Mortgage Loan Application and Acceptance Rates During the Coronavirus Pandemic*

Zolbayar Enkhbaatar†, Luis Alonso Ortega‡, Takanori Takeuchi§
(Group C)

Due on Dec 17, 2021

Abstract

The year 2020, marked by the coronavirus pandemic, was unique for housing prices behaviour and the dynamics of credit markets, something that recent literature is exploring. This paper conducts a first approach to estimate the effect of the coronavirus pandemic on loan acceptances rates, specifically for mortgage loans, with a particular focus on how the relationship of income on loan approval changed in 2020. Utilizing the Home Mortgage Disclosure Act Data, we conduct a two-year panel data regression to estimate the relationship between income and loan approval rates, and how these relationship changed in 2020. Our results suggest that income is still, in general, positively correlated with loan acceptance rates, but this relationship faded in 2020 amid the pandemic. Our results also suggest a positive level shift in loan acceptance rates as a whole in 2020, contrary to our initial expectations, and we provide some comments on this result. We believe there is still a considerable amount of further research to be done in this field (connection to the housing market, temporary or transitory effects in view of 2021, heterogeneous effects across income levels, and further characterization of a credit supply function), although the results here are valuable as a first study of the effect of the pandemic on the loan market.

1 Introduction

It's been almost two years since the Coronavirus pandemic hit the world unexpectedly. The pandemic has changed many economies, industries, norms forever. Some are expected while some are unprecedented. For us, the U.S. housing market phenomenon during the pandemic was interesting enough to do research.

1.1 U.S. housing market

The housing price in the U.S. has increased almost 20 percent year to year as of September, 2021 according to S&P/Case-Shiller U.S. National Home Price Index (Figure 1). This phenomenon is unprecedented. The housing price is even growing at a higher rate than the 2006-07 period of the housing price boom. Another interesting fact was that housing price has increased more in lower income area[1] (Figure 2).

*This paper is the final document for the term project in ECON-GA-1001, 2021 fall semester.

†Net ID: ze2026

‡Net ID: lao8318

§Net ID: tt2292

[1] [1]Zhao (2020) points out this fact focusing on the early stage of outbreak of Covid-19. We confirmed that this trend is robust in more recent periods, as discussed formally in the Appendix.

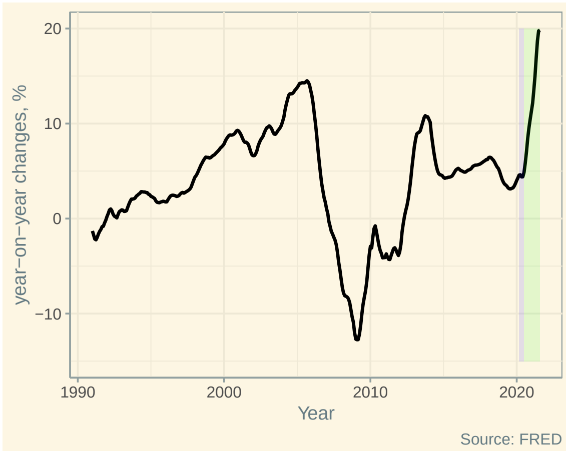


Figure 1: Case-Shiller Index in the U.S.

Figure 2: Housing Price Growth by Income

Articles and some papers point out many possible factors behind this growth[2]. (i) Policy actions played significant role. The Federal Reserve took a broad array of actions to limit the economic damage from the pandemic. They cut the interest rate considerably which lowered the cost of borrowing on mortgages, home equity loans, and other loans. The U.S. government also responded with several vast amounts of fiscal stimulus to the economy and relief including stimulus checks and unemployment benefit. (ii) The housing market has tightened considerably. Supply of homes for sale has fallen to historically low levels. Mortgage forbearance programs and the foreclosure moratorium might be answers here or people just did not want to sell their home during the pandemic. (iii) During the pandemic, the number of home mortgage loan applications have increased and also acceptance rate has increased too. This is very interesting since mortgage loan is one of the key drivers in the housing market.

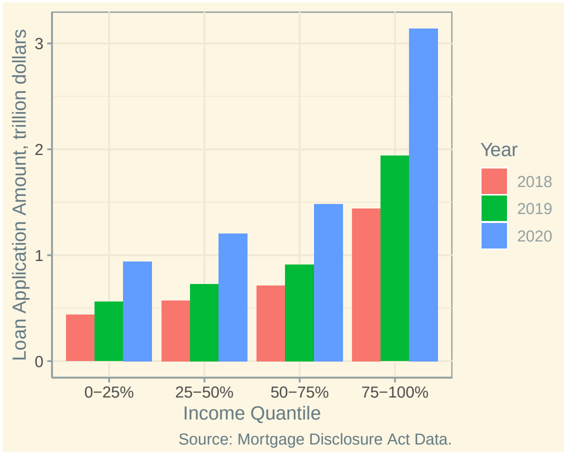
Moreover, mortgage loan is absolute key driving factor on housing market demand side because the purchase of a house usually requires a loan, which means its strong linkage to the lending attitude of banks. In other words, the mortgage loan market is the starting point for home purchases. From this point of view, we will pay particular attention to fact (iii) and analyze the trends in the mortgage loan market before and after the pandemic.

1.2 Home mortgage loan acceptance

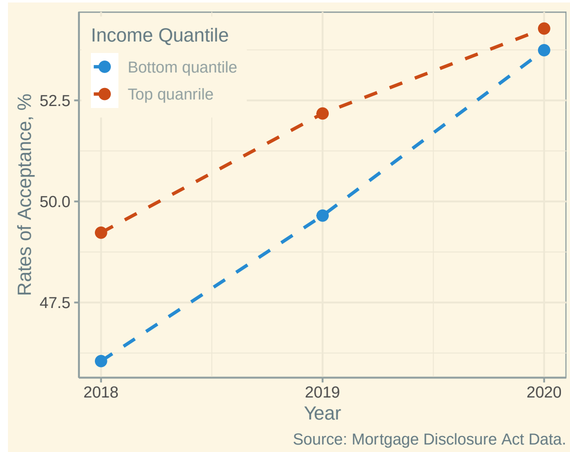
By the numbers, the number of mortgage loan applications has increased to 24.4 million in 2020 from 16.6 million in 2019 and 14.3 million in 2018 (Figure 3a). This is a significant growth. However, a more interesting fact is that banks, lenders accepted more mortgage loans in this period despite millions of more loan applications. The mortgage loan acceptance rate was 52.1% and 54.8% respectively in 2018 and 2019. In 2020, this rate reached 58.1% (Figure 3b).

Our initial assumption was that mortgage lenders would be more hesitant to grant mortgage loans during pandemic due to the uncertain environment and risk factors. However, it was not the case. From mortgage lenders' perspective, lenders usually consider following factors to determine loan acceptance: (i) higher the credit score, the more likely a loan will be approved for a mortgage, (ii) amount of debt relative to income including mortgage payments, (iii) down payment, (iv) employment history, (v) value and condition of the home. For some reasons, mortgage lending behaviour should have been changed and determining factors differed considerably during pandemic. Therefore, we wondered why and how

[2] [2]Balemi et al. (2021) provides a comprehensive literature review of the effect of Covid-19 shock on real estate markets including outside of the U.S.



(a) Application Amount



(b) Acceptance Rates

Figure 3: Loan Application and Acceptance By Income Quantile

mortgage lending behaviour has changed during the pandemic? So we seek out answers to this question in this paper and are primarily interested in the mortgage loan acceptance rate and its factors.

2 Related Literature

Based on the previous described phenomena on the mortgage loan application behaviour, we delve into related literature. The related literature is vast and heterogeneous, so we attempt to narrow down the topics into two main categories: Asymmetric Information and Adverse Selection, and Credit Supply factors.

2.1 Asymmetric Information and Adverse Selection

The first branch of literature revolves around the topic of asymmetric information and adverse selection in the financial system. As mentioned by the work of [3]Clancy (2014), following the seminal work of [4]Akerlof (1970), [5]Rothschild and Stiglitz (1976) and [6]Stiglitz and Weiss (1981), a large theoretical literature has stressed the important role of asymmetric information in financial markets. In financial market, this situation is prone to arise when, ex-ante, lenders have less information than the borrowing counterpart in terms of the creditworthiness (quality of the borrower), or even the borrower's propensity for risk. In such scenarios, fixing the interest rate such that it is commensurate with the average risk of the borrowers generates low-risk borrowers to not enter into transactions with the institutions. In this context, the interest rate serves a dual role: (i) it matches the demand and supply in the credit market, and (ii) it is an instrument that determines the quality of the borrower, and hence, the quality of the bank's loan portfolio ([7]Besley, 1994). Aware of this situation, the banks can then resort to credit-rationing rather than raising interest rates ([6]Stiglitz and Weiss, 1981). More generally, we can also encounter situations of risk miss-pricing and, in the extremes, market collapse.

Regarding the existing literature, we find both sides of the spectrum in terms of approach. On one hand, in a more case-study approach [3]Clancy (2014) examined how the adverse selection problem was the underlying factor behind the failure of New Bank Tokyo (a Government launched bank with the objective of providing finance to the region's small businesses), along with the introduction of borrower's reputation risk in a context of a single lending institution. On the other hand, [8]Marquez (2002) develops a fully theoretical model to describe the implications of competition in the adverse selection problem in terms of information dispersion. The fully theoretical model illustrates that information asymmetries help determine a bank's profit and its ability to enter new market, and that these effect is compounded as competition across banks intensifies. In particular, as information asymmetries are reduced, both the possibility of realizing positive profits on new loans and of entering an existing market are increased.

On a more empirical approach, the work of [9]Crawford, Pavanini and Schivardi (2018) should be highlighted. The authors study the interaction between imperfect competition and asymmetric information in the Italian market for small businesses lines of credit using a rich dataset with information about credit contracts between firms and banks. A structural model of firm's demand for credit is estimated, along with loan use and default, and these are joint with a model of bank pricing to individual firms. The authors find evidence of adverse selection in the form of a positive correlation between unobservables determining both demand and default and loan use and default. The authors then focus on a counterfactual analysis which shows that an increase in adverse selection and in the banks' cost of capital causes prices to increase, demand and loan use to fall, and defaults to rise. Nonetheless, the authors remark that higher market power moderate these effects. Another relevant piece of work is carried out by [10]Ambrose, Conklin and Yoshida (2016), which study credit rationing, income exaggeration and adverse selection in the mortgage market, motivated by the 2009 Global Financial Crisis. The authors examine the role of borrower concerns about future credit availability in mitigating the effects of adverse selection and income missrepresentation in the mortgage market. Using a national data set of subprime mortgages originated by a major financial institution during the housing price boom, the authors document the following regarding borrower income misrepresentation: (i) income misrepresentation is concentrated among borrowers who originated low-documentation loans but could have originated full-documentation loans instead, (ii) the majority of additional risk associated with low-doc mortgages was due to adverse selection and income falsification on the part of the borrowers with verifiable information, and (iii) lenders took action designed to counter potential borrower income falsification. All together, the authors empirical result suggest that borrowers concerns about future credit access can mitigate the effects of adverse selection in limited-information documentation mortgage contracts.

Finally, within the topic of adverse selection we note the work of [11]Weill and Godlewski (2009), which study collateral's and adverse selection in transitioning countries. The authors explore whether collateral's help solving adverse selection problems in transition countries. Even though the research is carried on a more macroeconomic approach, dealing with countries, the findings provide insights about the role of collateral's and risk premiums. In particular, using data on approximately 400 loans in 16 transitional countries, the authors find a positive relationship between the presence of collateral and risk premiums, which conflicts with the argument that collateral is a signalling instrument. However, the finding is in accordance with the observed risk hypothesis, according to which banks can sort borrowers by information they have on quality.

In the context of our research, we will not focus on the effects of the coronavirus pandemic in the information asymmetry between borrowers and lenders. In such sense, we then move to the topic labelled credit supply.

2.2 Credit Supply

The second branch of literature is a collection of papers that explore, in a general sense, the determinants of credit supply and, by contrast, are not specifically related to asymmetric information and adverse selection.

The first work to be mentioned is that of [12]Megbolugbe and Cho (1993), which provides an analysis of mortgage lending patterns for various policy-relevant geographic areas (i.e. central cities and areas with high concentration of minorities population) and borrower groups (such as low income or minority households).The authors estimate models of types of loan origination (FHA, Low Conventional, Medium

Conventional, and High Conventional) on several factors that capture specific effects: housing demand (i.e. homeownership, owner or renter cost, household size, median income and demographics), housing supply (factor costs, local governments, vacancy rate), and mortgage supply (credit availability, bank branches, percent minority). The authors then use a full model as an unrestricted specification, and test for joint significance on the mentioned factors groups to evaluate the overall relevance of the mentioned factors on loan origination. We highlight that only in the High Conventional Loan Origination Regression the authors find a robust effect of income on loan origination. That aside, the authors demonstrate that distinct factors influence the origination of different types of mortgage loans. In particular, the authors remark the robustness of two results: FHA loan origination shows a positive correlation between demographic factors such as proportion of young population and level of household mobility, and median household income is the most significant in both magnitude and statistical significance. This first work bring interesting insights on determinants of loan origination as a whole.

A second paper to be mentioned is that of [13]]Epley and Liano (1999), which focuses on developing a residential supply function for approved fixed-rate mortgages in US commercial banks as a first step to explain differences in origination patterns among groups of borrowers. Specifically, it models the lender's decision to offer the borrower a risk-adjusted loan bundle relative to the terms of the credit application. Concretely, they utilize vectors of individual loan characteristics that compromise the borrower's total risk profile, which include: creditworthiness of the borrower, factors that influence default risk, borrower's personal characteristics, and cost of fund of the lender. Importantly, a canonical correlation factor analysis is used to capture the lender's simultaneous decision. The authors find that the model appropriately predicts that local lenders must adjust concurrently the discount points, contract rate, loan-price ratio and loan maturity to account for the default risk that was uncovered in the loan application. In this paper, income along with other indicators of creditworthiness serve result as controls in the regression model.

Another interesting work is carried out by [14]Roszbach (2004), which actually turns into the use of credit scores as part of the loan evaluation process. Roszbach argues that as credit score models typically are used to minimize default rates or the number of incorrectly classified loans, they fail to take into account the multi-period nature of loans. This is particularly relevant for banks as not only knowing if but also when a loan will default is important. The author here estimates a bivariate tobit model with a variable censoring threshold and a sample selection effects for: (i) the decision to provide a loan or not, and (ii) the survival time of the granted loan. The results of the author suggest that income and owning a house play an important role in obtaining a loan, and that the sample selection bias should be taken into account when analyzing the duration of loans. These results are insightful both in terms of factors to take into account on the loan evaluation process, as well as the potential bias that could be generated should we not include both rejection and acceptances in the analysis.

Two other pieces of work are to be mentioned here, not due to the direct exploration of credit supply determinants, but to the important relations to the topic. First, [15]Murfin (2012) which explores the supply-side determinants of loan strictness. Naturally, the loan contract origination and the strictness of the contract are related topics, even though they can be examined separately. Murfin investigates how lender-specific shocks impact the strictness of the loan contract that a borrower receives. The evidence suggests that defaults inform lender's perception of their own screening ability, thereby impacting their contracting behavior. Second, [16]Peek, Rosengren and Tootbell (2003) focus on a more macroeconomic approach and attempt to identify the macroeconomic effect of a loan supply shock. In this case, we mention the work to note the importance of appropriate understanding of the loan market and how this branches out to even macroeconomic analysis via a credit channel. The authors use a variable that measures bank health (in particular, one that is unknown to the general public and is orthogonal to the state of the economy) in order to more clearly identify a loan supply shock. The authors then identify that the strongest effect among the components of GDP is on business inventories, and to a lesser extent to the other components of GDP, and this in turn reassures that the variable derives its nature as a loan supply shock.

2.3 Hypothesis

Based on the literature review, we note that the scope and interest of this paper is more in line with the "Credit Supply" literature review, as we intend to explore the behavior in the financial system amid the coronavirus pandemic. Despite the interesting nature of potential changes in asymmetric information and its effect on the mortgage market, we will focus more specifically in the determinants of loan origination amid the pandemic. Specifically, we will look at income as one of our main explanatory variables as it more often than not found to be a relevant factor that captures individual characteristics of the borrower.

Specifically, we work with two hypothesis derived from the literature review. First, in general terms, we expect that loan origination declined in 2020 amid the negative effects of the coronavirus pandemic on economic agents. The idea here is one revolving around risk: the pandemic starts to evolve, curfews start to be implemented as a response in order to mitigate contagion, the curfews generate a contraction of economic activity with negative effects on economic agents (loss of jobs, disruption of production, among others), which generates an environment less prone for bank lending amid higher systematic risk. Importantly, we acknowledge that a policy response exists, but we still expect the net effect to be negative given the magnitude of the event. Second, on a more specific note, we expect that the positive effect that income has shown on loan origination is lessened in the context of the coronavirus pandemic. We focus specifically in income given its recurrent mention in the literature, but we include other variables mentioned as controls.

3 Data Description

Given our interest and literature studies, our single best available data is The Home Mortgage Disclosure Act (HMDA). The HMDA data are the most comprehensive source of publicly available information on the U.S. mortgage market. Depository and nondepository institutions report directly to the Federal Financial Institutions Examination Council (FFIEC) which then gets publicly published.

The HMDA is loan-application level data and it includes a total of 48 data points providing information about the applicants, the property securing the loan or proposed to secure the loan in the case of non-originated applications, the transaction, and identifiers. The 2019 and 2020 data include information on respectively 16.6 and 24.4 million home loan applications.

3.1 Data table

We collapsed the loan application level data into the county level data due to convenience. There are 3,002 counties and computed the county average for each indicator. Also, we will use only 2019 and 2020 HMDA data.

We are clearly interested in acceptance rate and median income. However, we also added some more variables such as loan amount, loan term, loan to value ratio, 1st lien ratio and also refinance ratio as shown in our data table.

4 Model & Methodology

Taking into account the literature review, we look at [10]Ambrose, Concklin & Yoshida (2016) and the several hypothesis that the authors explored. In particular, we look at the loan application rejection model (probit model) that the authors explored. A particular note is that the variables included in the

Characteristic	**2019**[1] N = 3,002	**2020**[1] N = 3,002	**p-value**[2]
acceptance rate	53 (7)	56 (8)	<0.001
# of application	4,588 (11,981)	6,749 (18,284)	<0.001
loan amount	1,093 (3,758)	1,789 (6,293)	<0.001
median income	68 (19)	74 (20)	<0.001
loan term	314 (18)	309 (16)	<0.001
Loan to value	102.1 (609.3)	83.1 (225.8)	0.11
1st lien ratio	91.0 (4.9)	94.9 (3.0)	<0.001
refinance ratio	35 (8)	49 (11)	<0.001

[1]Mean (SD)
[2]Welch Two Sample t-test

Table 1: Summary Statistics for Our Main Data

model are only those available to the lender at the time of the reject decision. The model controls for loan, property, borrower and area characteristics, interest rate environment, and includes Metropolitan Statistical Area Fixed Effects. For our research, we take this model estimated by the authors, and adapt it to out data availability.

Specifically, we run a fixed-effect panel-data regression as follows:

$$y_{cst} = \beta_0 + \beta_1 \ln(Income_{cst}) + \beta_2 D_t^{2020} + \beta_3 \ln(Income_{cst}) D_t^{2020} + \alpha c_{cst} + FE_s + \varepsilon_{cst},$$

where y_{cst} is the average loan acceptance rate per county as a percentage, $\ln(income_{cst})$ is the average income of the county in logarithmic terms, D_t is a dummy variable which takes the value of 1 for 2020 and 0 for 2019. c_{cst} are a set of control variables for our regression, which includes other factors taken into account by the lending institutions, and FE_s are state-level fixed effects to control for heterogeneous characteristics of the financial system across depths.

Importantly, we focus on the income variable as this is regularly mentioned in the literature review among one of the main variables taken into account by financial institutions when evaluation the loan application. We are particularly interested in the sign of the β_3 coefficient, as it represents the change in the sensitivity of loan acceptance to income between 2020 and 2019. In addition, the β_2 coefficient will also give us an estimate of the effect of 2020 (in our model assumed to be mostly capturing the pandemic effect) on loan acceptances.

5 Results

5.1 Baseline Model

Table 2 shows results of our baseline regressions to detect the change in the relationship between loan acceptance rates and income by county. Note that the specifications in column (2) through (5) include state-level fixed effects as well as some control variables, such as average loan to value, debt to income, and the ratio of refinancing needs in loan applications by county. This is because of the idea that the level of each variable is expected to vary from state to state, especially in 2020, due to the adoption of different policies (ex. lockdown policies) depending on the infection situation[3].

We can see some results consistent with our hypothesis. First, the coefficients for income are positive and significant in all specifications, which suggests that higher income is correlated with higher acceptance

[3]The meaning and validity of including state-level fixed effects are discussed in more detail in the Appendix.

Dependent Variable:	acceptance rate, %				
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
(Intercept)	17.66***				
	(4.839)				
log(income)	7.918***	6.903***	6.902***	6.399***	5.217***
	(1.069)	(1.137)	(1.137)	(1.106)	(0.8654)
2020 dummy	30.91***	19.78***	19.78***	17.78***	14.97***
	(5.439)	(6.142)	(6.144)	(6.026)	(4.578)
log(income) $\times$ 2020 dummy	-6.467***	-4.142***	-4.141***	-3.667***	-3.940***
	(1.093)	(1.305)	(1.305)	(1.276)	(0.9707)
Loan to Value, %			-2.38×10^{-5}	-2.36×10^{-5}	1.18×10^{-6}
			(6.96×10^{-5})	(6.97×10^{-5})	(6.94×10^{-5})
Debt to Income, %				2.36×10^{-10}**	-7.11×10^{-12}
				(9.08×10^{-11})	(5.5×10^{-11})
Composition of refinance, %					0.3570***
					(0.0375)
State Fixed Effect	No	Yes	Yes	Yes	Yes
Fit statistics					
Observations	6,004	6,004	6,004	6,004	6,004
R^2	0.10385	0.47556	0.47556	0.47910	0.55110
Within R^2		0.15609	0.15610	0.16179	0.27764

Clustered (state code) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 2: Results for Our Baseline Model

rates in general. This result is consistent with most of the findings in the literature review, where income is more often than not found to be positively related with loan acceptance as a measure of borrowers creditworthiness. However, the coefficients of cross terms are negative, which means coefficients of income are lower in 2020 than that in 2019, although they are still positive after adding up the two coefficients. This finding would be supporting evidence for our "specific" hypothesis that the effect of variables, especially income, on the loan approval process has changed before and after the pandemic. Note that these results are robust through (2) to (5), even after controlling for state-level fixed effect and other control variables. For the "general" hypothesis, however, coefficients of the 2020 dummy variable are positive, suggesting an upward level shift in acceptance rates. This result goes against our general hypothesis, as we would have initially expected a more restrictive environment for loan origination amid a decline of economic activity and higher systematic risk. One potential explanation for this result is that our Dummy variable for 2020 jointly captures both the effect of the pandemic as well as the overall economic policy response carried out by policymakers (i.e. unprecedented expansionary monetary policy response, multiple stimulus bills, regulatory forbearance, etc.). Separating these effects could give us more clear insight of the real effect of the pandemic on loan acceptances.

5.2 Validity of Our Baseline Model

We have interpreted the results of the baseline analysis, but the validity of these formulations should be tested separately. Points to be tested include, for example, the existence of multicollinearity and heteroskedasticity. In this subsection, these two points will be formally verified.

Table 3 shows variance inflation factors (VIFs) for our specification. Here, VIFs are small in all the specifications, which implies that the problem of multicollinearity is not serious in our model. On the other hand, BP tests reject the null hypothesis of homoskedasticity in Table 4. Note that because of this reason, we used state-level clustered standard errors to report the results of our baseline model in Table 2.

	VIFs
(1)	1.12
(2)	1.91
(3)	1.91
(4)	1.92
(5)	2.23

Table 3: Multicollinearity: VIFs

	p-values
(1)	< 0.001
(2)	< 0.001
(3)	< 0.001
(4)	< 0.001
(5)	< 0.001

Table 4: Heteroskedasticity: BP tests

5.3 Home Purchase Purpose Loan As a Robustness Check

We've seen results for the baseline model seem at least partly consistent with our hypothesis. However, we would need some robustness checks to preclude the possibility that other movements are driving the results we have seen. Specifically, in 2020, refinancing needs are much higher than in a normal year due to low interest rates, and it is possible that higher acceptance rates are driven by changes in loan characteristics such as purpose of the loan. In other words, it is intuitively possible that refinance application is more likely to be accepted than completely new applications due to less asymmetric information between banks and applicants. We need to make sure that such a kind of change in loan attributes is not the main driver of this movement. From this point of view, we excluded loan applications for refinancing purposes from our main dataset and narrowed the sample to loan applications just for home purchases.

Table 5 derived from this alternative dataset also shows similar results to the baseline regressions both in significance and the magnitude of the coefficients. It seems that the change in the determinants of acceptance rate is not driven by the change in the loan attributes.

6 Conclusion and Potential Future Topics

In this analysis, we have seen that higher income is correlated with higher mortgage loan acceptance rate in general. This relationship, however, fades in 2020 amid the coronavirus pandemic, but it remains positive in net terms. Our model also suggests a level shift in 2020 acceptance rates, although it is difficult to directly specify the cause of this movement from our model.

Using the relationships we found in our analysis as a starting point, there should be some topics to be discussed in potential future research. These are the points we did not cover in this analysis, mainly because of data limitations (or because the methodology is too far away from what we learned in this semester). To start with, for example, while we have seen changes in the relationship between income and acceptance rates, we have not been able to identify the direct factors that have led to these changes. One possible hypothesis for this is that the various fiscal policies implemented in 2020 have focused primarily on helping low-income people, which may have changed the relationship between income and acceptance rates. To illustrate, the stimulus package, a prime example of fiscal policy conducted in 2020, provides a fixed amount of benefits for multiple times to individuals who fall below certain thresholds based on marital and work status, totaling more than $850 billion. If banks had factored such redistributive fiscal

Dependent Variable:	acceptance rate, %			
Model:	(1)	(2)	(3)	(4)
Variables				
(Intercept)	14.45*			
	(8.110)			
log(income)	9.681***	7.892***	7.891***	7.462***
	(1.771)	(1.599)	(1.601)	(1.549)
2020 dummy	34.56***	21.51***	21.50***	19.69***
	(6.045)	(6.119)	(6.125)	(5.994)
log(income) $\times$ 2020 dummy	-7.917***	-5.130***	-5.129***	-4.692***
	(1.366)	(1.385)	(1.386)	(1.349)
Loan to value			-1.15×10^{-5}	-1.37×10^{-5}
			(3.53×10^{-5})	(3.5×10^{-5})
Debt to income				3.23×10^{-10}**
				(1.36×10^{-10})
State fixed effect	No	Yes	Yes	Yes
Fit statistics				
Observations	5,682	5,682	5,682	5,682
R^2	0.08449	0.48398	0.48399	0.48527
Within R^2		0.10992	0.10992	0.11214

Clustered (state_code) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 5: Results for Purchase Purpose Loan Application

policies for a period of time into their loan acceptance decisions, this might explain the lower correlation with income in 2020.

There are several possible ways to verify this point, depending on the availability of data. The first is to wait for the accumulation of data and examine the dynamics of changes in the relationship between income and acceptance rates. If these changes in the relationship are temporary, coinciding in time with large-scale fiscal policy expectations, this would be a corollary to support the hypothesis. A more direct means of examination would be using loan-level data. The HMDA data we have used so far originally records loan-level application information with detailed attribute information of the applicant. The idea of identification here is as follows: if fiscal policy is influencing banks' acceptance decisions, then there should be discontinuous changes in acceptance rates between individuals with annual incomes slightly above and slightly below the threshold level defined in the stimulus package. Then, we can use these differences in the context of difference-in-difference and regression discontinuity. Unfortunately, elaborate identification of whether or not an individual is eligible for the stimulus package requires data such as each individual's marital status and spouse's employment status. Unfortunately, the current version of HMDA data lacks this information, making it difficult to accurately assign treatment control groups. When these data limitations are resolved in the future, it will be possible to conduct more sophisticated analyses within the framework of causal inference as described above.

On this topic of taking into account the monetary policy response, we should also note the effect of the unprecedented monetary policy response on the financial system's liquidity, and the macro-prudential response regarding regulatory forbearance. These factors, amid data availability, could also help to narrow down the specific effects of the pandemic from the economic policy response.

In this context of policy response, one straightforward analysis that should be carried out is extending these evaluation to include HMDA 2021 data. We have carried out a discrete analysis in terms of the effects evaluated. However, it would be insightful to extend the evaluation in order to test whether this findings are deemed temporary or permanent in terms of the banking behaviour, which is translated into the loan acceptance rate. It should be noted that 2021 is a year in which the pandemic is mostly present, so close attention should be payed to its potential lasting effects across different sectors.

Directly related to the above mentioned further research topic is that of extending data availability in order to account for more bank-specific characteristics. Literature reviewed reveals how bank characteristics, which in practice refer to bank performance indicators which can be thought of as directly related to funding costs, are also a relevant factor to consider among credit supply. The HMDA database does not include such bank-specific indicators in order to properly characterize a supply function, so extending the database to account for this factors could provide other interesting insights when analyzing the effect of the pandemic on overall loan origination.

Another direction the analysis could go is to reconnect this exploration of mortgage loan with the housing price literature. This is a natural idea when considered in conjunction with the fact that housing price increases are more pronounced in low-income areas, as we saw in Section 1 (and Appendix 1 for a more detailed analysis). In moving in this direction, a major theme of the analysis will be to identify which of the rise in house prices and the change in loan supply was the fundamental cause of recent movements. There are two possible explanations for the observed facts. That is, (1) rising housing prices (or expectations of rising prices) in low-income areas may have led to an increase in the loan acceptance rate by increasing the value of properties as collateral, or (2) an increase in the loan supply due to an increase in the loan acceptance rate may have led to a rise in housing prices in low-income areas due to demand factors. This means we need to pay enough attention to the problem of reverse causality. In this context, [17]Mian and Sufi(2009), which analyzed the relationship between subprime loans and housing prices during the housing boom before the global financial crisis, is well known and their idea would be applicable to ours. Precisely speaking, when comparing a region with a series of flat lands and an elastic housing supply (e.g., Indiana State) to a region where housing supply is inelastic because it faces the sea or mountains (e.g., near San Francisco), if we observe an increase in loan acceptance rate even in a region where the housing price increases are relatively suppressed by increased supply, we may say that the increase is due to loan growth.

In the context of reconnecting to housing price literature, one interesting route of exploration would be the heterogeneous behaviour between low-income and high-income households across both the loan-acceptance and housing price market. A descriptive look at the data, such as that in figures 2 and 3 provide an interesting insight on the different behaviour between income levels which would be interesting to explore on a more micro-level data analysis.

References

[1] Yunhui Zhao. Us housing market during Covid-19: aggregate and distributional evidence. *IMF Working Paper Series*, 2020.

[2] Nadia Balemi, Roland Fuss, and Alois Weigand. COVID-19's impact on real estate markets: review and outlook. *Financial Markets and Portfolio Management*, 2021.

[3] Brett Clancy. Lending to 'lemons': Adverse selection and the failure of new bank tokyo. *Social Science Japan Journal*, 17(2):155–167, 2014.

[4] George A Akerlof. The market for" lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, 1970.

[5] Michael Rothschild and Joseph Stiglitz. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, 90(4):629–649, 1976.

[6] Joseph E Stiglitz and Andrew Weiss. Credit rationing in markets with imperfect information. *The American economic review*, 71(3):393–410, 1981.

[7] Timothy Besley. How do market failures justify interventions in rural credit markets? *The World Bank Research Observer*, 9(1):27–47, 1994.

[8] Robert Marquez. Competition, adverse selection, and information dispersion in the banking industry. *The Review of Financial Studies*, 15(3):901–926, 2002.

[9] Gregory S Crawford, Nicola Pavanini, and Fabiano Schivardi. Asymmetric information and imperfect competition in lending markets. *American Economic Review*, 108(7):1659–1701, 2018.

[10] Brent W Ambrose, James Conklin, and Jiro Yoshida. Credit rationing, income exaggeration, and adverse selection in the mortgage market. *The Journal of Finance*, 71(6):2637–2686, 2016.

[11] Laurent Weill and Christophe J Godlewski. Collateral and adverse selection in transition countries. *Eastern European Economics*, 47(1):29–40, 2009.

[12] Isaac F Megbolugbe and Man Cho. An empirical analysis of metropolitan housing and mortgage markets. *Journal of Housing Research*, pages 191–224, 1993.

[13] Donald R Epley and Kartono Liano. The residential mortgage supply function of commercial banks. *Urban Studies*, 36(11):1959–1971, 1999.

[14] Kasper Roszbach. Bank lending policy, credit scoring, and the survival of loans. *Review of Economics and Statistics*, 86(4):946–958, 2004.

[15] Justin Murfin. The supply-side determinants of loan contract strictness. *The Journal of Finance*, 67(5):1565–1601, 2012.

[16] Joe Peek, Eric S Rosengren, and Geoffrey MB Tootell. Identifying the macroeconomic effect of loan supply shocks. *Journal of Money, Credit and Banking*, pages 931–946, 2003.

[17] Atif Mian and Amir Sufi. The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis. *The Quarterly journal of economics*, 124(4):1449–1496, 2009.

Appendix 1: Housing Price and Neighborhood Income

As discussed in Section 1, housing price growth in 2020 is concentrated in low-income areas. Here we confirm that this relationship is robust even after controlling for other factors such as growth in housing supply and demand. We make use of three datasets. First, for the housing price, we use Federal Housing Finance Agency data[4], which is available from 1995/1Q to 2021/2Q covering house prices in every quarter by three-digit zip code. For other factors in the housing market, we use data from Realtor[5], which contains monthly supply scores calculated from new listings, demand scores based on web views, and market hotness indicators as the balance of demand and supply by five-digit zip code from July 2017 to September 2021. The other data source is the SOI Tax Stat[6], which is an official annual statistics recording household income by 5-digit zip codes. We examine the relationship between household income and housing prices by collating this data at the 3-digit level to create the main data set. Note that during this procedure, 5-digit level data is aggregated to 3-digit level data, which leaves us 605 zip codes per quarter in our final dataset (Table A1)[7].

	variables	Mean	Std. Dev.	Min	Max	Sample size
1	annual house price growth, %	6.11	1.86	0.59	11.10	605
2	median square feet	1921.49	435.61	20.85	3498.64	605
3	demand score	44.71	21.00	1.69	96.58	605
4	supply score	57.58	15.25	1.33	90.24	605
5	hotness score	51.15	16.08	2.56	89.84	605
6	annual household income, $1,000	72.88	24.09	37.20	254.87	605

Data as at 2020 Q4.

Table A1: Summary Statistics

In Figure 2, we found a negative correlation between household income and the growth rate of housing prices from this dataset, and we will use regression analysis to confirm that this relationship is robust even after controlling for other factors. Table A2 shows the regression results for the relationship between house price gwoth and household income. Here, the dependent variable is the growth rate of housing prices. In all formulations, the household income level and housing price growth show a significant negative relationship, even after controlling for supply and demand factors[8]. This suggests that what we have identified in the scatter plot is not due to other factors.

Appendix 2: Validity of State-Level Fixed Effects

In Table 2, we included state-level fixed effect in the regression. The intuitive meaning of including this fixed effect has been discussed in Section 5.1, but here we discuss its econometric meaning and relevance.

As discussed in Section 4, our model setting can be described as

$$y_{cst} = X_{cst}\beta + \sum_{s\in S} D_s\alpha + \varepsilon_{cst},$$

[4]This can be downloaded from `https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index.aspx`.

[5]This data is available at `https://www.realtor.com/research/data/`.

[6]Data available at `https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi`.

[7]The latest data for SOI data is 2019 and we merge 2019 housing income data for those after 2019.

[8]Demand shows significant positive correlation with housing price, consistent with economic theory, although the effect of supply is not significant here.

Dependent Variable:	annual house price growth, %				
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
(Intercept)	9.311***		-0.1000		
	(2.418)		(0.8058)		
log(household income)	-2.188***	-2.297***			
	(0.5080)	(0.4748)			
log(supply score)	0.3768	0.0929	0.2780	0.0142	0.0810
	(0.2705)	(0.1640)	(0.2619)	(0.1682)	(0.1626)
log(demand score)	1.261***	0.9338***	1.335***	1.036***	0.9915***
	(0.1419)	(0.1384)	(0.1522)	(0.1231)	(0.1281)
0-25% income quantile dummy			0.8307***	0.7737***	1.130***
			(0.1888)	(0.1752)	(0.2559)
25-50% income quantile dummy					0.6615***
					(0.2279)
50-75% income quantile dummy					0.4597***
					(0.1199)
state fixed effects	No	Yes	No	Yes	Yes
Fit statistics					
Observations	605	605	605	605	605
R^2	0.32862	0.61372	0.27243	0.56157	0.57812
Within R^2		0.28416		0.18751	0.21819

Clustered (state) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table A2: Household Income and House Price Growth where X_{cst} includes all explanatory variables except for state-level fixed effect (such as income and 2020 dummy), S is the set of states in the U.S.[9], and D_s is the dummy variable indicating 1 when county c belongs to state s. From Frisch-Waugh theorem, estimating the vector of coefficients β from this formulation is equivalent to the following procedure. First, regress the state-level dummy variables on both dependent and independent variables, that is,

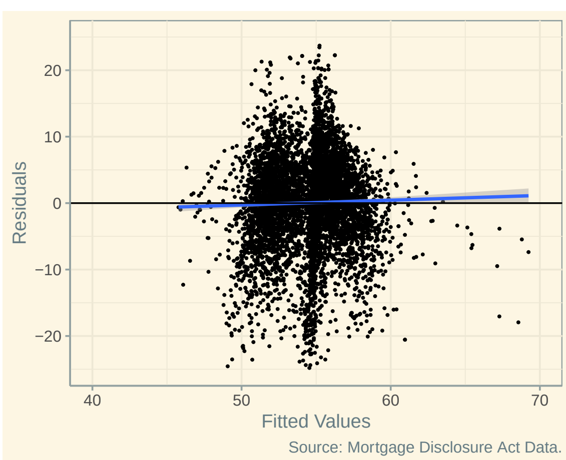
$$y_{cst} = \alpha_0^y + \sum_{s \in S} \alpha_s^y D_s + u_{cst}$$
$$x_{cst} = \alpha_0^x + \sum_{s \in S} \alpha_s^x D_s + v_{cst}.$$

And regress the residuals obtained from the first step as follows.

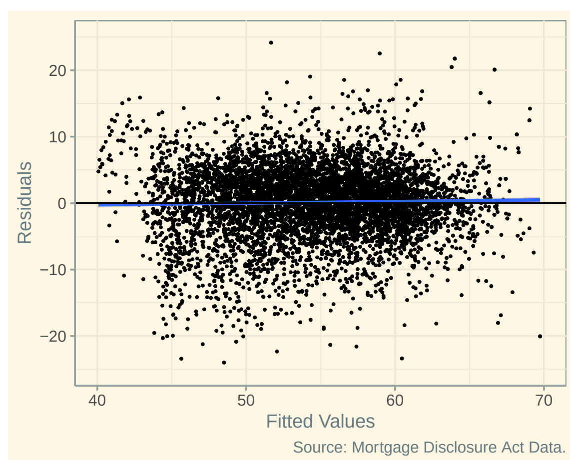
$$\tilde{y}_{cst} = \tilde{X}_{cst}\beta + \epsilon_{cst}$$

The coefficients α_s^y and α_s^x in the first step represent the mean of y and x in each county s, respectively. Residuals $\tilde{y}_{cst}$ and $\tilde{X}_{cst}$ obtained from the first step are, therefore, equal to the deviation from state-level mean. This implies that including the state-level fixed effect in the regression is equivalent to converting both the dependent and independent variables into deviations from the mean according to a certain attribute.

Now, we examine which formulation is actually more likely to be correct when fixed effects are included or not included in our case. Figure A1 shows the relationship between the residuals and the fitted values with and without the fixed effects, respectively (correspond to (1) and (2) of the regression model). The figure on the left suggests that the residuals are clustered around particular levels, indicating that this formulation does not fully eliminate differences in the level of particular groups. The right one, on the other hand, looks random and is closer to the white noise. This suggests that regression model with fixed effects is more appropriate in our analysis.



(a) Without fixed effect



(b) With fixed effect

Figure A1: Fitted Values and Residuals in the Baseline Model

[9]Strictly speaking, S has to exclude a state to avoid the problem of so-called dummy variable trap.