Internet Appendix

Refinancing Inequality During the COVID-19 Pandemic

Sumit Agarwal, Souphala Chomsisengphet, Hua Kiefer, Leonard C. Kiefer, and Paolina C. Medina

Appendix A Data Description: Freddie Mac Matched Transactions

We used a unique administrative loan-level dataset for conventional single-family loans funded by Freddie Mac. This dataset includes all outstanding single-family 30-year fixed-rate mortgages funded by Freddie Mac and active at the beginning of each refinance wave. We followed those loans through the entire duration of each wave and observed whether the loan was prepaid during the wave. Table A1 presents descriptive statistics for the data.

Table A1: Descriptive Statistics Freddie Mac Loans (Averages)

Wave	N	FICO	LTV	Loan Age (Months)	UPB	RATE
2015	715,363	726	0.77	79	$175,416	5.46%
2016	300,145	731	0.79	69	$198,219	5.04%
2017	232,306	714	0.78	121	$144,102	5.72%
2019	794,178	725	0.80	67	$221,727	5.24%
2020	1,351,845	733	0.80	57	$224,665	4.84%
All	3,393,837	728	0.79	69	$205,743	5.14%

Notes: This table presents descriptive statistics for the main variables considered in the analysis. The sample is restricted to all outstanding 30-year fixed rate mortgages on single-family properties that were not in-the-money for a refinance at the beginning of the wave, and became in-the-money during the wave. UPB = unpaid balance.

In addition, for a subset of loans newly in-the-month and prepaid, we matched a new loan originated at the same property address within a 45-day window of the closure of the prepaid loan. For those matched transactions, we collected loan-level attributes of the newly originated loan at the same address. Where the loan was refinanced, we observed the new loan product and loan attributes, including the new interest rate. We also identified cases where the prepayment was not for a refinance, but for a home purchase. Freddie Mac guarantees about one in five home loans in the United States. Consistent with that share, we found that we had matches for approximately 20 percent of the prepaid loans. The match rate varied by loan attributes: borrowers in the middle of the income distribution had slightly higher match rates than borrowers in the lowest and highest income quintiles. In the 2015 wave, we got a higher match rate (about 27 percent) due to including Home Affordable Refinance Program (HARP) loans. Table A2 contains a summary match rates across our sample.

Table A2: Refinancing Rate by Income and Wave

			Income Quintile (q1=low, q5=high)				
Wave	N	All	q1	q2	q3	q4	q5
2015	52,205	27.1%	30.7%	30.3%	28.3%	27.1%	22.1%
2016	29,842	21.2%	20.1%	22.9%	23.1%	21.9%	18.9%
2017	16,716	19.5%	19.6%	21.8%	21.0%	19.9%	15.4%
2019	78,550	19.6%	14.6%	19.5%	20.9%	20.9%	19.5%
2020	224,235	19.3%	18.6%	19.9%	19.7%	19.7%	18.7%
All	401,548	20.5%	19.5%	21.5%	21.4%	21.1%	19.2%

To assess the extent to which the matched loans broadly represent the full population of prepaid loans, we first compared the characteristics of matched loans to the unmatched loans across waves. Table A3 compares the origination FICO score, origination loan to value (LTV), origination debt to income ratio (DTI), interest rate, and unpaid balances (UPB) (at the beginning of the wave) for matched and unmatched loans. On these observables, the matched and unmatched loans are similar.

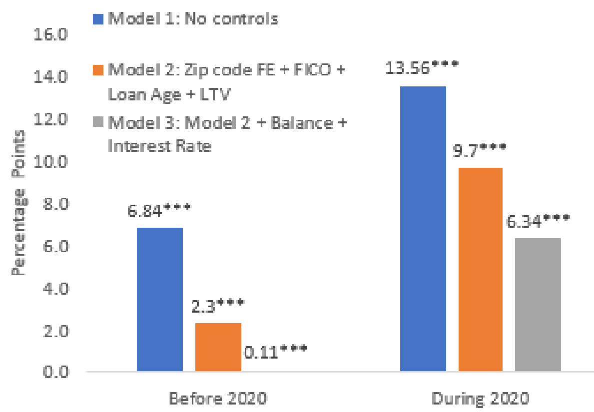
Table A3: Comparison of Matched and Unmatched Loans (Averages)

Wave	Matched?	N	FICO	LTV	DTI	Loan Age	UPB	Rate
before 2020	No Match	138,158	735	0.789	0.360	60	$242,226	5.18%
before 2020	Match	39,155	734	0.786	0.360	61	$238,389	5.18%
2020	No Match	180,960	744	0.811	0.360	37	$273,533	4.70%
2020	Match	43,275	745	0.819	0.358	33	$278,064	4.68%

The matched loans could differ in ways that the univariate distributions do not capture. To help mitigate this possibility in our analysis of the matched loans, we used sampling weights derived from our estimate of the likelihood of a prepaid loan being matched using observable characteristics. We coded each prepaid loan in our sample as one if there was a match and zero otherwise. We fit a linear probability model for the likelihood of a loan being matched using the wave, income quintile, FICO score, UPB, LTV, loan age, and potential savings from refinancing. We then used the inverse of the fitted probability from this regression as our sampling weight for matched loans (using a weight of 1 for all loans that were not prepaid). The regression showed that higher FICO loans, higher LTV loans, and younger loans had a slightly lower chance of having a match, but the difference was relatively small. For example, going from a FICO score of 741 to 739 increased the match probability by only 0.8 percentage points. A loan prepaid in less than one year (loan age under 12 months) was 1.7 percentage points less likely to be matched than a loan that was more than seven years old.

With this information, we estimate our main specification to study the evolution of refinancing inequality with and without controls. The results are presented in Table 2 and Figure A1.

Figure A1: Refinancing Income Gap Before and During 2020, Estimated with Different Sets of Control Variables (OLS model)



Note: The refinancing income gap is defined as the difference in refinancing activity between the top and bottom quintiles of the income distribution. It is represented by β_5 before 2020, and by $\beta_5 + \phi_5$ during 2020, based on the coefficients that result from estimating Equation 1 with different sets of control variables, using data from Freddie Mac.

Appendix B Refinancing Activity with McDash data: Basic Results, Alternative Time Periods and Heterogeneity Analysis

In the McDash data, we observe only when a mortgage is prepaid, but we cannot distinguish refinancing transactions from other types of prepayments. We thus proxy refinancing activity with prepayments and provide a variety of robustness tests to argue that our results with McDash data are also driven by rate-refinances (See). With McDash we do not observe income directly, but instead use an estimate of monthly income based on debt-to-income ratios reported at origination. Nevertheless, McDash data has broader coverage. We thus repeat some of the main features of our analysis for robustness purposes.

As before, to study the role of income on refinancing activity we regress a dummy variable indicating whether or not a mortgage was prepaid on a rich set of loan level covariates, income quintiles, and their interaction with a binary variable identifying observations corresponding to the pandemic period. The results are presented in Table B1. Column 1 shows the coefficients of interest without any control variable. The difference in refinancing activity between the top and bottom quintile of the income distribution before 2020 is 2.7 percentage points. After 2020, the difference in refinancing activity between the bottom and top quintiles of the income distribution increases to 7.4 percentage points (2.72 + 4.66). However, this change could be driven by changes in the composition of loans across zip codes that became newly in-the-money during the observation periods. To address this challenge, we gradually add a rich set of control variables to assess the sensitivity of our estimates. In column 2 we add zip code fixed effects flexible controls for borrower and loan attributes; namely dummy variables for FICO score bins, LTV bins, and bins of loan age.

Holding these characteristics constant, we find that the gap in refinancing activity between the bottom and top quintiles of the income distribution increased from 114 basis points (bps) to 554 bps. Finally, in column 3 we include original interest rates and unpaid balances in the set of control variables. Column 3 thus estimates the role of income on refinancing activity for individuals in the same zip code, with the same FICO score, loan age,LTV, interest rates and unpaid balances (which determine potential savings from refinancing). We find that the gap in refinancing activity between the top and bottom quintiles of the income distribution increases from -44 bps before 2020, to 832 bps during 2020 (-0.44 + 8.66).

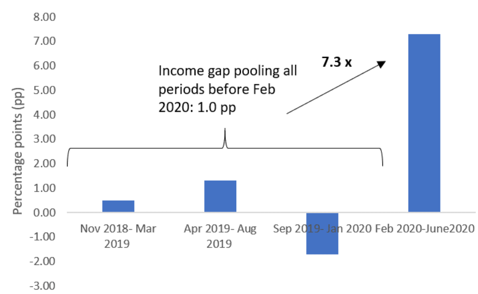
Table B2 studies heterogeneity in refinancing inequality over credit scores, loan to value rations and FICO scores. We estimate our preferred specification (Equation 1) splitting the sample across three selected variables of interest. Columns 1 and 2 split the sample based on original interest rates. Column 1 shows that for individuals with the higher original interest rates, the difference in refinancing activity between the top and bottom quintile of the income distribution was 87 bps. But during the first few months of 2020, the refinancing income gap grew to 331 bps (0.87 + 2.44). Individuals with lower interest rates also experienced an important increase in the refinancing income gap from 87 bps to 507 bps, as shown in column 2. This increase is larger than the increase in inequality experienced by individuals with the highest interest rate incentives. Similarly, in columns 3 and 4, we split the sample based on unpaid balances. Column 3 shows that for individuals with balances above the median, the refinancing income gap increased from 124 bps to 542 bps. For borrowers with balances below the median, column 4 shows that the refinancing income gap increased from 68 bps, to 274 bps. Finally, in columns 5 and 6 we split the sample by FICO score. For borrowers with FICO scores greater than 740 shown in column 5, the refinancing income gap increased from 112 bps to 491 bps between 2020 and previous periods of similar interest rate declines. The increase is comparable to the change experienced by borrowers with FICO scores below 740 shown in column 6, whose refinancing income gap increased from 72 bps before the pandemic to 446 bps in the first months of 2020.

We also perform a variety of tests to the data to confirm that when using McDash prepayments as an outcome variable we are capturing refinancing behavior. These validations include looking at the purpose of loans originated: new purchases, rate refinancing and cash-out refinancing transactions; as well as a geographic analysis of new purchases across the income distribution. These were all available in previous versions of the paper but are not removed for brevity. These are all available from the authors upon request.

Finally, we investigate the evolution of refinancing inequality over the last 15 months. As before, we considered five-month windows to allow a reasonable amount of time for refinancing. The results are presented in Figure B1. Panel (a) shows the refinancing income gap in each period ($\beta_5 + \phi_5$). For reference, panel (b) shows the evolution of mortgage rates during the period. While interest rates were consistently declining from their 2018 peak, refinancing inequality was not trending upward. Instead, a dramatic increase took place between February and June 2020, leading to inequality levels 7.3 times higher than in the 15 months immediately preceding the start of the pandemic. The results in Figure B1 are based on Table B3.

Figure B1: Pre-Pandemic Short Term Trends in Refinancing Inequality

(a) Refinancing income gap



(b) Fixed-term 30 year mortgage rate

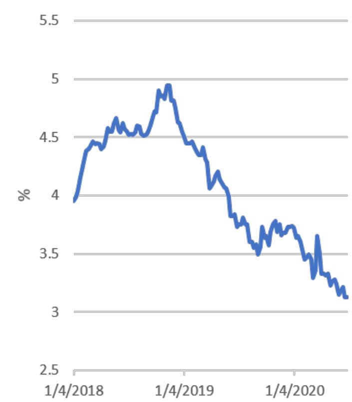


Table B1: Refinancing Inequality Before and During 2020: Probability of Refinancing (percentage points)

Dep.Var. Refinancing Indicator {0,1}			
	(1)	(2)	(3)
Income quintile 2	0.42^{***}	-0.18^{***}	-0.11^{*}
	(0.06)	(0.06)	(0.06)
Income quintile 3	1.24^{***}	0.08	-0.09
	(0.06)	(0.06)	(0.06)
Income quintile 4	1.97^{***}	0.39^{***}	-0.56^{***}
	(0.06)	(0.07)	(0.07)
Income quintile 5	2.72^{***}	1.14^{***}	-0.44^{***}
	(0.07)	(0.07)	(0.08)
Wave 2020	0.52^{***}	-0.90^{***}	0.86^{***}
	(0.08)	(0.08)	(0.09)
Income quintile 2:Wave 2020	1.28^{***}	1.02^{***}	1.69^{***}
	(0.10)	(0.11)	(0.11)
Income quintile 3:Wave 2020	2.63^{***}	2.36^{***}	3.75^{***}
	(0.11)	(0.11)	(0.11)
Income quintile 4:Wave 2020	4.26^{***}	4.09^{***}	6.40^{***}
	(0.11)	(0.12)	(0.13)
Income quintile 5:Wave 2020	4.66^{***}	4.40^{***}	8.66^{***}
	(0.12)	(0.13)	(0.14)
Mean of dep.var. in ref. cat.	2.46	2.46	2.46
Zip code fixed effect	No	Yes	Yes
Borrower controls	No	Yes	Yes
Controls for UPB and original interest rate	No	No	Yes
Observations	1775920	1775920	1775920
R2	0.011	0.04	0.049

Note: This table presents the results of estimating equation 1 with data from McDash. The dependent variable takes the value of one when a mortgage was prepaid, and zero otherwise. The full list of control variables is as follows: zip code fixed effects, loan age, FICO score, loan to value, original interest rate, investor type fixed effects (GSE, private label or portfolio), and unpaid balance (UPB) (continuous variables are split into discrete categories and controlled for as dummies). Income quintile 1 is the lowest income quintile.

Table B2: Heterogeneity Analysis: Sample Splits

	Dep. Var. Refinancing {0,1}					
	Interest Rate Split		Unpaid Balance Split		FICO Split	
	(1)	(2)	(3)	(4)	(5)	(6)
Income quintile 2	0.05	−0.95***	−0.24	0.09	0.05	0.02
	(0.06)	(0.29)	(0.26)	(0.06)	(0.11)	(0.08)
Income quintile 3	0.46***	−0.65**	−0.05	0.53***	0.33***	0.30***
	(0.07)	(0.27)	(0.24)	(0.07)	(0.11)	(0.08)
Income quintile 4	0.61***	−0.15	0.40*	0.56***	0.43***	0.33***
	(0.08)	(0.27)	(0.24)	(0.08)	(0.11)	(0.09)
Income quintile 5	0.87***	0.87***	1.24***	0.68***	1.12***	0.72***
	(0.09)	(0.27)	(0.23)	(0.12)	(0.12)	(0.10)
Wave 2020	−0.45***	−0.08	−0.68	−0.45***	−1.19***	−0.54***
	(0.08)	(0.85)	(0.64)	(0.09)	(0.14)	(0.11)
Income quintile 2:Wave 2020	0.90***	1.38	1.40**	0.95***	0.98***	1.08***
	(0.11)	(0.89)	(0.69)	(0.11)	(0.18)	(0.14)
Income quintile 3:Wave 2020	1.63***	2.99***	2.56***	1.56***	2.21***	2.05***
	(0.12)	(0.86)	(0.65)	(0.13)	(0.18)	(0.15)
Income quintile 4:Wave 2020	2.22***	5.06***	4.20***	2.19***	3.90***	3.12***
	(0.13)	(0.86)	(0.65)	(0.16)	(0.18)	(0.16)
Income quintile 5:Wave 2020	2.44***	4.20***	4.18***	1.96***	3.79***	3.74***
	(0.17)	(0.86)	(0.65)	(0.22)	(0.19)	(0.18)
Mean of dep.var. in ref. cat.	2.42	3.35	2.9	2.44	2.76	2.29
Sample filter	Rate incentive high	Rate incentive low	UPB high	UPB low	FICO ge 740	FICO lt 740
Zip code fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Controls for borrower attributes	Yes	Yes	Yes	Yes	Yes	Yes
Controls for UPB and interest rate	No	No	No	No	Yes	Yes
Observations	848,357	848,355	848,362	848,350	815,989	880,723
R2	0.063	0.036	0.051	0.040	0.054	0.039

Note: This table presents the results of estimating equation 1 with data from McDash. The dependent variable takes the value of one when a mortgage was prepaid, and zero otherwise. The full list of control variables is as follows: zip code fixed effects, loan age, FICO score, loan to value, original interest rate, investor type fixed effects (GSE, private label or portfolio), and unpaid balance (UPB) (continuous variables are split into discrete categories and controlled for as dummies). Income quintile 1 is the lowest income quintile.

Table B3: Refinancing Inequality Over the 20 Months to June 2020

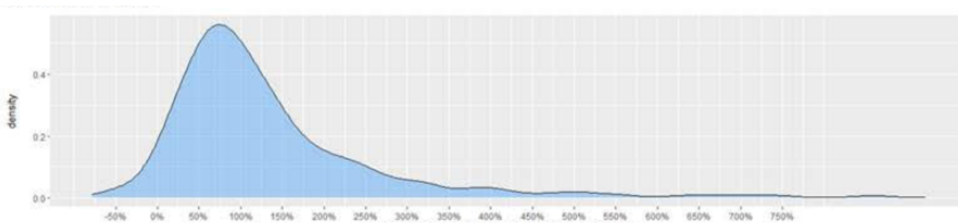
Dep.Var. Refinancing Indicator {0,1}			
	(1)	(2)	(3)
Income quintile 2	0.005***	-0.003***	-0.002***
	(0.001)	(0.001)	(0.001)
Income quintile 3	0.010***	-0.003***	-0.011***
	(0.001)	(0.001)	(0.001)
Income quintile 4	0.016***	-0.002*	-0.023***
	(0.001)	(0.001)	(0.001)
Income quintile 5	0.022***	0.005***	-0.026***
	(0.001)	(0.001)	(0.001)
Wave 2020	0.015***	0.004***	0.039***
	(0.001)	(0.001)	(0.001)
Income quintile 2:Wave 2020	0.012***	0.013***	0.026***
	(0.001)	(0.001)	(0.001)
Income quintile 3:Wave 2020	0.028***	0.031***	0.053***
	(0.001)	(0.001)	(0.001)
Income quintile 4:Wave 2020	0.047***	0.051***	0.083***
	(0.001)	(0.001)	(0.001)
Income quintile 5:Wave 2020	0.052***	0.056***	0.107***
	(0.001)	(0.001)	(0.002)
Mean of dep.var. in ref. cat.	0.015	0.015	0.015
Zip code fixed effect	No	Yes	Yes
Borrower controls	No	Yes	Yes
Controls for UPB and interest rate	No	No	Yes
Observations	1127525	1127525	1127525
R2	0.018	0.051	0.068

Note: This table presents the results of estimating equation 1 with data from McDash. We consider observations corresponding to the last 20 months before June 2020. The dependent variable takes the value of one when a mortgage was prepaid, and zero otherwise. The full list of control variables is as follows: zip code fixed effects, loan age, FICO score, loan to value, original interest rate, investor type fixed effects (GSE, private label, or portfolio), and unpaid balance (UPB) (continuous variables are split into discrete categories and controlled for as dummies). Income quintile 1 is the lowest income quintile.

Appendix C Applications by Lender

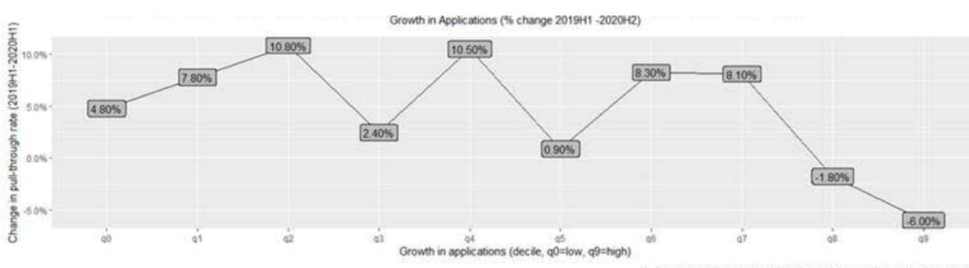
This section shows how applications received by lender changed between the first half of 2020 and the first half of 2019. Figure C1 shows the density of the growth rates in applications by lender. Figure C2 shows the change in funding rates by decile of lender-level growth rate during.

Figure C1: Lender-Level Growth Rates in Number of Applications: First Half of 2020 vs First Half of 2019 (density)



Note: The figure shows the distribution of the lender level growth rates in applications submitted to Freddie Mac's Loan Product Advisor tool, between the first six months of 2020 and the first six months of 2019.

Figure C2: Change in Funding Rates by Decile of Lender-Level Growth Rate in Applications: First Half of 2020 vs First Half of 2019



Note: The figure shows the change in the fraction of applications funded by Freddie Mac for lenders with different growth rates in applications. Changes and growth rates are calculated between the first six months of 2020 and the first six months of 2019.

To describe changes in applications' processing times across the income distribution, we focus on applications eventually purchased by Freddie Mac. Table C1 shows that, compared with previous waves of low interest rates, applications have been processed slightly faster during 2020 than before. Further, conditional on an application resulting in a loan purchased by Freddie Mac, income does not predict differences in processing times. In January 2019, it took nearly the same time to process applications of borrowers in the bottom decile of the income distribution (an average of 45 days) as it did to process applications of borrowers in the top decile (44.2 days). In June 2020, applications of the lowest-income borrowers took about the same time as before to process (44 days), and only applications of borrowers with the highest income were slightly delayed (processing time of 48.9 days). This is evidence against the hypothesis that refinancing inequality is driven by lenders prioritizing *the processing* of applications of high-income borrowers over applications of low-income borrowers received in the same period.

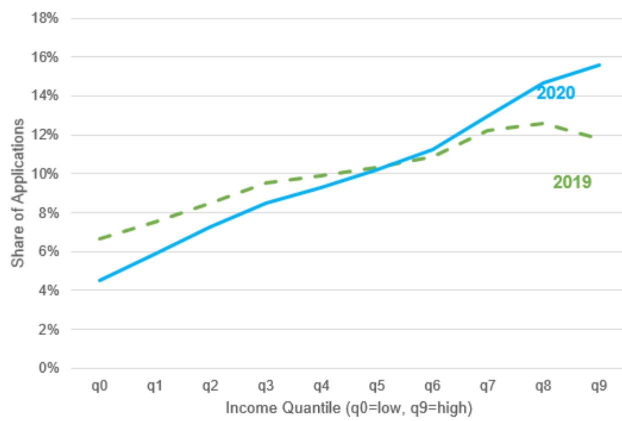
Table C1: Time to Process Applications, across the Income Distribution

Income Decile	2019 Jan	2019 Feb	2019 March	2019 April	2019 May	2019 Jun	2019 July	2019 Aug	2019 Sep	2019 Oct	2019 Nov	2019 Dec	2020 Jan	2020 Feb	2020 March	2020 Apr	2020 May	2020 Jun
1	45.0	41.6	40.7	37.1	40.5	40.8	40.0	41.0	40.7	44.4	45.5	46.2	46.5	36.4	34.4	39.8	43.3	44.0
2	46.9	39.9	38.2	35.6	39.6	39.7	39.0	39.6	41.3	44.7	44.7	44.8	45.5	35.4	33.2	39.5	43.5	43.6
3	47.8	40.5	39.7	34.5	39.6	37.9	38.6	39.4	39.9	43.9	45.2	45.2	46.1	35.1	32.8	40.1	43.9	43.1
4	46.7	40.8	40.0	33.9	40.6	38.5	38.7	40.3	39.5	43.6	45.2	44.8	44.3	33.9	32.9	40.0	44.1	43.4
5	45.8	41.3	41.4	34.0	41.1	40.8	38.1	38.6	39.6	43.6	46.1	44.2	44.2	34.0	33.0	40.4	45.4	44.3
6	47.4	40.1	41.8	32.7	39.3	40.2	37.4	40.6	39.5	44.1	45.7	47.3	45.3	35.4	33.2	40.4	44.2	44.0
7	41.5	44.5	38.8	35.0	40.0	39.9	38.8	40.2	39.8	44.7	46.0	47.5	44.8	33.9	32.8	40.1	44.6	45.0
8	43.3	38.5	44.0	34.0	40.7	38.0	40.0	39.9	39.5	44.2	46.8	45.9	45.9	34.4	32.9	40.6	46.0	45.2
9	47.4	43.2	43.0	36.0	39.3	38.8	40.1	41.5	40.2	45.9	47.1	48.3	45.8	38.0	33.2	41.8	47.2	45.2
10	44.2	44.2	38.7	35.3	43.9	44.2	40.0	43.0	42.7	47.6	48.0	49.5	51.4	36.7	35.3	42.0	48.8	48.9

Note: This table presents the average time between application and closing (in days) for conventional loans, across income deciles and over time. It considers only applications ultimately funded by Freddie Mac, and therefore approved by the lender. Each row corresponds to applications submitted on a given income decile. Each column corresponds to applications submitted on a given year and month. The data covers the period between January 2019 and June 2020.

Even if lenders didn't change their approval rates, capacity constraints at the operational stages of the application-funding process could lead lenders to re-direct their marketing efforts towards high-income borrowers. We calculate income deciles for the portfolio of active mortgages, and plot the distribution of applications across those deciles, for the 2019 and 2020 waves as defined above (see Figure C3).

Figure C3: Refinance Applications by Income Group



Note: This figure shows the fraction of refinancing applications submitted through Freddie Mac's Loan Product Advisor tool, that fall in each income group. We define ten income groups, based on the deciles of the income distribution observed in Freddie Mac's portfolio of active mortgages. The analysis is presented for the 2019 and 2020 waves of refinancing activity defined in Figure 1.

Appendix D Refinancing Inequality and COVID-19 Severity

D.1 Non-linear effects of COVID-19 Case Rates and other measures of severity

In Section 5 we study the relation between COVID-19 case rates and refinancing inequality with a summary variable that compares individuals in the bottom quintile of the distribution of COVID-19 case rates, to individuals in quintiles 2,3,4 and 5. Here we esimate a more flexible specification that justifies the choice of summary variable used in the main text.

We estimate the following equation:

$$\begin{aligned} y_{izct} = \alpha_z + \alpha_t + \sum_{j=2}^{5} \beta_j * Income\ quintile_{ji} + \sum_{k=2}^{5} \gamma_k * Severity\ Q_{kizct-1} \\ + \sum_{k=2}^{5} \sum_{j=2}^{5} \phi_{jk} * Income\ quintile_{ji} * Severity\ Q_{kizct-1} + \delta * X_{it} + \epsilon_{izcgt} \end{aligned} \tag{3}$$

where y_{izct} indicates whether mortgage i in zip code z and county or state c was refinanced in period t; Income quintile ji represents a set of dummy variables indicating whether mortgage i belongs to income quintile j; $Severity\ Q_{kizct-1}$ is a dummy variable indicating whether mortgage i in zip code z in county or state c belongs to the quintile k of the distribution of COVID-19 severity in month $t-1$; and X_{it} is a vector of loan-level controls. To reflect that refinancing applications take between 1 and 1.5 months to be processed, we use a one month lag of the variables to measure the severity of the crisis. This way, the refinancing activity after households increased their time at home in month t-1, is measured in month t.

This flexible specification allows us to identify non-linearities in the effect of the pandemic on refinancing inequality.

The coefficient β_5 represents the refinancing income gap in geographic area-months where the pandemic had the least impact. The coefficient ϕ_{5k} represents increases in the refinancing income gap for mortgages in geography-months that lie on the jth quintile of the distribution of COVID-19 severity, relative to those in the bottom quintile.

In addition of using COVID-19 case rates as our only measure of severity, we also use 3 alternative measures of severity: unemployment insurance claims, time spent at home and forbearance rates. The coverage of these four variables is imperfect and subject to availability by data providers. COVID-19 case rates are available for 3,023 counties, which cover 99.9 percent of our mortgage data. We refer to these counties and mortgages as our base coverage for the pandemic analysis. All four variables are available for a subset of 1.2 million observations at the mortgage-month level, representing 43 percent of our base coverage for the pandemic analysis.[24]

We estimate Equation 3 four times, each time using COVID-19 case rates, time spent at home, unemployment insurance claims and forbearance rates as alternative measures of severity of the pandemic.[25] This approach should be considered as complementary to the mediation analysis discussed in Section 6. The full set of coefficients is presented in columns 2,3,4, and 5 of Table D1. Column 1 considers all observations for which information about

[24] Mobility measures are available only for 26 percent of those counties covering 86 percent of observations in our base coverage. Unemployment insurance claims at the county level are available for 52.2 percent of mortgages in our base coverage. Forbearance rates (at the state level) are available for 87 percent of observations in our base coverage. For robustness, we also perform the analysis with unemployment insurance claims at the state level. This allows to increase our coverage to 79 percent of our original observations, however with a coarse measure of unemployment insurance claims. The results are qualitatively the same. The analysis with unemployment insurance at the state level is available upon request.

[25] For robustness, we also replicate the analysis with unemployment insurance claims measured at the state-month level. This increases our observations to 2.3 million mortgage-months. The results are qualitatively the same and are available from the authors upon request.

COVID-19 case rates is available, even if some of the other variables is missing. That column is included only for comparison purposes. We plot and interpret the coefficients of interest in Figure D1.

Figure D1 plots our estimates for the refinancing income gap across geography-months with different levels of COVID-19 severity, where severity is measured by case rates, time spent at home, forbearance rates or unemployment insurance claims. The black line represents projected levels of the refinancing income gap in county-months that fall in different quintiles of the COVID-19 severity distribution. These correspond to β_5 for the bottom quintile of COVID-19 severity and $\beta_5 + \phi_{5j}$ for quintiles j=2 to 5, respectively. As before, we can see a slight inverse U-shape when we measure severity with case rates (panel (a)). The same shape is present when we measure the severity of the pandemic time spent at home (panel (d)). Forbearance has a sustained positive correlation with refinancing inequality (panel (b)). In contrast, unemployment insurance claims have small effects across the board (panel (c)). The blue bars represent changes in the refinancing income gap relative to the bottom quintile of COVID-19 severity, as we move to higher quintiles of severity (that is ϕ_{5j} with j=2 to 5, in Equation 3). The largest increases in inequality result from moving from the bottom quintile to any of the other quintiles. The effect from moving across contiguous quintiles is smaller. Figure D1 is based on columns 3,4 and 5 of Table D1.

Motivated by these non-linear effects, we look at the impact of the severity of the pandemic on refinancing inequality by interacting income quintiles with different summary measures of severity that compare the bottom quintile to the remaining top quintiles. (See Figure D2). We confirm that the results are qualitatively the same as in the mediation analysis presented in the main body of the text.

D.2 Mediation Analysis

Figure D3 plots the results of the mediation analysis based on all 5 columns of Table 5.

The first bar of Figure D3 represents the omnibus effect of High COVID-19 case rates on refinancing inequality without additional controls for local economic conditions. Bars 2 to 4 include a separate set of control variables, as described in the horizontal axis. The last bar includes all of these controls simultaneously. The fraction of mortgages under forbearance and time spent at home explain, respectively, 30 ((1.26 - 0.88)/1.26) and 10 percent ((1.26 - 1.14)/1.26) of the increases in refinancing inequality tied to the local impact of the pandemic. Unemployment insurance claims do not have explanatory power. Together, these three variables explain around 44 percent (((1.26 - 0.71)/1.26) of the impact of local economic conditions on refinancing inequality.

Table D2 provides the coefficients used to produce Figure 3

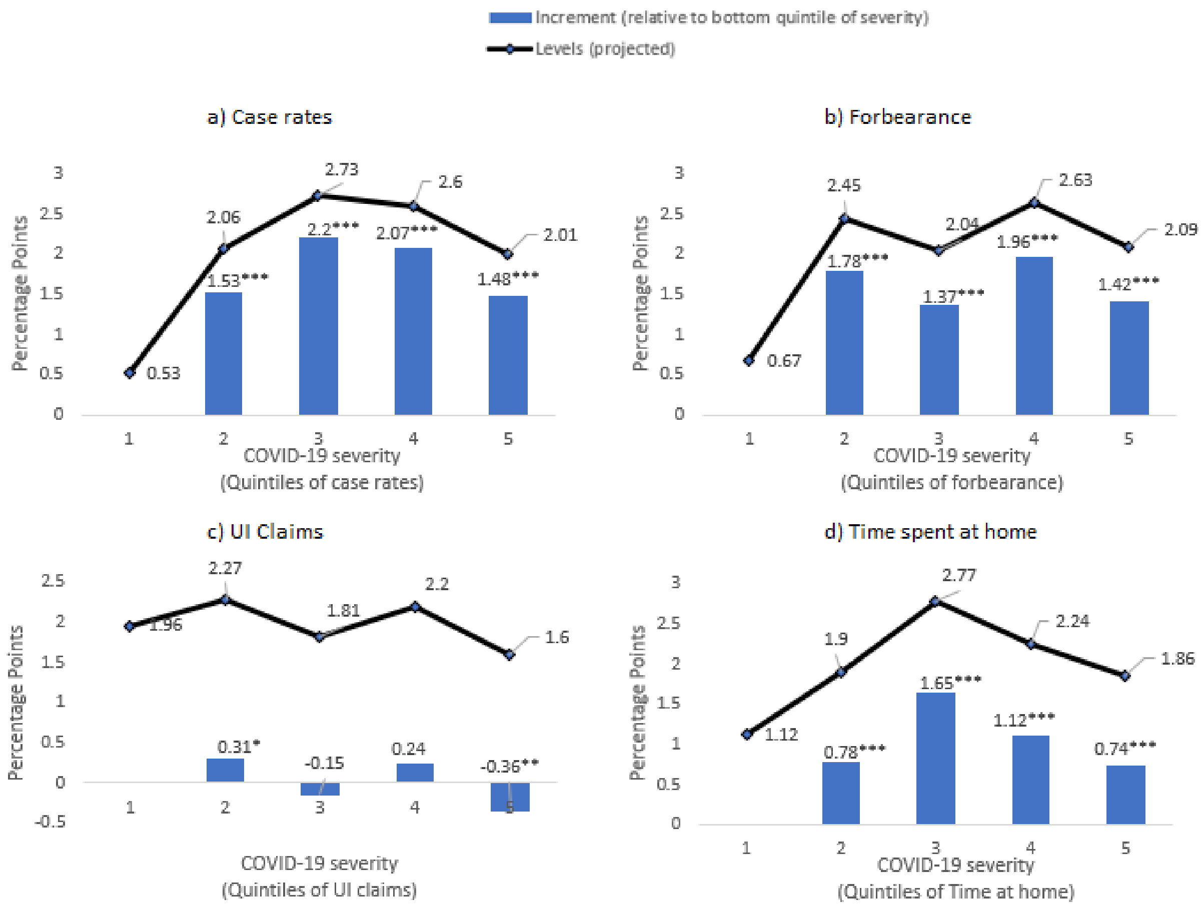
Table D1: Refinancing Inequality Across the Distribution of COVID-19 Severity by Severity Measure

	(1)	(2)	(3)	(4)	(5)	(6)
Severity Measure	Case rate	Case rate	Forbearance	UI Claims	Time at Home	Case rate
β_2 IncomeQuintile2	0.01	0.23***	0.40***	0.78***	0.58***	0.29**
	(0.05)	(0.09)	(0.08)	(0.09)	(0.08)	(0.14)
β_3 IncomeQuintile3	-0.12**	0.22**	0.46***	1.29***	0.78***	0.18
	(0.05)	(0.09)	(0.09)	(0.1)	(0.09)	(0.15)
β_4 IncomeQuintile4	-0.44***	0.16*	0.36***	1.77***	0.95***	0.16
	(0.06)	(0.1)	(0.1)	(0.11)	(0.1)	(0.16)
β_5 IncomeQuintile5	-0.39***	0.53***	0.67***	1.96***	1.12***	0.14
	(0.07)	(0.11)	(0.11)	(0.13)	(0.12)	(0.18)
γ_2 Severity2	-1.28***	-0.82***	-1.33***	0.06	-0.33***	-0.64***
	(0.07)	(0.12)	(0.13)	(0.11)	(0.11)	(0.16)
γ_3 Severity3	-1.85***	-1.13***	-1.48***	0.59***	-0.47***	-0.78***
	(0.09)	(0.13)	(0.14)	(0.11)	(0.12)	(0.17)
γ_4 Severity4	-1.97***	-1.16***	-1.25***	0.30***	-0.2	-0.77***
	(0.09)	(0.13)	(0.18)	(0.11)	(0.14)	(0.18)
γ_5 Severity5	-1.98***	-0.89***	-1.20***	0.23*	0.27	-0.45**
	(0.1)	(0.16)	(0.15)	(0.12)	(0.18)	(0.2)
$\phi_{2,2}$ IncomeQuintile2:Severity2	0.64***	0.44***	0.42***	-0.04	0.1	0.47***
	(0.07)	(0.13)	(0.12)	(0.13)	(0.12)	(0.18)
$\phi_{3,2}$ IncomeQuintile3:Severity2	1.09***	0.90***	0.84***	-0.08	0.39***	0.88***
	(0.08)	(0.14)	(0.13)	(0.14)	(0.13)	(0.19)
$\phi_{4,2}$ IncomeQuintile4:Severity2	1.65***	1.16***	1.49***	-0.15	0.55***	0.93***
	(0.08)	(0.14)	(0.14)	(0.15)	(0.14)	(0.2)
$\phi_{5,2}$ IncomeQuintile5:Severity2	1.98***	1.53***	1.78***	0.31*	0.78***	0.94***
	(0.09)	(0.15)	(0.16)	(0.16)	(0.16)	(0.22)
$\phi_{2,3}$ IncomeQuintile2:Severity3	0.84***	0.62***	0.32***	-0.33**	0.27**	0.61***
	(0.08)	(0.13)	(0.12)	(0.13)	(0.13)	(0.19)
$\phi_{3,3}$ IncomeQuintile3:Severity3	1.53***	1.18***	0.90***	-0.62***	0.70***	1.01***
	(0.08)	(0.13)	(0.13)	(0.14)	(0.13)	(0.2)
$\phi_{4,3}$ IncomeQuintile4:Severity3	2.33***	1.85***	1.46***	-0.95***	0.99***	1.32***
	(0.09)	(0.14)	(0.14)	(0.15)	(0.14)	(0.22)
$\phi_{5,3}$ IncomeQuintile5:Severity3	2.88***	2.20***	1.37***	-0.15	1.65***	1.26***
	(0.1)	(0.15)	(0.16)	(0.16)	(0.16)	(0.23)
$\phi_{2,4}$ IncomeQuintile2:Severity4	0.83***	0.59***	0.2	-0.14	0.08	0.63***
	(0.08)	(0.12)	(0.15)	(0.13)	(0.12)	(0.19)
$\phi_{3,4}$ IncomeQuintile3:Severity4	1.49***	1.08***	0.71***	-0.32**	0.31**	0.96***
	(0.08)	(0.13)	(0.16)	(0.14)	(0.13)	(0.21)
$\phi_{4,4}$ IncomeQuintile4:Severity4	2.47***	1.68***	1.21***	-0.38**	0.69***	1.13***
	(0.09)	(0.14)	(0.16)	(0.15)	(0.14)	(0.22)
$\phi_{5,4}$ IncomeQuintile5:Severity4	3.16***	2.07***	1.96***	0.24	1.12***	1.23***
	(0.1)	(0.15)	(0.17)	(0.17)	(0.16)	(0.24)
$\phi_{2,5}$ IncomeQuintile2:Severity5	0.70***	0.38***	0.19	-0.14	-0.02	0.40**
	(0.08)	(0.13)	(0.13)	(0.13)	(0.14)	(0.2)
$\phi_{3,5}$ IncomeQuintile3:Severity5	1.34***	0.88***	0.37***	-0.30**	-0.07	0.67***
	(0.08)	(0.13)	(0.13)	(0.14)	(0.14)	(0.21)
$\phi_{4,5}$ IncomeQuintile4:Severity5	2.08***	1.17***	0.81***	-0.67***	-0.1	0.52**
	(0.09)	(0.14)	(0.14)	(0.15)	(0.15)	(0.23)
$\phi_{5,5}$ IncomeQuintile5:Severity5	2.76***	1.48***	1.42***	-0.36**	0.74***	0.39
	(0.1)	(0.15)	(0.14)	(0.16)	(0.16)	(0.25)
Mean of dep.var. in ref. cat.	0.71	0.98	0.98	0.98	0.98	0.98
Zip code fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Clustered error	Zip	Zip	Zip	Zip	Zip	Zip
Controls for borrower attributes	Yes	Yes	Yes	Yes	Yes	Yes
Controls for rate and UPB	Yes	Yes	Yes	Yes	Yes	Yes
Controls for other severity measures	No	No	No	No	No	Yes
Filters	No filter	Coverage of all Controls for Local Economic Conditions				
Observations	2,895,722	1,242,204	1,242,204	1,242,204	1,242,204	1,242,204
R2	0.02	0.02	0.02	0.02	0.02	0.02

Note: The notes to this table are presented in the next page.

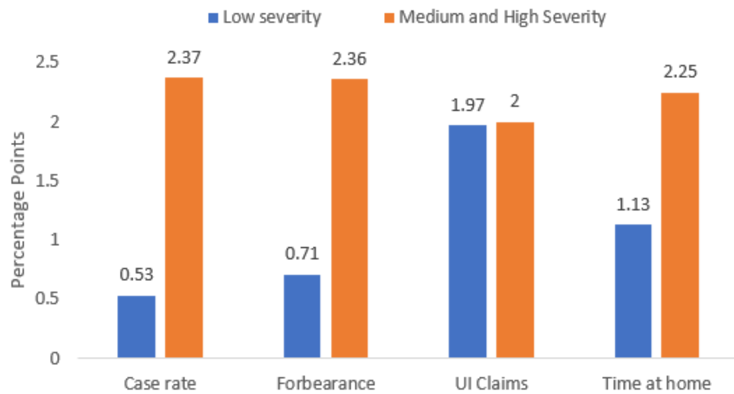
Notes to Table D1: This table presents the results of estimating Equation 3, using data from McDash. Column 1 considers all observations for which we have coverage of COVID-19 case rates. Columns 2-6 considers observations for which we have coverage of COVID-19 case rates, forbearance rates, time spent at home and unemployment insurance claims. The dependent variable takes the value of one when a mortgage was prepaid and zero otherwise. We consider mortgages that were not in-the-money in February 2020 and became in-the-money in any of the subsequent periods until July 2020. For these mortgages we have monthly observations between the first month in which they turn in-the-money and up until the month in which they are prepaid. The reference category for calculating mean prepay rates in columns 1 to 5 is defined as the bottom quintile of income and corresponding severity measure. The reference category for calculating mean prepay rates in column 6 is defined as bottom quintile of income and bottom quintile in all severity measures. All columns include time and geographic fixed effects, as well as controls for borrower attributes, interest rate and unpaid balance. In addition, Column 6 includes controls for all other severity measures and their interaction with income quintiles (coefficients for not shown for brevity but available from the authors upon request). The coefficient β_5 represents the refinancing income gap in geography-months where the pandemic had the least income. The coefficient ϕ_{5k} represents increases in the refinancing income gap for mortgages in geography-months that lie on the jth quintile of the distribution of COVID-19 severity, relative to those in the bottom quintile. Standard errors are clustered at the county level. UPB = unpaid balance.

Figure D1: Refinancing Income Gap Across County-Months With Different Levels of COVID-19 severity



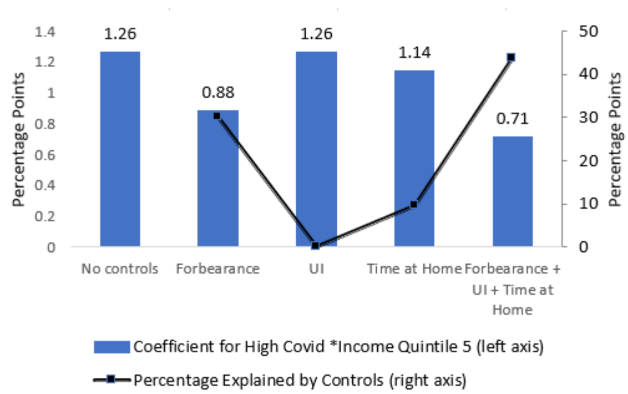
Note: COVID-19 severity is measured separately in each panel with four measures: case rates, forbearance, unemployment insurance (UI) claims and time spent at home. The black line represents projected levels of the refinancing income gap in geography-months that fall in different quintiles of the COVID-19 case rate distribution. The refinancing income gap is defined as the difference in refinancing activity between the top and bottom quintiles of the income distribution. For mortgages in county-months in the bottom quintile of the COVID-19 severity distribution, the refinancing income gap is represented by β_5. For mortgages in geography-months in quintiles k =2 to 5 of the COVID-19 severity distribution, the refinancing income gap is represented by $\beta_5 + \phi_{5k}$. Where β_5 and ϕ_{5k} result from estimating Equation 3 with the corresponding measure of COVID-19 severity. The blue bars represent changes in the refinancing income gap relative to the bottom quintile of COVID-19 severity, captured by ϕ_{5k}. Sample is restricted to observations for which all severity measures have coverage. This graph is based in the coefficients presented in Table D1, using data from McDash.

Figure D2: Estimates of the Refinancing Income Gap by Severity of the Disruption to Local Economies (Different Measures of Disruption)



Note: The refinancing income gap is defined as the difference in refinancing activity between the top and bottom quintiles of the income distribution. The severity of disruption to local economies is measured with four different variables: COVID-19 case rates, forbearance rates, initial unemployment insurance claims and time spent at home. We estimate Equation 3, using one of these four variables as a measure of severity each time. Low severity corresponds to the bottom quintile, medium to high severity corresponds to quintiles 2 to 5 of the corresponding severity distribution. Sample is restricted to observations for which all severity measures have coverage. All differences between low and medium to high severity estimates are statistically significant at standard levels, except when severity is measured with unemployment insurance (UI) claims. This graph is based in the coefficients presented in Table D1, using data from McDash.

Figure D3: COVID Case Rates and The Refinancing Income Gap with and without Controlling for Time at home, Forbearance and Unemployment Insurance



Note: This graph shows the coefficient for High COVID * Income Quintile 5 from an expanded version of Equation 2 (ϕ_5) that includes different controls for local economic conditions. The first bar does not include any additional control. Bars 2 to 4 control, separately,for forbearance, unemployment insurance and time at home and their interaction with income quintiles, respectively. The last bar includes all those controls at the same time. The black line represents the fraction of the coefficient without controls that is explained by each set of control variables. For example, the fraction explained by Forbearance is calculated as 1- 0.88/1.26= 30%

Table D2: The Mediator Role of Time at Home in Areas of High and Low Fintech Activity

Dep.Var. Refinancing Indicator {0,1}				
	(1)	(2)	(3)	(4)
IncomeQuintile2	0.01	0.03	0.15	0.14
	(0.12)	(0.13)	(0.1)	(0.1)
IncomeQuintile3	0.09	0.09	-0.02	-0.08
	(0.16)	(0.17)	(0.12)	(0.12)
IncomeQuintile4	0.09	0.23	0.07	-0.04
	(0.2)	(0.21)	(0.14)	(0.14)
IncomeQuintile5	-0.03	-0.12	-0.16	-0.21
	(0.22)	(0.25)	(0.17)	(0.17)
HighCOVID	-0.45***	-0.54***	-0.52***	-0.24*
	(0.17)	(0.2)	(0.11)	(0.13)
HighTimeHome		-0.12		-0.42***
		(0.2)		(0.14)
IncomeQuintile2:HighCOVID	0.75***	0.83***	0.37***	0.28**
	(0.16)	(0.22)	(0.11)	(0.14)
IncomeQuintile3:HighCOVID	1.02***	0.98***	0.81***	0.40**
	(0.19)	(0.28)	(0.14)	(0.18)
IncomeQuintile4:HighCOVID	1.31***	1.58***	1.05***	0.46**
	(0.21)	(0.3)	(0.16)	(0.2)
IncomeQuintile5:HighCOVID	1.84	1.64	1.23	0.85
	(0.28)	(0.34)	(0.19)	(0.25)
IncomeQuintile2:HighTimeHome		-0.12		0.11
		(0.24)		(0.13)
IncomeQuintile3:HighTimeHome		0.04		0.55***
		(0.27)		(0.16)
IncomeQuintile4:HighTimeHome		-0.47		0.80***
		(0.31)		(0.2)
IncomeQuintile5:HighTimeHome		0.33		0.50**
		(0.37)		(0.22)
Sample filter	Bottom Fintech	Bottom Fintech	Top Fintech	Top Fintech
Observations	275,254	275,254	275,372	275,372
R2	0.05	0.05	0.04	0.04

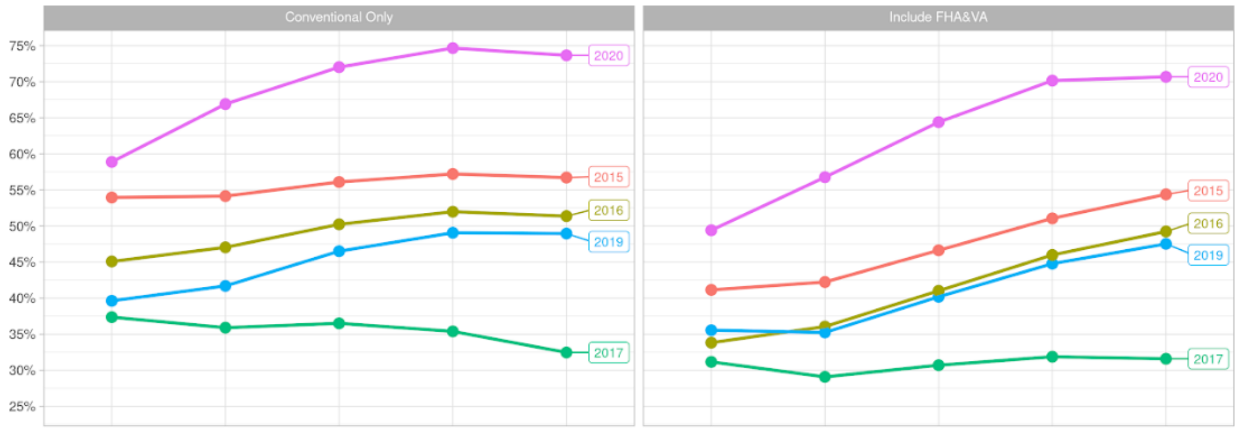
Notes: This table presents the results of re-estimating equation 2 with data from MacDash. Columns 1 and 2 consider observations in the bottom quartile of fintech activity. Colmns 3 and 4 consider observations in the top quartile of fintech activity. Fintech activity is measured aggregating HMDA data at the city level, counting loans origiated by lender flagged as fintech in Fuster et al. (2019). Columns 2 and 4 include controls for time spent at home and its interaction with income quintiles. For time spent at home we create a dummy that takes the value of one when a given geography-month is in quintiles 2-5 of the distribution of time at home. The dependent variable takes the value of one when a mortgage was prepaid, and zero otherwise. The full list of control variables is as follows: zip code fixed effects, loan age, FICO score, loan to value, original interest rate, investor type fixed effects (GSE, private label or portfolio), and unpaid balance (UPB) (continuous variables are split into discrete categories and controlled for as dummies. Income quintile 1 is the lowest income quintile.

Appendix E HMDA data: New originations, including FHA and VA loans

In this section, we perform several robustness analysis. First, we look at the perspective of refinancing inequality on the bases of new originations instead of propensity to refinance. Second, we extend the analysis to include FHA loans and VA loans. Third, we use data from HMDA. We calculate the fraction of new originations in each income quintile that are refinancing transactions.

For the pool of new originations including conventional loans as well as FHA and VA loans, we calculate income quintiles and, for each income quintile, the fraction of new originations for which the purpose of the loan is a refinancing transaction. We then repeat the analysis considering only conventional loans. Figure E1 presents the results separately for each year. We can see that the slope of refinancing activity by income quintile is significantly steeper in 2020, relative to previous years. We thus confirm that are results are robust to different approaches (new originations vs propensity to refinance), to further inclusion of FHA and VA loans, and to alternative datasets such as HMDA. We confirm that refinancing inequality is significantly larger in 2020 than in previous refinancing booms.

Figure E1: Fraction of New Originations that are Refinancing Transactions by Income Quintile Over Time



Note: This figure is based on HMDA data. The left panel considers only conventional loans. For this pool of loans we calculate income quintiles, and for each income quintile we calculate the fraction of new originations which are refinancing loans. The right panel considers conventional loans, FHA loans and VA loans. For this pool of loans we calculate income quintiles, and for each income quintile we calculate the fraction of new originations which are refinancing loans.

Appendix F Closed Form Solution for Optimal Refinancing

The square-root rule suggested by Agarwal et al., 2013 approximates their exact solution for the optimal refinancing differential which requires calls to the Lambert's W-function. It is written as

$$x^{\star}(k_n, M_n, \lambda_n) = -\sqrt{\frac{\sigma k_n}{M_n(1-\tau)}\sqrt{2(\rho+\lambda_n)}} \tag{4}$$

where σ is the annualized standard deviation of the mortgage interest rate, τ is the marginal tax rate, and ρ is the real discount rate. The values of them suggested by Agarwal et al., 2013 are $\sigma = 0.0109$, $\tau = 0.28$, and $\rho = 0.05$. Plugging these numbers back to Eq 4, we have

$$x^{\star}(k_n, M_n, \lambda_n) = -\sqrt{\frac{0.0109 k_n}{0.72 M_n}\sqrt{2(0.05+\lambda_n)}} \tag{5}$$

Their suggested value of σ was based on the standard deviation for monthly differences of the Freddie Mac 30-year mortgage rate from April 1971 to February 2004, which is 0.00315. An annualized standard deviation of $\sigma = \sqrt{12} \times 0.00315 = 0.0109$.

M_n, λ_n, and k_n are loan level (i.e., for the nth loan) measures on current remaining outstanding balance, expected real rate of exogenous mortgage prepay probability, and refinance cost. The value of M_n is straightforward. The value of λ_n is defined as

$$\lambda_n = \mu + (\frac{p_n}{M_n} - i_n^0) + \pi, \tag{6}$$

where p_n is the nominal annual mortgage payment amount of the existing loan n, i_n^0 is the mortgage interest rate for existing loan n, μ is the hazard of relocation, and π is the inflation rate. However, if the existing loan is a fixed rate mortgage, λ_n can be expressed as

$$\lambda_n = \mu + \frac{i_n^0}{exp(i_n^0 \Gamma_n) - 1} + \pi, \tag{7}$$

where Γ_n is the remaining life of the existing mortgage in years, and the rest parameters are defined as before. Agarwal et al., 2013 suggested $\mu = 0.1$, and $\pi = 0.03$. Thus, Eq (6) can be written as

$$\lambda_n = 0.1 + (\frac{p_n}{M_n} - i_n^0) + 0.03, \tag{8}$$

and Eq (7) can be written as

$$\lambda_n = 0.1 + \frac{i_n^0}{exp(i_n^0 \Gamma_n) - 1} + 0.03, \tag{9}$$

The value of k_n is defined as

$$k_n = F + fM_n[1 - \frac{\tau}{\theta + \rho + \pi}[\frac{1 - exp(-(\theta + \rho + \pi)N)}{N} \frac{\rho + \pi}{\theta + \rho + \pi} + \theta]] \tag{10}$$

where F is the fixed cost of refinance, f is the points divided by 100, τ is the marginal tax rate, θ is the expected arrival rate of a full deductibility event - a move or a subsequent refinancing,N is the number of years of the new mortgage, ρ and π are defined the same as before. Agarwal et al., 2013 suggested $F = 2000$, $f = 0.01$, and $\theta = \mu + 0.1 = 0.2$. If we plug these numbers (as well as those values previously suggested for ρ and π) into to Eq (10) and assume $N = 30$, we have

$$k_n = 2000 + 0.007905 M_n. \tag{11}$$

So, ultimately, to compute the optimal threshold, $x^\star(k_n, M_n, \lambda_n)$, we need to collect information from the an individual loan n on its current remaining outstanding balance (i.e.,M_n), the existing interest rate (i.e., i_n^0), the annual payment size (i.e., p_n) if the loan is not a fixed rate loan, and the remaining life of the existing mortgage in years (i.e.,Γ_n), and then use Eqs (8), (9), (11) and (5), we can calibrate $x^\star$ for any individual n.

In the following we explore the possibility that borrowers across the income distribution differ in their prepayment risk or in the closing costs they face, and thus may be incorrectly classified as in-the-money with the parameters we use in our main analysis.

Moving Patterns Across the Income Distribution The prospect of moving to a new home is one determinant behind the decision to refinance a mortgage: individuals who expect to move sooner have less of an incentive to refinance, holding everything else constant. We explore the possibility that increases in refinancing inequality during the first half of 2020 were driven by differences across the income distribution in the probability of moving to a new house. We test the hypothesis that low-income individuals would be more likely to move because of a negative shock to local economic conditions. Using the matched transactions data from Freddie Mac, we compare the probability of prepaying an existing mortgage and then buying a new home, across the income distribution and over time (for the five waves of low interest rates considered in the main analysis). Panel (a) of Figure F1 shows that, on average, the probability of prepaying to buy a new home in periods of low interest rate before 2020 was 0.93 percent for borrowers in the bottom quintile of the income distribution. This probability decreases slightly, by 0.19 percentage points during the first half of 2020. Similarly, for borrowers in the top quintile of the income distribution, the probability of prepaying an existing mortgage to buy a new home before 2020 was 1.27 percent. During 2020, the probability that this group would prepay their mortgage to buy a new home decreased 0.15 percentage points. Therefore, the difference over time in new purchases is about the same across the income distribution. In fact, the probability that high-income borrowers will prepay an existing mortgage to buy a new home is slightly smaller than for borrowers in other income groups. We thus conclude that the increases in refinancing inequality are unlikely to be driven by moving patterns across the income distribution and over time.

Delinquency Patterns Across the Income Distribution Another reason why low-income borrowers did not increase their refinancing activity at the same rate as high-income borrowers could be that low-income borrowers experienced a more than proportional increase in default rates during the period of analysis. For a refinancing transaction to make sense, borrowers need to retain their loans for a sufficiently long period of time. For individuals with a high probability of default, refinancing may not be optimal even if interest rate differentials suggest so. To explore this possibility, we compare trends in delinquency for borrowers in the top and bottom quintiles of the income distribution. Panel b) of Figure F1 shows the evolution of the fraction of newly in-the-money mortgages delinquent at any point during the corresponding refinancing waves considered in the analysis. Compared with periods of low interest rates before 2020, during the first few months of 2020 individuals in the bottom quintile of the income distribution increased the probability of delinquency by 3.06 percentage points, from a base of 4.34 percent. During the same period, individuals in the top quintile of the income distribution increased their probability of delinquency by 4.90 percentage points from a basis of 1.87 percent. Thus, while low-income borrowers generally have higher probabilities of delinquency, changes in delinquency patterns across the income distribution and over time cannot explain the sharp increases in refinancing inequality during the first few months of the pandemic. If anything, high-income borrowers with newly in-the-money mortgages are increasing their delinquency probabilities at a higher rate than low-income borrowers.[26]

Closing Costs Across the Income Distribution Another potential reason why low-income borrowers could be less likely to refinance is if they face higher closing costs than high-income borrowers. However, the structure of closing costs is inconsistent with this view and in general we would expect closing costs to be lower in absolute dollar terms for low-income individuals. While income does not affect closing costs per se, closing costs have a fixed component and a variable component. A portion of the closing costs is typically fixed irrespective of how high the loan balance refinanced would be. This would include certain taxes that come as a fixed dollar amount and part of the loan originators costs which are fixed. Another portion of the closing costs would scale with the loan size, this would include transfer taxes and some of the origination charges. Since high income individuals usually carry larger loan balances (and property values) the dollar value of closing costs (inclusive of variable and fixed component) will usually be higher for high income borrowers.

In our analysis, we take this into account in two ways. First, we classify a mortgage as being in the money, based on the closed form solution of Agarwal et al. (2013). As described in Appendix F, we say that a mortgage is in the money when the interest rate differential between the old and new loans are larger than a loan specific threshold, which in turn depends on closing costs, unpaid balances, loan age and the original interest rate. Closing costs are modeled with a fixed and variable component that depends on unpaid balances.[27].

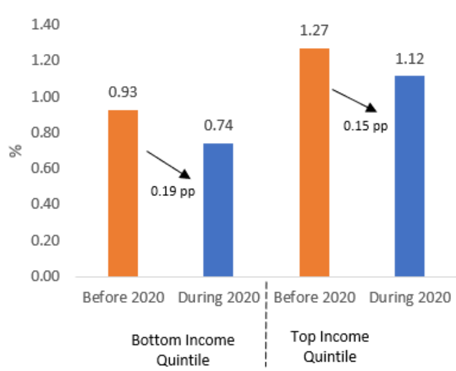
[26] In contrast to An et al. (2021) who study all active mortgages in the market, we focus only on mortgages that became in-the-money, according to the definition of Agarwal et al. (2013) during the first months of 2020 by studying whether our results about refinancing inequality are driven by differences in prepayment across the income distribution for mortgages that we classify as being in-the-money.

[27] Holding everything constant, higher closing costs require a higher interest rate differential. However, increases in closing costs due to higher unpaid balances are more than compensated by the direct impact

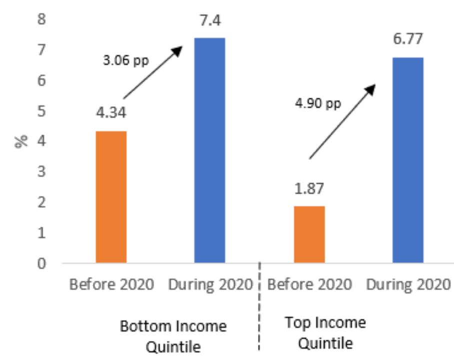
Thus, in contrast to uniform rules for being “in the money" (such as 100 bps across the board), individuals with different closing costs get a different threshold to be classified as “in the money". In addition, our main specification controls for unpaid balances (as well as original interest rate, LTV, loan age, thus neutralizing any mechanical difference on closing costs that could lead to different incentives to refinance across income groups.

Figure F1: Probability of Prepaying to Buy a New New Home or Entering Delinquent Status, Over Time and Across the Income Distribution

(a) Prepaying to Buy a New Home



(b) Delinquency



Note: This figure considers observations from mortgages in the matched transactions database of Freddie Mac, that became newly in-the-money during the five waves of low interest rates considered in the analysis. Panel (a) shows the probability that a borrower prepays a mortgage and enters a new mortgage property for a new purchase within 45 days. Panel (b) shows the probability that a mortgage goes delinquent. A mortgage is flagged as delinquent if it is delinquent at any point during the five-month window that defines each refinancing wave. In both cases, probabilities are calculated across the income distribution and over time. The orange bars consider observations corresponding to periods of low interest rates in 2014, 2015, 2016, and 2019. The blue bars consider observations corresponding to 2020.

of unpaid balances in the form of higher savings from refinancing and as a result higher unpaid balances require lower interest rate differentials.

Appendix G Mechanisms and mediators of refinancing inequality

In Section 6 we discuss the role of unpaid balances in affecting the increases of refinancing inequality observed during 2020. In that section, for brevity, we show the main coefficients of interest. Here we provide a table with the full set of coefficients for brackets of unpaid balance and FICO, as well as their interaction with the dummy variable for wave 2020.

Table G1: Refinancing inequality before and during the pandemic with flexible interactions

Dep.Var. Refinancing Indicator {0,1}			
	(1)	(2)	(3)
Income quintile 2	-0.04	-0.02	0.12*
	(0.07)	(0.07)	(0.07)
Income quintile 3	-0.12	-0.09	0.35***
	(0.08)	(0.08)	(0.08)
Income quintile 4	-0.15	-0.11	0.71***
	(0.1)	(0.1)	(0.09)
Income quintile 5	0.11	0.16	1.42***
	(0.12)	(0.12)	(0.12)
Wave2020	1.19***	3.04***	6.51***
	(0.11)	(0.14)	(0.29)
Income quintile 2:Wave2020	2.14***	2.03***	1.45***
	(0.17)	(0.17)	(0.17)
Income quintile 3:Wave2020	3.79***	3.66***	2.25***
	(0.18)	(0.18)	(0.19)
Income quintile 4:Wave2020	5.48***	5.35***	3.07***
	(0.2)	(0.2)	(0.21)
Income quintile 5:Wave2020	6.23***	6.06***	2.91***
	(0.22)	(0.22)	(0.25)
FICO: [720,740)	-2.41***	-1.72***	-2.42***
	(0.1)	(0.1)	(0.1)
FICO: [680,720)	-3.82***	-2.64***	-3.83***
	(0.08)	(0.08)	(0.08)
FICO: [640,680)	-4.81***	-3.26***	-4.81***
	(0.09)	(0.09)	(0.09)
FICO: (0,640)	-4.64***	-3.05***	-4.66***
	(0.09)	(0.09)	(0.09)
upb: (275k, 300]	-3.11***	-3.14***	-3.05***
	(0.19)	(0.19)	(0.21)
upb: (250k, 275k]	-3.55***	-3.58***	-3.12***
	(0.19)	(0.19)	(0.2)
upb: (225k, 250k]	-4.31***	-4.36***	-3.46***
	(0.18)	(0.18)	(0.19)
upb: (200k, 225]	-5.13***	-5.19***	-3.92***
	(0.18)	(0.18)	(0.18)
upb: (0, 200k]	-6.35***	-6.43***	-4.14***
	(0.16)	(0.16)	(0.16)
Wave2020:FICO: [720,740)		-1.63***	
		(0.21)	
Wave2020:FICO: [680,720)		-2.87***	
		(0.16)	
Wave2020:FICO: [640,680)		-4.14***	
		(0.19)	
Wave2020:FICO: (0,640)		-4.81***	
		(0.22)	
Wave2020:upb: (275k, 300]			-0.33
			(0.37)
Wave2020:upb: (250k, 275k]			-1.18***
			(0.37)
Wave2020:upb: (225k, 250k]			-2.28***
			(0.35)
Wave2020:upb: (200k, 225]			-3.12***
			(0.33)
Wave2020:upb: (0, 200k]			-5.60***
			(0.27)
Mean of dep.var. in ref. cat.	1.15	1.15	1.15
Zip fixed effect	Yes	Yes	Yes
Borrower controls in regression	Yes	Yes	Yes
UPB and original interest rate	Yes	Yes	Yes
UPB x Wave 2020	No	Yes	No
FICO x Wave 2020	No	No	Yes
Observations	3,001,491	3,001,491	3,001,491
R2	0.14	0.15	0.15

Note: The notes to this table are presented in the next page.

Notes to Table G1: This table presents the results of running specification 1 with data from Freddie Mac, adding flexible interactions as controls. Specifically, we include the interaction of FICO buckets and UPB buckets with a dummy for Wave 2020. This is the basis of the mediation analysis discussed in Section 6 and includes the coefficients for the relevant control variables omitted from Table 7 for brevity.

Appendix H Longer horizons for the 2020 refinancing wave

Table H1 presents the coefficients used to produce Figure 4. We note that the sample size exhibits a big jump between columns 1 and 2, but only very small changes as we move to columns 3 to 5. The reason for that is that our sample is comprised of mortgages that became in the money during the period of analysis. For the duration of 2020 and 2021 interest rates were continuously trending down, reaching historically low levels. This sample size pattern shows that, given the distribution of interest rates in the portfolio of active mortgages, by the end of 2020 basically all mortgages that would become in the money in 2020 or 2021 reached that status by the end of 2020.

Furthermore, while the sample size did not change much in 2021, refinancing inequality continued to increase. This is consistent with the broad take always of our analysis that emphasize that being in-the-money is not enough to reach a refinancing outcome: the interaction of borrowers and lenders characteristics, along with the incentives they face, determines who refinances and who doesn't. Both at the onset of the pandemic as through the extended period of analysis, high-income borrowers were significantly more likely to refinance than low-income borrowers.

Table H1: Refinancing Inequality Before and During 2020: Different Long Term Horizons

	Dep.Var. Refinancing Indicator {0,1}			
	(1)	(2)	(3)	(4)
Income quintile 2	0.42***	0.60***	0.60***	0.60***
	(0.06)	(0.05)	(0.05)	(0.05)
Income quintile 3	1.24***	1.33***	1.33***	1.33***
	(0.06)	(0.06)	(0.06)	(0.06)
Income quintile 4	1.97***	2.08***	2.08***	2.08***
	(0.06)	(0.06)	(0.06)	(0.06)
Income quintile 5	2.72***	2.81***	2.81***	2.81***
	(0.07)	(0.07)	(0.07)	(0.07)
Wave 2020	0.52***	6.68***	18.06***	28.20***
	(0.08)	(0.07)	(0.10)	(0.12)
Income quintile 2:Wave 2020	1.28***	3.86***	7.44***	8.29***
	(0.10)	(0.10)	(0.13)	(0.14)
Income quintile 3:Wave 2020	2.63***	6.75***	12.67***	13.16***
	(0.11)	(0.11)	(0.13)	(0.15)
Income quintile 4:Wave 2020	4.26***	8.44***	16.21***	16.02***
	(0.11)	(0.11)	(0.14)	(0.16)
Income quintile 5:Wave 2020	4.66***	7.94***	17.64***	17.03***
	(0.12)	(0.12)	(0.16)	(0.17)
Mean of dep.var. in ref. cat.	2.46	2.46	2.46	2.46
End period of Wave2020	June2020	Dec2020	June2021	Dec2021
Zip fixed effect	No	No	No	No
Borrower controls	No	No	No	No
Savings control	No	No	No	No
Observations	1,775,920	2,781,497	2,969,624	2,971,756
R2	0.01	0.04	0.14	0.2

Note: This table presents the results of estimating equation 1 with data from McDash. The dependent variable takes the value of one when a mortgage was prepaid, and zero otherwise. Wave 2020 is defined as the period starting in February 2020 and ending on: June 2020 for column 1; December 2020 for column 2, June 2021 for column 3, December 2020 for column 4.