

An analysis of the determinants of loan demand after the Great Recession

Joshua D. Ingram *New College of Florida*

The determinants of the demand for loans by individuals with a reported annual income of under \$200,000 in 2008 and 2019 are compared in this paper using the LendingClub loan data. The goal of this paper is to determine the effects of several individual-level variables on a borrower's demand for loans and to observe any significant changes in these effects during the Great Recession in 2008 and in 2019 before the Great Lockdown. We use the two-stage least squares method to avoid endogeneity bias and compare the significance, direction, and magnitude of the estimated coefficients to determine any fundamental changes. There is a need to use an alternative method to estimate the demand for loans if all levels of annual income are to be included, as the OLS assumptions are broken when including influential observations that have extremely high incomes. Our results suggest that the demand curve takes on a non-linear form. Interest rate has a negative relationship with demand, whereas income has a quadratic relationship with the demand for loans. Due to instability in the 2019 model, we can not fully confirm a strong difference in the effects on the demand. However, we observe an obvious change in the structure of the data between 2008 and 2019 that has an effect on the market for loans.

Introduction

Understanding how the market for consumer loans functions is essential for lenders to understand how they should price and market their services. By estimating the demand function, we can further understand the behavior of loan seekers and lenders can adjust their marketing campaigns accordingly. One such step to understanding this behavior is to observe if the market for loans changed in the U.S. after the Great Recession. LendingClub is a peer-to-peer lending platform and provides demographic data on the borrowers and information on every loan on the platform since its beginning in 2007. Having access to this data enables us to estimate a demand curve, as well as to observe any changes after 2007.

For the purposes of this project, two years are selected from the data, 2008 and 2019. These subsets are treated as cross-sectional data. In order to estimate the demand for loans, the two-stage least squares method is used to avoid endogeneity bias introduced by the interest rate as a determinant of loan demand. After the attempt to remedy several model assumptions being broken, we find that the demand curve takes on a non-linear trend in 2008. We find that interest rate is negatively associated with the demand for loans, as expected. Additionally, income of a borrower has a quadratic relationship with the demand for loans. In 2019, the models used begin to fail due to the assumption of homoscedasticity being broken. However, there is a noticeable structural change in the data and which could indicate a change in the behavior of the market after 2008 or in the way LendingClub operates. We continued to compare the models between the two years, but with caution. We found that the demand curves took on the same form, as well as the relationships with interest rate

and income. Although, the effect sizes were slightly different. Finally, we found that several demographic variables were significant in predicting demand in 2019 and not in 2008, such as region, employment length, and home ownership status. However, the comparisons of our models need to be taken lightly due to the 2019 model not being reliable. There is more work to be done, as moving forward would require models that take into account extremely high income individuals and can address the changes in the data in 2019.

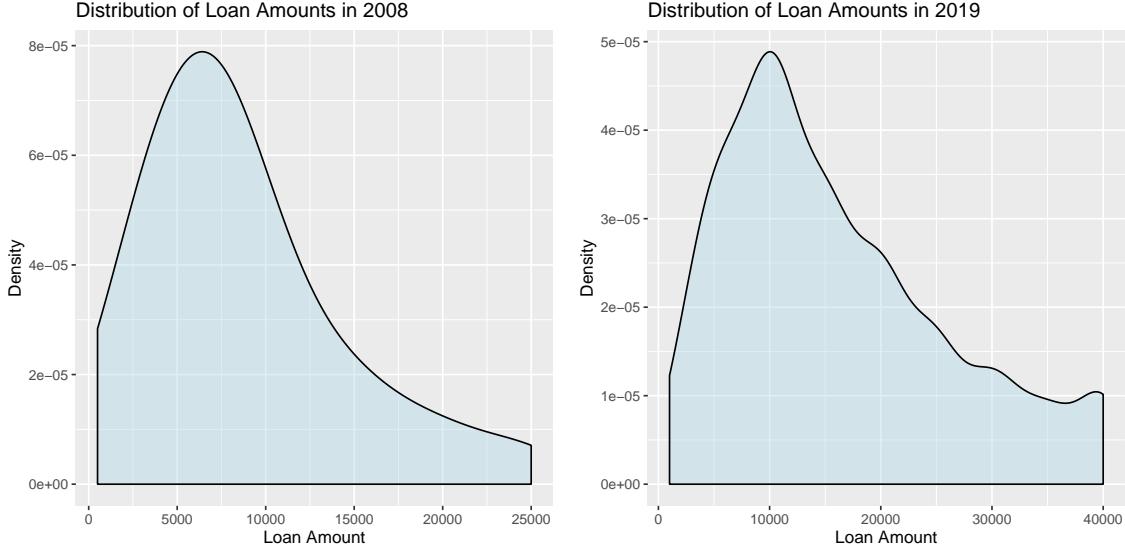
LendingClub and the Data

LendingClub

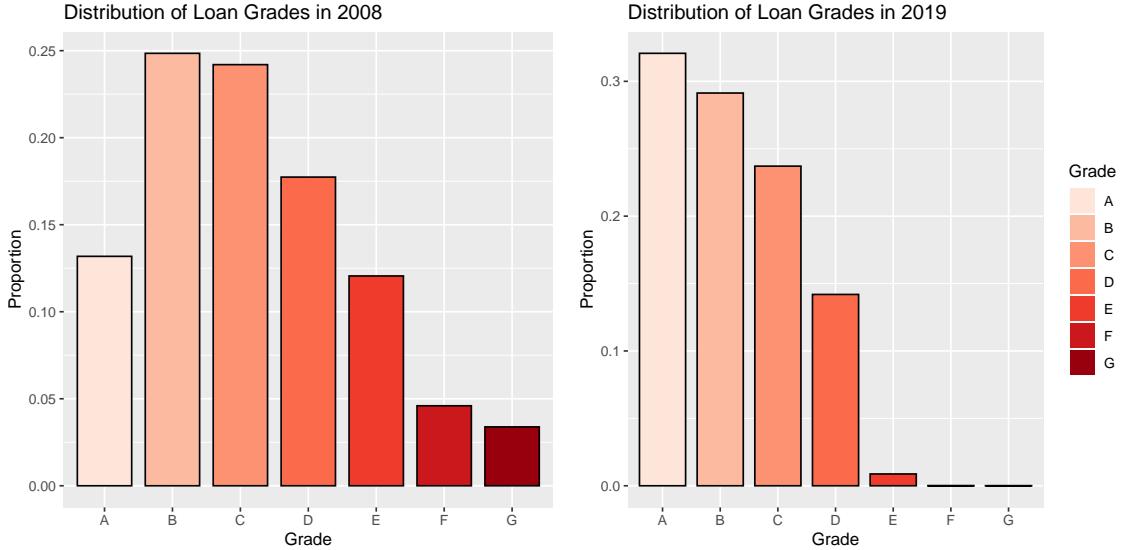
LendingClub is the world's largest peer-to-peer lending platform that launched as one of Facebook's first applications in 2007 (The Economist 2020). LendingClub has proven itself to be rather successful, with over 2.8 million loans listed on the platform since its beginning and over \$650 million in revenue in 2019 (Yahoo! Finance 2020). The platform acts as a market where lenders and loan-seekers come together to make transactions directly, entirely removing the need for banks to be the intermediaries. This generally gives borrowers the opportunity to find lower interest rates on a loan than what they would pay elsewhere. Users on the platform clearly take advantage of this opportunity because according to LendingClub, 67% of borrowers on the platform use these loans to refinance existing loans or to pay off their credit cards as of March 2020 (LendingClub 2020).

Currently, borrowers are able to request a maximum loan of \$40,000, but the limit was \$20,000 up until 2016. When borrowers decide to request a loan, they supply several details about themselves and the loan. These details include their FICO credit score, reason for the loan, employment information, and many more details. Once these details are received, LendingClub provides a loan grade and subgrade that determines the interest rate to be paid and that acts as a quick way for lenders to evaluate the riskiness of the loan. Once things are processed, the loan is listed on the platform and lenders are able to look at the details provided by the borrower to determine if they would like to fund the loan.

Below are graphs of the density of loan amounts in 2008 and 2019. It can be seen that most of the loans were around \$6,000 in 2008 and \$10,000 in 2019. The average loan amount in 2008 was \$8,740, but in 2019 it nearly doubled to \$16,187. This is likely due to the fact that borrowers have access to more funds on the platform after 2016.



The distribution of loan grades have noticeably changed between 2008 and 2019, seen below. The majority of loans in 2008 were granted loan grades of B, whereas in 2019 most loans were given a grade of A. Very few loans were within the E-G range in 2019, whereas a good proportion of loans could be found in this range in 2008. This change in the distribution of grades granted by LendingClub could be caused by a number of factors, perhaps a change in the way loan grades are determined by LendingClub.



The Data

The data used for the analysis performed in this paper was obtained directly from the LendingClub website. It contains details provided by the borrower and by LendingClub on every single loan listed on the platform since 2007. The raw data is comprised of 21 different datasets separated by year or quarter beginning in 2007 and ending in the first quarter of 2020. Once combined, there were 2.88 million loans with 150 variables for each loan, as well as an accompanying [data dictionary](#). Before any data cleaning, 23.74% of the data was

missing. Once the data was subsetted to include only our variables of interest, 0.3% of the data was missing. Due to the size of the data and the small percentage of missing values, any rows containing an NA value were outright removed. Two subsets of the data were made to be used in our analysis, one being a subset of all loans in 2008 and another of all the loans in 2019.

Additionally, any observations with reported annual incomes over \$200,000 were removed from the data due to these being a small number of cases that had incredible effects on the assumptions for the models used. There were also three new categorical variables created from the existing data. One being “region”, which is the region of where the address given by the borrower is located, another being “emp_length_cat”, which indicates which of the several employment length categories the applicant falls into. Finally, “real_estate_col” was created and it is a binary variable that indicates whether the borrower has some form of real estate collateral or not.

Table 1: Relevant Variable Definitions

Variables	Definitions
annual_inc	The self-reported annual income of the borrower
dti	Debt-to-income ratio of the borrower
emp_length_cat	Indicates the borrower’s category for length of employment
fico_range_low	The lower range of the borrower’s FICO credit score
home_ownership	The home ownership status provided by the borrower
int_rate	The interest rate on the loan
loan_amnt	The listed amount of the loan applied for by the borrower
real_estate_col	Indicates whether the borrower has any real estate collateral
region	The region of the address provided by the borrower

Theoretical Models

The framework used to estimate the models in this analysis is based on several papers studying the determinants of the demand and supply of loans, as well as the assumption that this market is in equilibrium. We first define our demand function and supply function. This is followed by redefining the interest rate in a such a way that we can avoid any endogeneity bias since the interest rate and loan amount are determined simultaneously. Without redefining interest rate, it has a non-zero covariance with the error term and breaks one of our model assumptions. Once we find the newly defined interest rate, this new form is replaced in the demand and supply functions.

The model used for the demand for loans is defined as:

$$Q_D = r\alpha_1 + X_1\alpha_2 + \epsilon_1$$

Where Q_D is the demanded loan amount, r is the interest rate of the loan, and X_1 is a vector of explanatory variables affecting the demand for loans. ϵ_1 is the i.i.d error term that follows a normal distribution centered at 0. We further define our model for the supply of loans as:

$$Q_S = r\beta_1 + X_2\beta_2 + \epsilon_2$$

Where Q_S is the supplied loan amount, r is the interest rate of the loan, and X_2 is a vector of explanatory variables affecting the supply of loans. Again, ϵ_2 is the i.i.d error term that follows a normal distribution centered at 0. Since we assume our market is at equilibrium, Q_D and Q_S are equal, as well as the interest rates. However, interest rate and the loan amount are determined simultaneously, meaning r is an endogenous explanatory variable. To avoid any bias, we replace r with an estimated \hat{r} . This defines \hat{r} in such a way that it is not correlated with the error term ϵ . \hat{r} is determined by both supply side and demand side explanatory variables:

$$\hat{r} = X_1\delta_1 + X_2\delta_2 + \epsilon_3$$

We then replace r with \hat{r} in our equations to correct for the endogeneity bias. Below is our new demand and supply equations:

$$Q_D = \hat{r}\alpha_1 + X_1\alpha_2 + \epsilon_1$$

$$Q_S = \hat{r}\beta_1 + X_2\beta_2 + \epsilon_2$$

Demand Variables

Now that we have corrected for endogeneity bias, we can focus on the explanatory variables in the models for the demand and supply of loans. First, we will discuss the variables that have effects on the demand for loans. We would expect interest rate to cause the demand for loans to decrease as the interest increases. However, there are several other demographic variables that may have significant effects on the demand for loans. One such variable is the income of borrowers. Empirical research conducted by Manrique and Ojah (2004) used both a household's transitory income and permanent income to predict the probability of a household holding consumer and real-estate debts. They found that the probability of holding consumer or real-estate debt increased as income increased. As an individual's income rises, the available money for that individual to pay off a loan would increase and lenders would be more willing to allocate more funds towards that borrower. In our model for the demand for loans, we would expect the demand for loans to increase as income of the borrower increases.

The employment status of an individual could also act as an indicator for income and income stability, so we would expect a longer length of employment to increase the amount of loans demanded. Suggested by Bernanke and Gertler (1986) and mentioned by Manrique and Ojah (2004), a borrower's ownership of collateral has a positive association with access to credit. For our demand model, we use the home ownership status of the borrower to indicate their ownership of real-estate collateral. This is separated into three categories, whether they own, rent, or have a mortgage. This approach allows more freedom to study the differences in a borrower's status. We would expect the ownership status and mortgage status to increase the demand for loans.

Our last explanatory variable for the demand model is region. There is not necessarily a specific expectation for the effects of the region in the U.S. a borrower may belong to, but it allows for the observation of a difference in levels of demand for loans, if it exists.

Supply Variables

We also need to define several variables that affect the supply for loans in order to find \hat{r} . We expect interest rate to affect the supply of loans, with an increase in supply as the interest increases. Studying the default rates of LendingClub loans, Polena and Regner (2018) found that a borrower’s debt-to-income ratio had a positive association with the default rate of a loan. Lenders would be taking on more risk as a borrower’s debt-to-income ratio increases, so as the borrower’s level of outstanding debt rises, we would expect the supply of loans to decrease.

The credit worthiness of a borrower would also be expected to have an effect on the supply of loans. Represented by the FICO credit score, as a borrower’s credit score increases, we would expect the supply of loans to increase as the lenders have an indicator of a borrower’s ability to pay back debt. As mentioned in the demand variables section, Bernanke and Gertler (1986) find that a borrower’s ownership of collateral has a positive association with their access to credit. With this in mind, whether a borrower owns real-estate collateral is used in the supply model. If a borrower owns collateral, we would expect the supply of loans to increase.

Econometric Models

The variables available to us can be found in Table 1 and will be utilized in estimating the demand models for both 2008 and 2019. We also specify what our supply model is, but only do this for the sake of showing the steps taken to reach the estimated interest rate. In our analysis, we do not estimate the supply function, but only use the supply side variables to estimate interest rate. For the demand model, the response variable will be “loan_amnt” and the explanatory variables will be “int_rate”, “annual_income”, “emp_length_cat”, “region”, and “home_ownership”. Because “emp_length_cat”, “region”, and “home_ownership” are categorical variables, there will be several dummy variables for each possible category. Below is a table of descriptions of the variables and their categories:

Table 2: Demand Model Variables

Variables	Definitions
annual_inc	The self-reported annual income of the borrower
emp_length_cat0-1	Borrower has 0-1 years of employment length
emp_length_cat2-5	Borrower has 2-5 years of employment length
emp_length_cat6-9	Borrower has 6-9 years of employment length
emp_length_cat10+	Borrower has 10 or more years of employment length
emp_length_catunknown	Borrower’s employment length is unknown
home_ownershipRENT	Borrower rents a house
home_ownershipOWN	Borrower owns a house

Variables	Definitions
home_ownershipMORTGAGE	Borrower has a mortgage
int_rate	The interest rate on the loan
loan_amnt	The listed amount of the loan applied for by the borrower
regionnorthwest	Borrower's address is in northwest U.S.
regionmidwest	Borrower's address is in west U.S.
regionnortheast	Borrower's address is in northeast U.S.
regionsouth	Borrower's address is in south U.S.
regionpacific	Borrower's address is in pacific U.S.

Without specifying a dummy variable for each category, our initial demand model will take the form of:

$$\begin{aligned} \text{loan_amnt}_i = & \alpha_0 + \alpha_1(\text{int_rate}_i) + \alpha_2(\text{emp_length_cat}_i) + \alpha_3(\text{region}_i) \\ & + \alpha_4(\text{home_ownership}_i) + \epsilon_{1,i} \end{aligned}$$

For the supply model, “loan_amnt” will be the response variable and the explanatory variables will be “int_rate”, “dti”, “fico_range_low”, and “real_estate_col”. Below is a table of descriptions of the variables and their categories:

Table 3: Supply Model Variables

Variables	Definitions
dti	Debt-to-income ratio of the borrower
fico_range_low	The lower range of the borrower's FICO credit score
int_rate	The interest rate on the loan
loan_amnt	The listed amount of the loan applied for by the borrower
real_estate_col/mortgage	Borrower owns a house or has a mortgage
real_estate_colnone	Borrower has no real estate

Like the demand model, the intital form of the supply model without specifying each dummy variable for the categorical predictors will take the form of:

$$\text{loan_amnt}_i = \beta_0 + \beta_1(\text{int_rate}_i) + \beta_2(\text{dti}_i) + \beta_3(\text{fico_range_low}_i) + \beta_4(\text{real_estate_col}_i) + \epsilon_{2,i}$$

However, as outlined in the theoretical section of this paper, interest rate is endogenous because it is simultaneously determined with the loan amount. Since this breaks the OLS assumption of a covariance of 0 with our error term and explanatory variables, the two-stage least squares method is used to estimate interest rate. We use both the supply and demand models' explanatory variables to estimate interest rate, denoting it “int_rate_hat”. The first-stage model to estimate interest rate is defined below, without specifying each dummy variable for all categorical predictors:

$$int_rate_i = \delta_0 + \delta_1(dt_i) + \delta_2(fico_range_low_i) + \delta_3(real_estate_col_i)$$

$$+ \delta_4(emp_length_cat_i) + \delta_5(region_i) + \epsilon_{3,i}$$

Notice that the “home_ownership” variable was removed from the model estimating interest rate. This is because “real_estate_col” was a variable created from “home_ownership” and they have near perfect collinearity. Once the interest rate model is specified, it is used to obtain estimates of the interest rate and a new variable “int_rate_hat” is created from these estimates. Since we are focusing on the demand model for this project, we move forward to redefine our demand model as:

$$\begin{aligned} loan_amnt_i = & \alpha_0 + \alpha_1(int_rate_hat_i) + \alpha_2(emp_length_cat_i) \\ & + \alpha_3(region_i) + \alpha_4(home_ownership_i) + \epsilon_{1,i} \end{aligned}$$

Once the final demand model is estimated for both 2008 and 2019, we will compare the effects of each variable and look for any structural differences in the models.

Results

2008

First-Stage

The first step to estimating our demand model in 2008 was to conduct the first-stage estimation of interest rate. The initial model outlined in the “Econometric Models” section for the estimated interest rate had no transformations, but after testing positive for heteroscedasticity and a non-linear relationship, a log transformation was performed on the response, “int_rate”. This solved the problem of non-linearity and gave us homoscedasticity, confirmed by both visual aid and a failure to reject the null hypothesis of the White Test.¹ Our residuals don’t quite follow a normal distribution, but since our sample size is rather large (over 2000 observations) we can disregard this assumption being broken due to the Central Limit Theorem. This model did an excellent job of estimating interest rate, with a multiple R^2 of 0.7309 and the overall model was statistically significant.

Second-Stage

The first stage-model was then used to estimate the interest rate of each loan that we trained our models on. The newly estimated interest rate is now used in the second-stage model to find the demand function. First, we estimated the demand model as defined in the “Econometric Models” section, but quickly found that we faced heteroscedasticity, non-linearity, and non-normality. Normality is not a worry due to our sample size, but to remedy the other issues a log transformation was made on the response. This, however, did not

¹See Appendix A for any visualizations, tests, or other output for the final 2008 models.

Table 4: First-Stage Estimating Interest Rate

<i>Dependent variable:</i>	
	log(int_rate)
dti	0.008*** (0.0003)
fico_range_low	-0.004*** (0.0001)
real_estate_colnone	-0.013** (0.005)
annual_inc	0.00000*** (0.00000)
emp_length_cat6-9	-0.00002 (0.008)
emp_length_cat0-1	-0.003 (0.007)
emp_length_cat2-5	0.001 (0.007)
regionwest	-0.0001 (0.007)
regionnortheast	0.006 (0.008)
regionsouth	-0.007 (0.007)
regionpacific	0.020 (0.028)
Constant	5.215*** (0.048)
Observations	2,306
R ²	0.731
Adjusted R ²	0.730
Residual Std. Error	0.109 (df = 2294)
F Statistic	566.394*** (df = 11; 2294)

Note:

*p<0.1; **p<0.05; ***p<0.01

help fix our model assumptions, but it did make improvements. After experimenting with different transformations on the explanatory variables, it was found that adding a squared “annual_inc” to our model had a significant effect. It fixed the issues of non-linearity, heteroscedasticity, and even increased the R^2 by over 0.05 units.²

Table 5: Second-Stage Demand Model

<i>Dependent variable:</i>	
	log(loan_amnt)
int_rate_hat	-0.029*** (0.007)
annual_inc	0.00002*** (0.00000)
I(annual_inc^2)	-0.000*** (0.000)
emp_length_cat6-9	0.061 (0.046)
emp_length_cat0-1	0.007 (0.040)
emp_length_cat2-5	0.059 (0.039)
regionwest	0.037 (0.042)
regionnortheast	-0.006 (0.045)
regionsouth	-0.037 (0.041)
regionpacific	-0.006 (0.163)
home_ownershipOWN	-0.071 (0.058)
home_ownershipRENT	-0.021 (0.031)
Constant	8.243*** (0.105)
Observations	2,306
R ²	0.195
Adjusted R ²	0.191
Residual Std. Error	0.638 (df = 2293)
F Statistic	46.417*** (df = 12; 2293)

Note: *p<0.1; **p<0.05; ***p<0.01

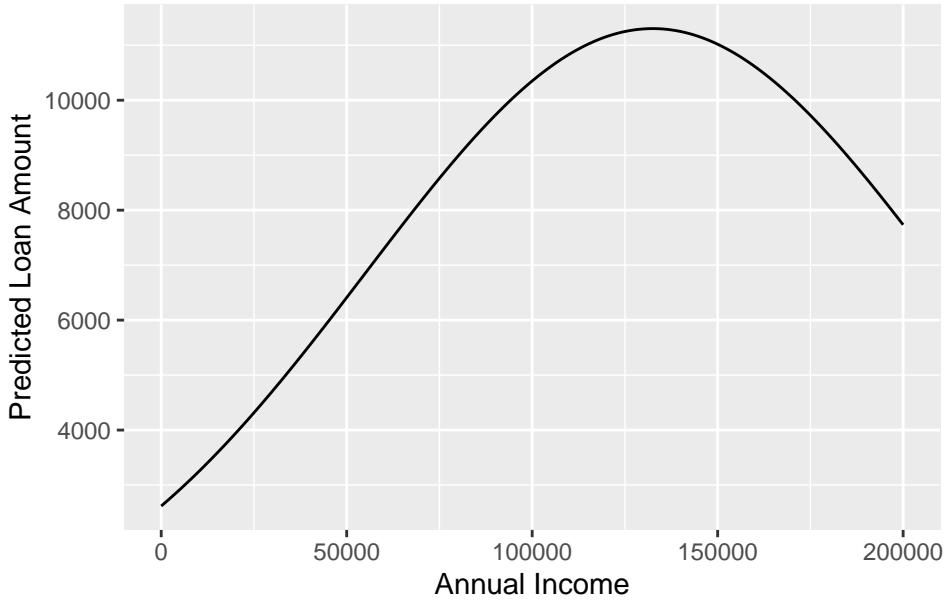
The results of our overall F-test show that our model is significant. However, after further investigation into the categorical variables possibly showing insignificance, only “int_rate_hat”, “annual_inc”, and “annual_inc²” were found to be statistically significant.³ This is rather surprising to see that region, employment length, and home ownership status of a borrower has no significant effect on the demand for loans, but the results for interest rate are expected. As expected, as interest rate increases, the demand for loans will decrease. However, because of the log transformation, the demand decreases by a factor of $e^{0.029}$ for every one percentage point increase in interest rate, meaning the direct effect on the demand is non-linear. The effect of annual income is difficult to verbally interpret, but the graph below shows the effect of income on loan demand, ceteris paribus.⁴

²Received a p-value of .1005 from the White Test for heteroscedasticity.

³This was determined using the incremental F-test, as well as R’s “step()” function that uses the AIC as its criterion.

⁴Interest rate was held constant at 15%, the employment length was “6-9” years, region was the “north-

Annual Income's Effect on Loan Demand



We can see that as the annual income of an individual rises, their demand for loans will increase until they hit a certain income (around \$130,000 in the graph above). After that point, the demand for loans begin to drop as income continues to increase. The income's effect on demand was expected to be positive, but it seems that as individuals' income levels rise to be high income, their demand lowers. This could be a result of several factors, but perhaps a higher level of access to money via their income has some effect on their need for a consumer loan. These higher income individuals may not be included in the typical demographic of borrowers that use the platform to refinance loans or pay off credit cards.

2019

First-Stage

The same approach for estimating the demand model was taken for 2019, though the stability of our models are questionable due to what seems to be a structural change in the data. Similar to the 2008 first-stage model estimating interest rate, a log transformation was performed on the response due to heteroscedasticity, non-linearity, and non-normality. This transformation did not fix any of our model assumptions, as well as several other attempts to remedy the issues.⁵ This subset of data is significantly larger than the 2008 subset, with over 370,000 observations, so the normality assumption being broken is not an issue. However, the issues of heteroscedasticity and non-linearity remain. Since the model assumptions were broken and there is further work to be done finding an alternative method to estimating the 2019 models, please continue to consider the following analysis with caution.

The final model used to estimate interest rate had a log transformation on the response and was statistically significant by the F-test. The R^2 was only 0.179, compared to a large

east", and home ownership status was "MORTGAGE"

⁵See Appendix B for any visualizations, tests, or other output for the final 2008 models.

Table 6: First-Stage Estimating Interest Rate

<i>Dependent variable:</i>	
	log(int_rate)
dti	0.002*** (0.00003)
fico_range_low	-0.004*** (0.00002)
real_estate_colnone	0.010*** (0.001)
annual_inc	-0.00000*** (0.00000)
emp_length_cat2-5	-0.020*** (0.001)
emp_length_cat0-1	-0.028*** (0.002)
emp_length_catunknown	-0.002 (0.002)
emp_length_cat6-9	-0.011*** (0.002)
regionmidwest	-0.007*** (0.002)
regionnortheast	-0.010*** (0.002)
regionwest	-0.016*** (0.001)
regionpacific	0.031*** (0.007)
Constant	5.536*** (0.011)
Observations	372,768
R ²	0.179
Adjusted R ²	0.179
Residual Std. Error	0.340 (df = 372755)
F Statistic	6,762.735*** (df = 12; 372755)

Note: *p<0.1; **p<0.05; ***p<0.01

R^2 of 0.7309 for the first-stage model in 2008. This shows our model is not doing a great job of estimating interest rate, though it had some ability to explain the variance in our data. It can already be seen that there is a difference in the structure of the data between 2008 and 2019.

Second-Stage

Now that the model to estimate interest rate has been found, the estimated interest rates are used in the demand model for 2019. Yet again, the model assumptions of homoscedasticity, linearity, and normality were broken. Non-normality of residuals is fine due to the sample size, but a log transformation was taken on the response to correct the other two. This seems to fix the issue of non-linearity, but heteroscedasticity was confirmed by both the White Test and the residuals versus fitted plot. Several actions were taken to address this, but only including the square of annual income in the demand model resulted in a noticeable change. This did not fix the issue of heteroscedasticity. Because of this, continue to proceed with caution.

Table 7: Second-Stage Demand Model

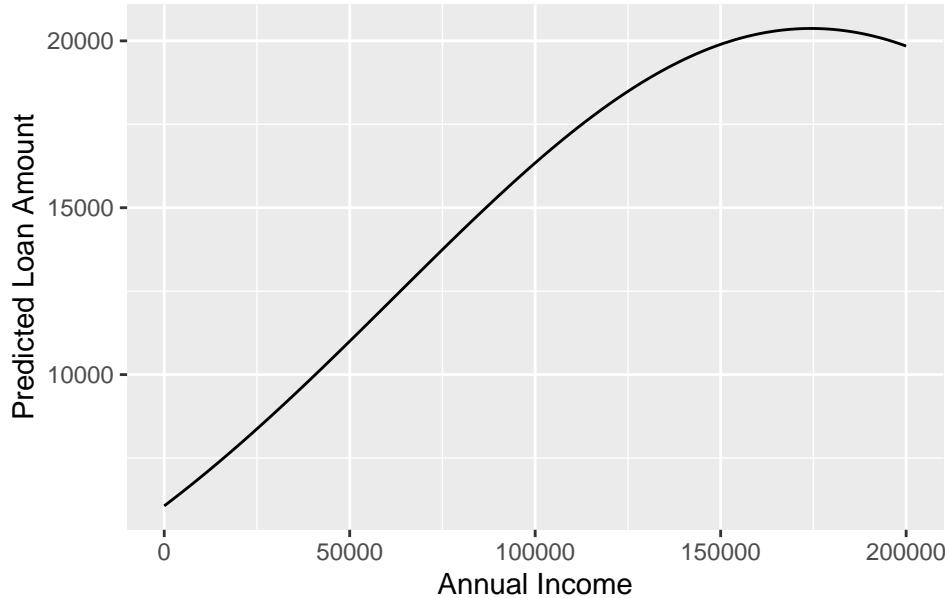
<i>Dependent variable:</i>	
	log(loan_amnt)
int_rate_hat	-0.011*** (0.001)
annual_inc	0.00001*** (0.00000)
I(annual_inc^2)	-0.000*** (0.000)
emp_length_cat2-5	-0.015*** (0.003)
emp_length_cat0-1	0.0002 (0.003)
emp_length_catunknown	-0.036*** (0.004)
emp_length_cat6-9	-0.003 (0.004)
regionmidwest	-0.012*** (0.003)
regionnortheast	-0.008** (0.003)
regionwest	0.007** (0.003)
regionpacific	0.086*** (0.014)
home_ownershipRENT	-0.112*** (0.003)
home_ownershipOWN	-0.062*** (0.004)
Constant	8.881*** (0.010)
Observations	372,768
R ²	0.149
Adjusted R ²	0.149
Residual Std. Error	0.679 (df = 372754)
F Statistic	5,016.184*** (df = 13; 372754)

Note: *p<0.1; **p<0.05; ***p<0.01

This model was found to be statistically significant by the F-test. The R^2 was 0.149, being lower than the 2008 model. Our model was significant and all the variables in the

model were found to be statistically significant in predicting the demand for loans, though these results could be caused by a poor model structure signalled by the residual versus fitted plot. Expectedly, as interest rate increases, the demand for loans decreases. Yet again, annual income has the same relationship with loan demand as in 2008. Although, the point at which loan demand begins to lower as income increases is around \$175,000 in annual income. See the graph below for a visualization of the relationship between loan demand and annual income, *ceteris paribus*.⁶

Annual Income's Effect on Loan Demand



The results of our model estimation for 2019 suggest that employment length, region, and home ownership status have a significant effect on the demand for loans. By our results, the demand for loans seems to be greater for those with 10 or more years of employment and those with 0 to 1 years of employment. Whereas individuals with an unknown employment length or 2 to 9 years of experience have lower demands for loans. These results are not exactly as expected. However, the results of home ownership status of the borrower were as expected. Borrowers that rented demand less loans than those that owned a house or had a mortgage. We also found that borrowers that had a mortgage had the greatest demand for loans.

Region was our last variable that had a significant effect on loan demand in 2019. We found that the demand for loans in the Midwest were the lowest, with it being $e^{0.012}$ times lower than the demand in the South. We also found that the demand for loans in the Pacific regions of the U.S. were the greatest, with the demand in the Pacific region being $e^{0.086}$ times greater than the demand in the South. The results of the 2019 demand model do suggest that all of our explanatory variables have a significant effect on predicting the demand for loans. However, due to the model assumptions being broken and from the appearance of the residual versus fitted plot, these results should be taken lightly.

⁶Interest rate was held constant at 15%, the employment length was “6-9” years, region was the “north-east”, and home ownership status was “MORTGAGE”

Comparisons

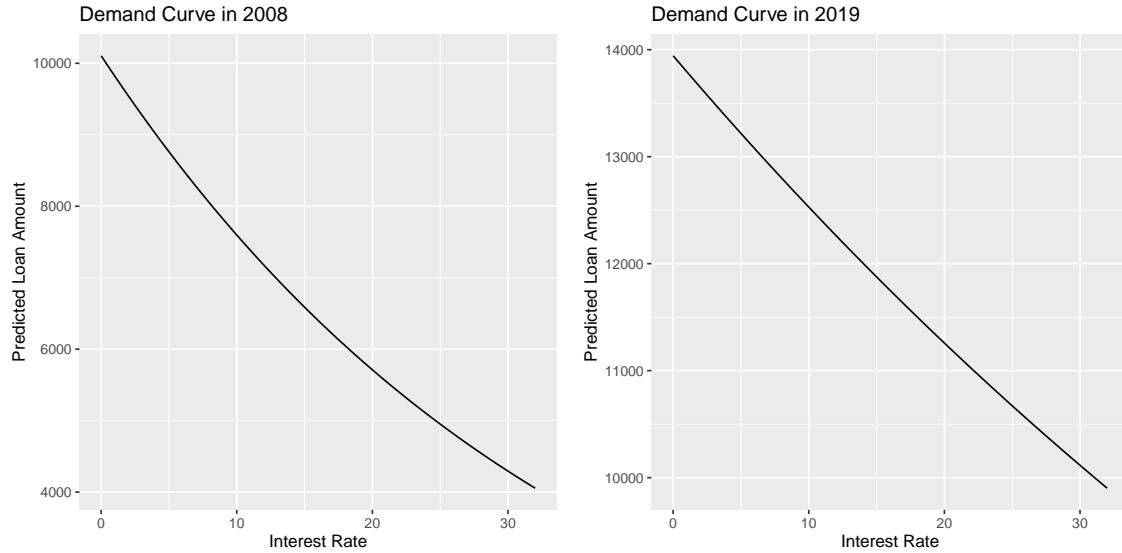
Both the 2008 and 2019 final demand models had the same form and were both significant, but the effects of the individual explanatory variables were different. Additionally and most importantly, the data seemed to have a structural change between 2008 and 2019. The 2019 models are not reliable due to the homoscedasticity assumption being broken, but we can still move forward with comparison if this is taken into consideration.

For both years, we found that the explanatory variables had a non-linear effect on the demand for loans. However, their magnitudes and significance levels were different. In 2008, we saw that for every one percentage point increase in the interest rate, the demanded loan amount decreased by a factor of $e^{0.03}$. Whereas in 2019, the demanded loan amount decreased by a factor of $e^{0.01}$. The form of our equations also allow us to easily observe the price elasticity of demand for loans. The price elasticity of demand in 2008 is 0.03 and in 2019 it is 0.01. This means the demand for loans is very price inelastic in both years, though it is even more inelastic in 2019.

We also found that the relationship between loan demand and income had a quadratic relationship in both 2008 and 2019. Once we take into account the difference in the loan limits between the two years, the rate at which income effects demand seems to be quite similar, though the threshold at which demand begins to decreases because of a high income is different. This annual income threshold is greater in 2019 than in 2008.

Finally, the significance of the categorical predictors were different in 2008 than in 2019. However, it cannot be said with confidence due to the instability of the 2019 model. The 2008 model does a noticeably better job at predicting the loan demand, with an R^2 of 0.1954 as compared to an R^2 of 0.149 for the 2019 model. Due to the structural change in the data and the difference in the magnitudes of effects on loan demand, there is an obvious change in the market for loans on the LendingClub platform from during the Great Recession in 2008 to the pre-Great Lockdown in 2019. Below is a visualization of the non-linear demand curves in 2008 and 2019 that shows the relationships between interest rate and the loan demand, *ceteris paribus*.⁷

⁷Holding interest rate at 15%, region is “northeast”, home ownership status is “MORTGAGE”, employment length is “6-9” years, and the annual incomes are the respective years average in the U.S. (\$52,000 in 2008 and \$58,000 in 2019).



Conclusion

The results of the analysis are not concrete due to the need of better methods for the 2019 model and to take into account extremely high income individuals (greater than \$200,000 per year), but we do observe a non-linear loan demand curve. Additionally, we find the expected effect of an increase in interest rate to decrease loan demand for both time periods. We also found a quadratic relationship of income with loan demand. Above all else, we clearly observe a structural change in the market for loans on the LendingClub platform from 2008 to 2019.

The findings from this project can be useful in understanding the relationship that interest rate and income have with demand in the market for loans. For this research to move forward, a new method to estimate the demand for loans needs to be used that will address the structural form of the data. This includes taking care of the issue of heteroscedasticity in the 2019 model, as well as including observations with very high annual incomes. Further research with this data could be conducted on the differences in the demand for loans for different income brackets and the different motivations for taking out a loan, as well as studying the factors that affect loan grades and subgrades. One could further investigate the default rates of LendingClub loans, as well as the change in demand for loans after the borrowers are less credit constrained after 2016 on the platform.

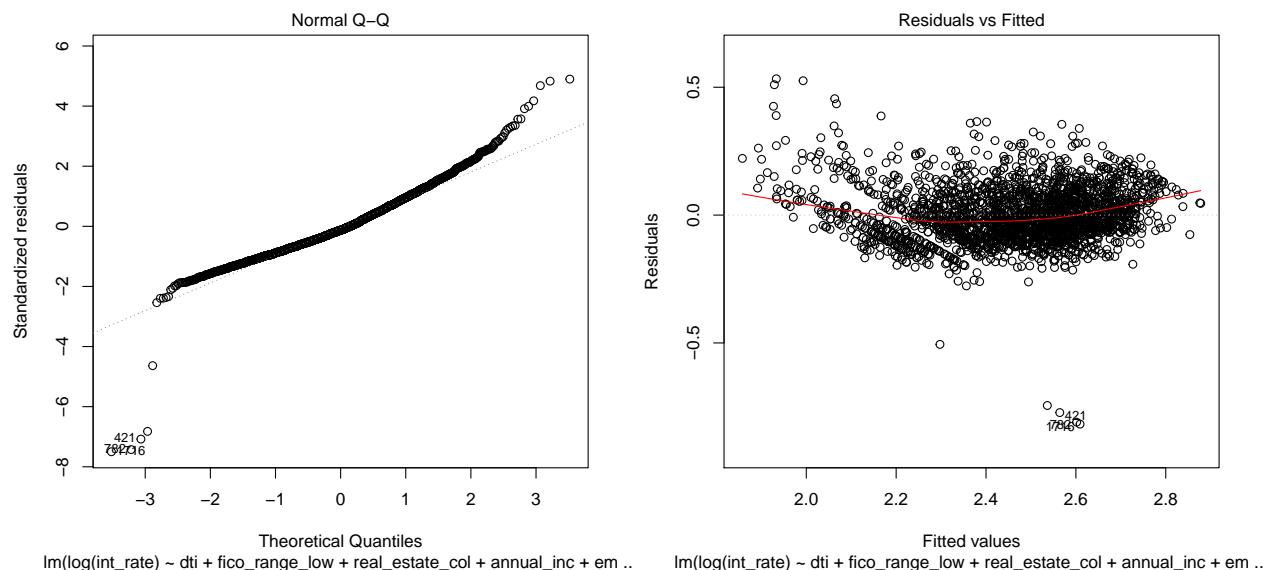
Appendix

Appendix A

2008 First-Stage Model Residual Diagnostics

Table 8: Variance Inflation Factor

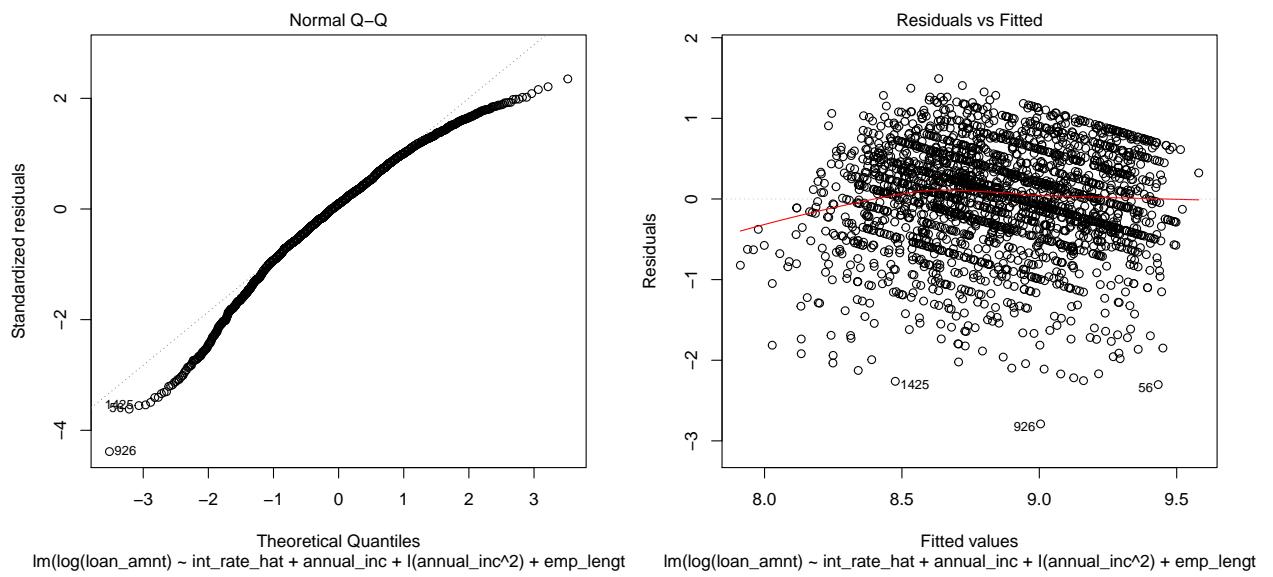
	GVIF	Df	GVIF^(1/(2*Df))
dti	1.098	1	1.048
fico_range_low	1.102	1	1.050
real_estate_col	1.203	1	1.097
annual_inc	1.132	1	1.064
emp_length_cat	1.078	3	1.013
region	1.055	4	1.007



2008 Second-Stage Model Residual Diagnostics

Table 9: Variance Inflation Factor

	GVIF	Df	$\text{GVIF}^{(1/(2*\text{Df}))}$
int_rate_hat	1.032	1	1.016
annual_inc	12.625	1	3.553
I(annual_inc^2)	12.103	1	3.479
emp_length_cat	1.083	3	1.013
region	1.056	4	1.007
home_ownership	1.243	2	1.056

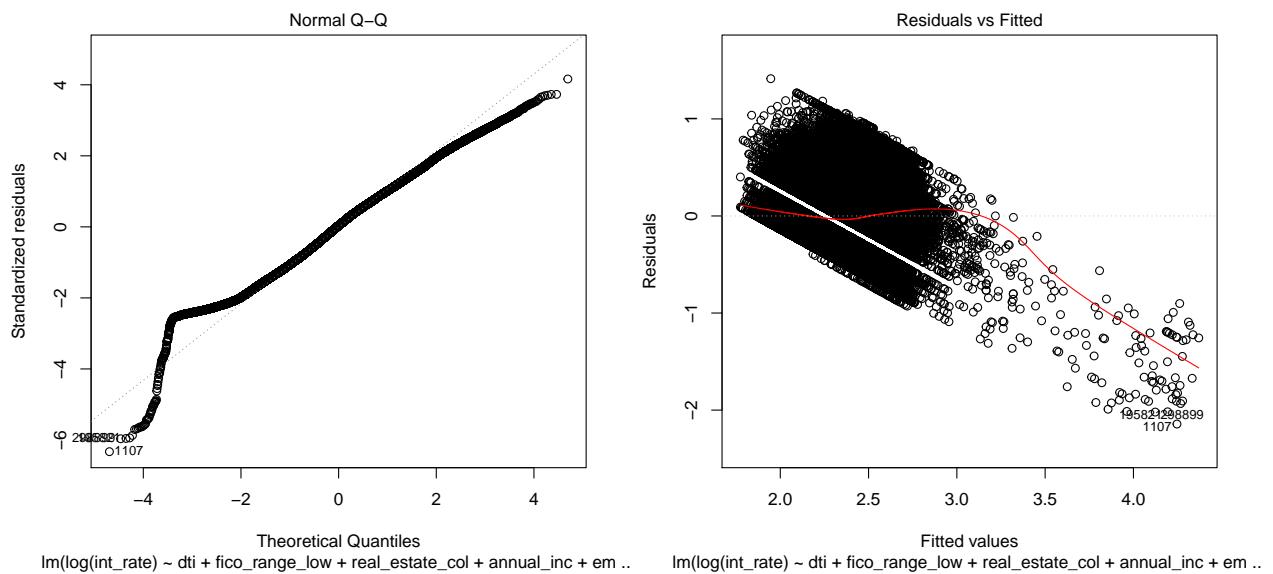


Appendix B

2019 First-Stage Model Residual Diagnostics

Table 10: Variance Inflation Factor

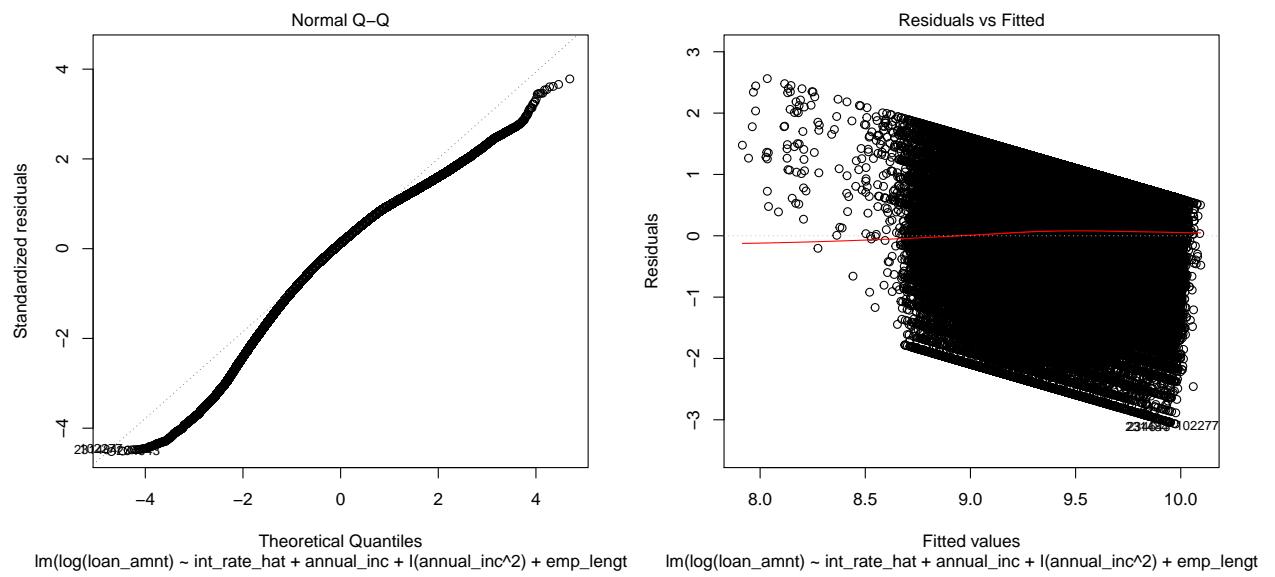
	GVIF	Df	GVIF^(1/(2*Df))
dti	1.039	1	1.019
fico_range_low	1.013	1	1.007
real_estate_col	1.127	1	1.062
annual_inc	1.134	1	1.065
emp_length_cat	1.096	4	1.011
region	1.047	4	1.006



2019 Second-Stage Model Residual Diagnostics

Table 11: Variance Inflation Factor

	GVIF	Df	GVIF^(1/(2*Df))
int_rate_hat	1.076	1	1.037
annual_inc	16.486	1	4.060
I(annual_inc^2)	15.803	1	3.975
emp_length_cat	1.146	4	1.017
region	1.052	4	1.006
home_ownership	1.157	2	1.037



References

- Bernanke, Ben, and Mark Gertler. 1986. "Agency Costs, Collateral, and Business Fluctuations." *NBER Working Paper Series; Camrbidge*.
- LendingClub. 2020. "LendingClub Website." <https://www.lendingclub.com/info/statistics.action>.
- Manrique, J., and K. Ojah. 2004. "Credits and Non-Interest Rate Determinants of Loan Demand: A Spanish Case Study." *Applied Economics* 36:8: 781–91. <https://doi.org/10.1080/0003684042000229514>.
- Perraudin, William R. M., and Bent E. Sørensen. 1992. "The Credit-Constrained Consumer: An Empirical Study of Demand and Supply in the Loan Market." *American Statistical Association*.
- Poghosyan, Tigran. 2011. "Slowdown of Credit Flows in Jordan in the Wake of the Global Financial Crisis: Supply or Demand Driven?" *Economic Systems* 35: 562–73. <https://doi.org/10.1016/j.ecosys.2011.04.004>.
- Polena, Michal, and Tobias Regner. 2018. "Determinants of Borrowers' Default in P2p Lending Under Consideration of the Loan Risk Class." *Games*. <https://doi.org/10.3390/g9040082>.
- Swain, Ranjula Bali. 2007. "The Demand and Supply of Credit for Households." *Applied Economics* 39:21: 2681–92. <https://doi.org/10.1080/00036840600749516>.
- The Economist. 2020. "Peer Review." <https://www.economist.com/schumpeter/2013/01/05/peer-review>.
- Yahoo! Finance. 2020. "LendingClub Corporation." <https://finance.yahoo.com/quote/LC/financials?p=LC>.