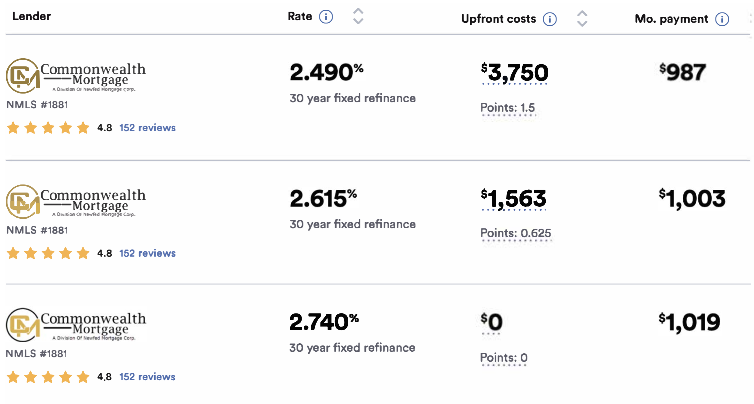
Internet Appendix

This appendix supplements the empirical analysis of this paper. Below is a list of the sections contained in this appendix.

Table of Contents

A.1 Additional Background About Rate and Upfront Closing Costs

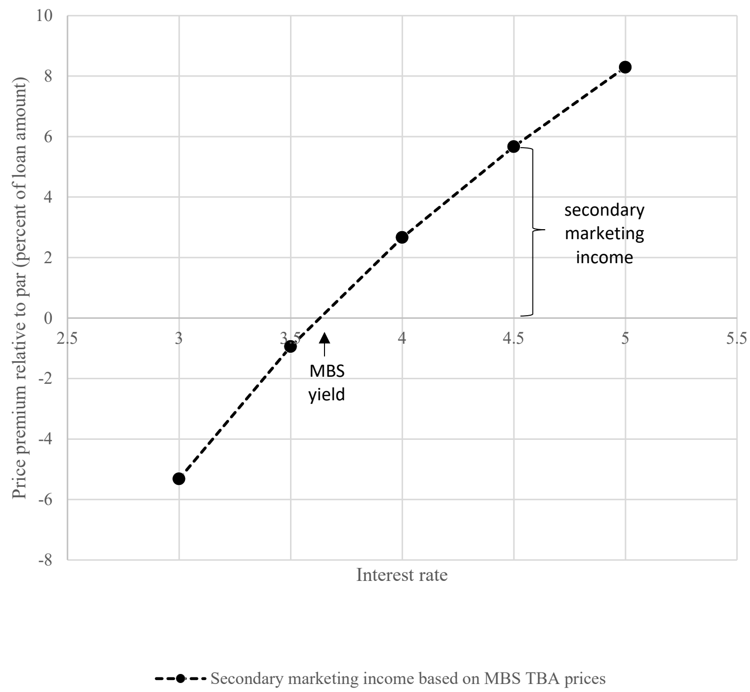
Figure A.1: Rate and upfront closing costs trade-offs facing mortgage borrowers



Lender	Rate	Upfront costs	Mo. payment
Commonwealth Mortgage A Division Of Newfed Mortgage Corp. NMLS #1881 4.8 152 reviews	2.490% 30 year fixed refinance	$3,750 Points: 1.5	$987
Commonwealth Mortgage A Division Of Newfed Mortgage Corp. NMLS #1881 4.8 152 reviews	2.615% 30 year fixed refinance	$1,563 Points: 0.625	$1,003
Commonwealth Mortgage A Division Of Newfed Mortgage Corp. NMLS #1881 4.8 152 reviews	2.740% 30 year fixed refinance	$0 Points: 0	$1,019

Note: Figure A.1 shows a screenshot obtained by the author from Bankrate.com for a $250,000 refinancing mortgage on September 18, 2021. It shows how a borrower may choose to pay 0 points for a 2.740% interest rate mortgage, 0.626 points for a 2.615% interest rate mortgage, or 1.5 points for a 2.490% interest rate mortgage.

Figure A.2: Secondary marketing income as a function of interest rates



Note: Figure A.2 plots the FNMA MBS TBA prices on January 2, 2014 expressed as a percentage point premium/discount over the loan amount on the y-axis for a variety of coupon rates on the x-axis. Secondary marketing income is the extent to which the secondary market value of the mortgage is above its principal balance.

A.2 Data Construction and Summary Statistics

A.2.1 Optimal Blue-HMDA sample

I constructed the Optimal Blue-HMDA sample by merging the Optimal Blue rate locks from 2018–2019 with the public HMDA data. Because Optimal Blue contains a lender identifier number but no lender names, the merge proceeds in two steps: (1) an initial match based on loan characteristics, and (2) a second filtering based on a correspondence between the lender ID in Optimal Blue and an anonymized version of HMDA lender IDs implied by the first step.

The initial match was made using loan amount, rate, year, loan type, loan purpose, loan term, ZIP code (with all ZIP codes corresponding to an HMDA census tract included), and up to a 5% difference in LTV with all matches kept in the data set. Then, for the second step I impose the requirement that the lender ID in Optimal Blue is matched to an anonymized version of HMDA lender ID at least 10% of the time.[1] Overall, this two-step procedure uniquely matches 1,186,906 out of 2,318,940 locks for 30-year, conforming fixed-rate mortgages, implying a match rate of 51%. The match rate is comparable to a 66% "lock pull-through rate," which is the percent of rate locks that turn into originated loans, that I understand to be reasonable based on conversation with representatives from Optimal Blue.

In terms of variable definitions, I construct a Black dummy equal to one if the mortgage has a HMDA-derived race variable of "Black or African American." The Hispanic dummy is equal to one if the mortgage has a HMDA derived ethnicity variable of "Hispanic or Latino." The Single Male and Single Female dummies are inferred from the HMDA-derived gender. Summary statistics for these samples are shown in the table below.

[1]The 10% requirement was set purposefully low to include cases where the Optimal Blue lender ID may not correspond to a HMDA reporter for example in the case of correspondent lending. It is sufficient to reduce the percent of matches that are non-unique from 49.6% to 3.9%.

Table A.1: Summary statistics for the 2018–2019 Optimal Blue-HMDA sample

	All		Black		Hispanic	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Loan amount ($'000s)	256695.6	117785.7	242574.6	117351.2	243938.2	112333.0
Origination cost ($)	1516.0	1807.2	1657.6	2062.3	1849.0	1969.4
Total loan cost ($)	3902.6	2362.4	4222.6	2713.2	4487.1	2547.2
Credit Score	747.9	44.5	728.7	47.3	732.7	45.5
LTV (%)	80.4	15.0	84.9	13.5	82.6	14.7
DTI (%)	34.973	9.681	37.363	8.849	38.239	8.600
Interest rate	4.544	0.579	4.692	0.612	4.674	0.603
Points paid	0.307	0.461	0.395	0.489	0.387	0.487
First-time home buyer (d)	0.252	0.434	0.436	0.496	0.268	0.443
Single Female (d)	0.330	0.470	0.356	0.479	0.441	0.497
# Observations	1,041,807		42,793		92,598	

Notes: This table reports summary statistics from the 2018–2019 Optimal Blue-HMDA merged sample. Loan amount is expressed in thousands of dollars, origination costs are expressed in dollars, credit score is the borrower's Optimal Blue credit score at origination, and LTV, interest rate are expressed in percentage points. The label (d) denotes dummy variables.

A.2.2 Optimal Blue-HMDA-CRISM sample

I also construct a merge between Optimal Blue, HMDA, and CRISM data sets for mortgages originated between 2013–2019, with loan performance until May 2022. The CRISM data set is an anonymous credit file match from Equifax consumer credit database to Black Knight's Mcdash loan-level Mortgage Data set. My Optimal Blue-HMDA-CRISM sample was constructed by joining together three merges, (i) the 2018–2019 Optimal Blue and HMDA merge described in Section A.2.1, (ii) a 2013–2017 Optimal Blue and HMDA merge, and (iii) the 2013–2019 Optimal Blue and CRISM merge.

Similar to the 2018-2019 Optimal Blue and HMDA merge, the 2013–2017 Optimal Blue and HMDA merge was also conducted in two steps, with an initial step based on loan characteristics, and a second step based on a correspondence between the Optimal Blue lender ID and an anonymized HMDA lender ID. A separate merge was conducted because the data fields in 2013–2017 HMDA are different than those in 2018–2019 HMDA: the interest rate, loan term, and LTV fields were not available, while loan amount was given in finer detail.

The first step for the 2013–2017 Optimal Blue to HMDA match was made using loan amount, year, loan type, loan purpose, occupancy, ZIP code (with all ZIP codes corresponding to an HMDA census tract included) with all matches kept in the data set. Then, for the second step I impose the requirement that the lender identifier in Optimal Blue is matched to an HMDA respondent ID at least 10% of the time.[2] Overall, this two-step procedure uniquely matches 1,382,057 out of 2,563,550 locks for 30-year, conforming fixed-rate mortgages, implying a match rate between locks to originated mortgages of 54%. The match rate is again comparable to a 66% "lock pull-through rate," which I understand to be reasonable based on industry sources.

The 2013–2019 Optimal Blue to CRISM match was made in one step. The variables used for matching are the loan amount, ZIP code, month of origination (which I require to lie within the date of the lock and the date of the lock plus the lock term), loan type, loan term, loan purpose, Equifax Risk Score (within 20 points of the Optimal Blue credit score), LTV (within 5%), and the rate. The more detailed loan-level information enabled the match to proceed despite not having lender information. Overall, I uniquely matched 617,058 out of 5,269,107 locks for 30-year, conforming fixed-rate mortgages, implying a match rate between locks to originated mortgages in the CRISM data set of 12%. The lower match rate is reasonable because neither the CRISM data nor the Optimal Blue data covers all

[2]The 10% requirement was set purposefully low to include cases where the Optimal Blue lender ID may not correspond to an HMDA reporter for example in the case of correspondent lending. It is sufficient to reduce the percent of matches that are non-unique from 75.2% to 11.8%.

US mortgage originations, so the overlap between the two must be smaller than the overlap between Optimal Blue and HMDA as the HMDA does provide essentially complete coverage of all US mortgage originations.

Combining the three merges, I get an Optimal Blue-HMDA-CRISM sample with 360,291 loans. In terms of variable definitions, I construct a Black dummy equal to one if the mortgage has a 2018–2019 HMDA derived race variable of "Black or African American." The Hispanic dummy is equal to one if the mortgage has a HMDA derived ethnicity variable of "Hispanic or Latino." In the case of 2013–2017 HMDA, these dummies are defined using the algorithm of Bhutta and Canner (2013). The Single Male and Single Female dummies are inferred from the 2018–2019 HMDA derived gender or the applicant gender when no co-applicant is present in the case of 2013–2017 HMDA. Finally, the Credit Card Revolver dummy is set equal to 1 if the primary borrower on the mortgage has a credit card balance of greater than or equal to $10,000 at the time of origination while also having a credit card utilization of greater than 40%.

Summary statistics on this sample is shown in Table 1.

A.2.3 The LoanSifter data

The LoanSifter data contains information about rate and upfront closing cost (i.e., points) trade-offs in rate sheets, which are prices that loan originators and mortgage brokers can offer to clients in locking the loan. Because these are actual available prices within a lender, they allow me to observe the rate and point menus that borrowers face. The sample period runs from September 9, 2009 to December 31, 2014 and consists of rate sheets from a sample of lenders from 50 metropolitan areas. Rate sheets observations are at the lender-day level, and in rare cases where a lender issues more than one rate sheet on a given day the observations with the best prices are kept. Linear interpolation was used to estimate the rate at various levels of points, following Fuster, Lo, and Willen (2022). To compare the rate and points menus in the lender rate sheets to the MBS TBA prices, I focus on rate sheets for conforming, 30-year, fixed-rate mortgages with a loan-to-value ratio of 80% and a loan amount of greater than or equal to $300k.

Summary statistics for this data are shown in Table A.2.

Table A.2: Summary statistics for the LoanSifter data

Year	No. of Lenders	Rate at -2 points	Rate at 0 points	Rate at 2 points	N lender-days obs
2009	93	5.42	5.01	4.65	3923
2010	93	5.10	4.70	4.44	16025
2011	83	4.82	4.46	4.25	16589
2012	86	4.07	3.67	3.41	18105
2013	126	4.42	4.07	3.80	19993
2014	103	4.52	4.21	3.97	19446

Note: This table contains information on the number of distinct lenders, mean rate at 0 points, mean rate at 2 points, and number of distinct lender-day observations by year. The data set comes from LoanSifter. The interest rates at 0 points and at 2 points are estimated through linear interpolation for lenders that do not offer mortgages at exactly those points.

A.3 Pass-through of TBA prices to lender rate sheets

I examine how the secondary marketing income-interest rate trade-off matches the retail interest rate and upfront closing costs trade-off on average in the cross-section, with results in Figure A.3. I use data LoanSifter matched with MBS TBA pricing data from 2009Q3 to 2014. Following the methodology of Fuster, Lo, and Willen (2022), I focus on borrowers with a $300k conforming mortgage, 700 LoanSifter credit score, 80% LTV, and 30% DTI. I estimate (i) the secondary marketing revenue generated by lenders in as implied by MBS TBA prices, and compare that with (ii) the sum of the secondary marketing revenue and the upfront closing costs they charge in the form of points. The secondary marketing revenue generated by lenders in as implied by MBS TBA prices is estimated using Equation (7), with the *Payup* set to zero due to the \$300k loan amount.

Then, with the interest rate spread to the Freddie Mac Primary Mortgage Market Survey (PMMS) rate[3] rounded to the nearest 1/8th $\tilde{c}$, I run a linear regressions of the form:

$$\phi_{ijt} = \sum_{l=1}^{N} \gamma_l \mathbb{1}(c = c_l) + \xi_{jt} + \epsilon_{ijt}, \tag{32}$$

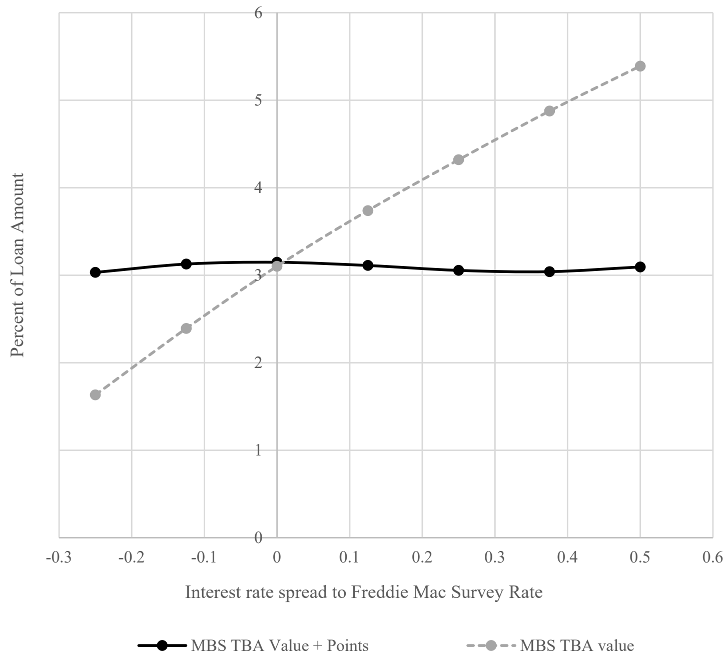
where c_l are the categorical variables of interest rate spread rounded to the nearest 1/8th, ξ_{jt} are lender-day fixed effects, and ϵ_{ijt} is the error term. ϕ_{ijt} is either the secondary marketing revenue generated the lender or sum of the revenue generated by lenders in the secondary market and the upfront closing costs in the form of points, both expressed as a percentage of the loan amount.

The predicted values of Equation (32) are plotted in Figure A.3, which shows that mortgages that are originated at a higher spread to the Freddie Mac Survey rate tend to command higher valuations in the secondary market but generate almost exactly the same lender total income. This suggests that higher secondary marketing income is almost entirely passed through to consumers in the form of lower upfront lender fees/points.[4] Given the near complete pass-through of secondary marketing income to primary market upfront closing costs on average, it is economically meaningful to say that mortgages with positive secondary marketing income have a part of their upfront closing costs "added to the rate" which is then subject to cross-subsidization.

[3] The Freddie Mac Primary Mortgage Market Survey rate is obtained from https://fred.stlouisfed.org/series/MORTGAGE30US.

[4] The same patterns also exist in the time series, as I illustrate in Appendix Figure A.4. In Figure A.4, there is some evidence that in more recent years the interest rate is slightly lower on low upfront closing cost mortgages than what would be implied by secondary marketing income, perhaps suggesting a role for markups. I abstract from markups that vary by points in this paper as the magnitude of the cross-subsidization I study is significantly larger than the differences shown in Figure A.4.

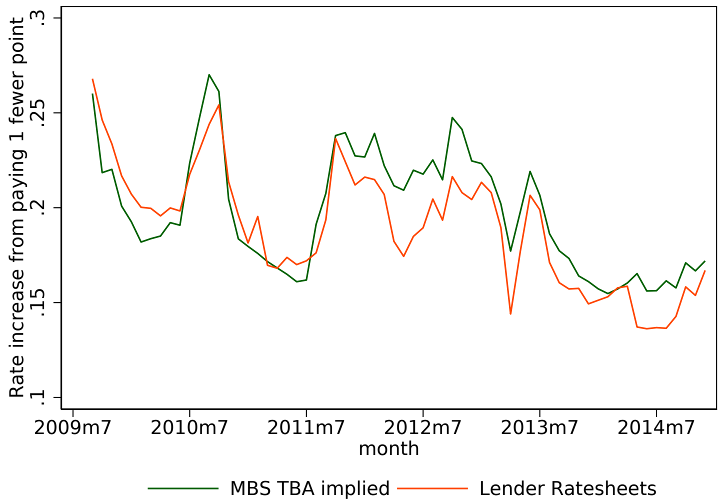
Figure A.3: Secondary marketing income and total lender revenues



Note: Figure A.3 presents estimates from a linear regression of (i) the estimated secondary marketing revenue as implied by MBS TBA prices and (ii) the sum of estimated secondary marketing revenue as implied by MBS TBA prices and upfront closing costs in the form of points on categorical variables of eighths of rate spreads with lender-day fixed-effects based on Equation (32). The grey dotted line plots the predicted values from the regression with estimated secondary marketing revenue as the regressor. The black solid line plots the predicted values from the same regression on the sum of estimated secondary marketing revenue and upfront closing costs in the form of points.

In addition to cross-section, I also examine the relationship between the rate and upfront closing cost trade-off in the time series in Figure A.4. Using the LoanSifter data, I estimate the rate increase from paying 1 less point (i.e., 1% of the loan amount less) in upfront closing costs as the interest rate increase from going from a mortgage with 1 point in upfront closing costs to a mortgage with 0 points within each lender rate sheet. To get the corresponding exchange rate in the MBS TBA data, I take the mortgage rate at 0 points from lender rate sheets and compute the increase in rate that would imply a 1% increase in the MBS TBA value of the mortgage, with interpolated values for coupon rates in between eighths. I then take the mean of the exchange rate implied by the LoanSifter data and the MBS TBA data by month, with results plotted in Figure A.4.

Figure A.4: The interest rate increase from paying 1 less point in upfront closing cost over time, lender ratesheets (red) versus MBS TBA implied (green)



Note: Figure A.4 presents estimates from taking monthly means of (i) the required increase in rate to make the mortgage value increase by 1% of the loan amount in the MBS TBA data (ii) the increase in rate going from 0 points in lender rate sheet to 1 point in lender rate sheet in terms of upfront closing costs paid. The data used is Morgan Markets for MBS TBA prices and LoanSifter for rate sheets. MBS TBA values are linearly interpolated in between eighths of interest rates and LoanSifter rates are linearly interpolated to arrive at the rate at 0 and 1 point in upfront closing costs.

Figure A.4 shows that the exchange rate implied by the the LoanSifter data and the MBS TBA data are fairly close to each other, with the MBS TBA implied exchange rate being slightly larger near the end of the sample. This is consistent with near complete pass-through of secondary marketing revenue to upfront closing costs, with a small discount to lower upfront closing cost mortgages in the retail market as compared to the secondary market near the end of the sample.

A.4 Additional motivating facts

In this section, I present some additional stylized facts that illustrate the existence of cross-subsidization of mortgage closing costs and its sizable distributional implications. First, I show in Section A.4.2 that almost all borrowers pay for most of their mortgage closing costs through a higher interest rate on their mortgage relative to mortgage-backed securities yields, rather than upfront. Second, I show that heterogenous borrower prepayment tendencies implies different borrowers with the same closing costs added to the rate end up with very different net present values (NPVs) of their extra interest rate payments, ex post, in Section A.4.3. Third, I assess magnitude of this difference by demographic groups in Section A.4.4.

A.4.1 Regression of choices of points as predicted by ex-post prepayment behavior

Table A.3: Choices of points as it relates to refinancing/prepayment behavior

	(1) Points		(2) Points	
Non-refi borrower	0.0659***	(5.32)		
5-year prepayment			-0.0841***	(-6.30)
Log(loan amount)	0.0511***	(2.71)	0.0497***	(2.73)
Credit score controls	Yes		Yes	
LTV controls	Yes		Yes	
DTI control	Yes		Yes	
Constant	-0.600***	(-2.83)	-0.519**	(-2.59)
Observations	25245		25245	
LenderXCountyXYear FEs	Yes		Yes	

Robust t statistics clustered by lender and county in parentheses.

* p<0.1, ** p<0.05, *** p<0.01

Note: The data used in this figure is the Optimal Blue-HMDA-CRISM data from January 2013 to May 2022, for 30-year, fixed-rate, conforming, primary residence mortgages originated in 2013–2019. The sample for (1) and (2) is further restricted to the set of borrowers whose mortgages originated before April 2016 and where the Freddie Mac Survey Rate decreased at least 1.2% since origination. Table A.3 presents OLS estimates of borrower choices of points on (1) an indicator variable for non-refinancing borrowers, defined as borrowers who did not refinance or otherwise prepay within five years despite facing a Freddie Mac Survey Rate decrease of at least 1.2%, and (2) borrowers who prepaid within five years.

A.4.2 Prevalence of mortgages with closing costs added to the rate

When borrowers take out a mortgage, they have a choice between adding closing costs to the rate of the mortgage or paying them upfront. In this section I assess the extent to which mortgage closing costs are added to the rate using the 2018–2019 Optimal Blue-HMDA data. The 2018–2019 HMDA data contains information about the upfront closing costs paid by the borrower in the form of loan origination charges and total loan costs, and the match to Optimal Blue data enables me to obtain information on when the rate was locked which then allows me to estimate the revenue that lenders expected to generate from the secondary market at the time the rate is set.

I first estimate the extent to which mortgage closing costs are added to the rate based on Equation (1). This statistic has a mean of 3.49%. I then compute the lender revenue from origination as the sum of y_{it} and the net loan origination charges from 2018-2019 HMDA data. The fraction of y_{it} to the sum of y_{it} and the net loan origination charges represents the share of lender revenue from secondary marketing income.[5]

The results of my analysis are shown in Figure A.5. The left panel in Figure A.5a shows that lenders make on average 4.6% of the mortgage balance as revenue for each mortgage they originate. This revenue compensates the lender for their costs. First, lenders need to pay for the upfront costs of mortgage insurance, also called loan-level price adjustments (LLPAs) by Fannie Mae and Freddie Mac. Second, lenders pay for loan originator compensation, which can be 1–2% of the loan amount. Third, lenders pay for the underwriting and processing costs associated with the origination. Relative to these expenses, the portion that is attributable to accounting profits are low: the Mortgage Bankers' Association (MBA) reports an average production profit of 0.14% of the loan amount in 2018 and 0.31% of the loan amount in 2017.[6]

The right panel of Figure A.5b shows that only a small fraction of lender revenue is paid as

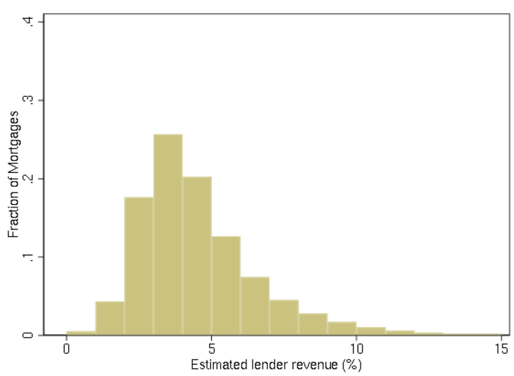
[5] A drawback of this approach is that it assumes that mortgage servicing has zero marginal cost, or that mortgage servicing rights have value equal to their expected stream of cash flows. Without explicit data on the value of mortgage servicing rights relative to their cash flow value, I compute a lower bound on the estimated lender revenues by looking at the MBS value of the net interest rate paid to investors by assuming counterfactually that mortgage servicing rights are worth zero. This lower bound is presented in Appendix Figure A.6, which still shows that the majority of mortgages still have most of their price of origination paid for through the rate.

[6] https://www.mba.org/2019-press-releases/april/independent-mortgage-bankers-production-volume-and-profits-down-in-2018. The MBA reports that average net production revenues in 2018 (excluding LLPAs) are 3.47% of the loan amount, which is consistent with my estimate of 4.6% with LLPAs.

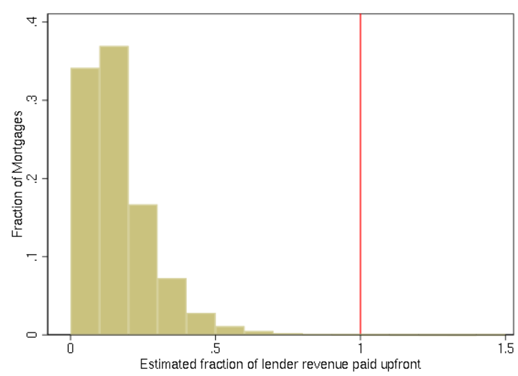
upfront net origination charges, with an average of 17.4%. That is, even though most of the lender costs of origination are incurred upfront, 82.6% of the price of origination is added to the rate of the mortgage and paid over time primarily by immobile and inactively refinancing borrowers. Hence, almost all mortgages being originated in the US can be considered "low upfront closing cost" mortgages whose price of mortgage origination are prone to cross-subsidization between borrowers with different refinancing speeds. Figures A.7 and A.8 repeats this exercise for purchase and refinance mortgages separately, and finds patterns that are broadly consistent.

Finally, Figures A.9, A.10, and A.11 repeats this exercise but with total loan costs in place of net origination charges. Total loan costs include appraisal and title search fees that may not go to the lender, though some lenders may generate revenue from part of these fees. But, even with this broader measure of upfront lender revenue only 31.4% of the price of mortgage origination is paid upfront on average.

Figure A.5: Lender revenue and percentage paid as upfront net origination charges



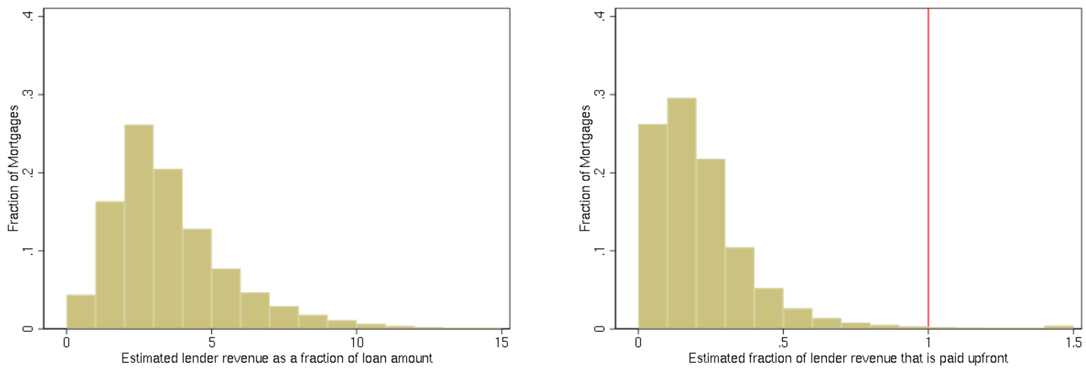
(a) Estimated lender revenue



(b) Fraction of lender revenue paid upfront

Note: The data used in this figure is the 2018–2019 Optimal Blue-HMDA data for 30-year, fixed-rate, conforming, primary residence mortgages. The data contains information on rates and net origination charges paid upfront and was linked to MBS TBA data following Fuster, Lo, and Willen (2022) to estimate secondary marketing revenue. Figure A.5a plots histograms of estimated lender revenue which consists of the sum of net origination charges and secondary marketing revenue. Figure A.5b then plots histograms of the fraction of lender revenue that is consists of net origination charges. The net origination charges used in this figure equals to the HMDA origination charges minus lender credit, and is set to zero for mortgages with less than zero net origination charges.

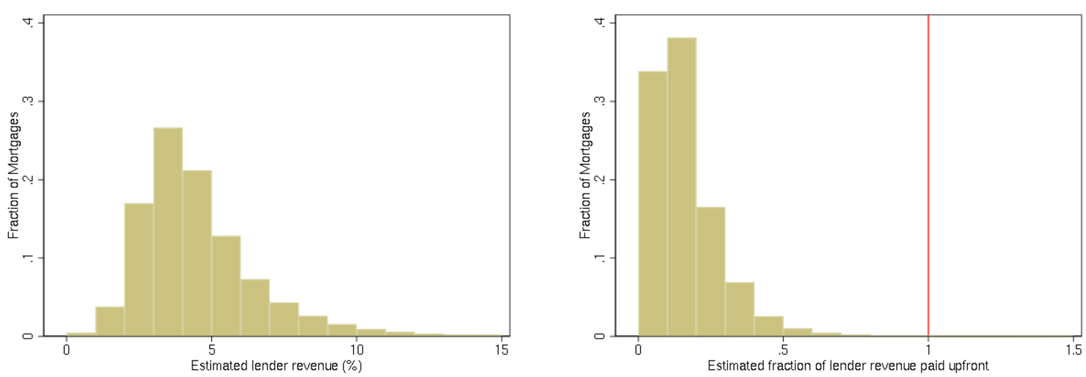
Figure A.6: Lender revenue and percent paid as upfront net origination charges, net of mortgage servicing revenue



(a) Estimated lender revenue (b) Fraction of lender revenue paid upfront

Note: The data used in this figure is the Optimal Blue data for 30-year, fixed-rate, conforming, primary residence mortgages originated in 2018-2019 matched to the 2018–2019 HMDA data. The data contains information on rates and net origination charges paid upfront and was linked to MBS TBA data following Fuster, Lo, and Willen (2022) to estimate secondary marketing revenue. A further 25 basis points was subtracted from the coupon rate for mortgage servicing. Figure A.6a plots histograms of estimated lender revenue which consists of the sum of net origination charges and secondary marketing revenue. Figure A.14b then plots histograms of the fraction of lender revenue that is consists of net origination charges.

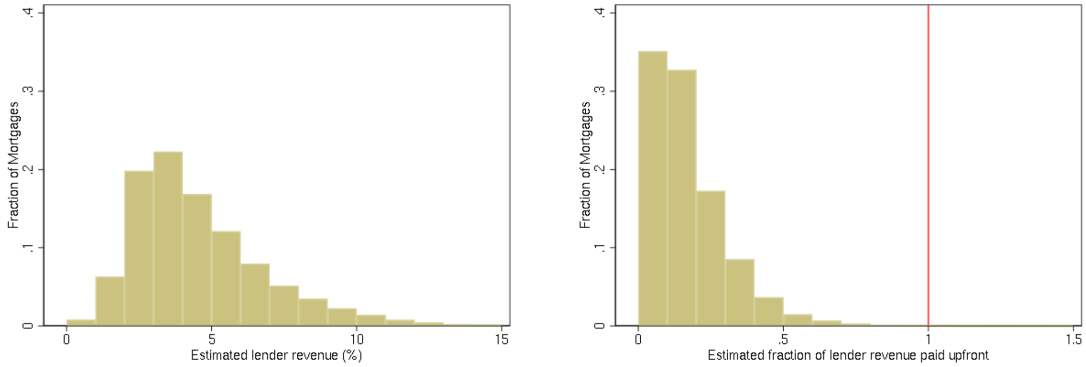
Figure A.7: Lender revenue and percentage paid as upfront net origination charges, purchase



(a) Estimated lender revenue (b) Fraction of lender revenue paid upfront

Note: The data used in this figure is the 2018–2019 Optimal Blue-HMDA data for 30-year, fixed-rate, conforming, primary residence purchase mortgages. The data contains information on rates and net origination charges paid upfront and was linked to MBS TBA data following Fuster, Lo, and Willen (2022) to estimate secondary marketing revenue. Figure A.7a plots histograms of estimated lender revenue which consists of the sum of net origination charges and secondary marketing revenue. Figure A.10b then plots histograms of the fraction of lender revenue that is consists of net origination charges. The net origination charges used in this figure equals to the HMDA origination charges minus lender credit, and is set to zero for mortgages with less than zero net origination charges.

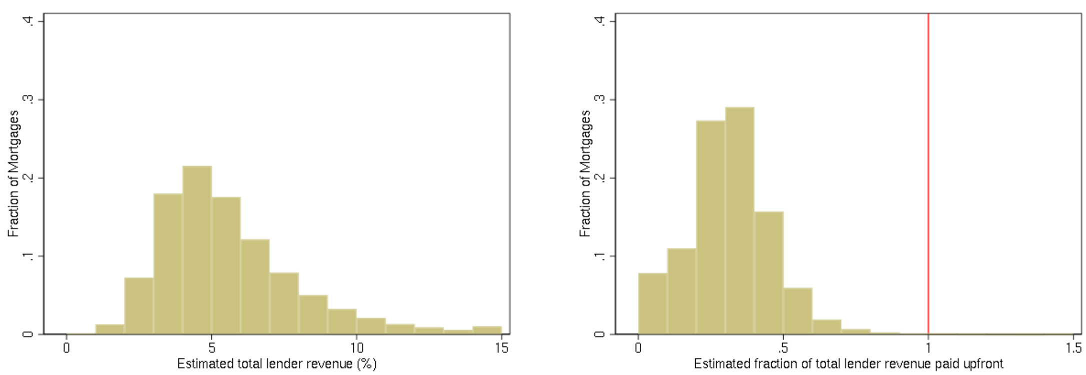
Figure A.8: Lender revenue and percentage paid as upfront net origination charges, refinance



(a) Estimated lender revenue (b) Fraction of lender revenue paid upfront

Note: The data used in this figure is the 2018–2019 Optimal Blue-HMDA data for 30-year, fixed-rate, conforming, primary residence refinance mortgages. The data contains information on rates and net origination charges paid upfront and was linked to MBS TBA data following Fuster, Lo, and Willen (2022) to estimate secondary marketing revenue. Figure A.8a plots histograms of estimated lender revenue which consists of the sum of net origination charges and secondary marketing revenue. Figure A.11b then plots histograms of the fraction of lender revenue that is consists of net origination charges. The net origination charges used in this figure equals to the HMDA origination charges minus lender credit, and is set to zero for mortgages with less than zero net origination charges.

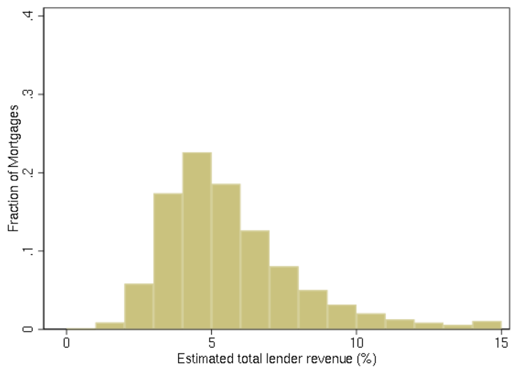
Figure A.9: Total lender revenue and percentage paid as total loan costs



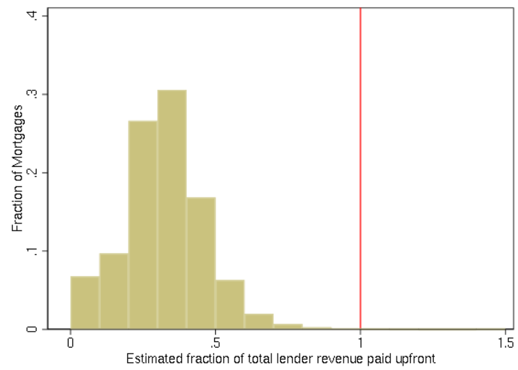
(a) Estimated lender revenue (b) Fraction of lender revenue paid upfront

Note: The data used in this figure is the 2018–2019 Optimal Blue-HMDA data for 30-year, fixed-rate, conforming, primary residence mortgages. The data contains information on rates and total loan costs paid and was linked to MBS TBA data following Fuster, Lo, and Willen (2022) to estimated secondary marketing revenue. Figure A.9a plots histograms of estimated lender revenue which consists of the sum of total loan costs plus secondary marketing revenue. Figure A.9b then plots histograms of the fraction of lender revenue that is paid upfront. The total loan costs used in this figure equals to the HMDA total loan costs minus lender credit, and is set to zero for mortgages with less than zero total loan costs.

Figure A.10: Total lender revenue and percentage paid as total loan costs, purchase



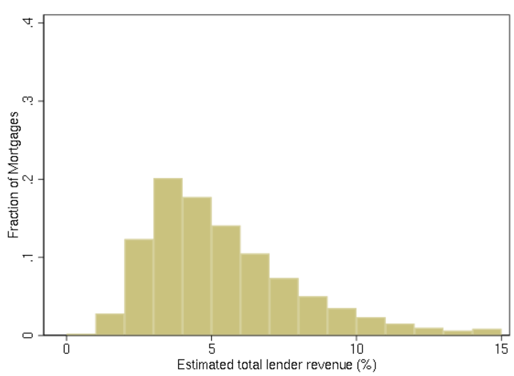
(a) Estimated lender revenue



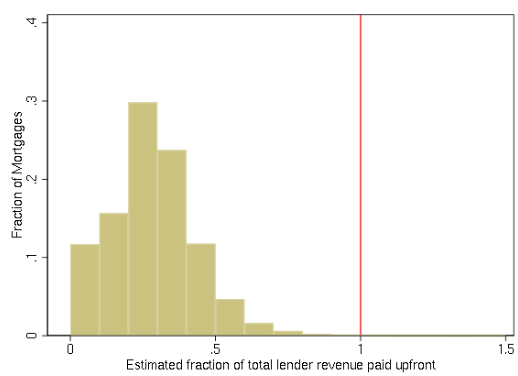
(b) Fraction of lender revenue paid upfront

Note: The data used in this figure is the 2018–2019 Optimal Blue-HMDA data for 30-year, fixed-rate, conforming, primary residence purchase mortgages. The data contains information on rates and total loan costs paid and was linked to MBS TBA data following Fuster, Lo, and Willen (2022) to estimated secondary marketing revenue. Figure A.10a plots histograms of estimated lender revenue which consists of the sum of total loan costs plus secondary marketing revenue. Figure A.10b then plots histograms of the fraction of lender revenue that is paid upfront. The total loan costs used in this figure equals to the HMDA total loan costs minus lender credit, and is set to zero for mortgages with less than zero total loan costs.

Figure A.11: Total lender revenue and percentage paid as total loan costs, refinance



(a) Estimated lender revenue



(b) Fraction of lender revenue paid upfront

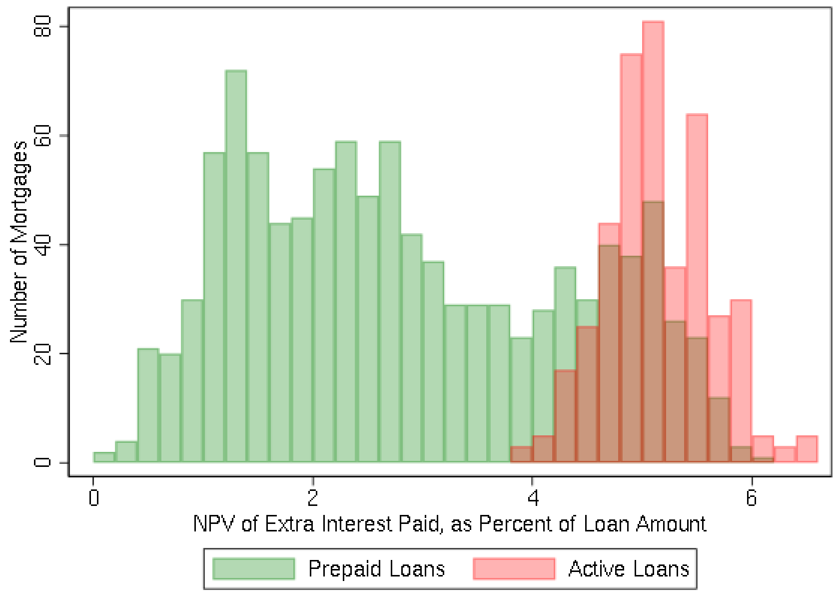
Note: The data used in this figure is the 2018–2019 Optimal Blue-HMDA data for 30-year, fixed-rate, conforming, primary residence refinance mortgages. The data contains information on rates and total loan costs paid and was linked to MBS TBA data following Fuster, Lo, and Willen (2022) to estimated secondary marketing revenue. Figure A.11a plots histograms of estimated lender revenue which consists of the sum of total loan costs plus secondary marketing revenue. Figure A.11b then plots histograms of the fraction of lender revenue that is paid upfront. The total loan costs used in this figure equals to the HMDA total loan costs minus lender credit, and is set to zero for mortgages with less than zero total loan costs.

A.4.3 Cross-subsidization of the price of mortgage origination when they are added to the rate

The interaction of the heterogeneity in refinancing tendencies and a component of the price of mortgage origination being added to the rate implies a cross-subsidization of the price of mortgage origination to the extent they are added to the rate. To illustrate this in my data, Figure A.12 looks at borrowers with similar amounts of the price of mortgage origination added to the rate, between 4.75-5.25% of the loan amount, in 2013 in my Optimal Blue-HMDA-CRISM sample and compares the NPV of the extra interest rate they paid as a percentage of their loan amount.[7] Due to differences in prepayment behavior, I find large differences in how much borrowers end up paying for the same 4.75-5.25% of the loan amount in the price of mortgage origination added to the rate, ranging from close to 0% to more than 6%.

[7] The year 2013 was chosen because it is the earliest year in my sample.

Figure A.12: NPV of extra interest paid, 2013 mortgages with 4.75–5.25% of the loan amount in the price of mortgage origination added to the rate



Note: The data used in this figure is the Optimal Blue-HMDA-CRISM data originated between January 2013 and December 2013, with performance to May 2022, for 30-year, fixed-rate, conforming, primary residence mortgages. The sample was further limited to mortgages with a secondary marketing revenue of 4.75–5.25% of the loan amount, as estimated based on MBS TBA prices following Fuster, Lo, and Willen (2022). The extra interest paid is relative to a mortgage with 0% secondary marketing revenue (i.e., originated at par) and is estimated as the difference between the mortgage interest rate at origination net of the fee for government guarantee (gfees) and the MBS TBA yields at the time of lock. The NPV of the extra monthly payment resulting from this difference in extra interest paid is then computed assuming a discount rate equal to the 10-year Treasury rate at the time of the rate lock and plotted in the histogram for loans that have prepaid (in green) and for loans that are still active (in red). CRISM data is attributed to Equifax Credit Risks Insight Servicing and Black Knight McDash Data.

The reason for the variance in outcomes in Figure A.12 is that, when the price of mortgage origination is added to the rate of the mortgage, lenders can only recover their price of mortgage origination over time through a higher interest rate payment. The principal balance of the mortgage remains unchanged. Therefore, borrowers who prepay earlier end up paying less, while borrowers who prepay later end up paying more. The transfers and deadweight losses studied in this paper come from the extent to which that borrowers who actively refinance pay less for their price of mortgage origination in expectation when they are added to the rate by receiving cross-subsidization from other borrowers. Appendix Section A.4.4

examines the variation in this cross-subsidization by borrower demographics.

A.4.4 The predictability of cross-subsidization by demographics

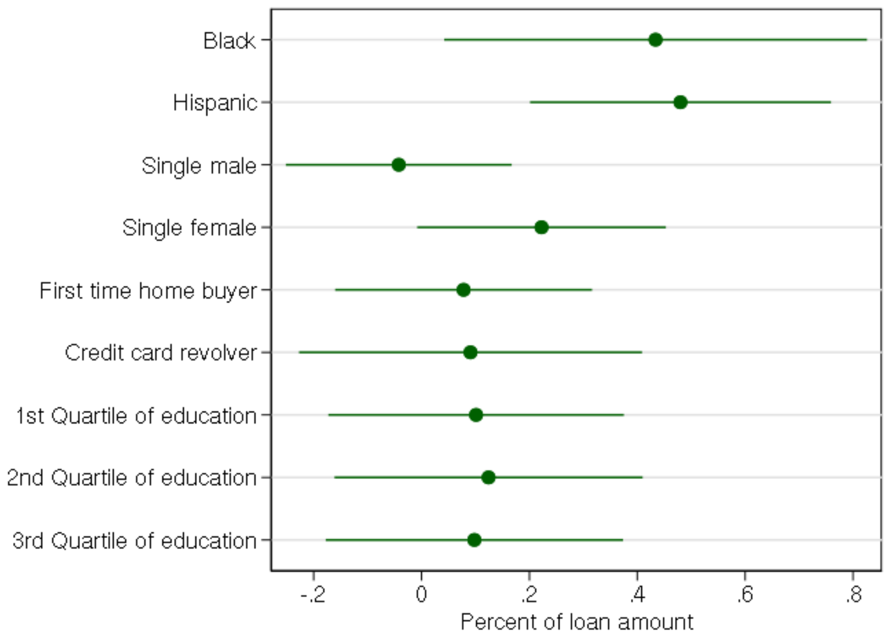
In this section I examine the extent of this ex-post cross-subsidization by demographics. To do so, I run the regression on loan level data in my Optimal Blue-HMDA-CRISM sample:

$$NPV_{i,t} = \beta X_i + \gamma Z_i + \xi_{\phi_{i,t} \times t} + \epsilon_{i,t} \tag{33}$$

where $NPV_{i,t}$ is the NPV of extra interest paid for their closing costs that are added to the rate over the observed life of the mortgage; X_i is a set of demographic and credit utilization variables including race (Black, Hispanic), gender (male and female), credit card revolver status, and quartiles of education; Z_i is a set of control variables including categories of credit scores at origination, LTV, DTI, and loan amount; $\xi_{\phi_{i,t} \times t}$ is the amount of closing costs added to the rate by time fixed effects.

The results of this analysis are shown in Figure A.13 and Table A.4. I find that Black and Hispanic borrowers paid an extra 0.5% of the loan amount for their closing costs added to the rate relative to other borrowers. For a $300,000 loan, the magnitude of this cross-subsidization is about $1500 per loan. Furthermore, single-applicant female borrowers paid an extra 0.24% of the loan amount for their closing costs added to the rate. A limitation of this analysis is that does not take into account the potentially unexpected decline in interest rate during this period, so a model is needed to get at the welfare effects ex ante.

Figure A.13: NPV of extra interest paid by demographic and borrower characteristics



Note: The data used in this figure is the Optimal Blue-HMDA-CRISM data from January 2013 to December 2013, for 30-year, fixed-rate, conforming, primary-residence mortgages originated in 2013. The graph plots regression coefficients from Column (2) of Table A.4. In particular, it shows that Black, Hispanic and single-applicant female borrowers pay more for their closing costs added to the rate than other borrowers. Other characteristics, such as single-applicant male borrowers, first-time home buyers, credit card revolvers (defined as someone with a more than 60% credit utilization and $10,000 in debt at the time of getting a mortgage), and quartiles by education are not statistically different from zero at the 5% level. CRISM data is attributed to Equifax Credit Risks Insight Servicing and Black Knight McDash Data.

Table A.4: Regression on NPV of extra interest paid by demographic and borrower characteristics

	(1)	
	NPV of Extra Interest Paid	
Black	0.434***	(2.17)
Hispanic	0.480***	(3.38)
Single male	-0.042	(-0.40)
Single female	0.223**	(1.89)
First-time home buyer	0.078	(0.64)
Credit card revolver	0.091	(0.56)
1st quartile of education	0.101	(0.72)
2nd quartile of education	0.124	(0.85)
3rd quartile of education	0.098	(0.70)
Log(loan amount)	-0.363***	(-2.92)
Credit Score controls	Yes	
LTV controls	Yes	
DTI control	Yes	
Constant	7.918***	(4.85)
Observations	1275	
ϕ by month FEs	Yes	

robust t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Note: The data used in this table is the Optimal Blue-HMDA-CRISM data from January 2013 to December 2013, for 30-year, fixed-rate, conforming, primary residence mortgages originated in 2013. This table contains regression results from estimating Equation (7). The dependent variable is the NPV of extra interest paid from the closing costs that are added to the rate. I include ϕ by month fixed effects, where ϕ refers to the amount of closing costs added to the rate rounded to the nearest percent of the loan amount. CRISM data is attributed to Equifax Credit Risks Insight Servicing and Black Knight McDash Data.

A.5 Model details

A.5.1 Exogenous states

The risk-free rate follows the Cox, Ingersoll, and Ross (1985) model which has a natural zero lower bound:

$$dr_{1t} = a(b - r_{1t})dt + \sigma\sqrt{r_{1t}}dW_t. \tag{34}$$

I estimate the evolution of exogenous states in the model via maximum likelihood[8] using the monthly three-month Treasury bill data from January 1987 to January 2021.[9] The results for the risk-free rate are as follows:

Table A.5: Estimation of the CIR model of interest rates

Parameter	Estimate	Standard Error
a	0.0910	0.0506
b	1.2649	0.7209
σ	0.4930	0.0175

Note: This table contains estimates from fitting the Cox, Ingersoll, and Ross (1985) model on the three-month Treasury bill data from January 1987 to January 2021. Estimation proceeds via the maximum likelihood, and standard errors are obtained from the inverse Hessian.

I model the average mortgage rate $\bar{c}_t$, changes in log real house prices ΔH_t, and changes in log real personal income ΔL_t and as a quarterly vector autoregression (VAR) with r_{1t} as an exogenous dependent variable. I use two lags in the VAR, with the constraint that the matrix of coefficients on first lag is identity and on the second lag is positive only for the

[8]The program was based on Kladivko (2021), with some modifications to obtain standard errors.
[9]Board of Governors of the Federal Reserve System (US), 3-Month Treasury Bill Secondary Market Rate [TB3MS], retrieved from FRED, Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/series/TB3MS.

house price coefficient to reduce dimensionality.[10] More specifically, with $\boldsymbol{s_t} = \begin{bmatrix} \bar{c}_t \\ 100 * \Delta H_t \\ 100 * \Delta L_t \end{bmatrix}$, the VAR equation is as follows:

$$\boldsymbol{s_t} = \boldsymbol{\mu} + r_{1t}\boldsymbol{\beta_{r_{1t}}} + \Phi_1 \boldsymbol{s'_{t-1}} + \Phi_2 \Delta H_{t-1} + \boldsymbol{e_t}, \tag{35}$$

where $\boldsymbol{e_t} \sim N(0, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{s}})$ and $\boldsymbol{\mu}, \boldsymbol{\beta_{r_{1t}}}, \Phi_2$ are the coefficients to be estimated. In terms of the state variables, data on $\bar{c}_t$ is obtained as the Primary Mortgage Market Survey (PMMS) rate,[11] H_t is obtained from the Case-Shiller National House Price Index,[12] and L_t is obtained from the US Personal Income[13] divided by the US population.[14] Furthermore, H_t and L_t are converted to real terms using the Consumer Price Index for All Urban Consumers.[15] The results of the VAR estimation are as follows:

[10]The second lag on the house price variable is added to capture momentum and mean reversion as in Glaeser and Nathanson (2017).

[11]Freddie Mac, 30-Year Fixed Rate Mortgage Average in the United States [MORTGAGE30US], retrieved from FRED, Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/series/MORTGAGE30US.

[12]S&P Dow Jones Indices LLC, S&P/Case-Shiller U.S. National Home Price Index [CSUSHPINSA], retrieved from FRED, Federal Reserve Bank of St. Louis;,https://fred.stlouisfed.org/series/CSUSHPINSA.

[13]U.S. Bureau of Economic Analysis, Personal Income [PI], retrieved from FRED, Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/series/PI

[14]U.S. Bureau of Economic Analysis, Population [POPTHM], retrieved from FRED, Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/series/POPTHM.

[15]U.S. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CPIAUCSL], retrieved from FRED, Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org/series/CPIAUCSL.

Table A.6: VAR estimates of state transitions

Parameter	$\boldsymbol{\mu}$	$\boldsymbol{\beta_{r_{1t}}}$	$\boldsymbol{\Phi_1}$			$\boldsymbol{\Phi_2}$	$\boldsymbol{\hat{\Sigma}_s}$		
$\bar{c}_t$	.418 (.135)	.075 (.027)	.892 (.031)	0	0	0	.110		
$100 * \Delta H_t$	.447 (.228)	-.079 (.072)	0	.513 (.088)	0	-.022 (.088)	.080	2.664	
$100 * \Delta L_t$	.277 (.320)	.086 (.104)	0	0	-.472 (.079)	0	-.019	.026	5.572

Note: This table contains estimates from fitting a constrained VAR described in Equation (35). Data on mean mortgage rates $\bar{c}_t$ is obtained from the Primary Mortgage Market Survey (PMMS), data on house prices H_t are taken from the Case-Shiller National Home Price index, and data on personal income Y_t are taken as the ratio of US aggregate personal income divided by the US population. House prices and income are divided by the CPI for urban consumers and then transformed into growth rates.

Furthermore, I assume that borrowers pay \$2000 + 1% of the loan amount to obtain the PMMS survey rate following Agarwal, Driscoll, and Laibson (2013), and that the upfront closing cost payment is a deterministic nonlinear function of the deviation from the Freddie Mac PMMS rate $c - \bar{c}_t$, as denoted by the nonlinear function $\bar{o}(c - \bar{c}_t)$. As Appendix Figure A.4 shows, the rate and upfront trade-off from lender rate sheets does vary over time and follows MBS TBA prices closely between the July 2009 to December 2014 period when I have rate sheet data. However, calibrating the model to the average trade-off as implied by MBS TBA prices over the period allows me to approximate the average borrower welfare while significantly reducing the computational burden.

With this assumption, the interest rate and upfront closing cost menu for refinances is given in Equation (37):

$$\psi_{it}(c, M_{it}) = \$2000 + 0.01 M_{it} + \bar{o}(c - \bar{c}_t(\bar{m}_t^l), t > 0 \tag{36}$$

where M_{it} is the remaining balance as given by the amortization formula starting from M_{i1}, and $\bar{o}$ is the average rate and upfront closing cost trade-off conditional on the market interest rate $\bar{c}_t(\bar{m}_t^l)$, which in turn incorporates time-varying lender costs $\bar{m}_t^l$. For new purchase originations, the rate and upfront closing cost menu faced by borrowers with a mortgage amount M_{i1} is the same as in Equation (37) plus an additional M_{i1} term that incorporates a lender markup:

$$\psi_{it}(c, M_{it}) = \$2000 + 0.01M_{i1} + \bar{o}(c - \bar{c}_t(\bar{m}_t^l) + \frac{M_{i1}}{M_{i1}}, t = 0 \tag{37}$$

where again $\bar{c}_t$ incorporates time-varying lender costs $\bar{m}_t^l$.

The estimates from Tables A.5 and A.6 are then used to simulate the transitions of the exogenous states in my model in Section 5.

A.5.2 OAS

An empirical model of prepayment behavior combined with my model of interest rates is needed to estimate the OAS in Section 5.1.2. For my empirical model of prepayment, I use my panel data to estimate a logit regression of an indicator variable for borrower prepayment on the spread of the mortgage interest rate to the Freddit Mac survey rate at origination (SATO) as well as categories of the interest rate incentive defined as the current mortgage interest rate minus the Freddit Mac survey rate. To maintain comparability to the TBA market from which I derive the market exchange rate between the interest rate and upfront closing costs, I further restrict my analysis to 30 year purchase mortgages with a balance above \$150k, FICO above 680, and LTV below 85% following Fusari et al. (2020). Results of this regression are shown in Table A.7, which is used for my model of $\hat{p}_{t'}$ as in Equation (21).

Table A.7: Logit model of prepayment

	(1)	
	Logit	
prepaid		
init_t	7.446***	(16.59)
init_t_sq	-4.169***	(-12.32)
sato	0.121	(0.62)
sato_sq	-0.765***	(-2.88)
refi_ratediff_gt0	0.348***	(4.55)
refi_ratediff_gtp25	0.345***	(4.08)
refi_ratediff_gtp5	0.599***	(8.31)
refi_ratediff_gtp75	0.322***	(4.69)
refi_ratediff_gt1	0.538***	(5.63)
refi_ratediff_gt1p25	0.144	(1.09)
burnout	-0.0549*	(-1.94)
burnout_sq	0.00182	(1.45)
Constant	-8.264***	(-54.48)
Observations	267603	

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: The data used in this regression is the Optimal Blue-HMDA-CRISM data from January 2013 to May 2022, for 30-year, fixed-rate, conforming, primary residence mortgages originated in 2013–2019. The sample is further restricted to "TBA likely" mortgages defined as mortgages with a loan amount of at least $150k, loan-to-value ratio less than or equal to 85%, and FICO at origination greater than or equal to 680. The independent variable is an indicator variable for whether the borrower prepaid their mortgage in a given month. The dependent variables include the spread of the mortgage interest rate to the Freddie Mac survey rate at origination (SATO) and its square, as well as categories of rate incentive (the current spread of the mortgage interest rate to the Freddie Mac survey rate). CRISM data is attributed to Equifax Credit Risks Insight Servicing and Black Knight McDash Data.

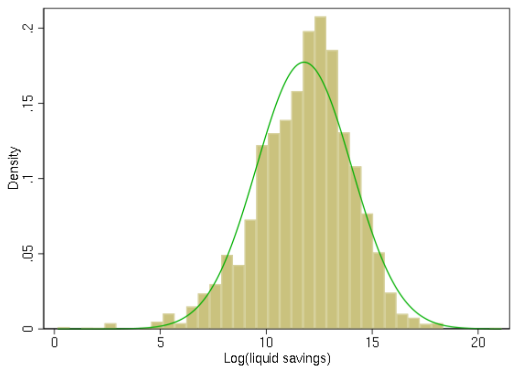
Using the prepayment model from Table A.7 and the interest rate model of Section A.5.1, with the risk-free rate r_{tf} being given as the implied 10 year rate under the Cox, Ingersoll, and Ross (1985) model, I estimate a $O\hat{A}S = 0.22\%$ by minimizing the equally-weighted difference between the observed MBS TBA price for the nearest two coupons above and below the Freddie Mac survey rate - gfees - servicing fees with the implied NPV given by Equation (21). The MBS TBA price is inclusive of the new production pay-up for a coupon (with data from Morgan Markets). The gfee is assumed to be 0.42% and servicing fee 0.25% following Fuster, Lo, and Willen (2022).

A.5.3 Distribution of liquid assets

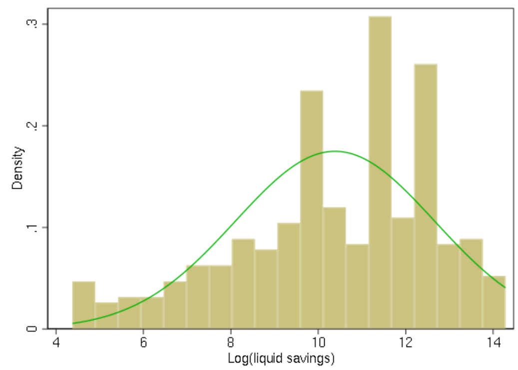
I estimate the distribution of liquid assets in the model using the Survey of Consumer Finances (SCF). To parsimoniously summarize the aggregate distribution of liquid assets, I fit log-normal distributions for liquid assets among Conventional mortgage holders (defined as households with X724=5) using the 2013–2019 SCF data. The data item for liquid assets is "Total Financial Assets," including bank accounts, CDs, mutual funds, stocks, bonds, liquid retirement savings, savings bonds, cash value of whole life insurance, and other managed/financial assets. Black households are defined as households with X6809=2, Hispanic households are defined as households with X6809=3.

Histograms of the those households' log liquid assets, along with the fitted log-normal distribution, are plotted in Figure A.14.

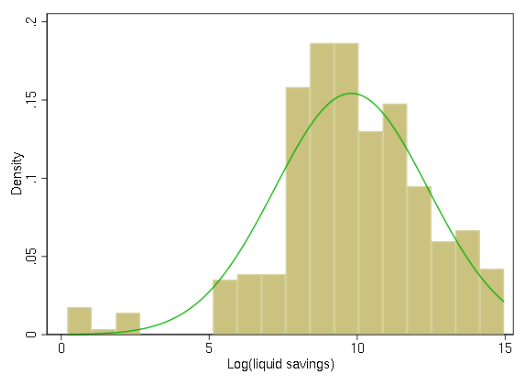
Figure A.14: Histograms of mortgage borrower's log liquid assets along with their fitted distributions



(a) Non-Black, non-Hispanic borrowers



(b) Black borrowers



(c) Hispanic borrowers

Note: The data used in this figure is from the 2013–2019 SCF. The log of the households' total liquid assets, including liquid savings, CDs, mutual funds, stocks, bonds, liquid retirement savings, savings bonds, cash value of whole life insurance, and other managed/financial assets, are plotted for households with a mortgage. In addition, a normal distribution is fit to the log liquid assets. For non-Black, non-Hispanic households with a mortgage, I estimate a mean of 11.78 and a standard deviation of 2.25. For Black households with a mortgage, I estimate a mean of 10.38 and a standard deviation of 2.28. For Hispanic households with a mortgage, I estimate a mean of 9.80 and a standard deviation of 2.59.

A.6 ADL

I investigate the cross-subsidization of low upfront closing cost mortgages from the perspective of lender rate-setting under the model of Agarwal, Driscoll, and Laibson (2013), hereafter referred to as ADL. ADL proposes a model of optimal refinancing based on a Brownian motion interest rate model as well as assumptions on the inflation rate, tax rate, and the probability of moving. The ADL parameters are as follows:

$\rho =$	0.03
$\sigma =$	0.004597
$\pi =$	0.0168
$\mu =$	0.074
$\tau =$	0.22

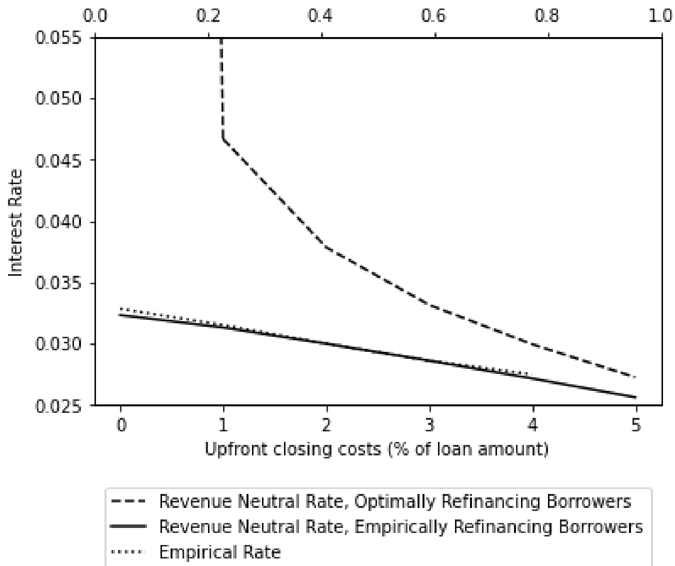
where ρ is the discount rate, σ is the standard deviation of the mortgage rate, π is the expected inflation rate, μ is the expected probability of moving, and τ is the tax rate.

I compute the counterfactual interest rates that the lenders would have to charge to remain revenue neutral if all borrowers behaved as in the ADL model. As an alternative model of optimal refinancing that is relatively easy to compute, this serves as a robustness check to my main calibration result in Figure 6. The result from the ADL model is shown in Figure A.15.

In particular, Figure A.15 shows suggests that optimally refinancing borrowers (in the sense of ADL) receive a substantially lower rate than what they would have received without cross-subsidization: if all borrowers were optimally refinancing, lenders would charge substantially higher interest rates particularly for low upfront closing cost mortgages, on the order of 1.49% more with a 1% upfront closing cost mortgage compared to only 0.13% with a 5% upfront closing cost mortgage.[16] In other words, the interest rates on lower upfront closing cost mortgages appear to be substantially discounted for optimally refinancing borrowers due to the presence of non-refinancing borrowers in the market, consistent with my main results.

[16] With zero upfront closing costs and optimally refinancing borrowers, I find that lenders would have to charge a rate of 91% to remain revenue-neutral.

Figure A.15: Pricing of mortgages mortgages by upfront closing cost choice, ADL Optimally Refinancing Borrower



Note: Figure A.15 presents the equilibrium rate and upfront closing costs trade-off from the model and compares it to the empirical rate and upfront closing costs trade-off that I estimate from the data. The “Market rate, model implied” solid line refers to the equilibrium rate and closing cost trade-off given the logit prepayment hazard function and an estimated OAS. The “Market rate, empirical” dotted line is the implied rate and upfront closing cost options from MBS TBA prices combined with the PMMS survey rate. I show that the rate and upfront closing cost options implied by MBS TBA prices is consistent with the options actually presented to borrowers in rate sheets from a regression with rate sheet fixed effects in Appendix A.3. Finally, the “No cross-subsidization counterfactual” dashed line refers to the model-implied equilibrium rate and closing cost trade-off in a world where the lender is pricing their mortgages for the calibrated ADL borrower with perfect information on their type.