

Economic Determinants of Foreclosures

Amanda Freche

December 9, 2017

ECO4935

I. INTRODUCTION

Although mortgages are essential to providing majority of the population looking to purchase a house with the ability to purchase a house, mortgages can also cause significant financial trouble as seen with the housing crisis in the early 2000's. Mortgage lenders require applicants to complete credit checks, look at outstanding debt in comparison to assets, and compare annual and monthly income to ensure that the applicant will be financially sound enough to continue with payments until the mortgage has been released and all the payments have been fulfilled. However even after a successful inspection of an applicant, deeming them financially stable enough to not default on their mortgage, there are still variables that can affect whether a homeowner will default on their mortgage and have to go through foreclosure. This paper will study some of these "outside factors" that mortgage lenders do not take into consideration when viewing applicants, and look at their effect on the foreclosure rate in counties in Florida.

II. SAMPLE

The sample used for this study are 53¹ counties in Florida from 2007 to midyear 2008. The housing market during this period within the United States was very tumultuous due to the financial crisis, causing significant variation and larger levels of foreclosure rates to study.

III. DEPENDENT VARIABLE

Estimated Foreclosure Rate by County

est_foreclosure_rate

¹ Calhoun County, DeSoto County, Dixie County, Gilchrist County, Glades County, Hamilton County, Hardee County, Holmes County, Jefferson County, Lafayette County, Liberty County, Madison County, Union County, and Washington County are all excluded due to lack of Single Family Home Sales Prices data.

The dependent variable for this study is the estimated foreclosure rate by county, or the rate of estimated number of foreclosures divided by the estimated number of mortgages in each county. These data are obtained through the U.S. Department of Housing and Urban Development (HUD)² throughout the entire 2007 calendar year through October 2008. To collect these data, the HUD synthesized data from The Mortgage Bankers Association, Equifax, and information from the United States Postal Service.

IV. INDEPENDENT VARIABLES

Estimated High Cost Loan Rate

est_hicost_loan_rate

The Estimated High Cost Loan Rate is the estimated percentage of loans made between 2004 and 2006, that are shown to be “high cost”. Data averaged from 2004 through 2006 are used because it is estimated that the average household will own their home between two to four years before possibly foreclosing, this would be the appropriate lag time before the using 2008 foreclosure rates. These loans are considered high cost based off the Home Mortgage Disclosure Act, where the rate spread is 3 percentage points above U.S. Treasury securities with comparable maturity. These high cost loans differ from traditional conventional loans because they are riskier and often have special protective measures to ensure that lenders are not being predatory against homeowners. The Estimated High Cost Loan Rate is gathered from data from the U.S. Department of Housing and Urban Development² within their local level data.

With an increase of high cost loan rates, it is hypothesized that there should be a strong positive correlation with the increase of foreclosure rates within a county. These high-cost, riskier loans are more likely to default, often because the homeowner has lower credit, irregular

² U.S. Department of Housing and Urban Development,
https://www.huduser.gov/portal/datasets/nsp_foreclosure_data.html

income, or other factors that make them less likely to be financially sound. It is also hypothesized that there should be a negative correlation between the high cost loan rates and the per capita income because these high cost loans are usually given to people with lower income.

Unemployment Rate **bls_unemployment_rate // bls_unemployment_rate_2004**

The unemployment rate is the level of unemployment during June 2008 using the official U3 unemployment rate. This counts people who currently do not have a job, have looked for employment within the last four weeks, and are currently available to work. These data are taken from the U.S. Department of Housing and Urban Development² gathered through the U.S. Bureau of Labor Statistics. In the second regression, annual unemployment rate from 2004 is used to account for a possible lag time in losing jobs and income to foreclosing. It is hypothesized that as unemployment rates increase, there should be an increase in foreclosure rates. When a homeowner is unemployed, they will be less likely to afford payments on their home, therefore increasing the possibility of foreclosing on their home.

Median Household Income **med_household_income**

Median household income represents the middle point of the income distribution for that county. Half of the income in the county is above this point, and half is below this point. These data are obtained through the Bureau of Economic and Business Research, collected by the Bureau of Economic Analysis (BEA)³ using data from the Small Area Income and Poverty Estimates (SAIPE) program. SAIPE provides single year estimates of median household income for school districts and federal purposes and is known as one of the most accurate estimates of

³ U.S Census Bureau and Bureau of Economic Analysis, <https://www.bebr.ufl.edu/data/localities/207/county>

income⁴. Median household income provides a more accurate measure of income, not skewing an average for large earners or those who are significantly below the poverty line⁵.

With an increase in per capita personal income, it is hypothesized that there should be a negative correlation with the increase of foreclosure rates. This is assuming that with a higher personal income, a homeowner is more likely to afford their mortgage loan and continue with their payments. If the median single-family home value is a higher proportion compared to the per capita personal income, the homeowner in that area may be more likely to default on their mortgage loan, therefore increasing the foreclosure rate in that county⁶.

New Construction Value

building_permits_pop // building_permits_pop_2004

The annual building permits for single family units is the number of building permits handed out within a year for single family units. Although these data include fully detached, semi-detached, row houses, and town houses, it can still be used as a proxy for new construction of single family homes during the 2007 and 2008 period⁷. These data are then divided by the population of each county⁸, to account for larger amounts of construction activity in more populated counties and vice versa. These data are obtained through the Bureau of Economic and Business Research collected by the U.S. Census Bureau⁹. In the second regression, single family building permits divided by population during 2004 is used to account for construction activity

⁴ U.S Census Bureau, <https://www.census.gov/programs-surveys/saipe/about.html>

⁵ Originally Per Capita Personal Income was used, but was eliminated due to the inaccurate averages for extremely high or low earners within the county.

⁶ This does not consider mortgage lenders offering disproportionately larger loans due to the increased income of individuals, assuming they are financially sound enough to cover these larger loans. Although issuing faulty loans was a major cause of the housing crisis, it is too complicated with a short amount of time to fully analyze this.

⁷ Previously, a percentage of houses built still in existence was used, but this was not effective in finding the number or percentage of construction activity during the time period.

⁸ Gathered through Bureau of Economic Business Research, <https://www.bebr.ufl.edu/data/localities/9085/county>

⁹ U.S Census Bureau and Bureau of Economic Business Research, <https://www.bebr.ufl.edu/data/localities/345/county>

occurring before the foreclosure period. These data are obtained through the U.S Department of Housing and Urban Development¹⁰.

There is significant research pointing to high amounts of construction and influx of new homes as being one of the causes of the housing bubble and eventually the economic downturn. Hedberg and Krainer from the San Francisco Federal Reserve are producing a series of working papers on the effects of housing supply on foreclosures¹¹. They state that “it is generally true that construction fell more in states where foreclosure rates were higher, but construction activity dropped to historically low levels in virtually every region” (Hedberg & Krainer, 2012, p. 2). According to the paper, one of the reasons why new construction homes were particularly affected by foreclosures is that young, new homebuyers that are more susceptible to unstable income and employment were more likely to purchase new homes (Hedberg & Krainer, 2012, p. 2). If this is assumed to be true, it is hypothesized that there will be a positive correlation with an increase in percentages of new home construction, an increase in high cost loan rates and unemployment rates, causing an increase in foreclosure rates within each county.

Median Sales Price and Home Value for Single Family Homes

med_home_value

The median sales price and home value for single family homes represents the middle point of the prices of single family homes sold within each county of Florida. Half of all the homes sold are above this point, and half of all the homes sold are below this point. These median sales price and home value data are collected from the Bureau of Economic and Business Research (BEBR)¹².

The median value is used instead of an average because extremely high or extremely low

¹⁰ U.S Department of Housing and Urban Development, <https://socds.huduser.gov/permits/index.html>

¹¹ Hedberg, W., & Krainer, J. (September 2012). Housing Supply and Foreclosures. FEDERAL RESERVE BANK OF SAN FRANCISCO. Retrieved from <http://www.frbsf.org/economic-research/files/wp12-20bk.pdf>

¹² Bureau of Economic and Business Research Housing and Real Estate Database, <https://www.bebr.ufl.edu/data/localities/3555/county>

sale prices or home values could skew the data and push the average home value in the county up or down. A single-family home is a detached, free-standing residential building. This paper specifically observes data of single-family homes, because they have the highest rate of homeownership compared to condominiums, duplexes, and other owned forms of multi-family housing. With an increase in single family home values and sales prices, it is hypothesized that there will be an increase in the size of a mortgage loan to cover the value of the house. With an increase in the mortgage loan and higher payment amounts, it is more likely that a homeowner will default on their loan payments, correlating to an increase in foreclosure rates.

V. RESULTS

Correlation Matrix – Regression 1

	<i>est_hicost_loan_rate</i>	<i>bls_unemployment_rate</i>	<i>med_household_income</i>	<i>building_permits_pop</i>	<i>med_home_value</i>
<i>est_hicost_loan_rate</i>	1				
<i>bls_unemployment_rate</i>	0.505732285	1			
<i>med_household_income</i>	-0.498774371	-0.203481361	1		
<i>building_permits_pop</i>	-0.2627743	-0.213234513	0.021897743	1	
<i>med_home_value</i>	-0.39424248	-0.214230447	0.593486497	-0.078848738	1

Correlation Matrix – Regression 2

	<i>est_hicost_loan_rate</i>	<i>bls_unemployment_rate_2004</i>	<i>med_household_income</i>	<i>building_permits_pop_2004</i>	<i>med_home_value</i>
<i>est_hicost_loan_rate</i>	1				
<i>bls_unemployment_rate_2004</i>	0.347177335	1			
<i>med_household_income</i>	-0.498774371	-0.095764337	1		
<i>building_permits_pop_2004</i>	-0.298577931	-0.133240853	0.127466004	1	
<i>med_home_value</i>	-0.39424248	-0.185299693	0.593486497	0.06185908	1

To account for collinearity issues, a correlation matrix was made to ensure that no independent variables were colinear. No two variables were highly correlated¹³, meaning they could all be used within the regression.

¹³ In Regression 1, High Cost Loan Rate and Unemployment were largely correlated since unemployment is one of the criteria of the loans. Median Household Income and Median Home Value were also largely correlated since an increase in income can lead to an increase in what one can afford for housing. The threshold for determining whether something is too correlated is the absolute value of 0.6.

Impact Summary – Regression 1

est_hicost_loan_rate	0.01171917
bls_unemployment_rate	0.013605646
med_household_income	0.006364206
building_permits	-0.00181438
med_home_value	0.001373702

Impact Summary – Regression 2

est_hicost_loan_rate	0.01989
bls_unemployment_rate	0.00234
med_household_income	0.00734
building_permits	0.00372
med_home_value	0.00145

Descriptive Statistics of Data – Regression 1

	Mean	Standard Error	Median	Standard Deviation	Range	Minimum	Maximum
<u>est_foreclosure_rate</u>	0.068789721	0.003124746	0.070728623	0.022748494	0.098974789	0.019659839	0.118634628
<u>est_hicost_loan_rate</u>	0.287279521	0.010090168	0.274550784	0.073457529	0.334319914	0.142648016	0.47696793
<u>bls_unemployment_rate</u>	0.057905186	0.001806747	0.055838119	0.01315332	0.069466389	0.033865676	0.103332065
<u>med_household_income</u>	45601.58491	1005.771166	45783	7322.124609	30446	33282	63728
<u>building_permits_pop</u>	0.005783278	0.000659538	0.004415972	0.004801508	0.028567385	0.000668188	0.029235574
<u>med_home_value</u>	220500	11481.43956	201300	83586.14166	492750	85000	577750

Descriptive Statistics of Data – Regression 2

	Mean	Standard Error	Median	Standard Deviation	Range	Minimum	Maximum	Observations
<u>est_foreclosure_rate</u>	0.068789721	0.003124746	0.070728623	0.022748494	0.098974789	0.019659839	0.118634628	53
<u>est_hicost_loan_rate</u>	0.287279521	0.010090168	0.274550784	0.073457529	0.334319914	0.142648016	0.47696793	53
<u>bls_unemployment_rate_2004</u>	0.047468358	0.001291678	0.046074872	0.009403559	0.054243218	0.031516443	0.085759661	53
<u>med_household_income</u>	45601.58491	1005.771166	45783	7322.124609	30446	33282	63728	53
<u>building_permits_pop_2004</u>	0.015421396	0.001967804	0.011529326	0.014325829	0.063272889	0.001158661	0.06443155	53
<u>med_home_value</u>	220500	11481.43956	201300	83586.14166	492750	85000	577750	53

Regression 1 with All Variables

$$\text{est_foreclosure_rate} = \beta_0 + \beta_1(\text{est_hicost_loan_rate}) + \beta_2(\text{bls_unemployment_rate}) + \beta_3(\text{med_household_income}) + \beta_4(\text{building_permits_pop}) + \beta_5(\text{med_home_value}) + \text{error}$$

Regression 2 with All Variables

$$\text{est_foreclosure_rate} = \beta_0 + \beta_1(\text{est_hicost_loan_rate}) + \beta_2(\text{bls_unemployment_rate_2004}) + \beta_3(\text{med_household_income}) + \beta_4(\text{building_permits_pop_2004}) + \beta_5(\text{med_home_value}) + \text{error}$$

Excel Regression Output – Regression 1

Regression Statistics	
Multiple R	0.91264975
R Square	0.832929565
Adjusted R Square	0.815156115
Standard Error	0.009780368
Observations	53

ANOVA					
	df	SS	MS	F	Significance F
Regression	5	0.022413874	0.004482775	46.86369514	3.89983E-17
Residual	47	0.004495813	9.56556E-05		
Total	52	0.026909687			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-0.078012573	0.015155215	-5.147572812	5.09872E-06	-0.108500934	-0.047524213	-0.108500934	-0.047524213
est_hicost_loan_rate	0.159536677	0.025059025	6.366435819	7.48947E-08	0.10912442	0.209948934	0.10912442	0.209948934
bls_unemployment_rate	1.034388732	0.120598276	8.577143584	3.55431E-11	0.791776293	1.27700117	0.791776293	1.27700117
med_household_income	8.69175E-07	2.47474E-07	3.512193062	0.000993208	3.71322E-07	1.36703E-06	3.71322E-07	1.36703E-06
building_permits_pop	-0.377876629	0.300838265	-1.256079006	0.21529536	-0.983085154	0.227331896	-0.983085154	0.227331896
med_home_value	1.64346E-08	2.07005E-08	0.793919835	0.431232412	-2.52095E-08	5.80787E-08	-2.52095E-08	5.80787E-08

Excel Regression Output – Regression 2

Regression Statistics	
Multiple R	0.766576227
R Square	0.587639112
Adjusted R Square	0.543770933
Standard Error	0.015365409
Observations	53

ANOVA					
	df	SS	MS	F	Significance F
Regression	5	0.015813184	0.003162637	13.39556641	3.96159E-08
Residual	47	0.011096502	0.000236096		
Total	52	0.026909687			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-0.07434546	0.023535089	-3.158919824	0.002767472	-0.121691953	-0.026998967	-0.121691953	-0.026998967
est_hicost_loan_rate	0.270754432	0.037039389	7.309905425	2.7743E-09	0.196240792	0.345268072	0.196240792	0.345268072
bls_unemployment_rate_2004	0.249342937	0.244673475	1.019084466	0.313380135	-0.242876604	0.741562479	-0.242876604	0.741562479
med_household_income	1.00176E-06	3.91248E-07	2.560433466	0.013727255	2.14675E-07	1.78885E-06	2.14675E-07	1.78885E-06
building_permits_pop_2004	0.259870187	0.156281122	1.662837988	0.103000479	-0.054526878	0.574267252	-0.054526878	0.574267252
med_home_value	1.73581E-08	3.2289E-08	0.537585206	0.593399856	-4.7599E-08	8.23152E-08	-4.7599E-08	8.23152E-08

Estimated High Cost Loan Rate

est_hicost_loan_rate

In the first regression, the t-stat for estimated high cost loan rate is 6.3664, meaning that it is statistically significant at the 99.9% confidence level. In the second regression, the t-stat is 7.3099, still statistically significant at the 99.9% confidence level. The impact summary is 0.011 in the first regression, meaning that for every standard deviation increase in the loan rate, there is a 0.011 increase in the foreclosure rate. In the second regression, the impact summary only increases by 0.008. This is consistent with the hypothesis that an increase in riskier loans would cause an increase in the foreclosure rate. When looking at the correlation matrix, this is also consistent with the hypothesis that high cost loan rates would be negatively correlated with

median income, because these loans are specifically given to people with lower than average incomes.

Unemployment Rate

bls_unemployment_rate // bls_unemployment_rate_2004

In the first regression, the t-stat for unemployment rate is 8.5771, meaning that it is statistically significant at the 99.9% confidence level. The impact summary for this regression is 0.0136 meaning that for every standard deviation increase in the unemployment rate, there is a 0.0136 increase in the foreclosure rate. This is consistent with the hypothesis that an increase in the unemployment rate can be associated with increase in the foreclosure rate.

However, in the second regression using data from 2004, the t-stat is 1.0190, meaning that it is not statistically significant at any confidence level. This is not consistent with the hypothesis that there could be a lag time between loss of job or income, and foreclosing. It is possible that losing income would have a more immediate effect on mortgage payments and foreclosures.

Median Household Income

med_household_income

In the first regression, the t-stat for median household income is 3.5121, meaning that it is statistically significant at the 99% confidence level. The second regression's t-stat is 2.5604, still statistically significant at the 98% confidence level. The impact summary for the first regression is 0.00636, meaning that for every standard deviation increase in the median household income, there is 0.00636 increase in the foreclosure rate. In the second regression, the impact summary only increases by 0.0009. This is not consistent with the hypothesis that an increase in income would mean that one is more likely to be able to afford their loans, and therefore not default. A possible explanation for this is the larger the income, the larger loan an applicant would be approved for, making it riskier if something would happen to their income, and therefore more likely to default on their loan.

New Construction Value

building_permits_pop // building_permits_pop_2004

In the first regression, the t-stat for the number of building permits representing new construction is -1.2560, meaning that it is not statistically significant within any confidence levels. There is no linear relationship between construction levels and the foreclosure rate. The coefficient and impact summary were negative however, indicating that as the foreclosure rate increased, the number of building permits, or value of new construction decreased during 2008.

In the second regression using the 2004 levels of building permits, the t-stat is 1.6628, statistically significant at the 90% confidence level, indicating that there is a linear relationship between 2004 construction levels and 2008 foreclosures. This impact summary is 0.00372, meaning that for every standard deviation increase in the 2004 building permit levels, there is 0.00372 increase in the 2008 foreclosure rate. The coefficient is also positive, indicating that as the construction levels increased, the foreclosure rate increased as hypothesized. This confirms that there is a lag time for the construction levels to have an effect on foreclosure rates.

Median Sales Price and Home Value for Single Family Homes

med_home_value

The t-stat for median sales price and home value for single family homes is 0.7939 in the first regression and 0.5375 in the second regression, meaning that it was not statistically significant at any conventional levels in either regression. There is no linear relationship between the price of homes and the rate of foreclosures.

VI. CONCLUSION

The results of this study show that outside variables that would not be accounted for during screening for loans by banks and other financial institutions are highly influential on the rate in which people foreclose on their homes. The first study achieved an adjusted R-Squared of 0.815, indicating that 81.5% of the variation of the estimated foreclosure rate can be explained by this model and the independent variables. The second study achieved an adjusted R-Squared of 0.5437, indicating that 54.4% of the variation is explained by the independent variables.

The strongest independent variable was high cost loan rate, having the strongest t-statistic in both regressions. This is mostly because these “high cost” loans are deemed risky and therefore high cost to the financial institutions, due to their likelihood of default. Therefore, there would be a strong correlation between these high risk, high cost loans and the foreclosure rate. This is also something that financial institutions would gauge for when administering loans, but could prove that these institutions administer these loans knowing they will fail and default due to their strong correlation with the foreclosure rate. The next strongest t-statistic in both regressions was Median Household Income. Although statistically significant, the positive coefficient and impact summary contradicted the original hypothesis that an increase in income will result in less likelihood of default. This could be because a larger income might lead to a larger mortgage loan, and with a loss of income or financial issues, more likely to foreclose.

The addition of the second regression had conflicting results with the unemployment rate and new construction values. In the first regression, unemployment was the strongest t-statistic and was the most statistically significant. This was consistent with the hypothesis that if someone does not have a job or steady income, they will not be able to afford their loan on their home and therefore default and foreclose. Although unemployment would be something that financial institutions would check for at the time of administering the loan, this variable can account for people changing their employment status during the life of the loan, after they have received the initial approval. However, in the second regression using 2004 unemployment rates, it was no longer statistically significant. This does not support the hypothesis that 2004 unemployment would have a larger effect on 2008 foreclosure rates. Because unemployment was statistically significant during the first regression and not statistically significant in the second regression, this proves that unemployment affects foreclosure rates immediately and there is not a possible lag effect. Construction values also had a different outcome when adding in the second regression with older 2004 data. In the first regression, new construction levels were not

statistically significant, meaning that an increase in construction did not affect foreclosure rate. However, when using 2004 construction levels in comparison to 2008 foreclosure rate, this t-stat was statistically significant. This could be because there is a possible lag time with the effects of increased construction levels on foreclosure rates. This statistical significance not only supports the hypothesis, but also confirms the research that increased construction can be attributed as one of the causes of the housing bubble and financial crisis.

When looking at the housing bubble and foreclosures during this time, the increase in foreclosures can mostly be attributed to macroeconomic forces such as unemployment and median income, rather than market specific variables such as median sales price of homes. It is possible that there would be a change in these variables if studying other locations, such as areas that were severely affected by the recession, however most likely the same results would occur. Overall, these large outside forces that financial institutions could not account for or prepare for, had the greatest effect on the foreclosure rate.

Works Cited

- BEBR Population Estimates - All Races - Total. (n.d.). Retrieved December 08, 2017, from <https://www.bebr.ufl.edu/data/localities/9085/county>
- Hedberg, W., & Krainer, J. (September 2012). Housing Supply and Foreclosures. FEDERAL RESERVE BANK OF SAN FRANCISCO. Retrieved from <http://www.frbsf.org/economic-research/files/wp12-20bk.pdf>
- Home Value and Home Prices - Single Family Homes - Median Sales Price (\$). (n.d.). Retrieved December 08, 2017, from <https://www.bebr.ufl.edu/data/localities/3555/county>
- U.S. Census Bureau. (n.d.). Housing - Monthly Building Permits Single-Family Units. Retrieved December 08, 2017, from <https://www.bebr.ufl.edu/data/localities/345/county>
- U.S. Census Bureau. (n.d.). Income - Median Household Income (SAIP). Retrieved December 08, 2017, from <https://www.bebr.ufl.edu/data/localities/207/county>
- U.S. Census Bureau. (n.d.). Small Area Income and Poverty Estimates (SAIPE) Program. Retrieved December 09, 2017, from <https://www.census.gov/programs-surveys/saipe/about.html>
- U.S. Department of Housing and Urban Development. (2008, October 20). HUD Provided Local Level Data. Retrieved December 08, 2017, from https://www.huduser.gov/portal/datasets/nsp_foreclosure_data.html
- U.S. Department of Housing and Urban Development. (2009). State of the Cities Data Systems (SOCDS). Retrieved March 18, 2018, from <https://socds.huduser.gov/permits/index.html>
- United States, U.S. Department of Housing and Urban Development. (2008, October 20). *Neighborhood Stabilization Program: Methodology and Data Dictionary for HUD*

Provided Data. Retrieved December 8, 2017, from

https://www.huduser.gov/portal/datasets/nsp_foreclosure_data.html