), along with statistics on the corresponding loan sizes.

The table shows that interest rates would equal 354 basis points in a counterfactual market with constant rates. Moreover, borrowers respond to the change in the profile of

Figure 4: Change in Initial Mortgage Amount



Notes: This figure reports the distribution of the changes in loan sizes between the counterfactual market with constant interest rates and the baseline case with discounted and reset rates.

interest rates by adapting their loan size, which on average decreases by £4,752, or 2.30 percent of the baseline average loan size.

The average change combines borrowers who increase their mortgage amounts with borrowers who decrease them. Figure 4 displays the full distribution of the change in mortgage amount, highlighting the heterogeneity of the change. Most notably, borrowers with the lowest k pay an interest rate equal to approximately 333 basis points in the observed mortgage market, because they always refinance, but in the counterfactual market with a single interest rate, they pay 354 basis points. As a result, they reduce their loan sizes when faced with this higher interest rate. In contrast, borrowers with the highest k pay an interest rate approximately equal to 383 basis points in the baseline market, because they never refinance, but pay 354 basis points in the counterfactual market. As a result, these borrowers increase their loan sizes. We calculate that borrowers who decrease their loan size have an average cost k equal to £1,991, whereas borrowers who increase their loan size have an average cost k equal to £6,069. The low k borrowers dominate overall, since their average loan sizes were higher to begin with.

The fourth row of Table 5 reports that the standard deviation of the initial loan size declines quite substantially, by £$6,734$, or 5.74 percent of the baseline standard deviation of the initial loan size. The reason is that one dimension of household heterogeneity, namely k, contributes to the determination of the loan size in the baseline model with refinancing. However, this dimension of heterogeneity becomes irrelevant when interest rates are constant. More specifically, the previous arguments suggest—and Figure 4 shows—that borrowers with larger loans in the model with refinancing decrease their loan sizes under the counterfactual, whereas borrowers with smaller loans in the model with refinancing increase their loan sizes under the counterfactual with constant interest rates and no refinancing. A constant, common interest rate thus pushes loan sizes to be more similar.

The last row of Table 5 reports that the number of mortgages increases, by $56,454$, or 1.39 percent of the baseline number of mortgages. The reason is that many marginal borrowers with valuation v and high cost k close to the threshold $v^*(k)$ in the baseline now buy a house, and take a mortgage. This effect, on net, is greater than the total exit of borrowers with low k, who exit the market because they prefer to rent rather than buy under the counterfactual. In sum, while the intensive margin effect of low k borrowers dominates and makes the average loan size lower under the counterfactual, the extensive margin effect of high k borrowers entering the market under the counterfactual dominates, making the total number of loans greater in the single-rate economy.

The decline in initial loan size and the increase in the number of mortgages together combine to reduce aggregate mortgage debt by 0.78 percent relative to the model which allows refinancing. While cross-subsidies are eliminated in the counterfactual, this shows that one consequence of this change is that the mortgage market shrinks in size. The shrinking of the mortgage market is a result of the reduced size of the average loan under the counterfactual, even though there is an increase in the total number of loans issued.

The Case of Multiple Groups. The case with multiple groups allows us to tie some of the heterogeneity in preferences v and in costs k with the observable heterogeneity in refinancing rates observed across income groups as well as across regions and devolved administrations of the U.K. This helps us to understand how the shift to a single mortgage rate structure could lead to heterogeneous outcomes for households in these groups.

To begin with, Table 6 and Table 7 show how the moments vary across income groups

Table 6: Aggregate moments (means), by inc. quantiles

	Inc. level	Prop. (Disc.)	Disc. rate	Reset rate	Bal.
0-10	23,436	0.59	3.47	4.00	51,012
10-20	29,408	0.59	3.47	3.94	68,455
20-30	34,525	0.59	3.46	3.92	80,064
30-40	39,555	0.59	3.43	3.88	90,117
40-50	44,973	0.59	3.40	3.84	100,747
50-60	51,309	0.60	3.36	3.81	112,381
60-70	59,463	0.61	3.32	3.77	126,325
70-80	71,390	0.62	3.26	3.73	145,410
80-85	80,375	0.63	3.20	3.71	166,038
85-90	94,323	0.64	3.14	3.70	187,160
90-95	123,280	0.65	3.05	3.67	223,873
95-100	210,304	0.66	2.90	3.60	340,415

The table above shares the proportion of mortgages under discounted rates, the average discount rate, the average reset rate and the average balance across the income-bins used for the multiple groups cross-subsidy calculation.

and regions in the UK. Table 6 shows that the proportion of households on the discounted rate increases significantly and monotonically with household income. Moreover, the average discounted rate (and indeed, reset rate) that households pay also falls as income rises, though the spread between the two rates remains roughly similar for all income groups. As might be expected, the average loan balance also rises with income.

Table 7 presents the regions and devolved administrations of the UK in order of the share of mortgages on the discounted rate. What is evident from the table is that relatively poorer regions and devolved administrations have lower shares of mortgages on the discounted rate (clearly, London has the highest share). Once again, the average discounted rate paid is also higher in relatively less wealthy regions. We next turn to analyzing how these data moments translate into estimates of the distribution of cross-subsidies across these groups in the UK population.

Model Calibration for Multiple Groups. Table 9 reports summary statistics of the calibrated parameters of the model when we estimate separate group-specific parameters for households grouped together by income or U.K. geographical region.

Panel A of the table refers to groups based on income, and Panel B to groups located in

Table 7: Aggregate moments (means), by UK regions and devolved administrations

	Prop. (Disc.)	Disc. rate	Reset rate	Bal.
Northern Ireland	0.47	3.46	4.11	91,047
North East (England)	0.54	3.51	3.88	87,369
Scotland	0.56	3.42	3.89	94,392
Wales	0.57	3.45	3.86	94,054
North West (England)	0.57	3.46	3.91	97,418
West Midlands (England)	0.57	3.41	3.73	104,215
Yorkshire and The Humber	0.59	3.46	3.93	94,577
East Midlands (England)	0.60	3.43	3.77	100,694
South West (England)	0.64	3.32	3.66	122,209
East of England	0.66	3.26	3.78	140,859
South East (England)	0.66	3.21	3.72	158,957
London	0.67	3.01	3.89	202,091

The table above shares the proportion of mortgages under discounted rates, the average discount rate, the average reset rate and the average balance across the UK regions and devolved administrations used for the multiple groups cross-subsidy calculation.

Table 8: Interest Rates and Loan Sizes, Groups

	Model	Counterfactual Income Groups	Counterfactual Regions
Discounted Rate	3.33	3.51	3.51
Reset Rate	3.83	3.51	3.51
Mean Initial Loan Amount	197,069	192,175	200,028
Standard Deviation Initial Loan Amount	131,260	122,459	115,068
Mean Loan Balance	118,330	115,569	120,292
Standard Deviation Loan Balance	102,913	97,599	94,843
Number of Mortgages	4,143,171	4,289,959	4,199,703

Notes: This table reports the statistics on the interest rates and on the initial loan sizes for the model estimated on data for each region separately.

Table 9: Calibrated Parameters, Groups

Panel A: Income Groups				Panel B: Regional Groups			
μ_{v_j}	0.277	σ_{v_j}	0.067	μ_{v_j}	-0.490	σ_{v_j}	0.140
	(0.384)		(0.027)		(0.264)		(0.020)
$\mu_{k_{1j}}$	4.493	$\sigma_{k_{1j}}$	0.575	$\mu_{k_{1j}}$	4.621	$\sigma_{k_{1j}}$	0.341
	(0.451)		(0.236)		(0.380)		(0.711)
$\mu_{k_{2j}}$	8.428	$\sigma_{k_{2j}}$	0.769	$\mu_{k_{2j}}$	37.949	$\sigma_{k_{2j}}$	12.985
	(0.480)		(0.267)		(26.544)		(16.939)
$\bar{u}_j$	2,303	M_j	16,605	$\bar{u}_j$	1,421	M_j	18,254
	(597)		(6,220)		(332)		(7,868)
α_j	0.692			α_j	0.755		
	(0.041)				(0.025)		

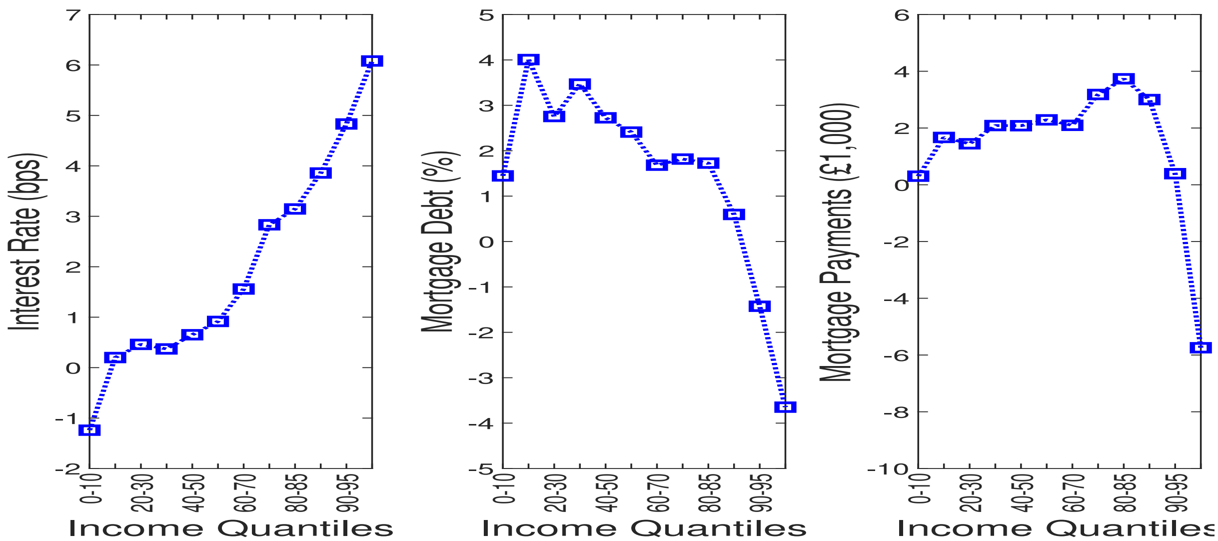
Notes: This table reports the average and the standard deviation (in parentheses) of each calibrated parameter. Panel A refers to the income groups, Panel B to the regional groups.

different regions and devolved administrations of the U.K. For each parameter, we report the average in the population, weighted by the size of each market M_j, and the (weighted) cross-group standard deviation in the population.

Table 9 shows that most parameters in Panel B (regional groups) are similar on average to those reported in Table 3 for the single group case. The parameters that perhaps display the most meaningful heterogeneity in the population are μ_{k1_j} and μ_{k2_j}. This fact is particularly interesting for our purposes because this heterogeneity is precisely what is needed to explain regional heterogeneity in refinancing activity—it contributes directly to our quantitative assessment of the equilibrium cross-subsidy across regions. The parameters in Panel A (income groups) displays some differences with those in Panel B (as well as those in Table 3) because the heterogeneity across and within income groups differs from the heterogeneity across and within regional groups, thereby affecting the average and the standard deviations of some of the parameters.

While we do not report measures of goodness-of-fit across regions, we do note that the model visually appears to fit the regional data well. This is perhaps not surprising given that Table 4 shows that the single-group model fits the aggregate data well; the same model might therefore be expected to fit as well or better at a lower level of aggregation.

Figure 5: Changes in Market Outcomes by Income Groups



Notes: The left panel displays the change in interest rates (in bps), the central panel displays the percent change in mortgage debt, and the right panel displays the change in mortgage payments for each income group in the counterfactual case with a constant interest rate relative to the baseline case.

Table 8 reports aggregated counterfactual estimates when the model is estimated with multiple groups. When we compare these aggregate statistics with those obtained from a single, heterogeneous group in Table 5, the differences are minimal. We go on to discuss how the results differ group-by-group.

Figure 5 shows how selected changes to mortgage market outcomes look for each of the different income groups that we consider. First, interest rates (shown in basis points) increase on average for all income groups under the counterfactual, showing that intra-group differentials also exist, and that the low-k borrowers dominate the group-specific averages. Second, the interest rate increases in the counterfactual single-rate economy are particularly pronounced for the highest income groups, rising by close to 10 bp per annum for the groups at the very top of the income distribution. This is a reflection of the fact that k is estimated to be lowest for the highest income groups, and is consistent with the regressive nature of the cross-subsidies. Third, these changes in interest rates translate into lower mortgage debt for all groups, but substantially lower mortgage debt for the very top income groups, who adjust their loan sizes downwards in the face of higher lifetime mortgage rates in the new single-rate world. Fourth, these changes in mortgage debt, even when combined with the higher interest rates, translate into lower mortgage payments for the highest income groups, an endogenous response to the single-rate structure—we describe the model mechanism in more detail when discussing the regional results below. Fifth, there is a mild increase in mortgage payments for those in lower income groups, who pay slightly higher rates on average, but also, for the higher k borrowers within each group, increase their mortgage sizes to take advantage of the more beneficial rate structure in the new single-rate world without the need to engage in costly refinancing. Overall, the picture emerges of cross-subsidies being regressive, and mortgage uptake and size being greater for lower income populations in the single-rate world.

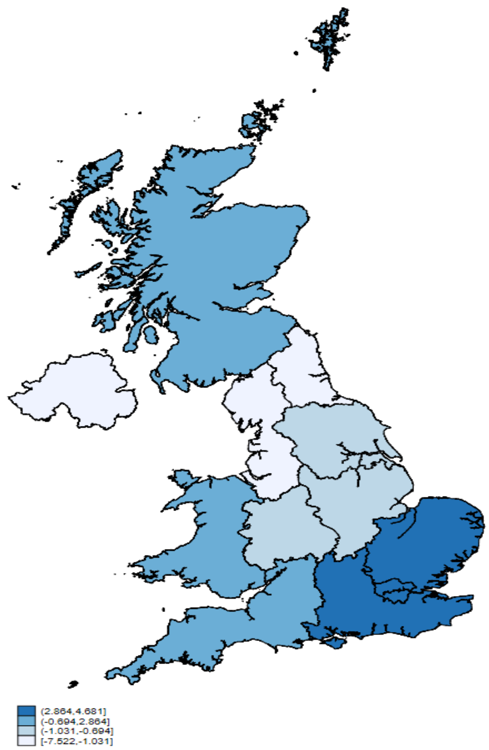
Turning to regional variation, Figure 6 presents maps that display some of the changes to mortgage market outcomes across different UK regions. In each panel, darker colors indicate more positive differences between a counterfactual market with constant interest rates and the baseline case with discounted and reset rates.

The top-left map displays the change in average interest rates paid on mortgages, reported in basis points. These changes are unevenly distributed. Households in Northern Ireland would experience a decrease in paid rates (of approximately 8 basis points on average), whereas households in Greater London would experience the largest increases of approximately 5 basis points on average. These patterns are consistent with a currently regressive cross-subsidy across regions of the U.K..

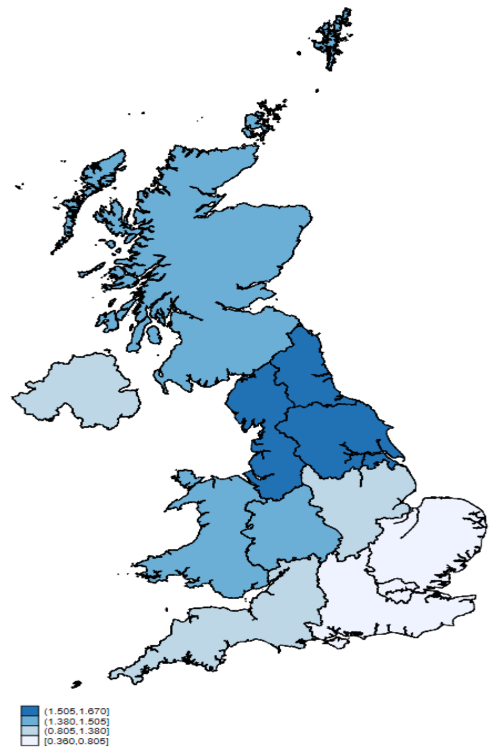
In the counterfactual equilibrium, as seen earlier in the case of income, households endogenously adjust their homeownership decisions as well as their mortgage debt conditional on entering the market to the new rate. The top-right map displays the changes in the homeownership rate across regions, which shows that there are increases in all regions. Greater London experiences the smallest increase (of 0.36 percentage points), whereas the East Midlands experiences the largest increase (of 1.67 percentage points). Overall, the North-East of the country and Scotland see the largest increases in homeownership in the new equilibrium.

The bottom left-hand plot shows the intensive margin response, namely the percent change in aggregate mortgage debt in each region. Overall, the changes are small, at most 3 percent. All regions experience a decline in mortgage debt given the rise in rates experienced by most market participants, with the largest declines (indicated by the lightest colors) in Greater London and the South East of England, and the smallest declines (indicated by the darkest colours) in Scotland, Wales, and Yorkshire and the Humber.

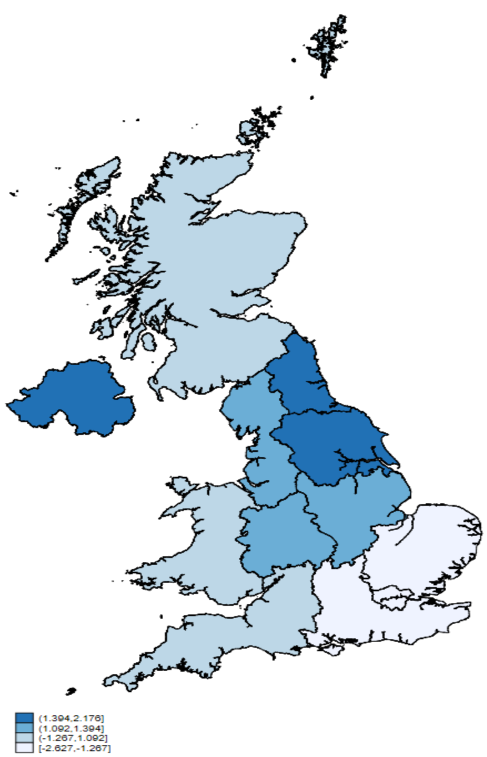
Figure 6: Regional Changes



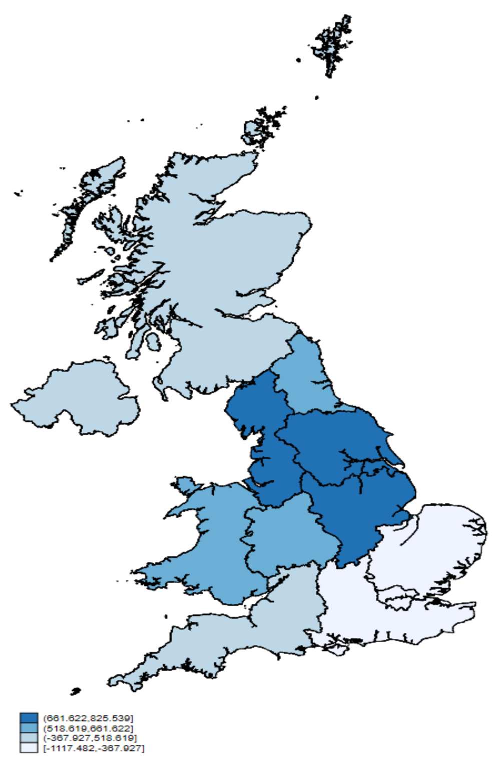
(a) Interest Rate



(b) Homeownership Rate



(c) Mortgage Debt



(d) Mortgage Payments

The bottom right-hand plot puts together the loan size and interest rate effects into an all-in change in household mortgage payments in each region. We calculate household annual interest payments in the baseline market as the sum of interest payments of mortgages on the discounted rate and of mortgages on the reset rate normalized by market size—formally as $\frac{r(Q_{0j}(r)+Q_{1j}(r))+RQ_{2j}(R)}{M_j}$, where the debt amounts Q_{0j}, Q_{1j}, and Q_{2j} are defined in equations (15), (17), and (19), respectively. Correspondingly, household annual interest payments in the counterfactual market equal $r_f Q_j(r_f)/M_j$, where $Q_j(r_f)$ is the aggregate mortgage debt of region j, defined in equation (20). The change in household annual interest payments exhibits greater heterogeneity across regions than the other outcomes. Most notably, households in Greater London would experience a decrease of approximately £1117, whereas households in the West Midlands would experience the largest increase in annual mortgage interest payments of approximately £826. Adjustments in loan amounts accounts for most of these changes in household interest payments, although of course the change in the profile in interest rates is the cause of these adjustments. Removing the regressive cross-subsidy counter-intuitively generates a *decline* in mortgage payments in the relatively more prosperous Greater London and East of England regions, and countervailing increases in total mortgage payments in the North-East, Yorkshire, and the East Midlands which arise as a result of endogenous increases in homeownership and mortgage debt occasioned by the single-rate structure.

5 Conclusion

In this paper, we structurally estimate refinancing cross-subsidies in the U.K. mortgage market using data from 2015. The U.K. is a particularly well-suited country for such analysis given the availability of high-quality and granular administrative data on the stock of all outstanding mortgages, which permits analysis of aggregate mortgage revenues in combination with a model of household refinancing. In addition, the U.K. setting features rich variation in mortgage refinancing behavior across the dimensions of household income and across geographical regions, which permits analysis of how financial cross-subsidies can vary across different groups of households.

Our model permits us to match broad features of the data, and the parameters reveal that there is considerable heterogeneity in mortgage refinancing costs across households, echoing findings in prior literature (Agarwal et al., 2016; Keys et al., 2016; Andersen et al., 2020). We push the literature further by quantifying cross-subsidies in this market. Using our parameters estimated using the 2015H1 data, we find that rates in the counterfactual single-rate equilibrium lie roughly 20bp above the discounted "teaser" rate on average, an increase of roughly 6%, and roughly 30bp below the reset rate that borrowers are routinely rolled on to at the expiration of the discounted rate fixation period, a decrease of roughly 8%. These are material changes given the importance of mortgages to household budgets.

We also find that these changes are unevenly distributed across income groups, as well as across regions and devolved administrations of the U.K. during the sample period. Relatively higher income households and more wealthy regions experience a heightened increase in rates on average in the counterfactual, while relatively poorer households and regions experience a smaller increase. That said, we find that average rates increase across the board given the high prevalence of discounted rates observed in the data. The more nuanced finding is that these changes to rates translate into increased homeownership across the board, as well as into increased mortgage uptake and average payments, especially for relatively poorer households, as well as households located in relatively poorer areas of the U.K.. In contrast, under the counterfactual, we see reduced debt for relatively richer households and areas as households endogenously respond to the change in rates. Put differently, elimination of the cross-subsidy essentially "democratizes" mortgages, making them more appealing to relatively poorer households, and those living in less well-off regions.

Our work has both methodological and economic contributions beyond the specific context that we study. We believe that our structural approach to estimating financial cross-subsidies is a useful way to provide a money-metric assessment of the impacts of heterogeneity in household inaction, with potentially wider implications for the field of household finance, where such heterogeneity is widely prevalent. And our results on the distribution of financial cross-subsidies in this important market show that studying household finances can be helpful for the broader goal of identifying the sources and consequences of wealth inequality, a continuing concern for society.

Finally, we reiterate that our inferences and estimates are derived from the stock of mortgages in the UK in 2015. As previously noted, both the spread between different types of rates and the composition of the mortgage stock has changed significantly since then. Any inferences about more recent periods will thus require additional work, a task we intend to take up going forward.

References

Abel, A. B., J. C. Eberly, and S. Panageas (2013): "Optimal inattention to the stock market with information costs and transactions costs," Econometrica, 81(4), 1455–1481.

Agarwal, S., J. C. Driscoll, and D. I. Laibson (2013): "Optimal mortgage refinancing: a closed-form solution," Journal of Money, Credit and Banking, 45(4), 591–622.

Agarwal, S., R. J. Rosen, and V. Yao (2016): "Why do borrowers make mortgage refinancing mistakes?," Management Science, 62(12), 3494–3509.

Alvaredo, F., L. Chancel, T. Piketty, E. Saez, and G. Zucman (2017): "Global inequality dynamics: New findings from WID. world," American Economic Review, 107(5), 404–09.

Andersen, S., J. Y. Campbell, K. M. Nielsen, and T. Ramadorai (2020): "Sources of inaction in household finance: Evidence from the Danish mortgage market," American Economic Review, 110(10), 3184–3230.

Armstrong, M., and J. Vickers (2012): "Consumer Protection and Contingent Charges," Journal of Economic Literature, 50(2), 477–93.

Badarinza, C., J. Y. Campbell, and T. Ramadorai (2016): "International comparative household finance," Annual Review of Economics, 8, 111–144.

——— (2018): "What calls to ARMs? International evidence on interest rates and the choice of adjustable-rate mortgages," Management Science, 64(5), 2275–2288.

Belgibayeva, A., T. Bono, P. Bracke, J. Cocco, and T. Majer (2020): "When Discounted Rates End: The Cost of a Taking Action in the Mortgage Market," Occasional Paper 54, Financial Conduct Authority.

Benetton, M. (2021): "Leverage regulation and market structure: An empirical model of the uk mortgage market," Journal of Finance, forthcoming.

Benetton, M., A. Gavazza, and P. Surico (2021): "Mortgage Pricing and Monetary Policy," Discussion Paper 16456, Centre for Economic Policy Research.

Benhabib, J., and A. Bisin (2018): "Skewed wealth distributions: Theory and empirics," Journal of Economic Literature, 56(4), 1261–91.

Beraja, M., A. Fuster, E. Hurst, and J. Vavra (2019): "Regional heterogeneity and the refinancing channel of monetary policy," The Quarterly Journal of Economics, 134(1), 109–183.

Best, M. C., J. S. Cloyne, E. Ilzetzki, and H. J. Kleven (2020): "Estimating the elasticity of intertemporal substitution using mortgage notches," The Review of Economic Studies, 87(2), 656–690.

Brunnermeier, M. K., and S. Nagel (2008): "Do wealth fluctuations generate time-varying risk aversion? Micro-evidence on individuals," American Economic Review, 98(3), 713–36.

Calvet, L. E., J. Y. Campbell, and P. Sodini (2009): "Measuring the financial sophistication of households," American Economic Review, 99(2), 393–98.

Campbell, J. Y. (2006): "Household finance," The journal of finance, 61(4), 1553–1604.

Campbell, J. Y., T. Ramadorai, and B. Ranish (2019): "Do the rich get richer in the stock market? Evidence from India," American Economic Review: Insights, 1(2), 225–40.

Caplin, A. (2016): "Measuring and modeling attention," Annual Review of Economics, 8, 379–403.

Choi, J. J., D. Laibson, B. C. Madrian, and A. Metrick (2002): "Defined contribution pensions: Plan rules, participant choices, and the path of least resistance," Tax policy and the economy, 16, 67–113.

Cloyne, J., K. Huber, E. Ilzetzki, and H. Kleven (2019): "The effect of house prices on household borrowing: a new approach," American Economic Review, 109(6), 2104–36.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2020): "Heterogeneity and persistence in returns to wealth," Econometrica, 88(1), 115–170.

Financial Conduct Authority (2019a): "Mortgage Market Study: Final Report," Discussion paper, Discussion paper.

——— (2019b): "PS19/27: Changes to mortgage responsible lending rules...," https://www.fca.org.uk/publication/policy/ps19-27.pdf, Accessed: 2021-10-19.

——— (2020a): "FCA issues research on mortgage switching," https://www.fca.org.uk/news/news-stories/fca-issues-research-mortgage-switching, Accessed: 2021-10-19.

——— (2020b): "PS20/11: Removing barriers to intra-group switching...," https://www.fca.org.uk/publication/policy/ps20-11.pdf, Accessed: 2021-10-19.

Gabaix, X., and D. Laibson (2001): "The 6D bias and the equity-premium puzzle," NBER macroeconomics annual, 16, 257–312.

——— (2006): "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," The Quarterly Journal of Economics, 121(2), 505–540.

Gomes, F., M. Haliassos, and T. Ramadorai (2021): "Household Finance," Journal of Economic Literature.

Greenwald, D. L., M. Leombroni, H. Lustig, and S. Van Nieuwerburgh (2021): "Financial and Total Wealth Inequality with Declining Interest Rates," Discussion paper, National Bureau of Economic Research.

Hubmer, J., P. Krusell, and A. A. Smith Jr (2020): "Sources of US wealth inequality: Past, present, and future," in NBER Macroeconomics Annual 2020, volume 35. University of Chicago Press.

Hurst, E., B. J. Keys, A. Seru, and J. Vavra (2016): "Regional redistribution through the US mortgage market," American Economic Review, 106(10), 2982–3028.

Keys, B. J., D. G. Pope, and J. C. Pope (2016): "Failure to refinance," Journal of Financial Economics, 122(3), 482–499.

Matějka, F., and A. McKay (2015): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," American Economic Review, 105(1), 272–98.

Miles, D. (2004): "The UK mortgage market: Taking a longer-term view," .

Robles-Garcia, C. (2020): "Competition and Incentives in Mortgage Markets: The Role of Brokers," Mimeo, Stanford University.

APPENDIX

A Overall Composition of UK Mortgage Market and Details on PSD007 Data

A.1 Overall Composition of UK Mortgage Market

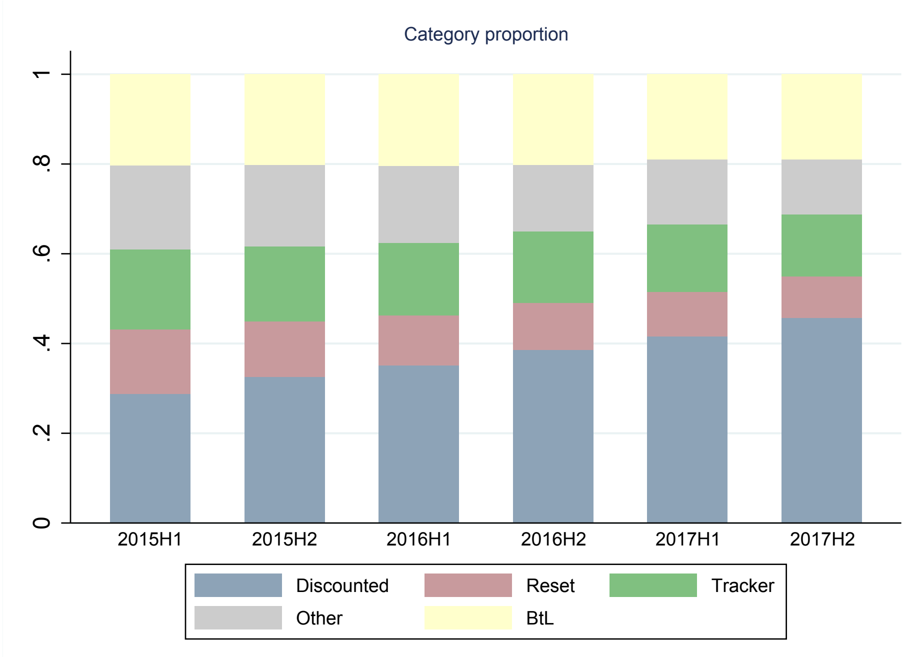
We discuss the composition of the UK mortgage market under two broad segments—the owner-occupier and the buy-to-let segment. The owner-occupier segment includes borrowers who buy a property for residence, and includes first-time-buyers, home-movers and refinancers. The buy-to-let segment includes mortgages by landlords who let their property out for rental earning. The PSD007 mortgage data used by us provides a snapshot of the universe of the owner-occupier segment, reported half-yearly starting 2015H1.

The overall size of the UK mortgage market is reported in a quarterly basis in the MLAR tables reproted by the FCA and accessible here. The report shows a breakdown of the UK mortgage market into owner-occupier (OO) and buy-to-let (BtL) segments. We combine data from the PSD007 and the MLAR tables to show the decomposition of the UK mortgage market into the OO and BtL segments from 2015H1 to 2017H2 in Figure A.1. The OO segment is further decomposed into mortgages under discounted rates, standard variable rates (reset rate), tracker rates, and an unclassified 'other' category. The overall share of the BtL segment is fairly stable during the sample at roughly 20%. The share of mortgages in our sample (discounted and reset rate mortgages) ranges from 47% of total in 2015H1 to 60% in 2017H2.

The MLAR tables highlight a significant difference between the OO and BtL segments of the mortgage market. While mortgages in the OO segment are issued by regulated institutions such as deposit taking institutions and building societies, the BtL mortgage are primarily issued by non-banks. For instance, in 2017H1, non-banks issued 99% of all BtL loans in the UK. Thus, these segments differ in both the types of properties (residential vs rental) and the issuing lender.[20]

[20]Mortgage lending by non-banks is reported under '*Residential loans to individual: Non-regulated*' in the MLAR report.

Figure A.1: **Composition of Mortgage Stock Over Snapshots**



The figure above shows the decomposition of the UK mortgage market into the owner-occupier and buy-to-let (BtL) segments. The owner-occupier segment is further broken down into mortgages under discounted, reset rate, tracker and an unclassified 'other' interest rate categories.

A.2 Details on PSD007 Data : Universe of UK Owner-Occupier Mortgages

As mentioned in the preceding sub-section, PSD007 data is collated by the FCA and includes loan-level information on the universe of mortgages in the owner-occupier or residential segment of the mortgage market. FCA is the conducts authority in the UK and all regulated financial institutions are mandated by law to share this data at a semi-annual frequency.

We have data on 6 PSD007 snapshots, reported half-yearly from mid-2015 to end-2017. Table A.1 provides a brief description of the loan-level variables reported in PSD007 relevant to our study. In each snapshot, we observe the loan balance, original size of the loan, term to maturity, original maturity, and interest rate for each mortgage as on the date of reporting. The data includes an indicator variable on whether a mortgage is incentivised (i.e., under a discounted rate), and if yes, the remaining period under the incentivised or discounted rate. We use the reported interest rate to calculate a spread against the yield on a nominal zero coupon bond maturing over a horizon over which the interest rate is fixed.[21]

Table A.1: **Description of Variables**

Variable	Description
Balance	Balance as on the date of reporting
Interest rate	Interest rate charged on the mortgage
$\mathbb{D}(Discounted)$	Indicator variable equalling 1 for mortgages under discounted rates; 0 for mortgages under reset rate
Spread	Spread over the yield on a nominal zero coupon bond maturing over a horizon comparable to the fixation period for interest rates (0 for mortgages under reset rate).
Original size	Original size at the time of mortgage account opening date.
Original term	Original term to maturity at the time of mortgage account opening date.
Term to maturity	Remaining term to maturity.
Remaining discounted period	Remaining period under discounted rates.
Borrower age	Borrower age as on the date of reporting.

The table above provides a brief description of mortgage level variables reported in PSD007 data relevant to our study.

In addition to the value of the interest rate, the database also includes information on type of interest rate and, as mentioned earlier, whether the mortgage is incentivised or discounted. The types of interest rates reported in the dataset are teaser, discounted, capped, standard variable rate, tracker and an unclassified other category. Of these, mortgages under teaser, discount, and capped interest rates are under incentivised/discounted rates.

Table A.2 shows the total number of mortgages in 2015H1 by interest rate type and incentivised status. The table shows that a vast majority of the mortgages reported as being incentivised are also reported to be under discounted rates. Most mortgages under discounted and capped interest rates are also reported as being incentivised. However,

[21] discounted mortgages are fixed-term mortgages with a specified period under the discounted rates. As in our model, mortgages move to the reset rate rate at the end of the discounted period. For a mortgage with a year remaining under discounted rates, spread is calculated against the yield on a nominal zero coupon bond maturing in a year. reset rate mortgages are variable rate mortgages; spread for reset rate mortgages is calculated based on the yield on short-term (6 months) UK Government bonds.

there are few discounted, discounted and capped mortgages which are reported as being non-incentivised and appear to have anomalous interest rates (explained shortly). We exclude such mortgages from our sample.

Table A.2: **Mortgages in 2015H1: Total Number by Interest Rate Type and Incentivised Status**

	Incenstivised		
	No	Yes	Total
Teaser	179,513	3,269,984	3,449,497
Discount	14,221	64,703	78,924
Capped	284	4,143	4,427
SVR	2,153,832	56,193	2,210,025
Tracker	934,176	708,123	1,642,299
Other	427,689	1,221	428,910
Total	3,709,715	4,104,367	7,814,082

The table above shows the total number of mortgages by type of interest rate, and whether the mortgage is reported as being incentivised in the mortgage snapshot for 2015H1.

Table A.3 shows the average interest rate by interest rate type and incentivised status in the 2015H1 snapshot. We observe that mortgages under discounted, discounted, and capped interest rates have overall lower average interest rates, bar the few mortgages in these categories anomalously reported as being non-incentivised. Mortgages under reset rate (or Standard Variable Rate, SVR) have higher average interest rates than these categories as well. The average rate interest of the 56,193 mortgages under reset rate reported as being incentivised is comparable to the interest rate of the non-incentivised reset rate mortgages.[22] We treat all instance of mortgages under reset rate as being non-incentivised and without any remaining discounted period.

Tracker mortgages are the remaining large category of mortgages whose interest rates are benchmarked to the contemporaneous Bank of England base rate or LIBOR. Table A.3 shows that the average interest rate of this category is the lowest when compared to the

[22]This is a data issue only in the 2015H1 snapshot.

other types. However, this category is quite distinct from the discounted rate mortgages and not 'incentivised' in the way modelled in our paper.

First, almost all mortgage origination with discounted periods in our sample (2015H1-2017H2) is classified under the discounted category. Second, at the end of a discounted period, mortgages from the discounted category transition to the reset rate category; and mortgages under reset rate, when refinanced under a discounted scheme, emerge in the discounted category in subsequent snapshots. However, there are no cross-flows between the discounted to tracker or the reset rate to tracker categories. Thus, the tracker category is relatively isolated from the other two categories, and we restrict our study on cross-subsidies to mortgages under the discounted and reset rate categories.

Table A.3: **Mortgages in 2015H1: Avg. Interest Rate by Interest Rate Type and Incentivised Status**

	Incenstivised		
	No	Yes	Total
Teaser	5.83	3.35	3.48
Discount	3.04	3.31	3.26
Capped	4.02	2.91	2.99
SVR	3.79	3.63	3.79
Tracker	2.22	2.16	2.19
Other	2.88	2.80	2.88
Total	3.39	3.15	3.26

The table above shows the average interest rate by type of interest rate, and whether the mortgage is reported as being incentivised in the mortgage snapshot for 2015H1.

Table A.4 shows the outstanding balance by interest rate type and incentivised status in the 2015H1 snapshot. Given that we exclude the discounted, discounted and capped mortgages reported as non-incentivised, tracker mortgages and the unclassified 'other' mortgages from our study, our filtered database comprises incentivised discounted, discounted and capped mortgages, and mortgages under reset rate.[23]

[23]The total number of mortgages, average interest rates, and outstanding balance reported in Tables

Table A.4: **Mortgages in 2015H1: Total Balance by Interest Rate Type and Incentivised Status (in £ billions)**

	Incenstivised		
	No	Yes	Total
Teaser	11.4	442.6	454.0
Discount	1.3	7.3	8.7
Capped	0.0	0.5	0.5
SVR	208.5	6.1	214.6
Tracker	121.7	90.6	212.3
Other	39.0	0.1	39.1
Total	381.9	547.3	929.2

The table above shows the total balance in £ billions by type of interest rate, and whether the mortgage is reported as being incentivised in the mortgage snapshot for 2015H1.

A.3 Merging across Mortgage Snapshots

The high quality disaggregated information in our database allows us to track mortgages across snapshots. In particular, we use the loan-level information on borrower date of birth and the 6-digit postcode to track mortgages across snapshots since these variables, when combined, provide a unique identifier for each mortgage.

We start with the 2015H1 snapshot as the base, and merge data from subsequent snapshots using this unique identifier. Thus, for each mortgage we track whether it is discontinued between specific snapshots and whether it originated in any of the snapshots. Exploiting our ability to observe mortgages across snapshots, we also track whether a mortgage transitions across categories (discounted-to-reset rate or reset rate-to-discounted) between snapshots, or whether it continues in the same interest rate category.

Figure A.2 provides a complete picture of mortgage origination, closure (i.e. mortgage account closure) and category flows across mortgage snapshots. The first column shows the breakdown of mortgages in the 2015H1 snapshot into discounted rate and reset rate

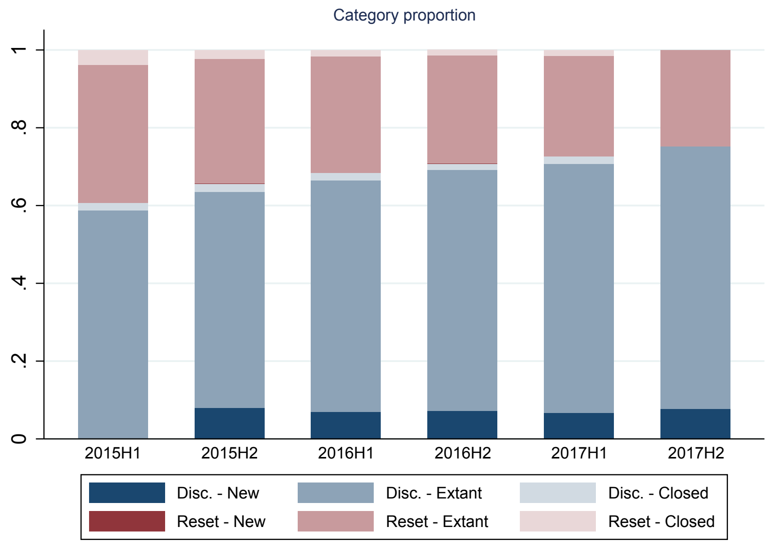
A.2-A.4 are before the data filtering and cleaning steps described in the current and the subsequent section.

mortgages, with the bars in lighter shade highlighting the mortgages of each category that are closed (or absent) by the next snapshot.[24] The second bar for 2015H2 has 6 components—'Discounted - Extant' - discounted mortgages in 2015H2 snapshot that were in the discounted category in the preceding snapshot; 'Discounted- Closed' - discounted mortgages in 2015H1 that are closed by the 2015H2 snapshot; 'Discounted - New' - discounted mortgages newly originated between 2015H1 and 2015H2; and similarly for reset rate mortgages. We observe all these 6 components for the snapshots in 2015H2, 2016H1, 2016H2 and 2017H1. For 2017H2, while we observe flows across interest rate categories (when compared to 2017H1), given lack of data for the subsequent snapshot, we do not observe which mortgages from the 2017H2 snapshot are closed by 2018H1.

Figure A.3 shows the value-weighted average interest rate for mortgages in the discounted and reset rate categories in a given snapshot based on their source (new mortgage, and if not, whether in the same or distinct category in the preceding snapshot). The figure shows that the average interest rate of new discounted mortgages is lower than that for new reset rate mortgages, a gap that has gone up across the snapshots. Given the lower interest rates for discounted mortgages, it is not surprise that mortgages undergoing a discounted—>reset rate transition between snapshots experience a sharp increase in the observed average interest rate, where as mortgages undergoing discounted—>discounted, or reset rate—>discounted transitions have lower average interest rate in comparison.

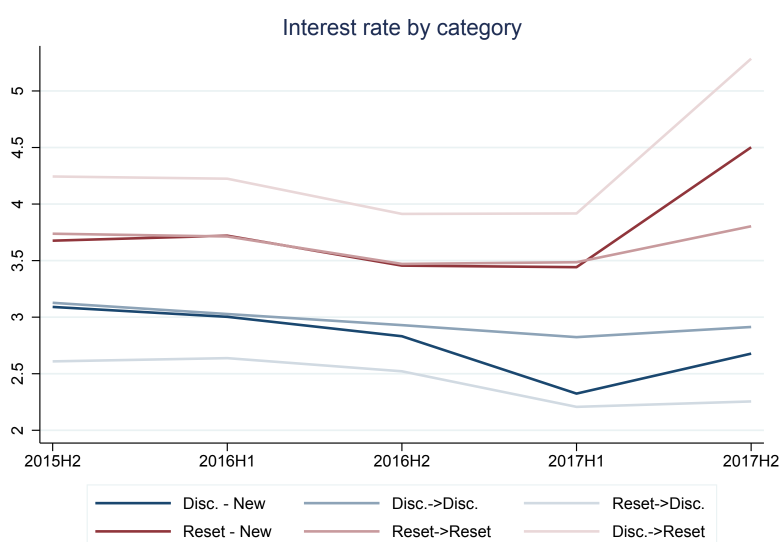
[24]There is a mass of mortgage reported with 0 balance in each snapshot. The identifiers of these mortgages are almost always absent in the next snapshot, indicating that the mortgage account is closed sometime in between the two snapshots. We treat mortgages with zero balance as being closed in the snapshot in which the zero balance is reported, and the characteristics of zero-balance mortgages are not reported in the summary statistics or used to generate the data moments used for estimation.

Figure A.2: **Proportion of Mortgages Under Discounted or Reset Rates**



The figure above shows the proportion of mortgages under discounted rates (dark blue) and the Standard Variable Rate in the mortgage stock reported at a half-yearly period from 2015H1 to 2017H2. Mortgages that are new in a given snapshot, and those that are discontinued in the next snapshot are shown in a darker and lighter shade, respectively.

Figure A.3: **Value-weighted Interest Rate**



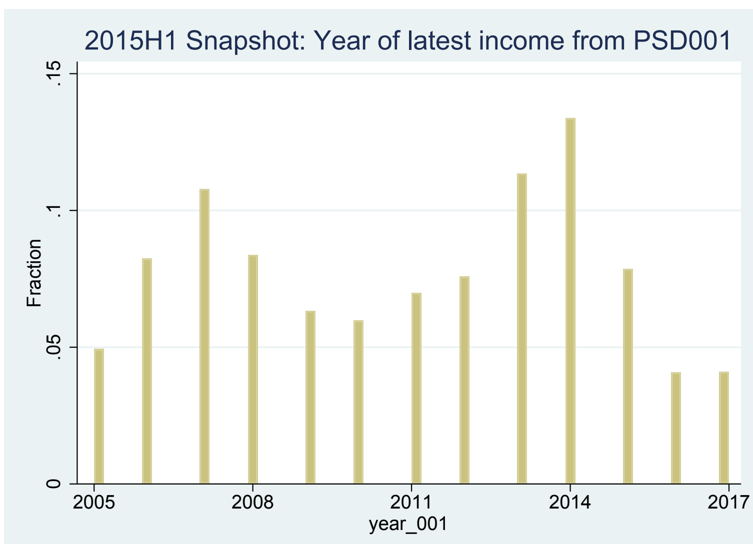
The figure above shows the value-weighted interest rate for mortgages under discounted rates and the Standard Variable Rate in the mortgage stock reported at a half-yearly period from 2015H1 to 2017H2.

A.4 Merging data on borrower income

The PSD 007 dataset on mortgage stock does not include information on borrower incomes. However, borrower incomes are reported to lenders at the time of loan origination, which is available in the PSD 001 dataset. We use the same variable used to merge information across stock snapshots (since it uniquely identified a mortgage) to merge the stock data with loan originations data. The resulting income for each loan in the mortgage stock is the latest income reported to the lender at the time of origination (the first instance of the mortgage being issued, or in a subsequent refinancing round), and is adjusted to the snapshot under consideration using local-area level income indices to obtain comparable income levels across the borrowers in a given snapshot. The local-area income indices are obtained from the Office of National Statistics.

Figure A.4 shows the latest year for which the borrower income is available for the mortgages in the 2015H1 snapshot. The year for the reported income at the time of origination does not vary across regions (see Table A.5) or across income bins.

Figure A.4: Latest year of reported income from PSD001



The figure above shows the latest year for which income data is available for mortgages in the 2015H1 snapshot. The latest available income is restated to the 2015H1 level using local-area indices to get comparable income levels across borrowers in the same snapshot.

Table A.5: 2015H1 Snapshot: Year of latest income, by region

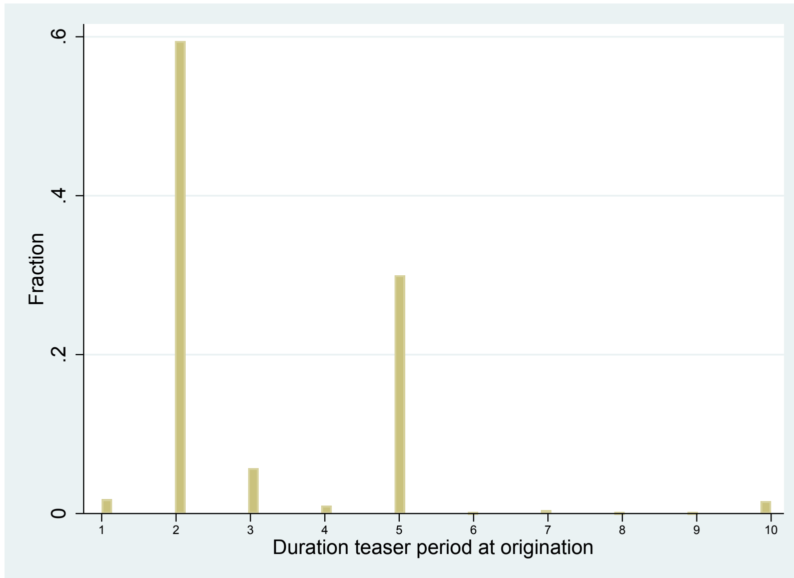
	count	mean	p10	p25	p50	p75	p90
East Midlands (England)	316,984	2,011	2,006	2,008	2,011	2,014	2,015
East of England	409,269	2,011	2,006	2,008	2,012	2,014	2,015
London	463,648	2,011	2,006	2,008	2,012	2,014	2,015
North East (England)	180,846	2,010	2,006	2,007	2,011	2,014	2,015
North West (England)	489,632	2,011	2,006	2,007	2,011	2,014	2,015
Northern Ireland	102,644	2,010	2,006	2,007	2,009	2,013	2,015
Scotland	428,209	2,011	2,006	2,008	2,011	2,014	2,015
South East (England)	608,188	2,011	2,006	2,008	2,012	2,014	2,015
South West (England)	359,110	2,011	2,006	2,008	2,012	2,014	2,015
Wales	199,310	2,011	2,006	2,007	2,011	2,014	2,015
West Midlands (England)	356,806	2,011	2,006	2,007	2,011	2,014	2,015
Yorkshire and The Humber	359,067	2,011	2,006	2,007	2,011	2,014	2,015

A.5 Discounted period at the time of mortgage origination

Table 1 showed that the average remaining discounted period is around 2 years across our mortgage snapshots at a half-yearly frequency from 2015H1 to 2017H2. Discounted mortgages in the UK pay a fixed interest rate during the discounted period, and the fixed-rate period at the time of origination for most mortgages fall between 2-5 years. Figure A.5 shows the distribution of discounted periods at the time of origination for all discounted mortgages issued in the UK from 2015-2017.[25] Consistent with the stock data, the modal discounted period at the time of origination is 2 years, followed by fixation period of 5 years.

[25]Source: database on mortgage originations, PSD001.

Figure A.5: **Discounted Period at Origination**



The figure above shows the discounted period for mortgages at the time of origination for all discounted mortgages issued from 2015H1 to 2017H2 in the United Kingdom.

B Notes on Basic Data Cleaning

In the preceding section, we discussed filtering out mortgages with anomalous interest rate types, tracker mortgages and mortgages under an unspecified 'other' category. This section discusses additional data cleaning steps undertaken to filter out observations with anomalous or inconsistent data on remaining discounted period, balance, interest rate, remaining term and borrower age. While discussing each filtering step, we show summary statistics for the raw database from PSD007 for discounted and reset rate mortgages, and the filtered database following the data cleaning steps described in the following sub-sections.

B.1 Reported Remaining Discounted Period

Table A.6 shares the summary statistics for the remaining period on discounted rates (in months) for discounted mortgages across the six snapshots. The mean and standard deviation of remaining discounted period is consistent across the snapshots, except 2015H1. This is driven primarily by mis-specification of reset rate mortgages as discounted mortgages in 2015H1. The histogram for remaining discounted period in 2015H1 (Figure A.6, with year based categories along the x-axis) clearly shows this misreported data in the large mass of mortgages with remaining discounted periods greater than 10 years, where the remaining term of mortgages is reported as the remaining discounted period. This is not seen in the remaining discounted periods for other snapshots. We restate reset rate (or SVR) mortgages as being not incentivised, and restate any reported discounted period for reset rate mortgages as missing data.

In addition, across all snapshots, there are few mortgages with remaining discounted periods less than -1 years, and greater than 11 years. All such observations are dropped from the sample. Figure A.6 shows the histogram for the remaining discounted period for the raw and filtered database for the 2015H1 snapshot, with remaining discounted periods in years along the x-axis.

Table A.6: **Remaining Discounted Period**

(a) Raw database

	mean	sd	p10	p25	p50	p75	p90
Remaining discounted period in 2015H1	35	64	5	12	22	40	55
Remaining discounted period in 2015H2	27	27	4	11	21	36	53
Remaining discounted period in 2016H1	26	27	4	12	21	35	52
Remaining discounted period in 2016H2	25	26	5	11	20	34	51
Remaining discounted period in 2017H1	25	26	4	10	18	33	52
Remaining discounted period in 2017H2	25	27	4	10	19	36	54

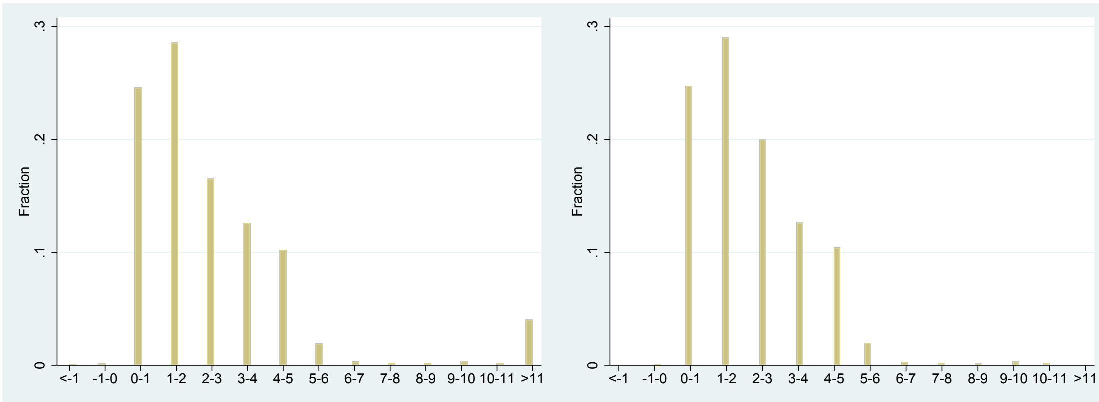
(b) Filtered database

	mean	sd	p10	p25	p50	p75	p90
Remaining discounted period in 2015H1	25	18	5	12	22	37	51
Remaining discounted period in 2015H2	25	18	4	11	21	36	52
Remaining discounted period in 2016H1	25	18	5	12	21	35	51
Remaining discounted period in 2016H2	24	18	5	11	20	33	51
Remaining discounted period in 2017H1	24	19	4	10	18	33	51
Remaining discounted period in 2017H2	24	20	4	10	19	35	54

The above tables shows summary statistics for the remaining discounted period for discounted mortgages across the PSD007 snapshots. Panel (a) shows the summary statistics for the raw database; panel (b) shows the summary statistics after the filtering steps described in Section B.1.

Figure A.6: **Mortgages in 2015H1: Histogram of Remaining Discounted Period**

(a) Raw database (b) Filtered database



The above figure shows a histogram for the remaining discounted period for discounted mortgages in the 2015H1 snapshot. Remaining discounted period is expressed in years along the x-axis. Panel (a) shows the summary statistics for the raw database; panel (b) shows the summary statistics after the filtering steps described in Section B.1.

B.2 Reported Balance

We discussed briefly in Section A.3 the mass of zero balance mortgages across snapshots. These mortgages are discontinued in subsequent snapshots, and we treat such instances of zero balance mortgages as a case of delayed reporting of account closure. To be consistent across snapshots, a mortgage with a zero balance in a given snapshot is treated as being discontinued in the same snapshot rather than subsequent one. Further we treat all the data for zero balance mortgages (balance, remaining term, interest rate) as missing and, therefore, do not include such mortgages in the moments used for estimation and the summary statistics reported in the paper proper.

The summary statistics for loan balances across snapshots (in Table A.7, and particularly that for discounted mortgages (in Table A.8) shows that all the moments (including mean and s.d.) for loan balances in 2017H2 are higher than that for other snapshots. However, the loan balance moments for reset rate mortgages are stable across the snapshots.

We find that the high mean and s.d. for discounted mortgages in 2017H2 is driven by misreported loan balances for two lenders. Specifically for the discounted mortgages issued by these two lenders in 2017H2, we replace the reported loan balance in 2017H2 with the estimated amortized loan balance based on the reported loan balance, remaining term, and discounted interest rate of 2017H1.[26] Figure A.7 shows the average balance for discounted-to-discounted flows across the snapshots: panel (a) shows the average loan balance before restatement, panel (b) shows the average loan balance after restating the loan balance for discounted mortgages in 2017H2 for the two aberrant lenders as described above.

Finally, there are very few instances of mortgages with negative loan balances. We drop all such observations from the sample. Figure A.8 shows a histogram of loan balances in 2015H1 (with categories based on £ across the x-axis) for both the raw database, and the filtered database.

[26] We estimate an amortized loan balance for 2017H2 only for discounted mortgages with at least 6 months on discounted periods in 2017H1. Further, we do this estimation only for mortgages which are on a capital and interest payment plan; i.e. we do not restate the 2017H2 loan balance for the small balance of interest only discounted mortgages in the stock of the two aberrant lenders.

Table A.7: **Balance over Mortgage Snapshots**

(a) Raw database

	mean	sd	p10	p25	p50	p75	p90
Balance in 2015H1	118,143	108,109	29,300	59,534	98,043	149,398	219,929
Balance in 2015H2	119,800	115,850	25,000	57,743	98,198	151,763	227,112
Balance in 2016H1	124,175	121,525	28,302	59,246	99,952	155,683	235,932
Balance in 2016H2	128,213	126,975	29,279	60,250	101,966	160,238	244,876
Balance in 2017H1	130,608	127,003	30,000	60,775	103,092	162,999	250,191
Balance in 2017H2	143,369	148,222	29,357	61,902	108,069	178,562	286,897

(b) Filtered database

	mean	sd	p10	p25	p50	p75	p90
Balance in 2015H1	123,325	98,092	38,770	64,821	101,620	152,765	223,988
Balance in 2015H2	127,332	105,483	38,758	65,237	103,424	157,122	233,834
Balance in 2016H1	130,092	111,117	38,309	65,061	104,278	160,214	241,596
Balance in 2016H2	133,558	116,289	38,336	65,680	106,060	164,432	250,009
Balance in 2017H1	134,998	117,715	37,984	65,622	106,807	166,905	254,782
Balance in 2017H2	140,451	125,369	37,953	66,386	109,479	173,774	269,100

The above tables shows summary statistics for the outstanding balance for *discounted* mortgages across the PSD007 snapshots. Panel (a) shows the summary statistics for the raw database; panel (b) shows the summary statistics after the filtering steps described in Section B.2.

Table A.8: **Balance for Discounted Mortgages**

(a) Raw database

	mean	sd	p10	p25	p50	p75	p90
Discounted Balance in 2015H1	132,914	118,277	41,060	69,340	109,398	164,822	243,679
Discounted Balance in 2015H2	137,404	126,456	41,204	70,310	111,618	169,675	253,591
Discounted Balance in 2016H1	141,611	131,665	42,089	71,260	113,446	173,908	263,006
Discounted Balance in 2016H2	145,534	137,818	42,032	72,205	115,718	178,832	272,083
Discounted Balance in 2017H1	147,971	136,726	42,466	72,871	117,125	181,872	276,900
Discounted Balance in 2017H2	164,083	159,831	43,273	75,735	124,746	202,136	320,884

(b) Filtered database

	mean	sd	p10	p25	p50	p75	p90
Discounted Balance in 2015H1	135,620	105,843	45,415	71,955	111,280	166,743	245,907
Discounted Balance in 2015H2	140,685	113,953	46,150	73,397	113,917	172,054	256,634
Discounted Balance in 2016H1	144,411	119,978	46,314	74,071	115,561	176,247	265,964
Discounted Balance in 2016H2	148,361	125,132	46,642	75,134	117,917	181,124	274,872
Discounted Balance in 2017H1	149,805	125,805	46,347	75,384	119,006	183,947	279,104
Discounted Balance in 2017H2	155,343	132,907	46,725	76,584	121,860	190,882	292,558

The above tables shows summary statistics for the outstanding balance for *discounted* mortgages across the PSD007 snapshots. Panel (a) shows the summary statistics for the raw database; panel (b) shows the summary statistics after the filtering steps described in Section B.2.

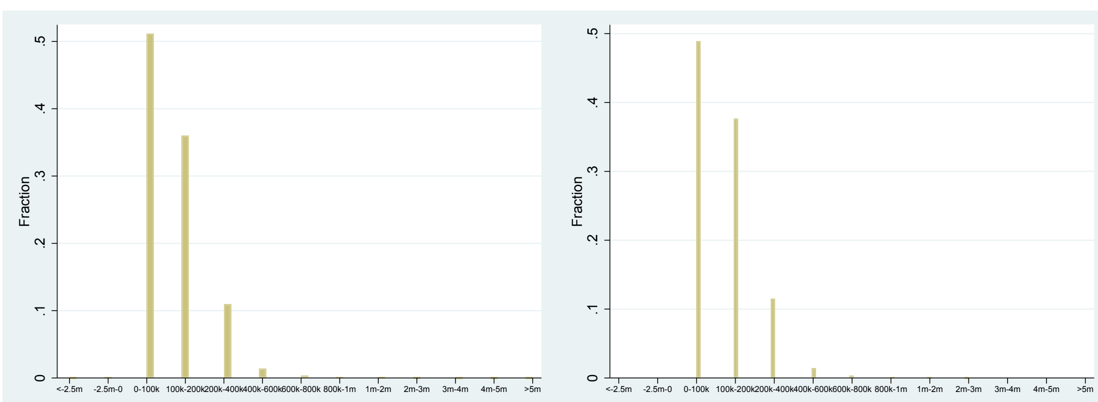
Figure A.7: **Balance for Discounted Flows**

(a) Raw database (b) Filtered database

The above tables shows the average balance of discounted mortgages based on their source (same category, cross-category) in a given snapshot. Panel (a) shows the summary statistics for the raw database; panel (b) shows the summary statistics after the filtering steps described in Section B.2.

Figure A.8: **Mortgages in 2015H1: Histogram of Balance**

(a) Raw database (b) Filtered database



The above figure shows a histogram for the outstanding balance for mortgages in the 2015H1 snapshot. Outstanding balance is expressed in categories in £ along the x-axis. Panel (a) shows the summary statistics for the raw database; panel (b) shows the summary statistics after the filtering steps described in Section B.2.

B.3 Reported Interest Rate, Remaining Term and Age

The filtering done based on the reported interest rate, remaining term, and borrower age is described below.

Interest rate: We drop all instances of negative interest rates, and winsorize at the 99.9% for each snapshot to address outliers which clearly appear to be a case of misreporting (for instance, interest rates of >1000%).

Remaining term: We drop all instances of negative remaining terms, and winsorize at the 99.9% for each snapshot to address outliers which clearly appear to be a case of misreporting (for instance, remaining term of of 9999 months).

Borrower age: We drop all instances of reported -ve age of borrowers.

Summary statistics of interest rates, remaining term, and borrower age in the raw and filtered databases will be shared separately in the Online Appendix.

C Description of local area variables

Demographic variables for local authority districts (LAD) are taken from the 2011 census, which can be accessed via the Office for National Statistics' Nomis service. Summary statistics for these variables can be found in table A.9; we consider measures of health, the relative size of the black population, the unemployment rate, education levels, long term economic activity, occupations, as well as levels of deprivation and social grade.[27] Figures A.9 to A.16 illustrate the geographical inequalities in these variables. We will explore the association of cross-subsidies across granular UK local areas with the characteristics described above in subsequent work related to this project.

Variable definitions. Bad health is defined as the proportion of individuals in a LAD reporting their health as either bad or very bad. Analogously, black is defined as the percentage of individuals who report to be either African, Caribbean, or other black. The unemployment rate is defined as the percentage of individuals who report as unemployed out of those asked who are economically active net of full time students. Long term inactive is defined as the proportion of people who are either long term unemployed or have never worked. An individual is considered to work in an elementary occupation if they report to work in a level 1 occupation of Standard Occupational Classification. A household is defined as deprived if they report to be deprived along one dimension of employment, education, health and disability, and household overcrowding. Social Grade is a socioeconomic classification used to inform the analysis of spending habits and consumer attitudes; we define low social grade as individuals falling below the AB level.

[27]Note that deprivation and social grade data from the 2011 census is only available for England and Wales.

Table A.9: Census 2011 Summary Statistics

Statistic	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
Bad health	0.0540	0.0147	0.0431	0.0518	0.0622
Black	0.0194	0.0403	0.0022	0.0048	0.0150
Unemployed	0.0645	0.0230	0.0459	0.0614	0.0770
No graduate degree	0.1202	0.0172	0.1120	0.1193	0.1253
Long term inactive	0.0489	0.0231	0.0311	0.0426	0.0600
Elementary occupation	0.1105	0.0240	0.0953	0.1116	0.1243
Deprived	0.3263	0.0176	0.3146	0.3238	0.3371
Low social grade	0.7694	0.0790	0.7217	0.7822	0.8269

Figure A.9: **Map: percentage deprived**

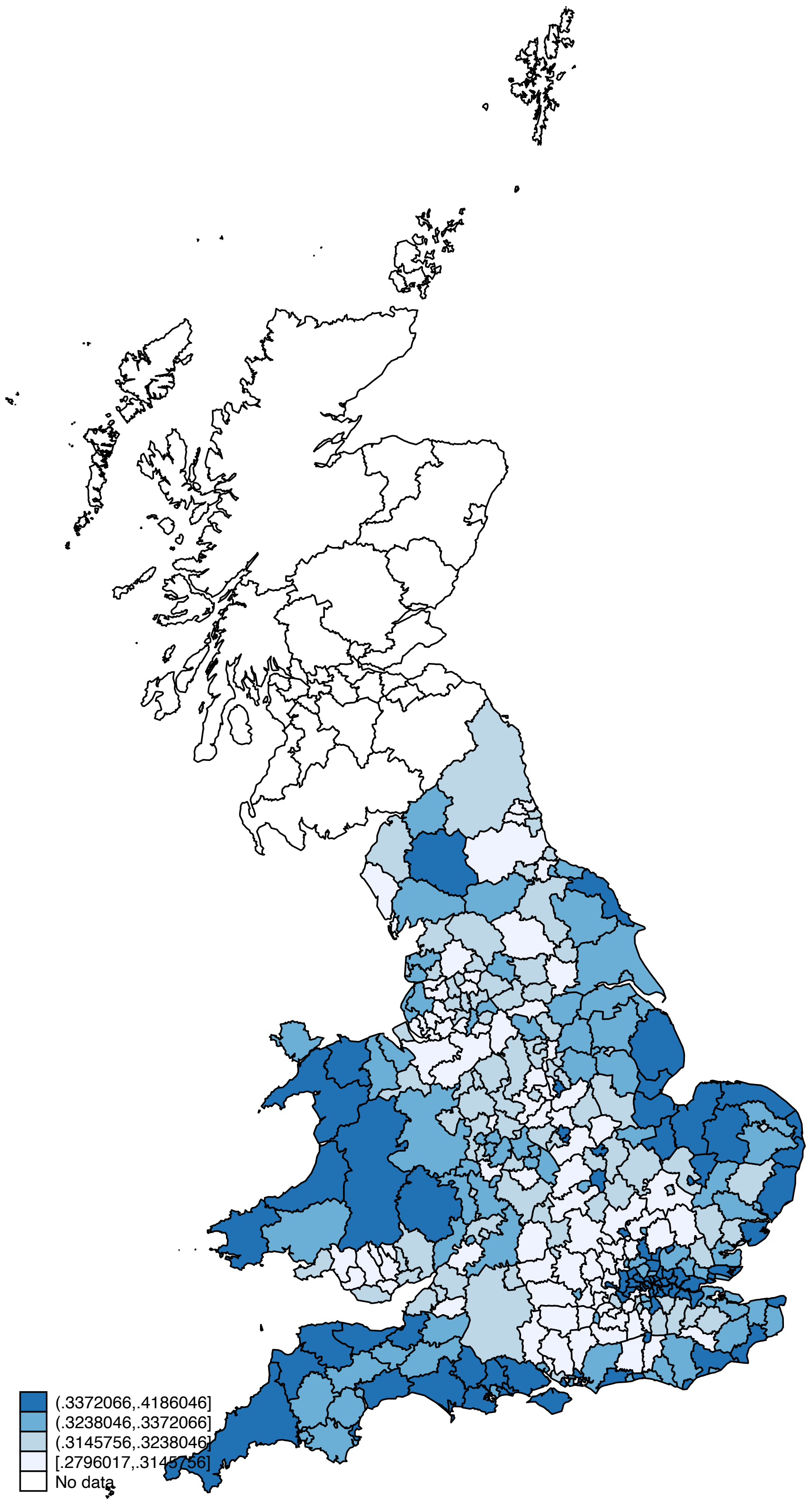


Figure A.10: **Map: percentage with bad health**

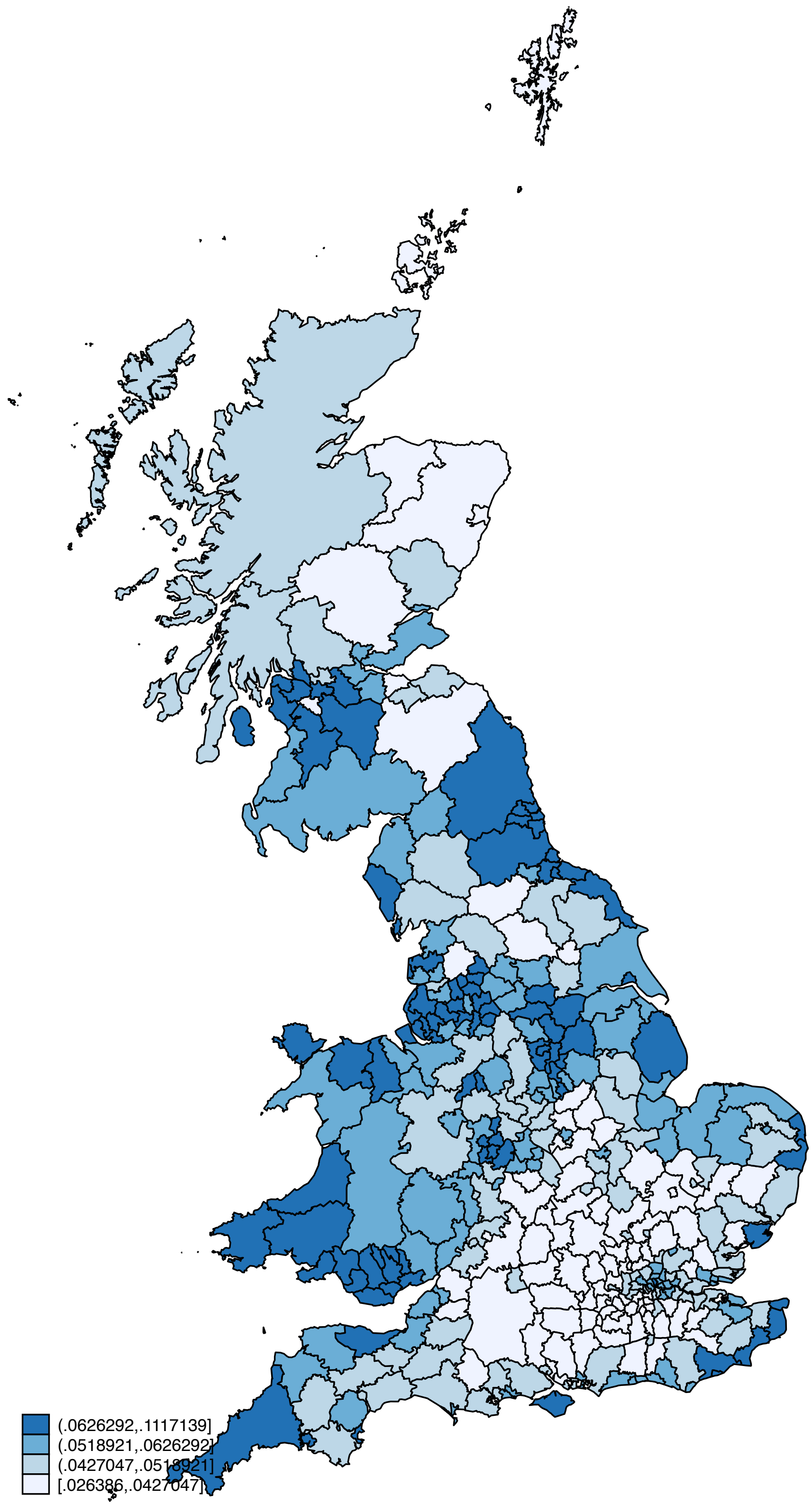


Figure A.11: **Map: percentage black population**

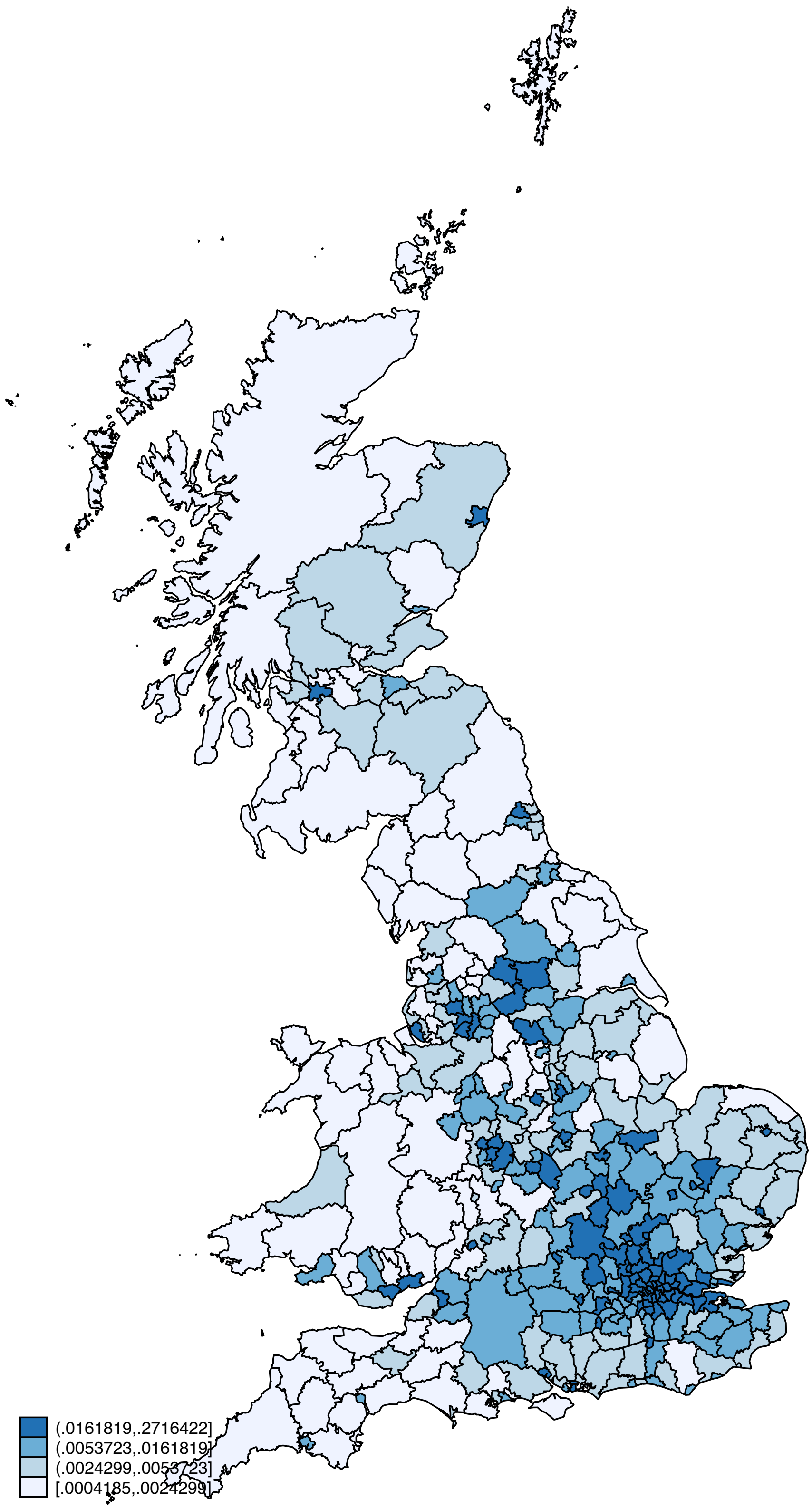


Figure A.12: **Map: percentage with elementary occupation**

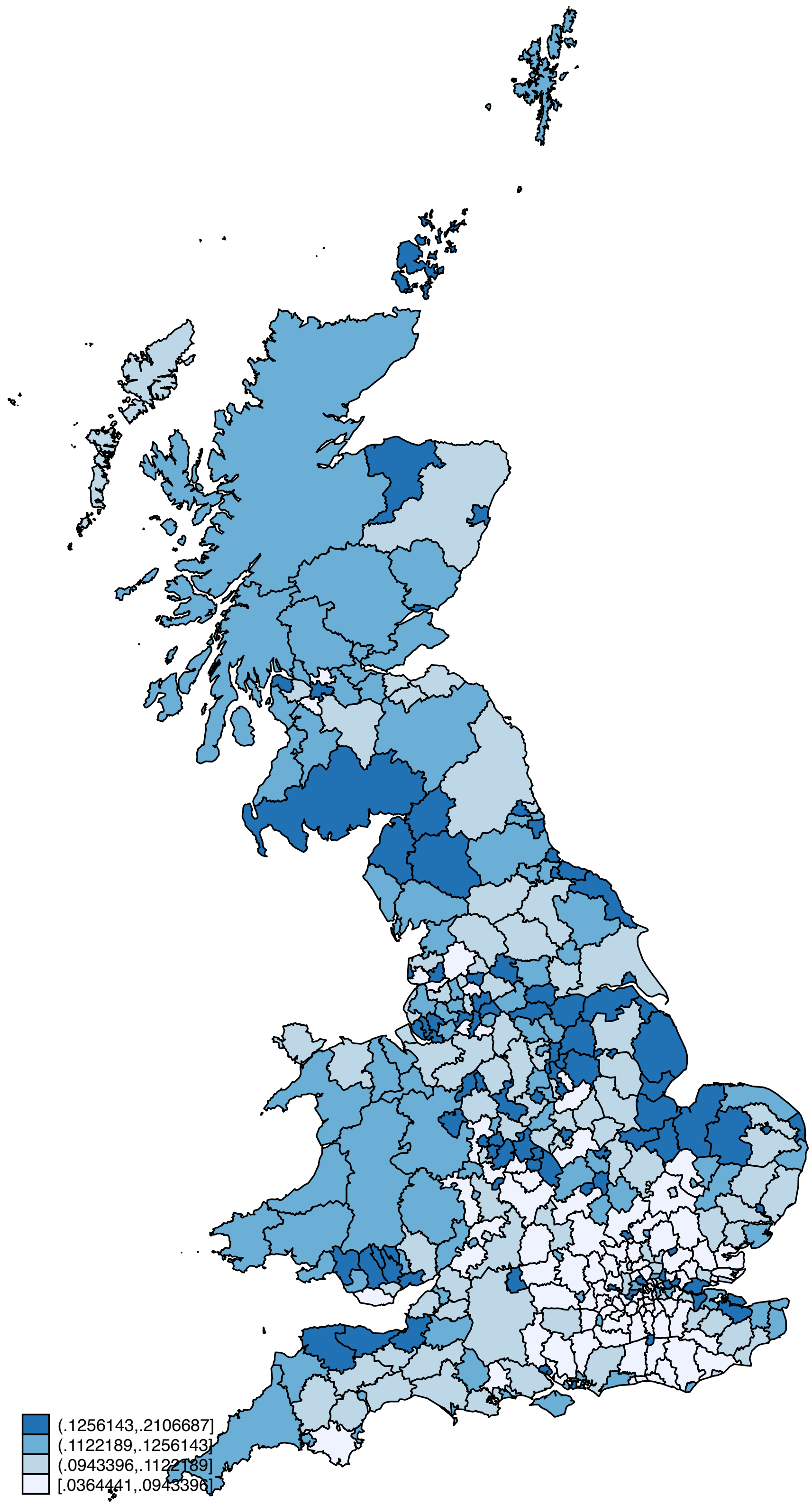


Figure A.13: **Map: percentage with long term inactivity**

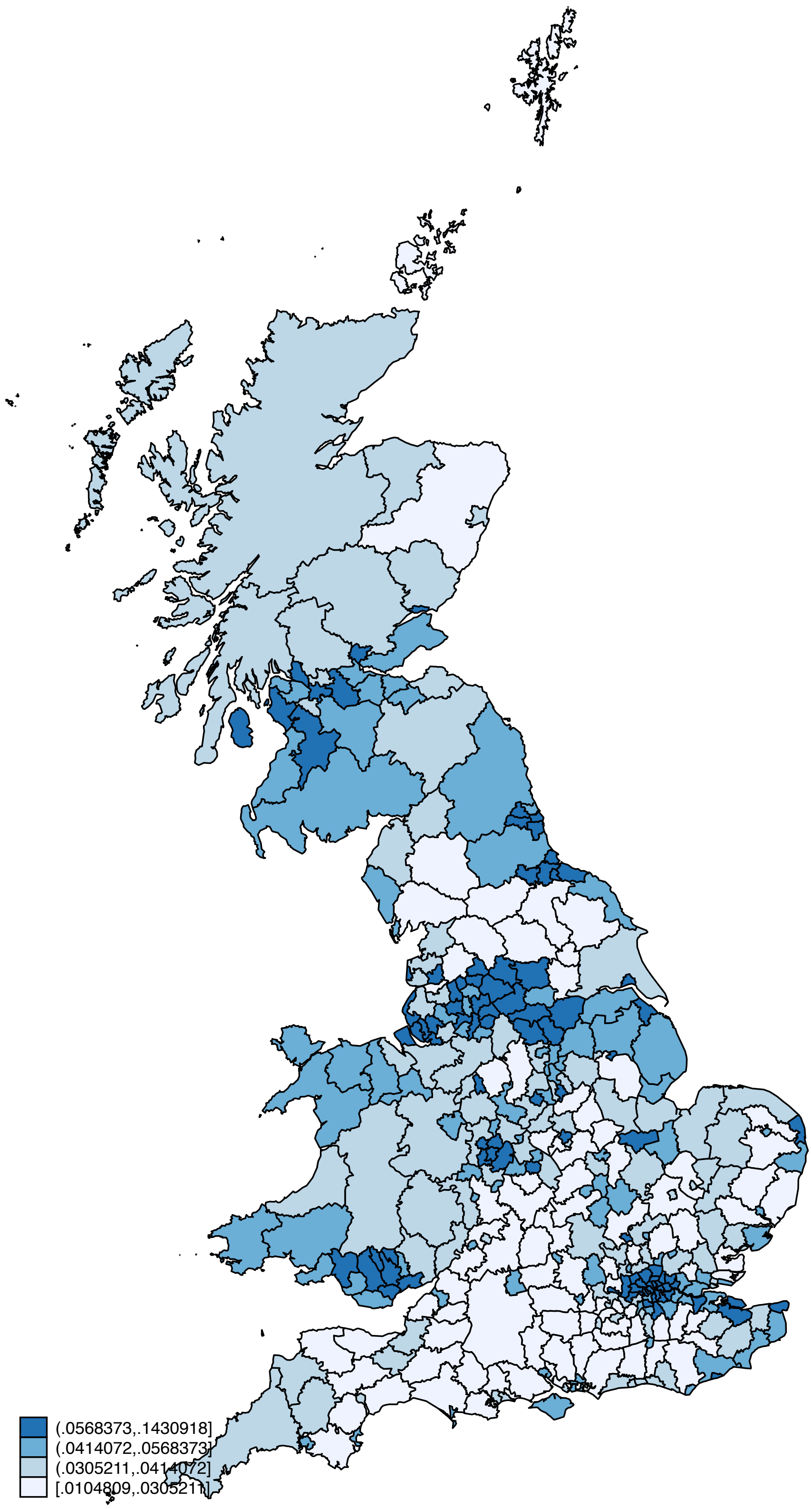


Figure A.14: **Map: percentage with low social grade**

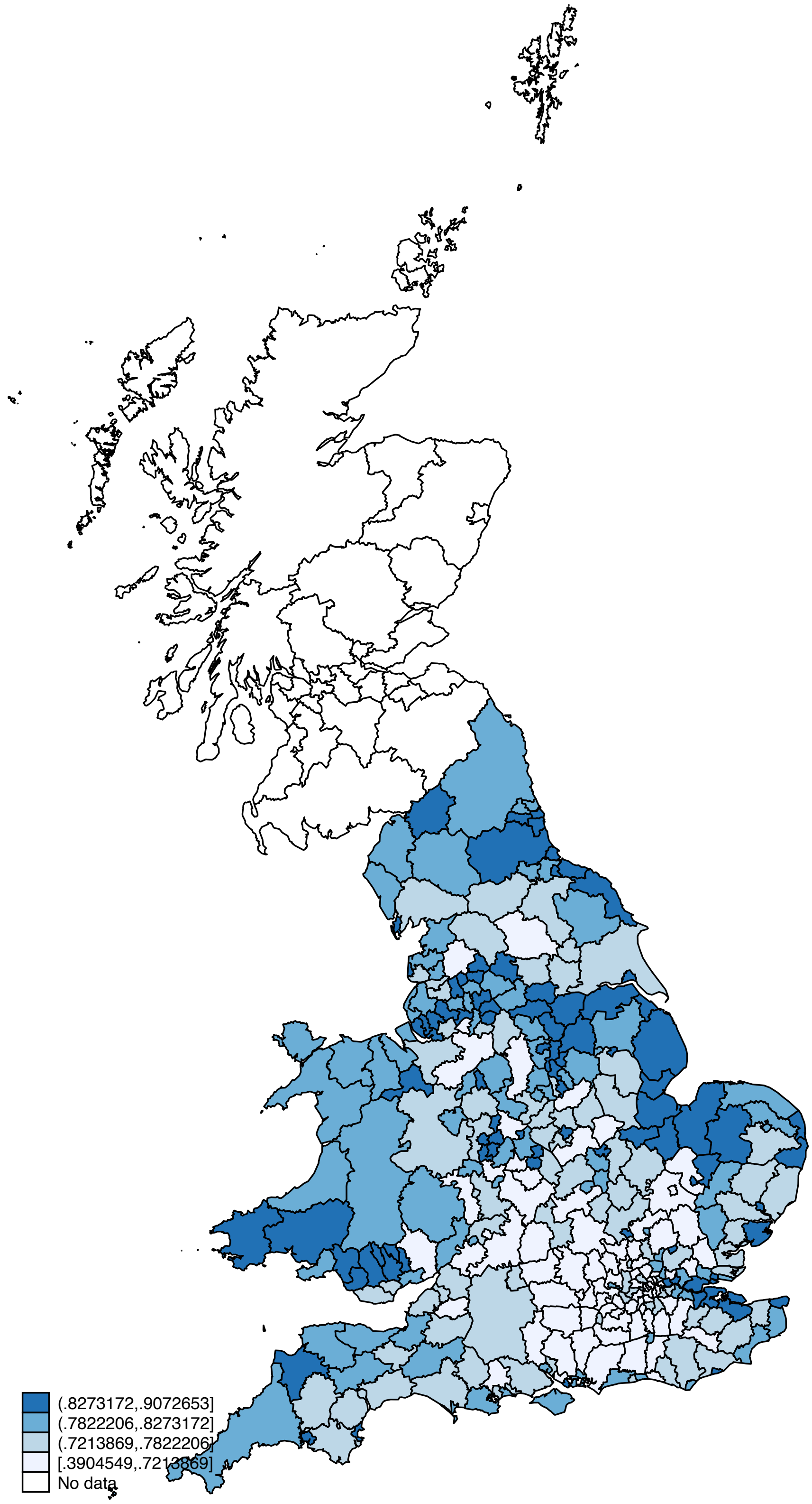


Figure A.15: **Map: percentage with no graduate degree**

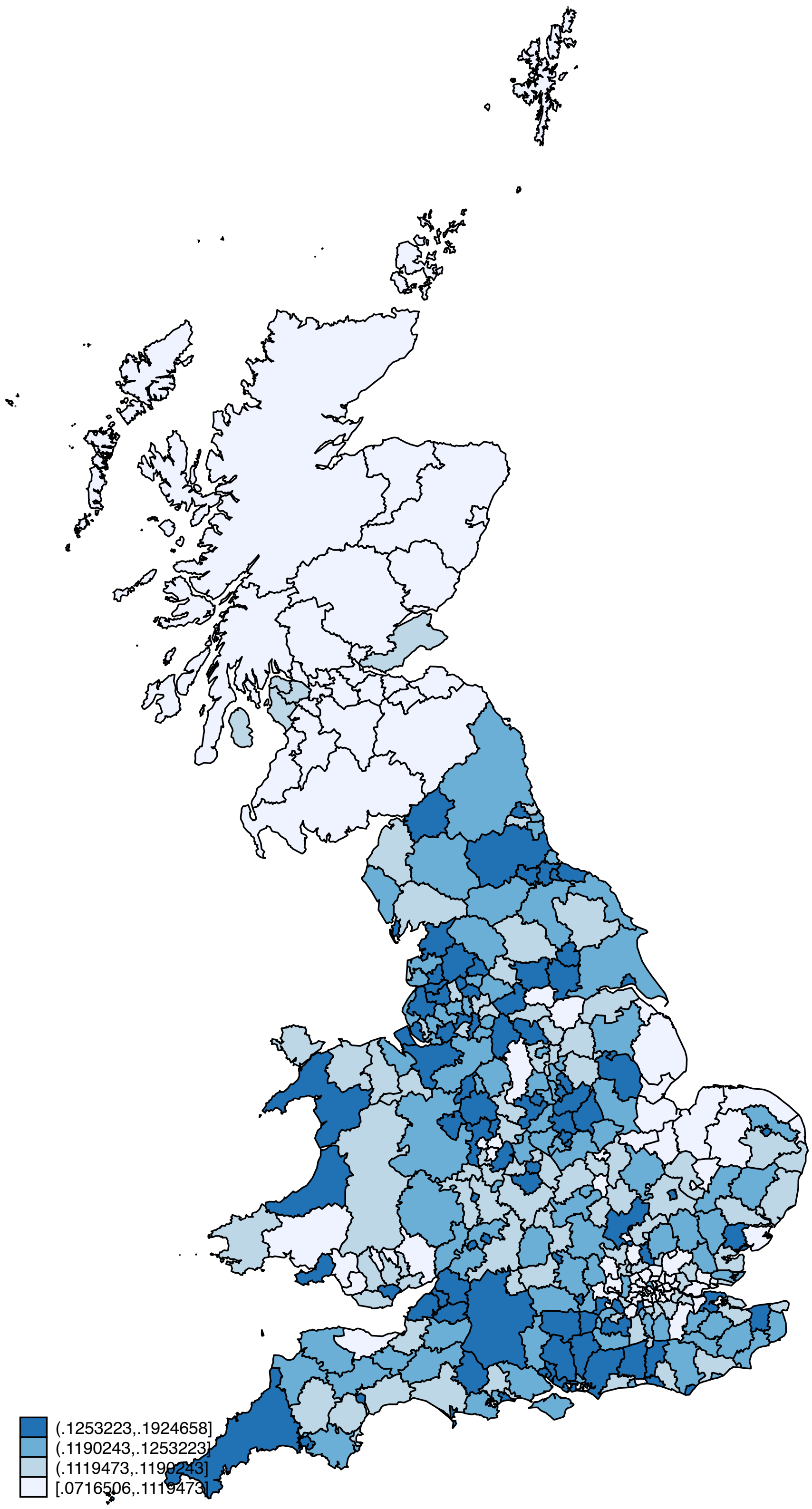


Figure A.16: **Map: unemployment rate**