Federal Deposit Insurance Corporation • Center for Financial Research

WORKING PAPER SERIES

Refinancing Inequality During the COVID-19 Pandemic

Sumit Agarwal
National University of Singapore

Souphala Chomsisengphet
Office of the Comptroller of the Currency

Hua Kiefer
Federal Deposit Insurance Corporation

Leonard C. Kiefer
Freddie Mac

Paolina C. Medina
Texas A&M University

This Version: February 2023

Published as: "Refinancing Inequality During the COVID-19 Pandemic" *Journal of Financial and Quantitative Analysis,* August 2024, Volume 59, Issue 5, 2133-2163. Available online.

FDIC CFR WP 2021-08

fdic.gov/cfr

Refinancing Inequality During the COVID-19 Pandemic

Sumit Agarwal[*], Souphala Chomsisengphet[†], Hua Kiefer[‡], Leonard C. Kiefer[§]

and Paolina C. Medina[¶]

February 3, 2023

Abstract

During the first half of 2020, the difference in savings from mortgage refinancing between high- and low-income borrowers was ten times higher than before. This was the result of two factors: high-income borrowers increased their refinancing activity more than otherwise comparable low-income borrowers and, conditional on refinancing, they captured slightly larger improvements in interest rates. Refinancing inequality increases with the severity of the COVID-19 pandemic and is characterized by an under-representation of low-income borrowers in the pool of applications. We estimate a difference of $5 billion in savings between the top income quintile and the rest of the market.

Keywords: Mortgage Refinancing, COVID-19, Wealth Inequality, Monetary Policy.

[*]National University of Singapore, ushakri@yahoo.com

[†]Office of the Comptroller of the Currency, souphala.chomsisengphet@occ.treas.gov

[‡]Federal Deposit Insurance Corporation, hkiefer@fdic.gov

[§]Freddie Mac, leonard_kiefer@freddiemac.com

[¶]Texas A&M University, pmedina@tamu.edu

We thank Erica Jing, Kim Cramer, Lu Liu, Tingyu Zhou, Scott Frame, John Mondragon and seminar participants at AFA, EFA, FIRS, AREUEA National Meeting, ITAM Finance Conference, CEBRA, SFS Cavalcade, NUS RE Seminar, American Enterprise Institute's Housing Collaborative, FHFA's Fall 2020 Econ Summit, OCC, Freddie Mac, FDIC and Texas A&M University for valuable comments. The views and opinions expressed in this paper are those of the authors and do not necessarily reflect the views of Freddie Mac or its management, the Federal Deposit Insurance Corporation, the Office of the Comptroller of the Currency, the United States Department of the Treasury, the United States, or any federal agency and do not establish supervisory policy, requirements, or expectations.

1 Introduction

Mortgage refinancing is one of the main channels through which expansionary monetary policy affects individual consumption (Di Maggio et al., 2017, 2020; Agarwal et al., 2022, 2017; Eichenbaum et al., 2018). However, the magnitude of the consumption response depends on the characteristics of those who take advantage of refinancing opportunities, including their marginal propensity to consume (Auclert, 2019; Wong, 2019). In this paper, we document that during the COVID-19 pandemic, savings from refinancing were concentrated in the top segments of the income distribution, and this concentration was higher than in previous refinancing booms. This finding is important for evaluating the effectiveness of monetary policy in times of economic turmoil, since high-income individuals have lower marginal propensities to consume (Baker et al., 2020; Karger and Rajan, 2020).

We study refinancing decisions of individuals for whom market interest rates were sufficiently low to justify refinancing efforts. Our main analysis uses a rich dataset of mortgages originally funded by Freddie Mac that, when refinanced and funded by Freddie Mac, are matched to the new (i.e., refinancing) loan. For these loans, we observe the contract terms of both the old and new loan, and detailed origination records for both mortgages, including borrower' income and loan purpose. In contrast to previous work using market average interest rates from the Primary Mortgage Market Survey (PMMS) to approximate savings from refinancing, we observe the interest rates that specific borrowers receive when they refinance.

We find that refinancing reduce monthly payments (principal and interest) by an average of \$272, leading to \$8,800 in savings over the expected life of the loan. These savings are 33% smaller than the \$11,700 from the standard PMMS-based calculations. Further, they are highly concentrated in the top segments of the income distribution. While savings from refinancing naturally vary across the income distribution (as they depend on unpaid balances and interest rate differentials), we find that most of the differences in savings from refinancing before 2020 were explained by off-the-shelf control variables (the borrower's FICO score, unpaid balance, original interest rate, loan-to-value (LTV), loan age). However the same analysis for 2020 reveals that, controlling for observable characteristics, the differences in savings from refinancing between the top and bottom quintiles of the income distribution increased ten times.

The increased inequality in the distribution of savings from refinancing is the result of two factors: individuals in the top quintile of the income distribution increased their refinancing activity more than their counterparties in the bottom quintile, and conditional on refinancing, they captured slightly larger improvements in interest rate differentials. Before 2020, individuals in the top and bottom quintile of the income distribution had basically the same probability of refinancing projected at 1.15 percent, holding observable characteristics fixed at the level of the bottom income quintile. During 2020, the bottom quintile of the income distribution increased its refinancing activity by 1.19 percentage points, whereas the top quintile of the income distribution increased its refinancing activity by 7.42 percentage points. Higher-income individuals also captured the largest improvements in interest rate differentials. Before 2020, individuals in the bottom quintile of the income distribution received a 1.66 percentage point reduction in interest rates, conditional on refinancing. This reduction reached 1.82 percentage points in 2020 (a 0.16 percentage point improvement). In contrast, individuals in the top quintile of the income distribution who refinanced their mortgages received reductions of only 1.50 percentage points before the pandemic but an average 1.86 percentage point reduction in 2020 (a 0.36 percentage point improvement).

If individuals in lower segments of the income distribution received the same savings from refinancing as individuals in the top quintile of the income distribution they would capture an additional $5 billion in refinance savings over the expected life of the loan. Our results survive several robustness tests, including expanding the horizon of analysis to December

2021, the use of different data sets and market definitions (including FHA and VA loans from HMDA), different definitions of refinancing waves, and the analysis of delinquency and moving/relocation patterns across the income distribution.

We then link these increases in refinancing inequality to the COVID-19 pandemic. The first half of 2020 was a period of historically low interest rates and unique macroeconomic conditions. We ask if increases in refinancing inequality observed during this period came about solely as a result of lower than ever interest rates, or if instead increases in refinancing inequality were related to the impact of the pandemic on local communities. We show that idiosyncratic variation in the severity of the pandemic is correlated with refinancing inequality at the local level. We argue that disruptions to local communities triggered by the pandemic exacerbated refinancing inequality, through their impact on borrowers behavior.

We discuss potential mechanisms behind the increases in refinancing inequality observed in 2020. For exposition we split the discussion into borrower or lender related mechanisms. On the borrower side low-income borrowers may fail to refinance if they are not eligible to do so, if they are not aware of the benefits of refinancing or if they don't know how to do so. Furthermore, the psychological worries of dealing with new health risks, and new work and family environments could turn refinancing into less of a priority, specially among lower-income households who don't have resources to smooth negative shocks, and who were more severely affected by the pandemic. We expand our geographic analysis to include measures of time spent at home, unemployment insurance, mortgages under forbearance and fintech activity to explore the relevance of these explanations.

We find no evidence that general unemployment explains increases in refinancing inequality, but we find suggestive evidence that financial hardship specific to homeowners, as proxied by the fraction of mortgages under forbearance on a given state-month, does, at least partially. We also find suggestive evidence that refinancing inequality was mitigated by familiarity with fintech platforms and exacerbated by limited financial literacy, and we argue that the psychological worries brought about by the pandemic can be considered as a residual channel.

On the lender side, we note that given the unprecedented surge in refinancing applications observed in 2020, lenders with limited resources may have prioritized the most profitable applications either at the approval/funding stage (conditional on applying), or at the pre-application stage by targeting marketing efforts towards high income borrowers. The former is more likely to occur when funding capacity is limited, the later is more likely when operational capacity is limited.

On average, lenders received more than twice as many applications during the pandemic than before the pandemic, but the fraction of applications eventually funded remained relatively constant (decreasing less than 1 percentage point from a base of 25 percent). Lenders with the largest growth rates in applications, saw larger decreases in the number of applications eventually funded by Freddie Mac. However, this effect is present across the income distribution and does not explain the increases in refinancing inequality observed.

We also find that low-income borrowers are underrepresented in the pool of applications received during the 2020 refinancing wave. Almost 16 percent of refinancing applications come from borrowers in the top decile of the income distribution of portfolio mortgages, and only 4.5 percent come from the bottom decile. This pattern is sharper than in previous years. On the lender side, operational constraints could lead lenders to re-direct their marketing efforts in 2020 more so than in previous years, towards borrowers with more profitable characteristics. Mediation analysis reveals that the effect of unpaid balances on refinancing activity is significantly larger in 2020 relative to previous years. The special role of unpaid balances in 2020 explains 53% of increases in refinancing inequality.

The evidence suggests that operational constraints were more relevant than funding constraints. A stronger than usual targeting of high-income borrowers, motivated by operational constraints, is one potential reason behind the under-representation of low-income borrowers observed in the pool of applications.

Refinancing inequality does not disappear in the second half of 2020 or in 2021, highlighting that the lessons of our analysis are not exclusive to pandemic periods. Instead, our results suggest that operational constraints in mortgage origination have distributional consequences, specially when financial or psychological barriers limit low-income borrowers ability to actively seek refinancing opportunities. The COVID-19 pandemic was a time in which these factors coalesced, but we would expect similar outcomes in future periods when some or all of these factors reoccur.

This paper contributes to a large literature studying mortgage refinancing and its consequences for the economy. Previous work has documented the importance of the mortgage refinancing and the mortgage-debt-service channel to stimulate spending (Di Maggio et al., 2020; Eichenbaum et al., 2018; Agarwal et al., 2022; Agarwal, 2007; Agarwal et al., 2022, 2017). Beraja et al. (2017); Wong (2019) and Laibson et al. (2020) show how the distribution of savings from refinancing across areas with different local economic conditions or across borrowers with different characteristics matters for the transmission of monetary policy. More generally, Auclert (2019) highlights the role of redistribution for the transmission of monetary policy to consumption. To our knowledge, our paper is the first to characterize the distribution in savings from refinancing across income groups. This measure arguably captures variations in marginal propensities to consume (Karger and Rajan, 2020).[1]

[1] To our knowledge, the only other paper using matched refinancing transactions with information on old and new interest rates for every transaction is Berger et al. (2019). Their focus differs from ours, as they study path dependent effects of monetary policy.

Our results build on the work of Agarwal et al. (2013); Keys et al. (2016); Johnson et al. (2015); Agarwal et al. (2017); Andersen et al. (2020) and DeFusco and Mondragon (2020), who discuss how limited financial literacy, behavioral biases, strict documentation requirements, or other frictions in the mortgage market explain low refinancing activity despite sufficiently low interest rates. We expand on this work by focusing on differences in refinancing activity across the income distribution, in both the extensive and intensive margins (i.e., propensities and dollar savings). Our results are consistent with Nothaft and Chang (2005); Gerardi et al. (2020, 2021); and Goodstein (2014) who find that propensities to refinance vary with income and race.

Finally, our paper contributes to a fast-growing literature studying the economic impact of the COVID-19 pandemic. Recent work documents strong decreases in consumption (Baker et al., 2020) and disruptions to credit and labor markets (Coibion et al., 2020). In all cases, the impact has disproportionally affected low-income individuals (Chetty et al., 2020; Kinder and Ross, 2020) thus increasing income inequality. In the mortgage market, Fuster et al. (2021) document that frictions in the labor market and operational bottle-necks led to binding capacity constraints which is consistent with our results. An et al. (2021) and Cherry et al. (2021) study the role of forbearance as policy response to the crisis which is another important and complementary dimension of analysis.

2 Data Description

Our analysis is based on several data sources. First, we use 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 during the period of analysis. We follow those loans through time and observe whether the loan was prepaid during the refinancing wave. In addition, for a subset of loans that were prepaid, we match a new loan also funded by Freddie Mac that was originated at the same property address within a 45-day window of the closure of the prepaid loan. For those matched transactions, we collect loan-level attributes of the newly originated loan at the same address. Where the loan was refinanced, we observe the new loan product and loan attributes, including the new interest rate. We also identify cases where the prepayment was not for a refinancing transaction but rather a home purchase, and we observe actual income instead of only debt-to-income ratios.[2]

The second data set consists of loan-level information provided by residential mortgage servicers and collected by Black Knight (commonly known as McDash data). This dataset provides extensive information on loan, property, and borrower characteristics at the time of origination as well as dynamically updated loan information after origination. The dataset also provides more comprehensive coverage of the mortgage market compared to Freddie Mac's data, since it includes conventional loans not sold to Freddie Mac. We restrict our sample to owner-occupied, single-family, first-lien loans. We focus on 30-year, fixed-rate conventional mortgages (that is, government-insured loans from the Federal Housing Administration, Veterans Administration and other entities are excluded). Exotic loans (such as loans with a balloon payment, negative amortization, or prepayment penalty) are excluded from our sample, as are loans that are in foreclosure, bankruptcy or real estate owned status, and those less than two months old. For each loan in the portfolio, we observe whether the loan was prepaid or not.

[2]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. 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 almost identical.

We complement the mortgage data with a rich set of variables tracking the impact of the pandemic on local economic conditions across different geographic areas. Specifically, we look at mobility restrictions, initial unemployment insurance claims, percentage of mortgages in forbearance, and COVID-19 case rates. Except for the percentage of mortgages in forbearance, we download the data from the public repository created by Chetty et al. (2020) to track the impact of the pandemic across the United States. For COVID-19 Case Rates, the data corresponds to seven-day moving average count of cases per capita at the county level, which we further average at the monthly level to match our mortgage data. For mobility, we use GPS-based mobility data released by Google. The data is expressed as an index based on the five-week period from January 3 through February 6, 2020. Mobility data are reported at the county-by-day level which we aggregate at the county-by-month level to match the frequency of the mortgage data. For unemployment, we use data on initial unemployment insurance claims at the county level, reported to the US Department of Labor. Weekly initial claims are averaged at the monthly level and expressed per 100 people in the 2019 labor force. Location is defined as the state liable for the benefits payment. To measure forbearance, we use data from the Monthly Industry Reports (TransUnion, 2020) that show the percentage of mortgages in hardship at the state level monthly between March 2020 and May 2020. [3]

Finally, we use a proprietary dataset with refinancing applications submitted to Freddie Mac's Loan Product Advisor (LPA) tool. This tool collects information on applications submitted to Freddie Mac's underwriting system, which includes both loans funded by Freddie Mac as well as loans that ultimately were not funded by Freddie Mac. These data broadly track trends seen in the Mortgage Bankers Association's Weekly Application Survey. About 18 percent of all new loans in the market are run through Freddie Mac's LPA tool. We do

[3] Hardship is defined as "affected by natural/declared disaster, accounts in forbearance, accounts deferred or for which the payment-due amount has been removed, or accounts whose account status and/or past-due amount has been frozen".

not observe when an application is approved by a lender and when it is not. Instead, we observe when an application ends as a mortgage ultimately purchased by Freddie Mac and when it does not.

3 Period of Analysis and Sample Selection

We focus our analysis on five periods characterized by interest rate reductions and high refinancing activity: 2015, 2016, 2017, 2019 and 2020. Figure 1 shows the evolution of 30-year fixed mortgage rates between 2014 and 2020 (left axis). The highlighted periods correspond to five-month windows with the largest declines in interest rates. The decline in interest rates between the high and low points of this period were 0.64 percentage points between October 2014 and February 2015, 0.25 percentage points between May and September 2016, 0.32 percentage points between May and September 2017, 0.71 between May 2019 and September 2019, and 0.77 percentage points between February and June 2020. The right axis of Figure 1 shows refinancing activity. Most of the periods with the largest declines in interest rates were also characterized by the largest spikes in refinancing activity. As a result, we refer to these periods as refinancing booms. We purposely start the analysis after the financial crisis of 2008 to focus on periods in which the mortgage market itself was not the driving force behind weakening economic conditions. Furthermore, we focus on the early stages of the pandemic since this was its most disruptive phase and because swift policy interventions can provide relief for immediate losses and mitigate the spiraling of the crisis.

In each period, we identify mortgages for which the refinancing option was not in-the-money before the window of observation and becomes in-the-money during the window of observation. We use the model of optimal refinancing proposed by Agarwal et al. (2013). Under some assumptions, this model identifies a threshold for which it is optimal to trade-in an old in-the-money refinancing option for a new out-of-the-money refinancing option that is acquired, taking into account closing costs, mortgage size, taxes, and the standard deviation of the mortgage interest rate. It also includes a measure capturing the combined effects of moving events, principal repayment and inflation-driven depreciation of the mortgage obligation. Following Keys et al. (2016), we use the parameter values calibrated in Agarwal et al. (2013) which include discount rate of 5% per year, a 28% marginal tax rate, and a probability of moving each year of 10% (See Appendix F for details). We consider these parameters to be conservative in that they suggest that individuals should refinance only when it is surely beneficial for them. Under these parameter choices, the optimal refinancing differentials range typically from 100 to 200 basis points (bps). When market interest rates relative to the interest rate on the borrower's current mortgage exceed the differential, we say that the borrower is in-the-money for a refinance. For robustness, in Appendix F we evaluate how refinancing incentives change when the probability of delinquency, the probability of moving events and closing costs differ by refinancing wave and income group.

Table 1 describes the set of newly-in-the-money mortgages before and after the pandemic in both the Freddie Mac and McDash data. Consistent with Agarwal et al. (2013) and Keys et al. (2016), potential savings from refinancing are defined as the present value of the savings from refinancing at the market rate, adjusting for the probability of moving, tax incentives, upfront costs, and discounting over time.

4 Refinancing Inequality Over Time

We describe the evolution of savings from refinancing across the income distribution using our matched-transactions data set. We study differences in the probability of refinancing across the income distribution and differences in actual savings from refinancing conditional on refinancing. We estimate the following equation with different outcome variables:

$$
\begin{aligned}
y_{it} = \alpha + \sum_{j=2}^{5} \beta_j * Income\ quintile_{ji} + \gamma * Wave\ 2020_{it} \\
+ \sum_{j=2}^{5} \phi_j * Income\ quintile_{ji} * Wave\ 2020_{it} + \delta * X_{it} + \epsilon_{it}
\end{aligned}
\tag{1}
$$

where y_{it} measures the outcome of interest for mortgage i in period t, $Wave$ 2020 is a dummy variable indicating whether the observation corresponds to the 2020 window of analysis. X_{it} is a vector of loan-level controls that will be added gradually across models. The reference category is the bottom quintile of the income distribution during periods before 2020.To capture refinancing activity for the entire portfolio of Freddie Mac loans, we weight matched prepayments by the probability of being matched, conditional on observable characteristics. Appendix A describes the matching process.

Depending on the outcome variable, this specification allows us to characterize refinancing activity or savings from refinancing across the income distribution, before and during the pandemic. This specification also provides direct estimates for differences across income quintiles and over time. For example, each coefficient β_j represents the difference in refinancing activity or savings from refinancing between the jth and the bottom quintiles of the income distribution, in periods before 2020. We refer to the difference between top and bottom quintiles of the income distribution as the *refinancing income gap*. We use the refinancing income gap as summary measure of inequality in refinancing activity and savings from refinancing, depending on the outcome variable. The coefficient β_5 is our estimate of the refinancing income gap before the pandemic and ϕ_5 represents the change in the refinancing income gap before and during the pandemic. $\beta_5 + \phi_5$ is our estimate for the refinancing income gap during the 2020 refinancing wave. This specification allows us to recover changes in refinancing activity and savings from refinancing within each quintile before and during the pandemic. For example, the coefficient γ represents the difference in refinancing activity or refinance savings for mortgages in the bottom quintile of the income distribution, and $\gamma + \phi_j$ represents our estimate of the change in refinancing activity or refinance savings in the jth quintile of the income distribution, holding everything else constant. We study how these parameters change with and without controlling for a set of off- the-shelf observable characteristics, namely zip code fixed effects, loan age, FICO score, LTV, original interest rates, and unpaid balance. Continuous variables are binned as follows: for credit score, 740+, [720,740), [680,720), [640,680),640-; for LTV: 95+, (90,95], (85,90], (80,85], (75,80],(70,75], (60,70], (0,60]; for age: 1 year or less - , 1-2 years, 2-3 years, 3-5 years, 5-7 years, 7+ years; for unpaid balance, (0,200k], (200k,225k], (225k,250k], (250k,275k], (275k,300k], (300k+.

The first outcome variable we use, is a dummy variable that takes the value of one when a mortgage is refinanced, and zero otherwise. The results are presented in Table 2. Column 1 does not use any control variable. The refinancing income gap increases from 6.84 percentage points before 2020 to 13.56 percentage points in 2020. Column 2 adds a first set of control variables: zip code fixed effects, loan age, FICO score and LTV which explain 66 percent of the refinancing income gap before 2020 (the estimate for the refinancing income gap before 2020 decreases to 2.3 percentage points, a 4.54 percentage point reduction from a base of 6.84). In contrast the same set of controls explain only 28 percent of the refinancing income gap in 2020 (the estimate for the refinancing income gap during 2020 decreases to 9.7 percentage points, a 3.86 percentage point reduction from a base of 13.56). Column 3 adds unpaid balances and original interest rates as controls and explains 98 percent of the refinancing income gap before 2020. In contrast, the same set of controls can explain only 53 percent of the refinancing income gap during 2020 (our estimate with the full set of controls is 6.34 percentage points, representing a 7.22 percentage reduction from a base of 13.56 in column 1). Including our full set of control variables, the difference in refinancing activity between the top and bottom quintiles of the income distribution during 2020 was significantly higher than before 2020 (6.34 vs 0.11 percentage points). This results are summarized graphically in Figure A1.

We then restrict the analysis to mortgages that were refinanced and are part of our matched transactions data. For these mortgages, we study the distribution of savings from refinancing conditional on refinancing. We calculate the dollar value of savings from refinancing as present value of the difference in interest costs under the old and new interest rates over the expected life of the loan. The expected life of the loan is parametrized by Agarwal et al. (2013). Savings from refinancing thus depend on loan age, unpaid balances, original interest rates and the interest rate of the new refinancing loan.

One important difference between high and low income borrowers that mechanically affects savings from refinancing is that the former group tends to carry larger unpaid balances since their property values are typically higher. For the same interest rate differential higher unpaid balances will carry larger savings from refinancing. To make sure that our analysis is not driven only by these mechanical effects, we continue to control non-parametrically for unpaid balances. In addition, we also look at interest rate differentials between the interest rate on the original (refinanced) loan and the new (refinancing) loan. This interest rate differential is a summary measure of savings from refinancing that is completely unaffected by unpaid balances. Actual savings are a non linear function of interest rate differentials. In that sense, both outcome variables are complementary because the former is reflective of lender-borrower behavior only whereas the later reflects how initial conditions (i.e. unpaid balances) interact with behavior resulting in different levels of savings.

We estimate Equation 1 with two outcome variables: interest rate differentials and the dollar value of savings from refinancing.[4] The results are presented in Table 3. In columns 1 to 3, we use interest rate differentials as the dependent variable. In columns 4 to 6 we use dollar savings as the dependent variable. Columns 1 and 4 present the results without controls. Columns 2 and 5 present the results with borrower controls. Columns 3 and 6 present the results with the full set of control variables, which importantly include loan age, original interest rates and unpaid balances. Consistent with our analysis on refinancing propensities we can see that before the pandemic, the vast majority of refinancing inequality is captured by off-the shelf observable characteristics.[5]

In column 3 we see that before 2020, conditional on refinancing, borrowers in the bottom quintile of the income distribution received a reduction of 166 basis points from their original interest rates (reference category). Relative to those borrowers, comparable borrowers in the top quintile of the income distribution received slightly lower interest rate reductions of 150 basis points (166 minus 16). This regression controls for original interest rates, unpaid balances, zip code fixed effects, and standard borrower-level controls. During 2020, all borrowers received large interest rate reductions (the coefficient for wave 2020 and its interaction with income quintiles are all positive and significant), but the improvement in contract terms for borrowers in the top quintile of the income distribution was larger than

[4] Note that we do not attempt to provide a causal interpretation of income in this analysis. Income clearly affects both average savings conditional on refinancing and the probability of refinancing. Thus, selection into refinancing is not random. Our goal is to describe average savings for individuals across the income distribution who go through a refinancing transaction. We do so comparing average savings conditional on refinancing over a discrete set of (income) categories (Angrist, 2001). We retain a linear model for this part of the analysis (instead of a two-step model or a conditional-on-positive Tobit estimate) to emphasize its descriptive nature: linear models are best suited for comparing means across different groups. Nevertheless, for robustness, we also consider a Tobit model. We find that our results are very similar in all cases.

[5] Columns 3 and 6 show an apparently contradicting story. After controlling for unpaid balances, loan age and original interest rate, the refinancing income gap in terms of interest rate differential is negative, but it is positive in terms of actual savings. This is the result of the non-linear mapping between interest rate differentials, and the fact that bins for unpaid balance and loan age are relatively wide. For example, the first bin includes all mortgages with unpaid balances under $200 thousand. A concentration of low income borrowers on the low end of the interval and high income borrowers in the high end of the interval would explain the sign reversion between these two columns.

for borrowers in the bottom quintile of the income distribution. Individuals in the bottom quintile improved their interest rate differentials by 16 basis points to reach a rate differential of 182 bps. Individuals in the top quintile improved their interest rate differentials by 36 bps to reach a rate differential of 186 bps. The slight edge of lower-income individuals refinancing before 2020 in terms of interest rate reductions disappeared in 2020. Similarly, in column 6, we see that before 2020, borrowers in the top quintile of the income distribution had $1,117.86 more in savings than comparable borrowers in the bottom quintile of the income distribution. This difference in savings increases to $3,532.16 in 2020. To benchmark these magnitudes, we note that average home values in the top and bottom quintiles of the income distribution correspond to $171,593 and $471,793. The 2020 gap in savings from refinancing represents 2.1% and 0.7% of their home values respectively. This further supports the argument that savings from refinancing are not reaching those with the largest marginal propensity to consume and would be more valuable for those in the lower income quintile.

We now describe average savings from refinancing on the entire portfolio of active mortgages, incorporating both the probability of refinancing and actual savings conditional on refinancing. We define savings from refinancing on the entire portfolio as a continuous variable that takes the value of zero for all mortgages that were not refinanced, or the corresponding value of savings from refinancing for mortgages that were refinanced. We estimate Equation 1 as before. We also estimate a Tobit model using the same equation as a latent linear index.[6] Our results are robust to these different functional form assumptions.

Table 4 shows the results of estimating Equation 1, as we gradually add control variables. Columns 1 and 4 show the coefficients of interest without loan-level controls. The difference

[6]The latter approach imposes functional form assumptions to explicitly model savings as a variable censored at zero: a latent linear index feeds into normal distribution censored at zero which is then estimated by maximum likelihood. In contrast, the former takes a more agnostic approach describing changes in average savings over a set of discrete categories, namely income quintiles before and after 2020. (Angrist, 2001).

in savings from refinancing between the top and bottom quintile of the income distribution before 2020 amounts to \$879 (or \$1,690 when measured with a Tobit model).[7] During 2020, the difference in refinancing activity between the bottom and top quintiles of the income distribution increases to \$2,288 (or \$5,009 when measured with a Tobit model). However, this change could be driven by changes in the composition of loans that became newly in-the-money during the observation periods.

To address this possibility, we gradually add a rich set of control variables to assess the sensitivity of our estimates. In columns 2 and 5 of Table 4 we add flexible controls for borrower and loan attributes, namely dummy variables for FICO score bins, LTV bins, and bins of loan age. We find that this basic set of controls explains a significant fraction of baseline inequality which now accounts for \$453 (\$1,087) when estimated with the ordinary least squates (OLS) model (Tobit model).

Finally, in columns 3 and 6 we include two additional controls that largely capture the potential savings from refinancing activity: baseline interest rate, and unpaid balance. Thus, columns 3 and 6 estimate the role of income on refinancing activity for individuals with the same FICO score, loan age, LTV, unpaid balance, and interest rate. In column 3, with OLS estimates, we find that the gap in refinancing activity between the top and bottom quintiles of the income distribution at baseline is fully explained, and even changes sign to reach a level of -\$131. In column 6, with a Tobit model, we find consistent results. Our full set of controls leads to a final difference of \$386. But, even among comparable mortgages, the difference in savings from refinancing across the income distribution increased significantly. In column 3, we see that the difference in savings accounts to \$1,313 with our OLS estimates,

[7] For the Tobit models, the savings gap expressed in dollar terms is calculated as $\Phi(\frac{x_1 * \beta}{\sigma}) + \sigma * \phi(\frac{x_1 * \beta}{\sigma}) - \Phi(\frac{x_0 * \beta}{\sigma}) - \sigma * \phi(\frac{x1 * \beta}{\sigma})$, where Φ/ϕ is the standard normal CDF/PDF, σ is the Tobit scale parameter, x_0 (x_1) refers to control variables evaluated at the baseline (reference level) and β is the full parameter vector. This calculation is reported in the bottom panel of Table 4.

an eleven-fold increase from pre-2020 levels ((1,313 + 131)/131). Similarly, in column 6, our Tobit estimates show that the difference in savings increased 9.8 times (3,798/386).

Panel (a) of Figure 2 plots our projections for savings from refinancing before and after the pandemic by income quintile, controlling for changes in the composition in the pool of newly in-the-money borrowers. We plot the coefficients β_j and $\beta_j + Wave_{2020} + \phi_j$ from column 3 of Table 4 for each quintile after summing in both cases the prepayment rate in the reference category (bottom quintile of the income distribution before 2020). Panel (b) shows the analogous results without controls.

Back-of-the-envelope calculations using our estimates for refinancing savings across the income distribution imply a gap in refinance savings over the expected life of the loan of \$5 billion between the top quintile of the income distribution and the rest of the market. That is, if individuals in lower segments of the income distribution (without controlling for observable characteristics) received the same savings from refinancing as individuals in the top quintile of the income distribution, they would capture an additional \$5 billion in refinance savings over the expected life of the loan.[8]

For robustness purposes we perform a similar analysis with data from McDash Analytics which we present in Appendix B. We confirm similar patterns of refinancing inequality before and during 2020, which we complement with heterogeneity analysis along FICO scores and LTV dimensions. We also confirm that our result is not driven by pre-existing trends in inequality over the 15-month period before the pandemic (See Table B3 and Figure B1).

[8]To calculate this number, we start from a market size of 30.9 million mortgages with fixed rates at 30 years maturity (American Housing Survey, with data as of 2017). We extrapolate our estimates for average savings for mortgages that become newly in-the-money in each income quintile j ($In\ Money\ IQ_j$) and difference in refinance savings for each quintile j relative to the top quintile of the income distribution ($Gap\ Q_{j5}$). To do so, we use the results for 2020 in column 1 of Table 4, also depicted in panel (b) of Figure 2. Specifically, we apply the following formula $\sum_{j=1}^{4} 30.4 * 0.2 * In\ Money\ IQ_j * Gap\ Q_{j5} = 4{,}964{,}017{,}020$. Savings from refinancing refer to the present value of savings over the expected life of the loan, accounting for the probability of prepayment as in Agarwal et al. (2013).

5 Intensity of the COVID-19 Pandemic and Refinancing Activity

The first half of 2020 was a period of historically low interest rates and unique macroeconomic conditions. For the second component of our analysis we ask if increases in refinancing inequality observed during this period came about solely as a result of lower than ever interest rates, or if instead increases in refinancing inequality were tied to the impact of the pandemic on local communities.

Our dataset consists of a monthly panel that follows mortgage refinancing activity between February 2020 and July 2020. Specifically, we consider mortgages that were not in-the-money in February 2020 and became in-the-money in subsequent months until July 2020. For these mortgages we consider monthly observations between the first month in which they turn in-the-money until the month in which they are prepaid, along with a vector of variables tracking the impact of COVID-19 at the county or state level. We use data from McDash Analytics due to its broader market coverage and we estimate the following equation:

$$\begin{aligned} y_{izct} = \alpha_z + \alpha_t + \sum_{j=2}^{5} \beta_j * Income\ quintile_{ji} + \gamma * High\ COVID_{izct-1} \\ + \sum_{j=2}^{5} \phi_j * Income\ quintile_{ji} * High\ COVID_{izct-1} + \delta * X_{it} + \epsilon_{izcgt} \end{aligned} \tag{2}$$

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; $High\ COVID_{izct-1}$ is a dummy variable indicating whether mortgage i in zip code z in county or state c belongs to one of the top four quintiles 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.[9] We use case rates as an omnibus measure for disruptions to local communities brought about by the pandemic. Higher case rates can lead to changes in local economies for many reasons, including more time at home, more financial distress, increases in unemployment, and many others. We bundle all those possible disruptions into one single measure to establish that refinancing inequality was affected by the impact of the pandemic on local economies.

The results are presented in Column 1 of Table 5. The coefficients for income quintiles 1 to 5 capture differences in refinancing activity when COVID-19 severity is low. The differences across the income distribution are not economically large, ranging from 0.13 to -0.13 percentage points. The refinancing income gap (β_5) is basically non-existent when COVID-19 severity is low (0.07 percentage points). In the bottom quintile of the income distribution, high COVID-19 severity leads to less refinancing activity (-0.59 percentage points). The refinancing income gap when COVID-19 severity is high reaches a level of 1.33 percentage points ($\beta_5 + \phi_5$) which is significantly larger than when COVID-19 severity is low. Since we have zip code and month fixed effects, the coefficients are identified by idiosyncratic variation in a particular location over time, that is, variation specific to a particular geography after controlling for aggregate time trends. Importantly, those aggregate time trends include the general worsening of the pandemic. Intuitively, our coefficients are identified out of idiosyncratic variation in case rates at the onset of the pandemic in New York, Houston, Florida etc which became hot-spots at different moments in time.

[9] In Appendix D, we present a more flexible specification using quintiles of the severity of the pandemic which justifies the choice of comparing the bottom quintile of the distribution of COVID-19 case rates to the remaining top four quintiles.

We conclude that refinancing inequality did not arise solely as a result of historically low interest rates and unique macroeconomic conditions affecting the economy as a whole. Instead, it was tied to the severity of the pandemic at the local level. Increases in case rates in a particular location over time, were correlated with increases in refinancing inequality, even after controlling for the general worsening of economic conditions through aggregate time trends.

We do not interpret this correlation as reflective of health conditions in the communities affected, but instead as reflective of the impact of the pandemic on local economic conditions. Specifically, we note that the location of the loan is more indicative of borrower behavior than of lender behavior. This occurs because lenders need not be located in the same place as the house they are financing.[10] In Columns 2 to 5 of Table 5 we expand the analysis to see how much of the variation in refinancing inequality captured by COVID-19 case rates is coming from changes in time spent at home, unemployment insurance claims and the fraction of mortgages in forbearance status. These results are interpreted in Section 6.

6 Mechanisms

We have shown that increases in refinancing activity were concentrated among individuals with higher income and that this result is attributable to disruptions to local communities brought about by the pandemic. For expositional purposes, we organize the discussion splitting potential mechanisms into those related to borrower behavior or characteristics and those related to lender behavior. The borrower side analysis is based on the impact of the

[10] Amel et al. (2018) shows that indeed more than half of mortgage lending is originated by lenders who do not have physical presence in the communities they serve.

pandemic on local economic conditions. The lender side analysis is based on application data.[11]

6.1 Borrowers

Borrowers may fail to refinance if they are not eligible to do so, if they are not aware of the benefits of refinancing their mortgage or if they don't know how to do so. Low income borrowers may not be eligible to refinance if they enter forbearance status or if they lose their jobs. They could also exhibit lower levels of financial literacy or sophistication leading them to miss on profitable opportunities to refinance.[12] In addition, the psychological worries of dealing with new health risks, and new work and family environments could make refinancing less of a priority, specially among lower-income households who don't have resources to smooth negative shocks, or who were more severely affected by the pandemic.

To study these possibilities, we investigate what fraction of the omnibus effect of COVID-19 case rates, documented in Section 5, can be explained by specific shocks to borrowers eligibility or to borrowers ability to internalize the benefits of refinancing their mortgage. To do so, we expand Equation 2 to include controls for unemployment insurance claims, fraction or mortages on forberearance and time spent at home, as well as their interaction with income quintiles. We then compare how the original coefficient for $High\ COVID{*}Income\ Quintile_5$ changes across specifications.

[11] We discuss the relevance of each potential mechanisms providing evidence consistent (or inconsistent) with it. However, we note that refinancing applications and refinancing activity are equilibrium outcomes that result of both borrower and lender behavior and we do not attempt to formally identify them. While the facts presented in each case may be consistent with a specific borrower- (lender-) side explanation, they could also be consistent with other lender-(borrower-) side explanations.

[12] While levels of financial education are unlikely to change during the pandemic, the local or macroeconomic environment can render differences in financial education more relevant than before for distributional outcomes. For example, if interest rates are high differences in individuals understandings of savings from refinancing are irrelevant (since no action is required), however when interest rates are low, the savings outcome for those with and without financial literacy who make different decisions, becomes relevant.

The results are presented in Columns 2 to 5 of Table 5. Unemployment insurance claims do not have explanatory power. 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. Together, these three variables explain around 44 percent ((1.26 - 0.71)/1.26) of the impact of local economic conditions on refinancing inequality. These results are also summarized in Figure D3.

Eligibility Borrowers may become ineligible for refinancing their loans if they cannot document their sources of income, which may happen if they are unemployed or furloughed, or if they experience income reductions or economic shocks. While the pandemic lead to large waves of unemployment specially among low income individuals (Adams-Prassl et al., 2020), we find no evidence that increases in refinancing inequality brought about by the pandemic are driven by increases in unemployment insurance claims (see Column 2 of Table 5). One potential explanation is that the pool of unemployment insurance claims is dominated by individuals with very low income, who do not own a home. As a benchmark, Chetty et al. (2020) emphasize the severity of unemployment for individuals with yearly salaries under 27 thousand dollars per year. The average yearly income of homeowners in the United states is more than twice as much.

While unemployment insurance claims may not capture economic shocks specific to homeowners, the fraction of mortgages under forbearance could. Column 3 of Table 5, shows that 30% of the increase in refinancing inequality brought about by the local impact of the pandemic is explained by the percentage of accounts under forbearance in the state-month. We think of this variable as a proxy of financial hardship for home owners: relative to unemployment insurance claims this measure is more likely to reflect hardship experienced specifically by home owners.

The lack of granularity in our forbearance data (state-by-month) limits our ability to accurately study substitution patterns between forbearance and mortgage refinancing.[13] Nevertheless, we note that quantitatively, this substitution is unlikely to be the main driver behind increases in refinancing inequality since the fraction of mortgages in forbearance was significantly lower than the fraction of mortgages who failed to refinance. Cherry et al. (2021) show that among borrowers in the bottom quartile of the income distribution, forbearance rates were the highest, reaching almost 7 percent of borrowers in that income group.[14] In contrast, more than 95 percent of newly in-the-money borrowers in the bottom quintile of the income distribution failed to refinance.[15]

We conclude that limits to eligibility are one potential factor behind increases in refinancing inequality, but they are unlikely to be the only borrower-side mechanism behind our main results.

Familiarity with online financial services and financial literacy During the pandemic mobility decreased significantly and individuals spent more time at home than before. If low income borrowers are less familiar with online financial services, then restrictions in mobility could lead to increases in refinancing inequality. Column 4 of Table 5 shows that restrictions in mobility explain 10% of the increases in refinancing inequality brought about by the pandemic.

[13] As per the CARES Act, borrowers are eligible to refinance their loans upon exiting forbearance, after making three consecutive payments on time but are not eligible to refinance their mortgages while they are in forbearance

[14] See panel a) of Figure A6 in Cherry et al. (2021).

[15] We note that the size of these two groups is not the same. However, even with the extreme assumption that the full 7% of mortgages identified by Cherry et al. (2021) became in-the-money during the first half of 2020, then a conservative upper bound for the fraction of mortgages in forbearance in our bottom income quintile would be $7\%*5 = 35\% << 95\%$. Furthermore, in general forbearance decreased over time, whereas the fraction of mortgages that became in the money increased over time (due to decreasing interest rates).

Furthermore, if the relation between mobility restrictions and refinancing inequality is related to familiarity with online financial services, then the explanatory power of time spent at home should be larger in areas where fintech lenders are more popular. To explore this possibility, we build a measure of fintech activity at the county level using HMDA data for the year 2019. First, we classify each application as corresponding to a fintech lender or not, based on the classification of Fuster et al. (2019). We then aggregate the counts of fintech and non-fintech applications at the county level and calculate the fraction of fintech applications in each county. We split the sample into quintiles of fintech activity and merge this information with our data from McDash. We then repeat the analysis in Columns 1 and 4 of Table 5 separately for mortgages with high and low fintech activity (i.e. top vs bottom quintiles of the fintech activity distribution). The results are presented in Figure 3, which is based on Table D2.

Each bar represents the coefficient for the omnibus effect of the pandemic (Income Quintile 5: Top COVID Case) with and without controls for time spent at home (orange and blue bars respectively). We make two observations based on Figure 3. First, adding controls for time time spent at home can explain 11% (1 - 1.64/1.84) of the omnibus effect of the pandemic on refinancing inequality in areas of low fintech activity, whereas in areas of high fintech activity, time spent at home explains 31% (1 - 0.85/1.23). This is consistent with the idea that familiarity with online financial services mitigated refinancing inequality. Second, the impact of the pandemic on refinancing inequality is lower in areas of high fintech activity, regardless of including or not our control variables. From the supply side, this is consistent with lesser operational bottle necks for fintech lenders. If fintech lenders are able to process refinancing applications faster, then they may be less likely to target their outreach efforts to high income borrowers only, and instead solicit more applications, even if from low-income borrowers.

In addition, limited financial literacy and limited awareness of policy rules could discourage low-income borrowers from applying to refinance their mortgage. Cherry et al. (2021) find that about 20 percent of borrowers in forbearance continue making their payments effectively using their forbearance status as ab open line of credit. In the context of our discussion, it is possible that some of these borrowers could be better-off refinancing their mortgage instead. For them knowing the detailed rules of policies like the CARES Act, the trade-offs between forbearance and refinancing, and the magnitude of their potential savings from refinancing, is more relevant than for borrowers far from the forbearance margin.

Mental bandwidth and cognitive taxes In addition to a number of material disruptions, as a residual channel, borrowers may have experienced psychological worries not captured by our data (Wallace and Patrick, 2020), which could in turn tax their mental bandwidth and make mortgage refinancing less of a priority.[16] These psychological worries were likely more severe among lower-income households who, as documented in a large medical literature, were more severely affected by the pandemic (Khatana and Groeneveld, 2020; Purtle, 2020; Khatana and Groeneveld, 2020). In addition, low-income borrowers were likely experiencing income reductions (due to furlough or reduced hours) while having lower emergency savings. Households without the means to afford child-care were probably more disrupted when dealing with new remote working environments while simultaneously adapting to online school for kids, trying to keep up with how to protect oneself and one's family, and trying to keep up with the world at large about new pandemic-related developments. This regressive cognitive tax could have depressed borrowers demand for mortgage

[16] In the psychology and economics literature, the term mental bandwidth is used as an umbrella term to capture to capture the brain's ability to perform basic functions Mullainathan and Shafir (2013). It encompasses cognitive capacity, which underlies our ability to solve problems and engage in logical reasoning, as well as executive control, which underlies our ability to plan, allocate attention and summon will power to overcome inertia. Furthermore, mental bandwidth is thought to be limited, and when people are mentally taxed by financial or emotional worries, they are left with less mental resources to attend to complex decisions with long term consequences. One such decision is mortgage refinancing, which requires both logical reasoning and the overcoming of inertia.

refinancing among low income borrowers more than among high income borrowers, and thus contributed to increases in refinancing inequality.

Overall we interpret the evidence as suggesting that some low-income borrowers refinanced less than high-income borrowers partly due to their forbearance status or ineligibility, but also due to limited financial literacy and sophistication, and due to the regressive tax on mental bandwidth introduced by the pandemic. These factors jointly contributed to an under-representation of low-income borrowers in the pool of refinancing applications.

6.2 Lenders

During the period of analysis, there was an unprecedented growth in refinancing applications. Figure C1 shows the density distribution of lender-level growth rates in the number of applications submitted to Freddie Mac's LPA tool between the first six months of 2020, and the first six months of 2019. The mode is close to 100 percent but the growth in applications was unequal across lenders, with some experiencing increases of more than 700 percent and others experiencing small decreases. [17]

If lenders have limited resources, they may prioritize the most profitable applications. While income does not directly affect profitability, income maybe correlated with characteristics that affect profitability, such as unpaid balances and credit quality. The prioritization of high-income borrowers can take place at the approval stage (conditional on applying), or at the pre-application stage by targeting marketing efforts towards high income borrowers. The former is more likely to occur when funding capacity is limited, the later is more likely when operational capacity is limited. In the following we explore these two possibilities.

[17]Mean of 145 percent, and standard deviation of 140 percent, from a base of about 2,000 applications per lender during the first six months of 2019.

Priority processing for high-income borrowers, conditional on applying We aggregate the application-level data at the lender-by-income-quintile level. We approximate approval rates with funding rates: i.e. the fraction of applications ran through Freddie Mac's LPA that resulted in loans that were eventually funded (purchased) by Freddie Mac. For each lender-by-income-quintile we calculate the change in funding rates between the first six months of 2019 and the first six months of 2020. First note that despite the large increase in applications, funding rates remained relatively constant: the average change in the funding rate during the period is -0.98 percentage points, from a basis of 25 percent. Then we explore whether changes in funding rates were different across applicants with different income levels. In column 1 of Table 6 we regress changes in funding rates between 2019 and 2020 on quintiles of borrower income. During this period, the funding rate of borrowers in the bottom quintile of the income distribution decreased by 0.37 percentage points. However, we do not find any evidence suggesting that the funding-rate gap between high- and low-income borrowers widened: all income quintile coefficients are small; only the coefficient for the top quintile of the income distribution is statistically significant, and it is negative. For individuals in the top quintile of the income distribution the funding rate decreased an average of 1.52 percentage points (-1.15 + -0.37), which is 1.15 percentage points larger than the reduction for borrowers in the bottom quintile of the income distribution. As noted before, 25 percent of applications during this period were funded. Thus, the results of Table 6 suggest that funding rates decreased slightly during the period, but there is no evidence suggesting that the funding-rate gap between high- and low-income borrowers widened.

In column 2 of Table 6, we allow for the possibility that funding rates may have evolved differently for lenders with large or moderate growth in applications. Specifically, we regress changes in funding rates on lenders' application growth rate. Consistent with Figure C2 we find that growth in applications is negatively correlated with changes in funding rates.

For every 100 percent increase in application volume, lenders saw a 3.2 percentage point decrease in funding rates, again from a base rate of 25 percent during the period. Finally, column 3 shows the interaction of lender-level growth rates and quintiles of applicant income. All coefficients are negative and, small in magnitude. The interaction of growth rates with high-income quintiles is negative and statistically significant. Among high-income borrowers, application growth is correlated with larger decreases in funding rates, compared with low-income borrowers. Specifically, a 100 percent increase in application growth is correlated with a 1.24 percentage point *decrease* in the funding-rate gap between high- and low-income borrowers. This is evidence against the hypothesis that refinancing inequality is driven by lenders prioritizing *the funding* of applications of high-income borrowers due to binding capacity constraints. To interpret the magnitude of these changes, we note that on average Freddie Mac funded about 25 percent of applications submitted through the LPA tool during the first half of 2019.

Targeting marketing efforts towards high-income borrowers (pre-application stage) 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). We find that, during 2020, 15.6 percent of mortgage refinancing applications come from borrowers with an income that places them in the top 10 percent of the income distribution of Freddie Mac's mortgage portfolio. Only 4.5 percent of applications come from borrowers in the bottom decile of the income distribution. In contrast, during the 2019 refinancing wave, 11.8 percent of applications came from borrowers in the top decile of the income distribution of Freddie Mac's portfolio at the time, and 6.7 percent came from borrowers in the bottom decile. This suggests that, while lower-income borrowers are generally under-represented in the pool of refinancing applications, the under-representation of low-income borrowers at the application stage, was stronger in 2020 than in previous periods of high refinancing activity. This under-representation of low income borrowers can be due to supply or demand considerations. Here we discuss the possibility that this is the result of lender's behavior.

Borrowers with high credit score or high unpaid balances could be more profitable for lenders since they have a higher probability of being approved and/or entail a higher gain on sale. Furthermore, high-income borrowers tend to carry larger balances and have higher credit scores. As a result, when capacity constraints became binding, lenders may have directed their marketing efforts towards high-income borrowers more so than in previous periods.

To explore this possibility we perform a mediation analysis of our results on refinancing propensities. We expand our main specification of Equation 1 to include a new term interacting FICO scores and unpaid balances with a dummy variable taking the value of one for observations corresponding to the 2020 wave, respectively. This term captures the differential relation between borrower characteristics and refinancing activity observed in 2020, relative to previous periods. We say that this differential relation is a mediator of increases in refinancing inequality if the significance of ϕ_5 from Equation 1 disappears (or decreases) when the new term is added. Quantitatively, we look at the ratio of ϕ_5 coefficients estimated with and without the new interaction term.[18] The results are presented in Table 7. Coefficient ϕ_5 corresponds to the term Income quintile 5:Wave2020. For ease of reference Column

[18]This is equivalent to the ratio of direct to total effects of a given variable used in (Das et al., 2020) because the coefficient estimated without the new interaction term is mathematically identical to the sum of its direct and indirect effects (i.e. the total effect), as defined in that paper.

1 shows our main estimate of ϕ_5 from Section 4 (taken from Column 3 of Table 2) which does not include the new interaction term.[19]

Column 2 shows the results of expanding Equation 1 to include the interaction of FICO scores and the 2020 dummy. Including the new interaction term in the estimation of Equation 1 reduces ϕ_5 from 6.23 to 6.06. In both cases ϕ_5 is highly significant. This implies that the differential effect of FICO scores on refinancing activity during the pandemic is not an important mediator of the increases in refinancing inequality observed during the period. Refinancing inequality is thus not explained by lenders soliciting high FICO borrowers more so than before. In Column 3 we expand Equation 1 to include the interaction of Unpaid balances and a 2020 dummy. Adding this term reduces ϕ_5 from 6.23 to 2.91. This implies that the differential effect of unpaid balances in 2020 is an important mediator of the increases in refinancing inequality observed during 2020.[20] The unusually high importance of unpaid balances observed during 2020, is consistent with lenders soliciting high-balance borrowers at a higher rate than before, due to capacity constraints. Finally, in the absence of borrower-specific information, lenders could also target based on fine geography and when capacity constraints bind this targeting could be more pronounced. In Column 4 we interact zip code fixed effects with a dummy for 2020. We do not find evidence consistent with this interpretation: our estimates for refinancing inequality remain practically unchanged when

[19]Our main analysis in Section 4 controls for unpaid balances and FICO scores assuming that their effect on refinancing activity is constant over time. In that section, our primary objective is to highlight that, high-income borrowers refinanced a lot more than what we would expect in normal times among borrowers with otherwise comparable characteristics. Here, our objective is explaining *why* high-income borrowers refinanced that much. Expanding our main specification to include the interaction of observable characteristics with a 2020 dummy is appropriate to study mechanisms because it allow us to perform a mediation analysis. It is not appropriate for the objectives of Section 4 because the resulting specification does not allow us to benchmark refinancing inequality to the levels we would expect in normal times: the new control variable absorbs the unusual features of early 2020 that ϕ_5 in Equation 1 intends to capture in the first place.

[20]The mediation effect of unpaid balances is mechanically explained by a large and significant interaction coefficient of unpaid balances x Wave 2020, and a high correlation between unpaid balances and income. Appendix G shows the full set of coefficients. Before the pandemic, individuals with balances over 300K were refinancing a 4.14 pp more than individuals with balances below 200K. During the pandemic, individuals with balances over 300K are refinancing a 9.74 pp more than individuals with balances below 200k. Furthermore, a 1% increase in income is correlated with a 0.78% in unpaid balances.

introducing this control. This is expected given regulatory constraints to prevent redlining strategies.

Overall, while this lender-side analysis does not rule out that lenders would change their approval rates if more low-income borrowers were to apply, the evidence suggests that operational constraints were more relevant than funding constraints. A stronger than usual targeting of high-income borrowers, motivated by operational constraints, is one potential reason behind the under-representation of low-income borrowers observed in the pool of applications.

7 Longer Horizons for the 2020 Refinancing Wave

Our main results focuses on refinancing inequality at the onset of the pandemic. This period is of interest in itself since swift policy interventions can provide relief for immediate losses and mitigate the spiraling of the crisis. We nevertheless provide additional insights into the evolution of refinancing inequality over the entire refinancing boom of which went on to the end of 2021. Our results show that increases in refinancing inequality were not a short lived phenomena, but persisted for the full duration of the refinancing wave that started in early 2020 and continued till the end of 2021.

We re-estimate Equation 1 four times, each time using a different end-point for the refinancing wave that started in 2020. In the first iteration we use the same definition as in the main analysis, with the refinancing wave ending in June 2020. In the second —separate— regression, the 2020 refinancing wave goes up to December 2020. In the third regression the refinancing wave that started in 2020 goes all the way to June 2021. In the fourth regression the last refinancing wave continues until December 2021. To facilitate comparisons across the four definition we estimate the model each time without controls. To get the largest market coverage we use data from McDash.

In Figure 4 we plot the levels of refinancing inequality under each of the different definitions. All refinancing waves previous to 2020 are pooled together as in the main analysis, and the definition (and estimate of refinancing inequality before 2020) is the same across all four regressions. Refinancing inequality before 2020 is given by the coefficient β_5, refinancing inequality for the wave that started in 2020 is given by the $\beta_5 + \phi_5$. The Table with the full set of coefficients is presented in Appendix H.[21]

We can see that refinancing inequality did not disappear in the second half of 2020 or in 2021. Instead, it increased. We argue this happened because, even when COVID-19 case rates and COVID-19 mortality decreased, the supply and demand factors identified in Section 6 were still present. During this extended time horizon, interest rates continued to decrease and refinancing activity continued to be strong peaking in March 2021 and coming back to pre-2020 levels only in early 2022.[22] On the lender side, operational capacity was still at the limit with record high applications. This incentivized lenders to prioritize the solicitation of the most profitable applications.

Our analysis reveals that the effect was not driven only by historically low interest rates since idiosyncratic variation in the severity of the pandemic was correlated with refinancing activity. Even though case and mortality rates stabilized, the impact of the pandemic on borrower related factors was long lasting. On the one hand, coming out of an economic shock is not immediate. While about 60% of individuals left forbearance status by May 2021 Cherry et al. (2021), low income individuals were more likely to continue on forbearance

[21]Note that it would not be appropriate to change the duration of the refinancing waves previous to 2020 because those were much shorter to begin with. In all cases, interest rates reach an upward trajectory a few months after the beginning of the corresponding wave.

[22]See https://fred.stlouisfed.org/series/MORTGAGE30US and https://www.consumerfinance.gov/data-research/research-reports/data-point-2021-mortgage-market-activity-trends/

keeping an accumulated debt overhang of $60 billion by that date. On the other, even in the summer of 2021 COVID-19 was centerpiece of the national discussion with the Delta variant becoming the dominant variant in the US, thus the threat of the pandemic continued to tax mental bandwidth with general worries about the economy, personal finances and health.[23]

Ultimately, our evidence is consistent with different refinancing experiences across the income distribution, with high income borrowers being sought after by lenders and low income borrowers carrying the burden of the initiative at times in which it was particularly hard to do so. The lessons of our analysis are not exclusive to pandemic periods. Instead, our results suggest that operational constraints in mortgage origination have distributional consequences, specially when financial or psychological barriers limit low-income borrowers ability to actively seek refinancing opportunities. The COVID-19 pandemic was a time in which these factors coalesced, but we would expect similar outcomes in future periods when some or all of these factors reoccur.

8 Robustness

Our main analysis is based on mortgages for which the option to refinance is in-the-money, according to the model for optimal refinancing of Agarwal et al. (2013). This model takes into account that individuals may not hold their loans to maturity and instead have a positive probability of prepaying their loans at any point, as well as an estimate for closing costs. In Appendix F, 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.

[23] https://www.cdc.gov/museum/timeline/covid19.html

In Appendix E we also verify the robustness of our results by looking at refinancing inequality with a third different data set. We use a different approach focused on new originations (instead of propensities to refinance) which does not impose filters like being newly in-the-money. We also including FHA and VA loans. The results are all consistent with our main analysis.

9 Final Comments

In this paper we introduce the concept of refinancing inequality, by which we refer to differences in savings from refinancing across the income distribution. We use the refinancing income gap, defined as the difference in refinancing activity between the top and bottom quintiles of the income distribution, as a summary measure to describe refinancing inequality over time, and as the severity of the pandemic increased. We find that during the COVID-19 pandemic, refinancing inequality increased considerably due to a variety of lender and borrower factors that coalesced during this period and lead to an unequal refinancing experience for high- and low-income borrowers.

Our findings suggest that there is room for targeted policies that promote refinancing activity of low-income borrowers, such as incentives for lenders to deliver loans to borrowers that meet a target profile, financial education efforts, and automation of refinancing processes to bypass the effects of behavioral biases.

References

Adams-Prassl, A., T. Boneva, M. Golin, and C. Rauh (2020). Inequality in the impact of the coronavirus shock: Evidence from real time surveys. *Journal of Public Economics 189*, 104245.

Agarwal, S. (2007). The impact of homeowners' housing wealth misestimation on consumption and saving decisions. *Real Estate Economics 35*(2), 135–154.

Agarwal, S., G. Amromin, I. Ben-David, S. Chomsisengphet, T. Piskorski, and A. Seru (2017). Policy intervention in debt renegotiation: Evidence from the home affordable modification program. *Journal of Political Economy 125*(3), 654–712.

Agarwal, S., G. Amromin, S. Chomsisengphet, T. Landvoigt, T. Piskorski, A. Seru, and V. Yao (2022). Mortgage refinancing, consumer spending, and competition: Evidence from the home affordable refinancing program. *Review of Economic Studies*.

Agarwal, S., I. Ben-David, and V. Yao (2017). Systematic mistakes of borrowers in the mortgage markets. *Journal of Financial Economics 123*, 42–58.

Agarwal, S., Y. Deng, Q. Gu, J. He, W. Qian, and Y. Ren (2022). Mortgage debt, hand-to-mouth households, and monetary policy transmission. *Review of Finance 26*(3), 487–520.

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

Amel, D., E. Anenberg, and R. Jorgensen (2018). On the geographic scope of retail mortgage markets. *FEDS Notes. Washington: Board of Governors of the Federal Reserve System*.

An, X., L. Cordell, L. Geng, and K. Lee (2021). Inequality in the time of covid-19: Evidence from mortgage delinquency and forbearance. *Available at SSRN 3789349*.

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

Angrist, J. D. (2001). Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *Journal of Business & Economic Statistics 19*(1), 2–28.

Auclert, A. (2019). Monetary policy and the redistribution channel. *American Economic Review 109*(6), 2333–67.

Baker, S. R., R. A. Farrokhnia, S. Meyer, M. Pagel, and C. Yannelis (2020). Income, liquidity, and the consumption response to the 2020 economic stimulus payments. Technical report, National Bureau of Economic Research.

Beraja, M., A. Fuster, E. Hurst, and J. Vavra (2017). Regional heterogeneity and monetary policy. Technical report, National Bureau of Economic Research.

Berger, D. W., K. Milbradt, F. Tourre, and J. Vavra (2019). Mortgage prepayment and path-dependent effects of monetary policy. Technical report, National Bureau of Economic Research.

Cherry, S. F., E. X. Jiang, G. Matvos, T. Piskorski, and A. Seru (2021). Government and private household debt relief during covid-19. *Brookings Papers on Economic Activity*.

Chetty, R., J. N. Friedman, N. Hendren, M. Stepner, et al. (2020). Real-time economics: A new platform to track the impacts of covid-19 on people, businesses, and communities using private sector data. *NBER Working Paper 27431*, 36–46.

Coibion, O., Y. Gorodnichenko, and M. Weber (2020). Labor markets during the covid-19 crisis: A preliminary view. Technical report, National Bureau of Economic Research.

Das, S., C. M. Kuhnen, and S. Nagel (2020). Socioeconomic status and macroeconomic expectations. *The Review of Financial Studies 33*(1), 395–432.

DeFusco, A. A. and J. Mondragon (2020). No job, no money, no refi: Frictions to refinancing in a recession. *Journal of Finance 75*(5), 2327–2376.

Di Maggio, M., A. Kermani, B. J. Keys, T. Piskorski, R. Ramcharan, A. Seru, and V. Yao (2017). Interest rate pass-through: Mortgage rates, household consumption, and voluntary deleveraging. *American Economic Review 107*(11), 3550–88.

Di Maggio, M., A. Kermani, and C. J. Palmer (2020). How quantitative easing works: Evidence on the refinancing channel. *The Review of Economic Studies 87*(3), 1498–1528.

Eichenbaum, M., S. Rebelo, and A. Wong (2018). State dependent effects of monetary policy: The refinancing channel. Technical report, National Bureau of Economic Research.

Fuster, A., A. Hizmo, L. Lambie-Hanson, J. Vickery, and P. S. Willen (2021). How resilient is mortgage credit supply? evidence from the covid-19 pandemic. Technical report, National Bureau of Economic Research.

Fuster, A., M. Plosser, P. Schnabl, and J. Vickery (2019). The role of technology in mortgage lending. *The Review of Financial Studies 32*(5), 1854–1899.

Gerardi, K., P. Willen, and D. H. Zhang (2020). Mortgage prepayment, race, and monetary policy. *Federal Reserve Bank of Boston*.

Gerardi, K. S., L. Lambie-Hanson, and P. S. Willen (2021). Racial differences in mortgage refinancing, distress, and housing wealth accumulation during covid-19. *Federal Reserve Bank of Philadelphia Payment Cards Center Discussion Paper* (21-2).

Goodstein, R. M. (2014). Refinancing trends among lower income and minority homeowners during the housing boom and bust. *Real Estate Economics 42*(3), 690–723.

Johnson, E., S. Meier, and O. Toubia (2015). Money left on the kitchen table: Exploring sluggish mortgage refinancing using administrative data, surveys, and field experiments. *Unpublished working paper*.

Karger, E. and A. Rajan (2020). Heterogeneity in the marginal propensity to consume: Evidence from covid-19 stimulus payments. Technical report.

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

Khatana, S. A. M. and P. W. Groeneveld (2020). Health disparities and the coronavirus disease 2019 (covid-19) pandemic in the usa. *Journal of general internal medicine 35*(8), 2431–2432.

Kinder, M. and M. Ross (2020). Reopening america: Low wage workers have suffered badly from covid-19 so policy makers should focus on equity. Technical report, Brookings Institution.

Laibson, D., P. Maxted, and B. Moll (2020). Present bias amplifies the household balance-sheet channels of macroeconomic policy. Technical report, National Bureau of Economic Research.

Mullainathan, S. and E. Shafir (2013). Scarcity. *Social Policy 46*(2), 231–249.

Nothaft, F. E. and Y. Chang (2005). Refinance and the accumulation of home equity wealth. *Building Assets Building Credit: Creating Wealth in Low-Income Communities*, 71–102.

Purtle, J. (2020). Covid-19 and mental health equity in the united states. *Social psychiatry and psychiatric epidemiology 55*(8), 969–971.

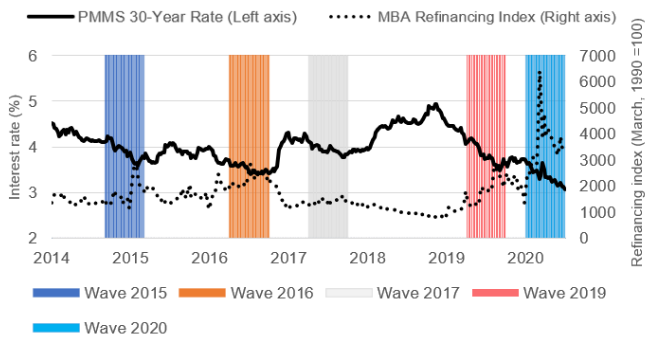
TransUnion (2020). Monthly industry snapshot: Financial services. Technical report, TransUnion.

Wallace, J. and V. Patrick (2020). Life in lockdown is testing parents' bandwidth, but there are ways to protect your mental energy. *The Washington Post*.

Wong, A. (2019). Refinancing and the transmission of monetary policy to consumption. *Unpublished manuscript*.

Figures and Tables

Figure 1: Evolution of Interest Rates and Refinancing Activity (2014–2020)

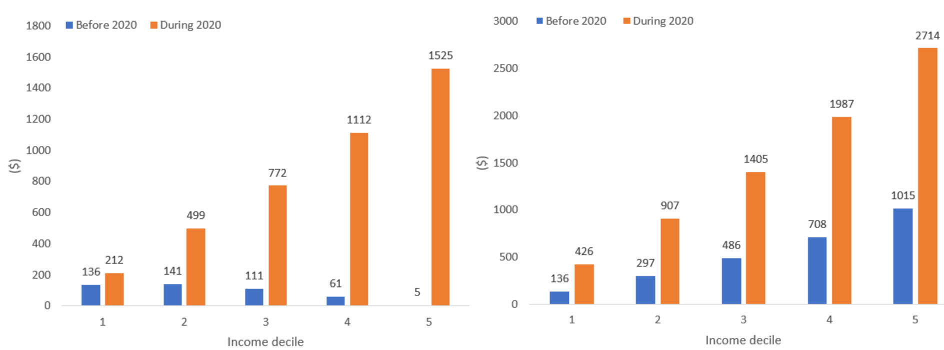


Note: This figure shows the evolution of mortgage interest rates and refinancing activity over time between 2014 and 2020. The left axis shows the 30-year fixed rate mortgage from the Primary Mortgage Market Survey (PMMS). The right axis shows the Mortgage Bankers Association (MBA) refinancing index, which takes the value of one on March 18, 1990. The highlighted periods correspond to five-month windows with the largest declines in interest rates and define five refinancing waves. The 2015 wave corresponds to October 2014 to February 2015, the 2016 wave corresponds to May 2016 to September 2016, the 2017 wave corresponds to May 2017 to September 2017, the 2019 wave corresponds to May 2019 to September 2017, and the 2020 wave corresponds to February 2020 to June 2020.

Figure 2: Savings from Refinancing for the Entire Portfolio, by Income Quintile, Before and During 2020

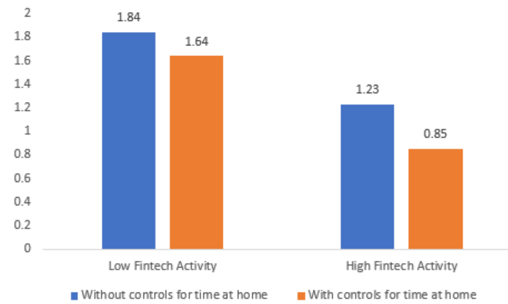
(a) Model with full set of control variables

(b) Model without control variables



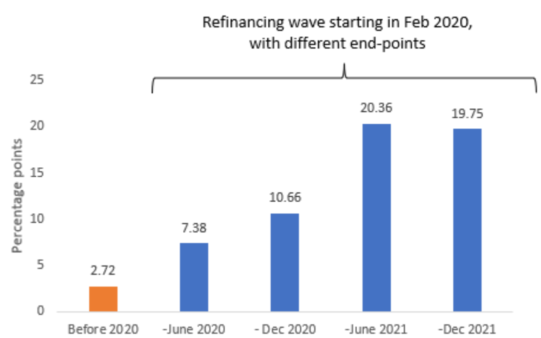
Note: Savings are projected based on the coefficients of Table 4 using data from Freddie Mac. The reference category captures refinancing levels in the bottom quintile of the income distribution before 2020. The projections in panel (a) are based on coefficients estimated with the full set of control variables, and as a result, the projection holds control characteristics fixed at the levels observed on individuals in the bottom quintile of the income distribution. The projections in panel (b) are based on coefficients without controls.

Figure 3: COVID Case Rates and The Refinancing Income Gap with and without Controls for Time at home in Areas of High or Low Fintech Activity



Note: This figure plots the coefficient for High COVID*Income Quintile 5 from Equation 2 which we estimate with or without controls for time spent at home and its interaction with income quintiles, using data from McDash. The first set of bars is estimated in areas of low fintech activity. The second set of bars is estimated in areas of high fintech activity. Low/high fintech activity corresponds to mortgages in the top/bottom quintile of the distribution of the county-level fraction of fintech applications in HMDA for the year 2019. The full set of coefficients is presented in Table D2.

Figure 4: Refinancing Income Gap for Before and During 2020 with Different End-points for the 2020 Refinancing wave. Regressions without controls.



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 without controls, using data from McDash. The orange bar corresponds to the refinancing income gap before 2020, pooling all original waves together. The blue bars correspond to the refinancing income gap observed during the refinancing wave that started in Feb 2020. Each bar is estimated on a separate regression in which the 2020 wave is defined with a different end period, as indicated in the horizontal axis.

Table 1: Descriptive Statistics of Newly in the Money Mortgages before and during 2020

	Freddie Mac Data				Mc Dash Data			
	Before 2020		During 2020		Before 2020		During 2020	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
FICO score	740	52	743	49	727	54	738	50
Income (monthly, thousands)	7.35	4.75	7.65	4.79	4.86	10.36	4.73	6.94
Loan Age (months)	55.41	49.89	58.55	52.20	68.49	45.94	56.86	46.05
Interest Rate (%)	4.58	0.93	4.39	0.77	5.14	0.30	4.68	0.22
LTV (%)	77.32	20.54	77.73	18.94	52.52	20.01	59.48	21.00
Unpaid Balance (thousands USD)	188.66	110.95	202.74	116.79	225.77	125.17	248.97	141.61
Potential Savings (thousands USD)	1.97	6.21	5.19	6.34	10.39	5.56	11.85	6.40
Rate Incentive (pp)	1.02	0.95	1.26	0.77	1.51	0.26	1.51	0.22
Number of newly in the money mortgages	2,041,992		1,351,845		1,095,038		680,882	
% of active mortgages that became newly in the money during each wave	9.03		20.14		7.30		15.16	
% of newly in the money mortgages that were prepaid during the period	8.68		16.59		3.92		7.31	

Note: This table presents descriptive statistics for the main variables considered in the analysis. The left panel corresponds to mortgages in the portfolio of Freddie Mac. The right panel corresponds to mortgages reported to McDash. In both cases, we include mortgages active at the beginning of each refinancing wave and newly in-the-money. We say that a mortgage is newly in-the-money during a refinancing wave when it was not in-the-money at the beginning of the refinancing wave but becomes in-the-money during the corresponding refinancing wave. We say that a mortgage is in-the-money when it satisfies the conditions outlined in Agarwal et al. (2013). For Freddie Mac data, income corresponds to the income reported at origination. For McDash data, income is estimated from debt-to-income ratios reported at origination. Potential savings are defined as the present value over the expected life of the loan of the difference in outflows calculated with the original interest rate of each loan and the Primary Mortgage Market Survey (PMMS) rate at the end of the corresponding period. The expected life of the mortgage is parametrized as in Agarwal et al. (2013). Rate incentive is defined as the difference between the original interest rate of each loan and the PMMS rate at the end of the corresponding refinancing wave. pp = percentage points. LTV = loan to value.

Table 2: Refinancing Inequality Before and During 2020: Probability of Refinancing

Dep.Var. Refinancing Indicator {0,1}			
	(1)	(2)	(3)
Income quintile2	1.50***	-0.1	-0.04
	(0.07)	(0.1)	(0.07)
Income quintile3	3.02***	0.2***	-0.12
	(0.08)	(0.1)	(0.08)
Income quintile4	4.68***	0.9***	-0.15
	(0.11)	(0.1)	(0.1)
Income quintile5	6.84***	2.3***	0.11
	(0.13)	(0.1)	(0.12)
Wave 2020	4.44***	0.9***	1.19***
	(0.12)	(0.1)	(0.11)
Income quintile2:Wave 2020	2.93***	2.4***	2.14***
	(0.18)	(0.2)	(0.17)
Income quintile3:Wave 2020	4.63***	4.4***	3.79***
	(0.2)	(0.2)	(0.18)
Income quintile4:Wave 2020	6.13***	6.4***	5.48***
	(0.21)	(0.2)	(0.2)
Income quintile5:Wave 2020	6.72***	7.4***	6.23***
	(0.24)	(0.2)	(0.22)
Mean of dep.var. in ref. cat.	1.15	1.15	1.15
Zip Fixed Effect	No	Yes	Yes
Borrower Controls in Regression	No	Yes	Yes
UPB and Original Interest Rate	No	No	Yes
Observations	3,001,491	3,001,491	3,001,491
R2	0.04	0.14	0.14

Note: This table presents the results of estimating equation 1 with data from Freddie Mac. We consider observations that were prepaid and matched to a new rate-refinancing loan during the period of analysis. The dependent variable is a dummy variable that takes the value of one when a mortgage goes through a rate refinancing transaction, 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, 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 3: Refinancing Inequality Before and During 2020: Savings Conditional on Refinancing

	(1)	(2)	(3)	(4)	(5)	(6)
	Rate difference (pp)	Rate difference (pp)	Rate difference (pp)	Savings ($)	Savings ($)	Savings ($)
Income quintile 2	-0.26***	-0.16***	-0.12***	1,621.93***	1,077***	604.26***
	(0.01)	(0.01)	(0.01)	(57.32)	(99)	(65.53)
Income quintile 3	-0.40***	-0.25***	-0.14***	3,207.19***	2,302***	722.41***
	(0.01)	(0.01)	(0.01)	(62.44)	(96)	(67.87)
Income quintile 4	-0.53***	-0.32***	-0.17***	4,551.42***	3,354***	762.02***
	(0.01)	(0.01)	(0.01)	(65.28)	(94)	(73.98)
Income quintile 5	-0.62***	-0.37***	-0.16***	6,362.34***	4,906***	1,117.86***
	(0.01)	(0.01)	(0.01)	(72.56)	(93)	(83.6)
Wave2020	-0.11***	0.07***	0.16***	2,326.31***	990***	2,485.52***
	(0.01)	(0.01)	(0.01)	(63.41)	(111)	(85.63)
Income quintile 2:Wave2020	0.12***	0.06***	0.13***	274.25***	598***	524.13***
	(0.01)	(0.01)	(0.01)	(90.58)	(137)	(99.97)
Income quintile 3:Wave2020	0.18***	0.07***	0.16***	386.59***	995***	989.87***
	(0.01)	(0.01)	(0.01)	(92.42)	(132)	(101.24)
Income quintile 4:Wave2020	0.25***	0.10***	0.19***	644.15***	1,515***	1,548.78***
	(0.01)	(0.01)	(0.01)	(94.58)	(129)	(105.47)
Income quintile 5:Wave2020	0.28***	0.10***	0.20***	1,061.15***	2,133***	2,414.30***
	(0.01)	(0.01)	(0.01)	(104.31)	(126)	(116.47)
Mean of dep.var. in ref. cat.	1.66	1.66	1.66	2,898	2,898	2,898
Zip fixed effect	No	Yes	Yes	No	Yes	Yes
Borrower controls	No	Yes	Yes	No	Yes	Yes
Original rate and UPB	No	No	Yes	No	No	Yes
Observations	76,955	76,955	76,955	76,955	76,955	76,955
R2	0.12	0.42	0.55	0.25	0.51	0.54

Notes: This table presents the results of estimating equation 1 with data from Freddie Mac. We consider observations that were prepaid and matched to a new rate-refinancing loan during the period of analysis. For columns 1, 2, and 3, the dependent variable is defined as the difference between interest rates of the old (refinanced) and new (refinancing) loans. For columns 4, 5, and 6, the dependent variable takes the value of the dollar savings from refinancing, as defined in Agarwal et al. (2013). The full list of control variables is as follows: zip code fixed effects, loan age, FICO score, loan to value, original interest rate, and unpaid balance (UPB) (continuous variables are split into discrete categories and controlled for as dummies). Income quintile 1 is the lowest income quintile. pp = percentage points.

Table 4: Savings from Refinancing for the Entire Portfolio, Across the Income Distribution, Before and During 2020

	Dep.Var. Realized Refi Savings ($US)					
	(1)	(2)	(3)	(4)	(5)	(6)
Income quintile 2	161***	14*	5	2,793***	1,362***	1,138***
	(7)	(7)	(7)	(71)	(69)	(69)
Income quintile 3	350***	103***	-25***	4,637***	2,293***	1,360***
	(7)	(7)	(8)	(69)	(68)	(68)
Income quintile 4	572***	231***	-75***	6,423***	3,283***	1,534***
	(7)	(8)	(8)	(67)	(67)	(70)
Income quintile 5	879***	453***	-131***	8,135***	4,277***	1,494***
	(7)	(8)	(8)	(66)	(66)	(71)
Wave 2020	290***	-75***	76***	4,602***	1,159***	1,528***
	(9)	(9)	(9)	(79)	(79)	(80)
Income quintile 2:Wave 2020	320***	298***	282***	1,428***	1,657***	1,260***
	(12)	(12)	(12)	(104)	(102)	(100)
Income quintile 3:Wave 2020	629***	663***	585***	2,172***	2,961***	2,338***
	-12	-12	-12	-100	-98	-98
Income quintile 4:Wave 2020	989***	1,089***	975***	2,713***	4,085***	3,377***
	(12)	(12)	(12)	(98)	(96)	(96)
Income quintile 5:Wave 2020	1,409***	1,573***	1,444***	3,075***	4,969***	4,245***
	(12)	(12)	(12)	(96)	(94)	(96)
Model	OLS	OLS	OLS	Tobit	Tobit	Tobit
Tobit scale				15,334	14,712	14,505
Borrower controls	No	Yes	Yes	No	Yes	Yes
Original rate and UPB	No	No	Yes	No	No	Yes
Omitted group	Low inc pre2020	Low inc pre2020	Low inc pre2020	Low inc pre2020	Low inc pre2020	Low inc pre2020
Savings for omitted group ($)	136	136	136	3,114	3,114	3,114
Savings gap before 2020	879	453	-131	1,690	1,087	386
Savings gap in 2020	2,288	2,026	1,313	5,009	5,073	3,798
Observations	3,002,394	3,002,394	3,002,394	3,002,394	3,002,394	3,002,394
Log likelihood				-2,237,309	-2,195,549	-2,188,227

Notes: This table the results of estimating equation 1 or a Tobit model using equation 1 as the linear index, with data from Freddie Mac. We consider observations that were not prepaid during the period of analysis or were prepaid and matched to a new rate-refinancing loan. Matched prepayments are weighted by the inverse of the probability of a match. When a mortgage is refinanced, the dependent variable takes the value of the dollar savings from refinancing according to the formula of Agarwal et al. (2013). When a mortgage is not refinanced, the dependet variable takes the value of zero. The full list of control variables is as follows: zip code fixed effects, loan age, FICO score, loan to value (LTV), original interest rate, 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 5: Refinancing inequality and the severity of the COVID-19 Pandemic

	Dep.Var. Refinancing Indicator {0,1}				
	(1)	(2)	(3)	(4)	(5)
Income quintile 2	0.13***	0.20***	0.09**	0.13***	0.15**
	(0.04)	(0.06)	(0.04)	(0.04)	(0.06)
Income quintile 3	0.01	0.21***	-0.07	-0.01	0.11
	(0.04)	(0.07)	(0.04)	(0.05)	(0.07)
Income quintile 4	-0.13***	0.22***	-0.31***	-0.16***	0.05
	(0.05)	(0.08)	(0.05)	(0.05)	(0.08)
Income quintile 5	0.07	0.23***	-0.09*	-0.06	0.002
	(0.05)	(0.08)	(0.06)	(0.07)	(0.09)
High COVID	-0.59***	-0.64***	-0.32***	-0.57***	-0.30***
	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)
Income quintile 2:High COVID	0.44***	0.44***	0.36***	0.43***	0.34***
	(0.05)	(0.05)	(0.06)	(0.07)	(0.07)
Income quintile 3:High COVID	0.84***	0.83***	0.67***	0.78***	0.57***
	(0.05)	(0.05)	(0.07)	(0.08)	(0.08)
Income quintile 4:High COVID	1.17***	1.15***	0.71***	1.11***	0.52***
	(0.06)	(0.06)	(0.07)	(0.08)	(0.09)
Income quintile 5:High COVID	1.26***	1.26***	0.88***	1.14***	0.71***
	(0.06)	(0.06)	(0.08)	(0.08)	(0.09)
Mean of dep.var. in ref. cat.	0.98	0.98	0.98	0.98	0.98
Controls for local economic conditions:					
UI Claims and					
UI Claims x Income Quintiles?	No	Yes	No	No	Yes
Forbearance and					
Forbearance x Income Quintiles?	No	No	Yes	No	Yes
Time home and					
Time home x Income Quintiles?	No	No	No	Yes	Yes
% explained by controlling for local economic conditions		0%	30%	10%	40%
Observations	1,970,835	1,970,835	1,970,835	1,970,835	1,970,835
R2	0.04	0.04	0.04	0.04	0.04

Notes: This table presents the results of re-estimating equation 2 with data from MacDash, sequentially adding as controls different measures of local economic conditions and their interaction with income quintiles. For each measure of local economic conditions we create a dummy that takes the value of one when a given geography-month is in quintiles 2-5 of the distribution of forbearance, UI claims or time at home, respectively. The dependent variable takes the value of one when a mortgage was prepaid, and zero otherwise. All columns include month and zip code fixed effects, as well as the following control variables: 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 6: Funding Rate, Application Growth and Borrower Income

	Change in Funding Rate*100 (2019-2020)		
	(1)	(2)	(3)
Growth in Applications		-0.0321***	-0.0249***
		(0.0013)	(0.0038)
Income quintile 2	-0.3145		0.0654
	(0.7617)		(0.9365)
Income quintile 3	-0.6754		0.2724
	(0.7496)		(0.9232)
Income quintile 4	-0.8812		0.9206
	(0.7303)		(0.8995)
Income quintile 5	-1.1517*		1.4925*
	(0.6907)		(0.8498)
Income quintile 2: Growth in Applications			-0.0011
			(0.0050)
Income quintile 3: Growth in Applications			-0.0042
			(0.0049)
Income quintile 4: Growth in Applications			-0.0094**
			(0.0047)
Income quintile 5: Growth in Applications			-0.0124***
			(0.0044)
Constant	-0.3773	3.4523***	2.7208***
	(0.5722)	(0.2598)	(0.6997)
Mean of dep.var. in ref. cat.	-0.38 pp	-0.38 pp	-0.38 pp
Observations	2,634	2,634	2,634
R2	0.0014	0.1953	0.1998
Adjusted R2	-0.0001	0.1949	0.1970

Note: This table uses observations at the lender-by-borrower income quintile level. The dependent variable is the change in the fraction of applications submitted through Freddie Mac's Loan Product Advisor (LPA) tool that were eventually funded by Freddie Mac. Application Growth Rate is calculated at the lender level. Changes and growth rates are calculated between the first half of 2020 and the first half of 2019. Income quintile 1 is the lowest income quintile and income quintile 5 is the highest income quintile. This table includes lenders with at least 1,000 submissions to LPA in 2020 and excludes income quintile x lender cells with less than 25 observations. Results weighted by number of LPA submissions. Standard errors in parenthesis. pp= percentage points.

Table 7: Refinancing Inequality Before and During 2020: Probability of Refinancing Estimated with Flexible Interactions

Dep.Var. Refinancing Indicator {0,1}				
	(1)	(2)	(3)	(4)
Income quintile 2	-0.04	-0.02	0.12*	-0.16**
	(0.07)	(0.07)	(0.07)	(0.06)
Income quintile 3	-0.12	-0.09	0.35***	-0.35***
	(0.08)	(0.08)	(0.08)	(0.07)
Income quintile 4	-0.15	-0.11	0.71***	-0.53***
	(0.1)	(0.1)	(0.09)	(0.09)
Income quintile 5	0.11	0.16	1.42***	-0.57***
	(0.12)	(0.12)	(0.12)	(0.11)
Wave 2020	1.19***	3.04***	6.51***	-0.07
	(0.11)	(0.14)	(0.29)	(0.06)
Income quintile 2:Wave 2020	2.14***	2.03***	1.45***	2.26***
	(0.17)	(0.17)	(0.17)	(0.17)
Income quintile 3:Wave 2020	3.79***	3.66***	2.25***	4.06***
	(0.18)	(0.18)	(0.19)	(0.19)
Income quintile 4:Wave 2020	5.48***	5.35***	3.07***	5.99***
	(0.2)	(0.2)	(0.21)	(0.21)
Income quintile 5:Wave 2020	6.23***	6.06***	2.91***	6.98***
	(0.22)	(0.22)	(0.25)	(0.24)
Mean of dep.var. in ref. cat.	1.15	1.15	1.15	1.15
Zip Fixed Effect	Yes	Yes	Yes	Yes
Borrower Controls	Yes	Yes	Yes	Yes
UPB and Original Interest Rate	Yes	Yes	Yes	Yes
FICO x Wave 2020	No	Yes	No	No
UPB x Wave 2020	No	No	Yes	No
Zip x Wave 2020	No	No	No	Yes
Observations	3,001,491	3,001,491	3,001,491	3,001,491
R2	0.14	0.15	0.15	0.20

Notes: This table presents the results of re-estimating equation 1 with data from Freddie Mac, adding the interaction of unpaid balances, FICO scores and zip codes with a dummy variable for the 2020 refinancing wave. We consider observations that were prepaid and matched to a new rate-refinancing loan during the period of analysis. The dependent variable is a dummy variable that takes the value of one when a mortgage goes through a rate refinancing transaction, 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, and unpaid balance (UPB) (continuous variables are split into discrete categories and controlled for as dummies). Income quintile 1 is the lowest Income quintile.

Internet Appendix

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.

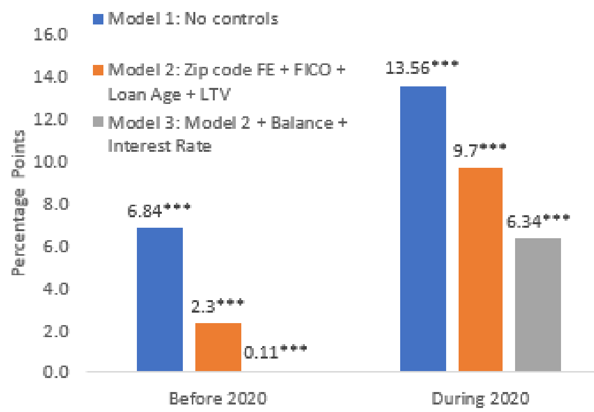
Table A3: Comparison of Matched and Unmatched Loans (Averages)

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).

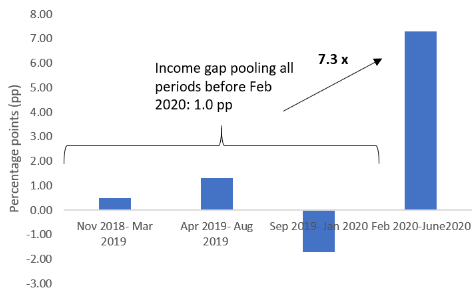
Table B2 studies heterogeneity in refinancing inequality over credit scores, loan to value rations and FICO scores. We estimate our preferred specification (Equation 1) splitting the sample across three selected variables of interest. Columns 1 and 2 split the sample based on original interest rates. 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

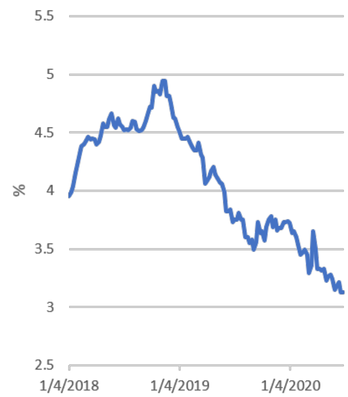


Table B1: Refinancing Inequality Before and During 2020: Probability of Refinancing (percentage points)

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

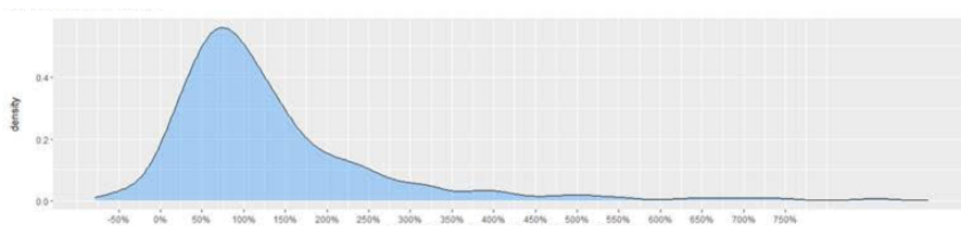
Dep.Var. Refinancing Indicator {0,1}			
	(1)	(2)	(3)
Income quintile 2	0.005***	-0.003***	-0.002***
	(0.001)	(0.001)	(0.001)
Income quintile 3	0.010***	-0.003***	-0.011***
	(0.001)	(0.001)	(0.001)
Income quintile 4	0.016***	-0.002*	-0.023***
	(0.001)	(0.001)	(0.001)
Income quintile 5	0.022***	0.005***	-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

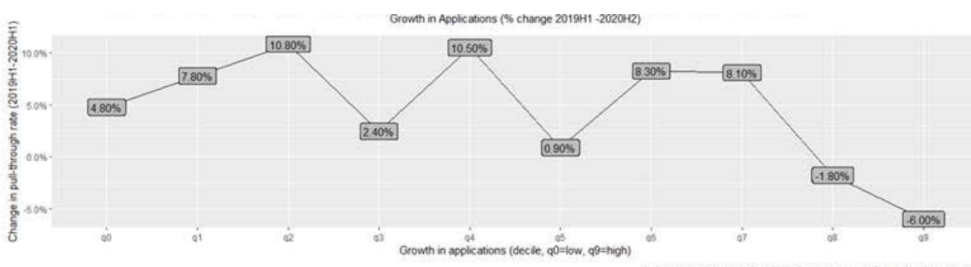
This section shows how applications received by lender changed between the first half of 2020 and the first half of 2019. Figure C1 shows the density of the growth rates in applications by lender. Figure C2 shows the change in funding rates by decile of lender-level growth rate during.

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.

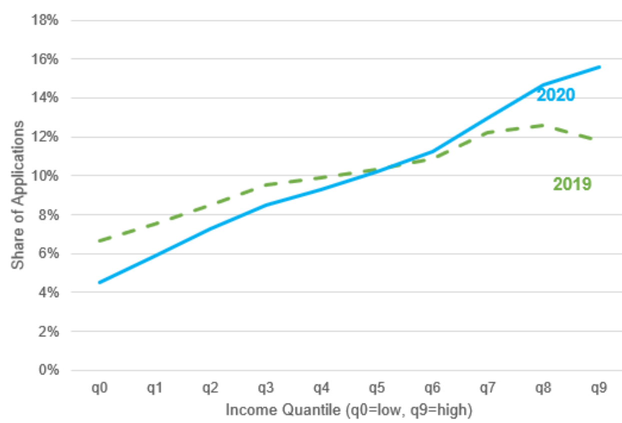
Table C1: Time to Process Applications, across the Income Distribution

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:

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

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; $\textit{Severity } Q_{kizct-1}$ 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

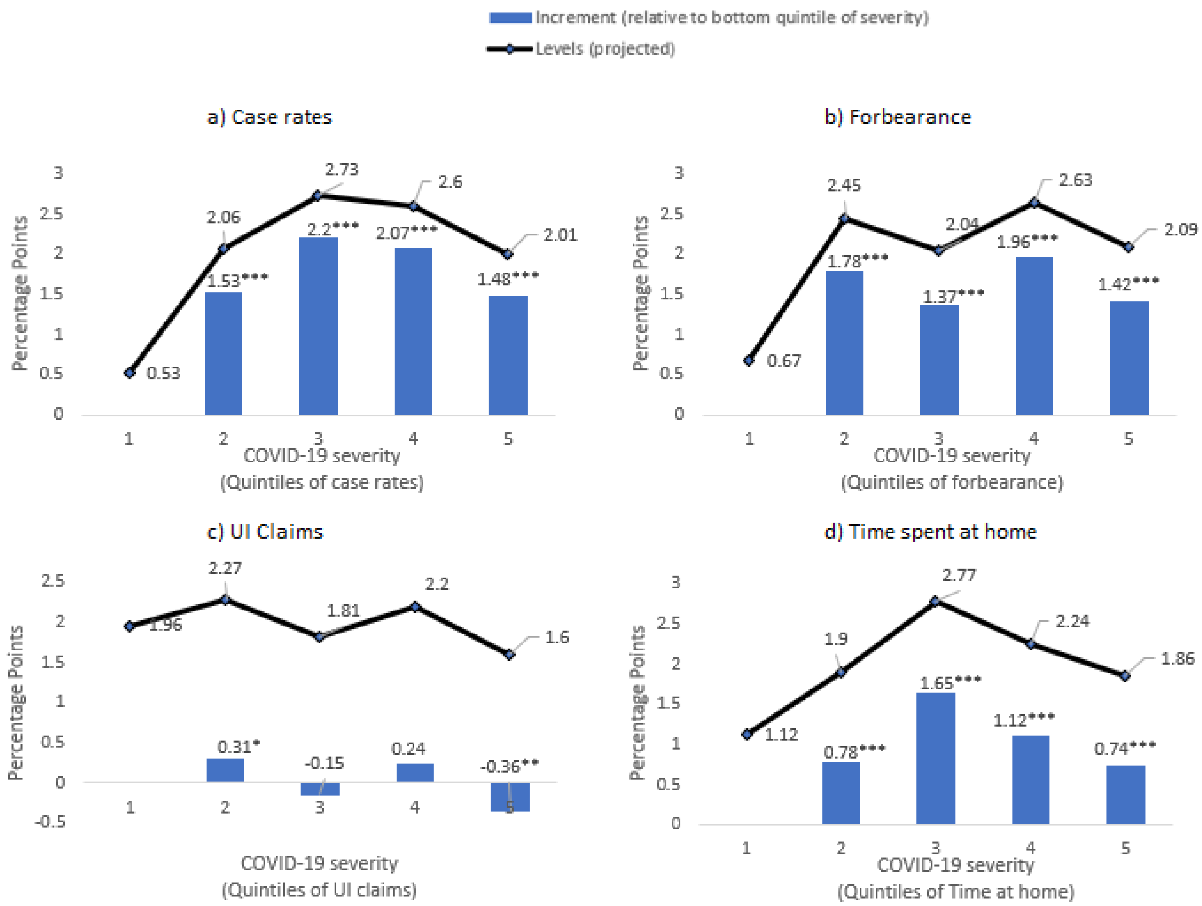
Table D1: Refinancing Inequality Across the Distribution of COVID-19 Severity by Severity Measure

	(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.

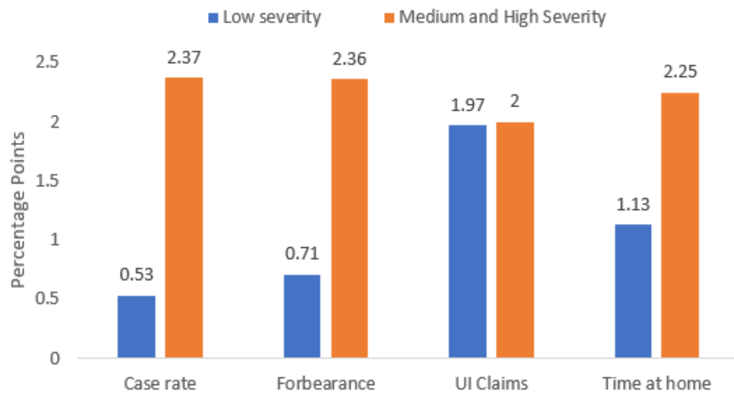
Notes to Table D1: This table presents the results of estimating Equation 3, using data from McDash. Column 1 considers all observations for which we have coverage of COVID-19 case rates. Columns 2-6 considers observations for which we have coverage of COVID-19 case rates, forbearance rates, time spent 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



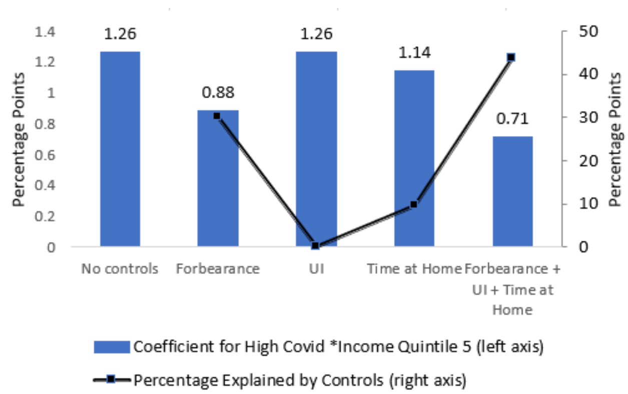
Note: COVID-19 severity is measured separately in each panel with four measures: case rates, forbearance, unemployment insurance (UI) claims and time spent at home. The black line represents projected levels of the refinancing income gap in geography-months that fall in different quintiles of the COVID-19 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

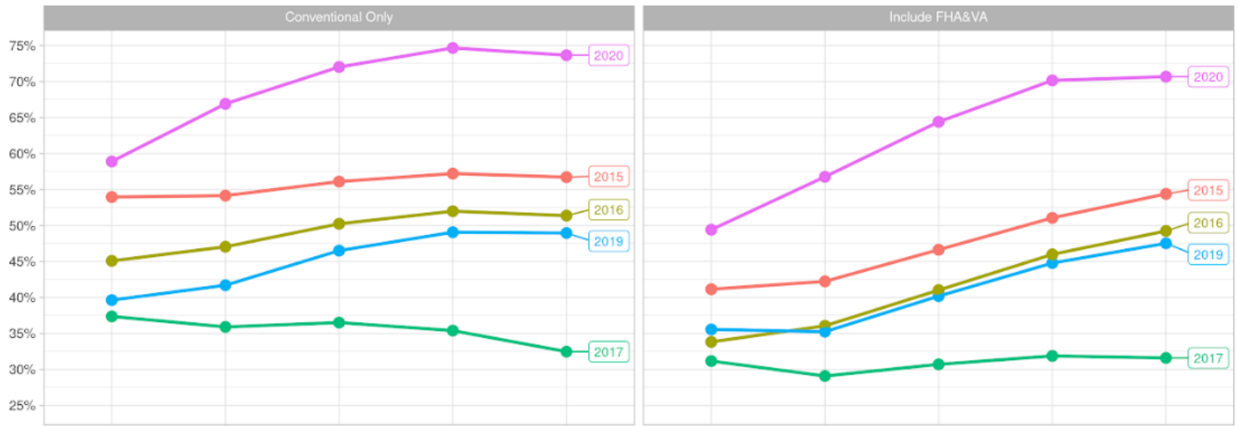
Notes: This table presents the results of re-estimating equation 2 with data from MacDash. Columns 1 and 2 consider observations in the bottom quartile of fintech activity. Colmns 3 and 4 consider observations in the top quartile of fintech activity. Fintech activity is measured aggregating HMDA data 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

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

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

$$k_n = 2000 + 0.007905 M_n. \quad (11)$$

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].

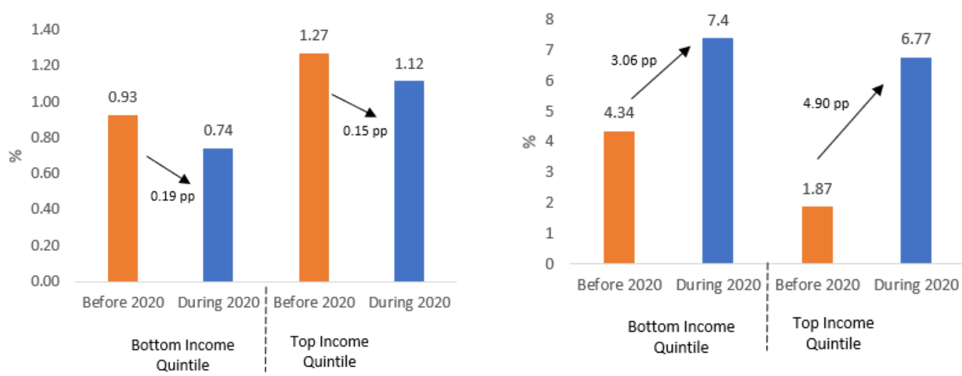
[26] In contrast to An et al. (2021) who study all active mortgages in the market, we focus only on mortgages that became in-the-money, according to the definition of Agarwal et al. (2013) during the first months of 2020 by studying whether our results about refinancing inequality are driven by differences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