Connor Boone
Research Practicum

Introduction

The US Department of Housing and Urban Development (HUD) has adopted mechanisms to promote housing affordability for all US residents within its jurisdiction and achieves this goal by a number of specific market manipulations and funding techniques. An example such tools used by HUD are the government-sponsored enterprises (GSEs) Fannie Mae and Freddie Mac, which are two publicly traded, financial service agencies under the conservatorship of the US Department of Treasury. They are tasked with engaging with mortgage originators in the secondary mortgage market to insure and securitize mortgages loans into financial assets known as mortgage-backed securities (MBS). In 2017, it was found that these agencies have issued or guaranteed 95 percent (or more) of MBS issued since 2008 (USGOA, 2019). In 2019, it was found that collectively the organizations issued debt instruments in the amount of some $1.302 trillion (Freddy Mac, 2019).

Another tool used by HUD for economic and housing development was authorized under the Housing and Community Development Act of 1974. This act represents one of the biggest federal policy directives since the Great Depression in where the government took a more direct role in promoting home ownership among the populace and also added more explicit types of housing protections for underprivileged or marginally discriminated borrowers and renters. Further, this program initiated a type of block grant program known as the Community Development Block Grant Program, or CDBG. CDBG is a block grant program that provides funds to underdeveloped areas throughout the US by providing direct funding to local governments to disperse directly to eligible participants (HUD).

The Homebuyer Assistance Program (HAP) uses CDBG funding as a type of loan or block grant given to potential homebuyers in the Northern Kentucky Market. This program is administered by municipalities that lie within its borders. This program is based on annual household income, attendance of HUD education programs, and the property adhering to specific coding requirements. The focus of this investigation is to examine whether there exists a measurable correlation between foreclosure rates and being a participant in HAP. Foreclosure is an event where a lender or civil authority initiates legal action against a homeowner to reclaim debt owed to them through repossession of a property.

This study will focus on the City of Covington, located in Northern Kentucky. Covington is a municipality that shares the county seat of Kenton County with Independence, Kentucky. In 2017, the City had a population of around 40,500 people, and is directly influenced by its proximity to the Cincinnati metropolitan region (US Census, 2018).

Literature Review

Extensive research exists regarding the intricate relationship between down payments, home purchasing and mortgages. But, as noted by Freeman and Harden, little research directly investigates the relationship between down payment assistance and default (Freedman, Harden, 2015). Their investigation focused on using low and moderate-income homeowners to gauge the reliability of mortgage bowers who have used down payment assistance programs. Their conclusions find that many socioeconomic factors such as race, education, age, and credit worthiness play important parts in assessing the health of these types of loan, but conclude that there is no difference in mortgage performance between those who used down payment assistance versus those who did not (Freeman, Harden, 2015).

Other studies into foreclosures such as an investigation through Case Western University in 2008, saw that certain factors, such as subprime lending, were the leading predictor of foreclosure, and that subprime mortgages accounted for 82% percent of foreclosures between 2005 and 2006. They also found that specific lenders were mostly responsible for initiating civil action via a *Lis pendens* filing with the intent of initiating foreclosure proceedings on a property (Coulton, et. all, 2008).

Data

The City of Covington is subdivided into some 15,997 individual land parcels, called Parcel Identification Numbers, or PIDNs, which are area plots owned by a specific entity. These PIDN values served as id values that organized three distinct cohorts. Parcels are grouped into 24 neighborhoods in Covington that vary widely in socioeconomic conditions. Each parcel has characteristics that are recorded by municipal and county governments for tax keeping purposes. The data for this analysis was collected via several municipal outlets, including, the City of Covington and the Kenton County Property Value Administration.

Data was organized into cross-sectional cohorts that were classified according to each specific parcel's year of sale. This project was severely hampered by the unavailability of data, and because of this, sales data by parcel was only available from 2014, onward. Further, the way in which these data are organized within the local government made data collection and value matching very difficult. For this reason, for each year of total sales, the cohorts only contained around half of their total values to due missing covariate characteristics such as building square footage and year of construction. If this data does exist, it is not readily available.

Cohorts only contained residential property transfers as only residential homebuyers are eligible under HAP guidelines to apply for the program. HAP homebuyers and their non-grant receiving peers were classified into three separate groups organized by PIDN and year of sale.

Prior research into this topic showed the effectiveness of combining a Cox proportional hazard model with that of an ad-hoc hedonic model, which resulted into the cross-sectional cohorts (Coulton, et. all, 2008). Hedonic characteristics used were building square footage (sqft) and age (year of construction minus end year of survey data, 2018), and value at time of property transfer (saleval). All of these housing characteristic values were logged. The years of data for the investigation were gathered between 2014 and 2018. Summary Statistics for the data are displayed in **Tables 1, 2,** and **3.**

[Table 1.] **[Table 2.]** **[Table 3.]**

The HAP variable is a dummy bivariate that takes the value of either 1 or 0, where 1 indicates if the parcel was part of the HAP program during the cohort year and takes a 0 value otherwise. The fdummy and sdummy are also bivariates that take the value of either 1 or 0, representing whether or not the parcel was either foreclosed upon or sold during time t_n. Spell is a variable that represents the number of months that have passed from when the parcel was originally sold until a censor event (foreclosure or sale) was triggered. When a censor wasn't triggered, the value for spell was t_n.

The relevance of using housing characteristics in the model were based on several studies that employed the Cox proportional survival model to estimate foreclosures likelihood, such as Towe and Lawley (2010), who used housing values in estimating the contagion effect of foreclosures among neighbors, and Larsen (2012), who used housing characteristics such age of house in years and lot size in square feet, to investigate whether lender experience of reselling a foreclosed property in a certain area is related to the probability of sale.

Survival Regression

In this section, the model used will be briefly described to explain the benefits of using a survival regression for this type of analysis. Survival regressions, which are typically used in fields such as health economics or biostatistics, are measures used to identify factors that affect the survival, or fitness, of an individual in reaching or lasting to point t_i, *ceteris paribus*. Fitness could be used to describe something such as the likelihood of survival, or chance of remission, when an individual takes a novel medicine to treat a fatal disease.

From another perspective, however, a researcher must view the positive event as also a negative one, for who's to say that the glass is always half full? Now, instead of looking at the probability of survival, she begins to look at the *hazard*, or natural risk, that exists for the patient to perish from their novel treatment. Larsen acknowledges that in the mid 1980s, research began to use these types of regressions to evaluate real estate transactions, such as Green and Shoven, who in 1986 used this type of regression to evaluate early mortgage loan payoffs (Larsen, 2012).

There is another useful benefit for using this type of regression, which allows for concurrent, secondary censor events to affect hazard rates calculation, but without causing a direct change on overall hazard for the population. The censor event in this particular model is foreclosure, but a secondary censor for sale or transfer of property was added that allows for a specific observation to be removed from the sample pool without affecting hazard calculation. The hazard function at a particular time, t, with covariate vector(s) is defined as:

$$hazard,\ h(t) = h_0(t) * \exp\left(ß_{HAP_i}x_1 + ß_{saleval_i}x_2 + ß_{age_i}x_3 + ß_{sqft_i}x_4\right) \qquad \text{(Eq. 1)}$$

Wherein each explanatory variable represents a static measure from time $t = 0$, until a censor event is triggered at $t = t_i$.

This model was used for each cohort to calculate a hazard rate for a home to foreclose from its sale date to time *t*, which was in this case the end of data collection (2018).

Methodology

The above discussion converges to form this null hypothesis for the investigation:

H_0	There exists no measurable difference between the calculated hazard of a HAP participant from foreclosing on their property versus someone who did not participate in HAP.

The preliminary models that were created we used to create a graphic known as a Kaplan-Meier non-parametric analysis, which allows for a product limit estimator to be used to calculate survival values for data sets that include failures and suspensions. This equation, where *S,* or survival probability, that life is longer than t_i is given by:

$$\hat{S}(t) = \prod_{i:t_i \leq t}(1 - \frac{d_i}{n_i}) , \qquad \text{(Eq. 2)}$$

Where time t_i represents a time in which at least one censor event has occurred, d_i the *number of events* (i.e., foreclosures or sales) that happened at time t_i and n_i the *individuals known to have survived* (indicating they have **not yet** been censored) up to time t_n. (Kaplan, Meier, 1958)

This model was used to construct survival curves for each cohort, which are shown below.

[Figure 1.] **[Figure 2]** **[Figure 3.]**

The resulting hazard that results from each of the cohorts follow similar patterns. However, certain characteristics distinguish themselves. For one, the number of HAP participants that foreclosed on their properties from t_i to t_n was different across each of the datasets, with values being: (2/46) for 2014, (2/42) for 2015, and (0/47) in 2016, respectively. This will be important when ascertaining the significance of the regression results.

Each covariate is essential in measuring the frequency of occurrence for a hazard ratio over time. This puts great importance on the hazard ratio, which is calculated using the beta coefficients of its survival, in this case Cox, regression. In the case of a continuous covariate, the effect change in the hazard function is proportional to the calculated value for the hazard ratio. This is effectuated through changes in the covariate selling rate functions. When the hazard ratio is greater than 1, it means there is an increase in the selling rate function, and when this value is less than 1, this would indicate a decrease in the selling rate function.

For example, a hazard ratio of 0.10 means that a one-unit increase in the covariate causes a 1% decrease in the selling rate function. Conversely, a 1.10 hazard ratio would indicate that for a one unit increase in the covariate value would lead to a 10% increase in the selling rate function. If the covariate is a dummy bivariate, the hazard ratio relates the hazards for the two different types of said covariate.

Empirical Results

[Table 4.]

In order to understand the co-interactions between HAP and the other variables used in this regression, it is imperative to understand the relationship between spell, or time from t_i to censor or to t_n. When examining the results of this table, there are two tests that are the most useful in determining the reliability of the data.

For one, the LR test or the likelihood-ratio test assesses the goodness of fit of two competing statistical models based on the ratio of their likelihood functions, which are statistical parameters for a given set of observations that are equal to a joint probability distribution. When examining the results of this test, it seems that only the 2016 cohort is significant at the 10% level. This would also be useful for explaining the Logrank test of these regressions. The Logrank Test is used to compare the survival distributions between two samples. The null hypothesis for this test states that the two groups that exist within the model have identical hazard functions, which is expanded mathematically as:

(Logrank Test, Eq. 3):

$$H_0: h_1(t) = h_2(t)\ ;$$

$$H_1: h_1(t) \neq h_2(t).$$

Now, when reexamining the testing values from **Table 4.** without their respective covariates, some conclusions can begin to be made regarding the logic behind these regressions. First, the only regression that seems to be significant when only regarding HAP vs non-HAP is the 2016 model. Additionally, the HAP variable, calculated to be -17.144, is insignificant at the 10% level meaning that any effect that this variable would have on the selling rate function is insignificant. Finally, the Logrank test identifies that the null hypothesis for this statistic can be rejected at the 5% level, indicating that there is a difference in the hazard functions between the two sample populations.

Because the 2016 HAP participants didn't experience *any* foreclosures between 2016 and 2018, it could be said that being a HAP participant in this year mean that a parcel was *less* likely to foreclose than one which didn't participate in HAP. The degree to which this difference exists however was unable to be ascertained as the data was tabulated. An important caveat to these assumptions is that these regressions, as they currently lack covariates, are very weak in their robustness, and in order for a clearer picture to be presented it is important to include other types of data.

[Table 5.] **[Table 6.]**

The final results of the regressions are displayed above. The logged variables, age and sqft, and exponentiated to calculate their proportional hazard. Some interesting takeaways from these results shows the relative uniformity between the hazard rates of saleval and sqft, which could be thought to be attributes that would cause a foreclosure, which had hazard ratios of $\approx$ 1 and 0.5, respectively. A hazard rate close to 1 indicates that the covariate has no effect, and a value of 0.5 indicates that about half of the individuals in a treatment group are experiencing a censor compared to the control group.

Because in 2015 and 2016, the hazard value for sqft was calculated to be $\approx$ 0.5, this would indicate that expected hazard for individuals who participated in HAP, relative to a 1 ft^2 increase in total square footage, was 50% lower when compared to that the control group. The age variable in the 2014 data was significant at the 5% level, which means that three times the number of events are expected to be triggered in the hazard of the treatment group compared to that of the control group for each one year increase in age, *ceteris paribus*. None of the hazard ratios for the HAP bivariate in any cohort were significant at the 10% level.

Conclusions

Regrettably, the data as it was ran provided insufficient results to clearly identify whether individuals who took HAP assistance were or were not more likely to foreclose upon their mortgage. The results did show some encouraging statistics, however. The number of people who receive HAP is increasing, with only 22 people receiving the assistance via the program in 2014 versus 48 in 2016. Also, the total number of foreclosures in the City is also decreasing from 81 in 2014 to 73 in 2016.

The most critical blow to the statistical significance of conclusions drawn was the lack of sales data from when this program was originated in Covington in 2008, until 2014. This sales data would have painted a vibrant picture of the relationship between HAP participation and foreclosure rates, especially because these sales years were important to the Great Recession. Because this data was incomplete, the results were negatively affected. It is imperative for municipal governments to prioritize safe-keeping policies for their data in order to properly be able to show higher-tiered levels of government the success of their social welfare programs. This is vital to secure steady access to revenue, in the case of Covington, from both Frankfort and Washington.

Continued research on this topic is important due to address negative connotations regarding those who receive financial assistant for the government in order to secure housing. Additional research and data collection are necessary to bolster the claim that increased, direct financial assistance is an effective tool for governments to ensure equal access to homeownership amongst all income strata.

Tables and Figures.

2014 Property Sales Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
hap	473	0.04	0.2	0	0	0	1
fdummy	473	0.1	0.3	0	0	0	1
sdummy	473	0.4	0.5	0	0	1	1
spell	473	44.6	20.5	1	30	60	60
saleval	464	10.9	0.8	7.7	10.3	11.4	13.3
sqft	473	7.4	0.4	6.4	7.1	7.6	8.4
age	473	4.4	0.6	1.1	4.3	4.7	5.4

Table 1.

This table shows the summary statistics for the 2014 Cohort. Saleval, sqft, and age are all logged variables.

2015 Property Sales Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
hap	525	0.05	0.2	0	0	0	1
fdummy	525	0.1	0.3	0	0	0	1
sdummy	525	0.3	0.5	0	0	1	1
spell	525	38.0	15.1	1	27	48	48
age	525	4.3	0.9	0.0	4.1	4.7	5.2
sqft	525	7.4	0.4	6.5	7.1	7.6	8.7
saleval	525	10.9	1.1	0.0	10.4	11.5	13.7

Table 2.

This table shows the summary statistics for the 2015 Cohort. Saleval, sqft, and age are all logged variables.

2016 Property Sales Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
hap	465	0.1	0.3	0	0	0	1
age	465	4.4	0.7	0.0	4.5	4.7	5.4
sqft	465	7.4	0.4	6.1	7.1	7.6	9.3
saleval	465	10.9	1.3	0.0	10.5	11.5	14.3
spell	465	27.9	11.4	2	16	36	36
fdummy	465	0.1	0.3	0	0	0	1
sdummy	465	0.3	0.5	0	0	1	1

Table 3.

This table shows the summary statistics for the 2016 Cohort. Saleval, sqft, and age are all logged variables.

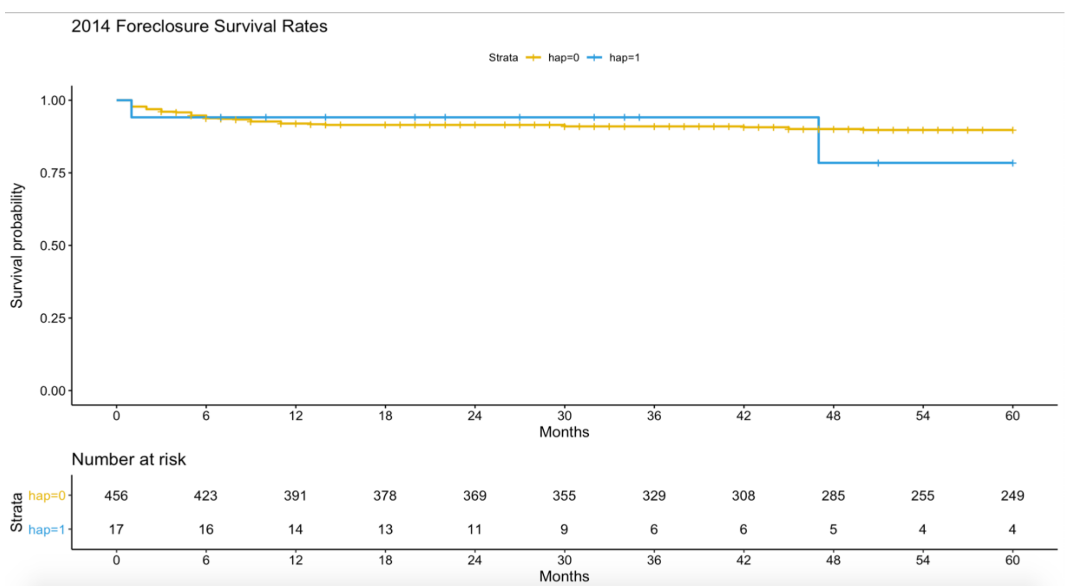

Figure 1.

This figure shows the graph for the calculated Kaplan-Meier non-parametric analysis for cohort 2014. The Yellow line represents homeowners who did not receive HAP benefits and the blue line represents those who did during FY2014. Underneath this graph, the risk table shows the number of individuals who were at risk of censoring a time *n*.

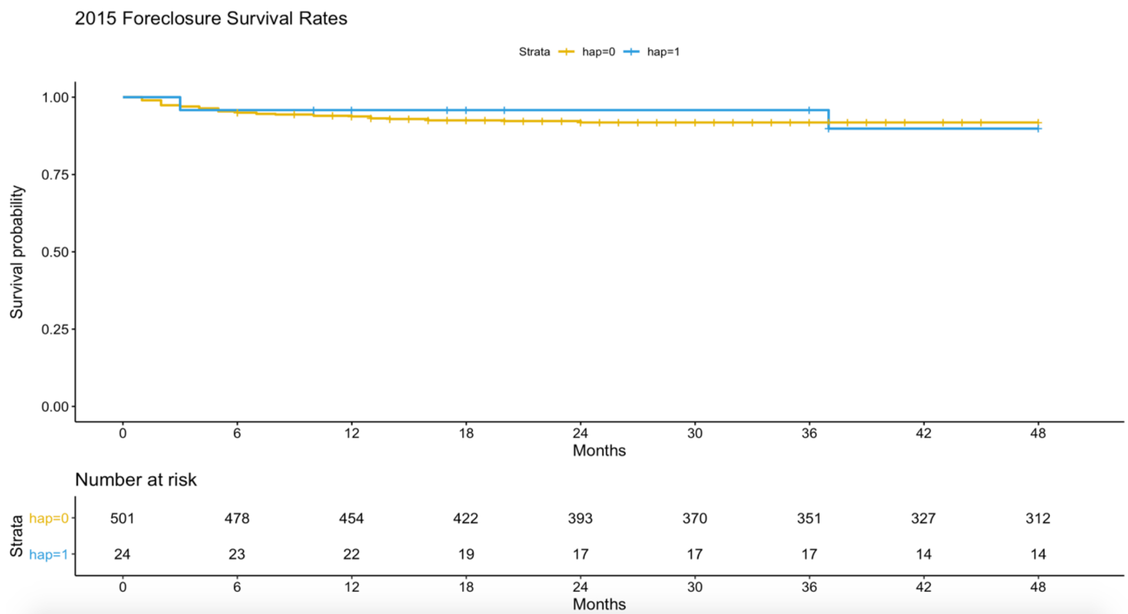

Figure 2.

This figure shows the graph for the calculated Kaplan-Meier non-parametric analysis for cohort 2015. The Yellow line represents homeowners who did not receive HAP benefits and the blue line represents those who did during FY2015. Underneath this graph, the risk table shows the number of individuals who were at risk of censoring a time *n*.

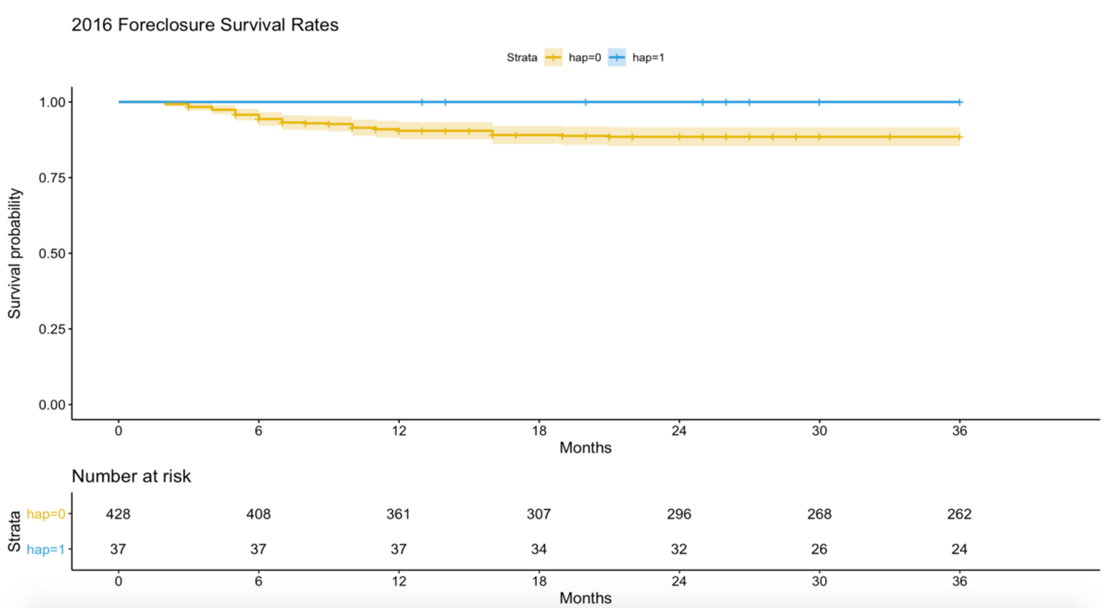

Figure 3.

This figure shows the graph for the calculated Kaplan-Meier non-parametric analysis for cohort 2016. The Yellow line represents homeowners who did not receive HAP benefits and the blue line represents those who did during FY2016. Underneath this graph, the risk table shows the number of individuals who were at risk of censoring a time *n*.

Cox Proportional Analysis of Foreclosure Rates

	Dependent variable:		
	spell		
Cohort	(2014)	(2015)	(2016)
hap	0.278	0.052	-17.144
	(0.723)	(0.725)	(2,511.273)
Observations	473	525	465
R^2	0.0003	0.00001	0.018
Max. Possible R^2	0.692	0.629	0.706
Log Likelihood	-278.299	-260.263	-280.254
Wald Test (df = 1)	0.150	0.010	0.000
LR Test (df = 1)	0.136	0.005	8.452***
Score (Logrank) Test (df = 1)	0.149	0.005	4.423**

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

Table 4.

This table shows the Cox Proportional Analysis of the HAP variable regressed exclusively on each of the cohorts' dependent variable, spell. The results are discussed in Empirical Results.

Cox Proportional Analysis of Foreclosure Rates

	Dependent variable:		
	spell		
	(2014)	(2015)	(2016)
saleval	0.107	-0.155	-0.070
	(0.206)	(0.115)	(0.098)
hap	0.458	-0.021	-17.262
	(0.726)	(0.727)	(2,543.237)
age	1.091**	0.365	-0.062
	(0.485)	(0.334)	(0.207)
sqft	-0.598	-0.688*	-0.726*
	(0.433)	(0.415)	(0.371)
Observations	464	525	465
R^2	0.021	0.018	0.029
Max. Possible R^2	0.690	0.629	0.706
Log Likelihood	-266.664	-255.530	-277.589
Wald Test (df = 4)	6.710	8.190*	5.180
LR Test (df = 4)	9.618**	9.471*	13.784***
Score (Logrank) Test (df = 4)	7.081	8.566*	9.769**
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 5. (left)

This table shows the total results of the Cox Proportional Analysis.

Table 6. (lower left)

This table used the values from the calculated betas of the regressions show in **Table 5.**, to create a table displaying the hazard ratios of each variable in the regression.

Hazard Table

Covariate (p-value)	Cohort: (2014)	(2015)	(2016)
saleval	1.11	0.86	0.93
	(0.603)	(0.177)	(0.473)
hap	1.58	0.98	3.19e-8
	(0.528)	(0.977)	(0.995)
age	2.98**	1.44	0.94
	(0.025)	(0.275)	(0.765)
sqft	0.55	0.50*	0.484*
	(0.168)	(0.097)	(0.051)

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

Works Cited:

Coulton, C., Chan, T., Schramm, M., & Mikelbank, K. (2008). Pathways to foreclosure: A longitudinal study of mortgage loans, Cleveland and Cuyahoga County, 2005–2008. Cleveland, Oh.: Case Western Reserve University, Mandel School of Applied Social Sciences, Center on Urban Poverty and Community Development.

Department of Housing and Urban Development. (n.d.). COMMUNITY DEVELOPMENT BLOCK GRANT PROGRAM - CDBG.

Freddie Mac. (2019, May). Freddie Mac Investor Presentation.

Freeman, A., & Harden, J. J. (2015). Affordable homeownership: the incidence and effect of down payment assistance. Housing Policy Debate, 25(2), 308-319.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282), 457-481.

Lawley, C., & Towe, C. (2010). The contagion effect of neighboring foreclosures on own foreclosures.

U.S. Census Bureau QuickFacts: Covington city, Kentucky. (2018, July 1). Retrieved from https://www.census.gov/quickfacts/fact/table/covingtoncitykentucky/PST045218