

# Economic Dimensions of the Foreclosure Crisis: A Focus on the New York City MSA

**Sean P. MacDonald\***

**Eric Doviak\*\***

## ABSTRACT

Employing Home Mortgage Disclosure Act data on loan originations from 2004 – 2010 and 2010 New York State pre-foreclosure filing notices, this study seeks to identify the correlation between some characteristics of census tracts and the distribution of pre-foreclosure filing notices and high cost loans within the New York City metropolitan area from 2006 through 2012. The findings are examined within the context of the census tracts within which borrowers resided. American Community Survey data on employment-population ratios, poverty rates, and median household income were then matched to our HMDA-PFF dataset to obtain a measure of the relationship of particular census tract variables to default rates. Our analysis of the data finds that differences in census tract characteristics have a statistically significant effect on default and foreclosure patterns.

## 1 INTRODUCTION

This study examines mortgage default rates across census tracts within the New York City metropolitan area from 2006 through 2012. Our investigation begins with the question of the extent to which community characteristics - employment to population ratios, poverty rates, the ratio of loan amounts to median home values by census tract, and median household income explain variances in default and foreclosure rates across the New York City metropolitan area's census tracts.

In an earlier study, "Who Defaults on their Home Mortgage?" (Doviak and MacDonald, 2012), the authors examined the significance of loan characteristics, including measures of loan amount, interest rate paid, loan type, and high cost vs. non-high cost loan, as well as borrower characteristics, including income, race, ethnicity, and the presence of a co-borrower on the probability of loan default. This study found a strong correlation between the receipt of a high-cost loan and/or a pre-foreclosure filing notice and the loan amount, borrower race and ethnicity, and borrower income.

A pre-foreclosure filing notice (PFF) is a formal notification that mortgage servicers are required to send to delinquent borrowers at least 90 days prior to filing for foreclosure on a primary residence in the State of New York. Under the law (enacted on December 15, 2009), servicers must inform homeowners that their loan is in default, indicate the amount necessary to cure the default and indicate measures that can be taken to avoid foreclosure, such as negotiating a loan modification with their lender and/or consulting with a non-profit housing counselor (New York State Division of Financial Services, 2009).

In an effort to identify loan and borrower demographic characteristics that may make a borrower more likely to default, we matched 2004 – 2008 loan origination data from the Home Mortgage Disclosure Act (HMDA) to 2010 pre-foreclosure filing (PFF) data from the New York State Department of Financial Services (DFS) and traced loans from origination to default.<sup>1</sup>

---

\*New York City College of Technology, City University of New York

\*\*Baruch College, City University of New York

Three key findings emerged from this study. One was that a strong predictor of default is the amount borrowed. This finding would not be surprising. Everything else equal, a borrower who took out a larger loan was left with less equity in the home (and often negative equity) after home prices tumbled during the recent economic and financial crisis.

Second, borrowers with incomes between \$80,000 and \$200,000 appeared to have received disproportionately more pre-foreclosure filing notices than borrowers in lower and higher income ranges. Black and Latino borrowers were disproportionately represented within this income range, suggesting another reason the foreclosure crisis had a greater impact on these borrowers.

Finally, the interest cost of the loan was found to be a strong predictor of default. Borrowers who obtained “high-cost” loans were more likely to receive a pre-foreclosure filing, and lower-income borrowers were more likely to have received a high-cost loan than borrowers with higher incomes.

While the study could not conclusively prove discriminatory lending practices, it did find that Black and Latino borrowers were more likely to have received a high cost loan, since the HMDA data captures the difference in the rate spread between loans originated to minorities and loans originated to whites.

However, the inability to prove discrimination stems from the limitations of the HMDA data itself. There is simply no basis upon which to conclude that borrowers paying higher interest rates were doing so because they were deemed by the lender to be “higher risk.” The HMDA data, lacking measures of borrower credit scores and loan-to-value, simply makes drawing such a conclusion difficult.

## **2 SUMMARY OF STUDY**

The present study seeks to expand upon this initial inquiry by examining information about the characteristics of the communities within which borrowers reside through the inclusion of American Community Survey data. Were high cost loans and higher default rates correlated with census tracts characterized by particular criteria? Did communities characterized by higher poverty rates, lower employment rates, and lower ratios of applicant income to median household income experience higher rates of default and a greater prevalence of high cost loans? How significant was loan amount to median home values in these same communities?

This study, therefore, continues the examination of variables that have contributed to default rates by employing the same matched PFF-HMDA data set together with American Community Survey census tract level data to study the influence of community factors on loan defaults and the percentage of high cost loans.

The review of the literature that follows briefly examines studies that have examined some of the characteristics of default and foreclosure patterns based upon the characteristics of the communities in which loans were originated. Some of these studies employ either Census tract or MSA level data in conjunction with HMDA data to obtain measures of the racial/ethnic composition of metropolitan areas to examine which communities had the highest concentration of high cost loans. Others examine differences in mortgage interest rates by race and ethnicity by census tract using either the Census Bureau’s American Housing Survey or American Community Survey data. Studies that have been able to access a proprietary data set of actual loans have been able to make the connection between loan and borrower characteristics more convincingly, and have been able to point to some evidence of different lending

practices based on race and ethnicity. Overall, however, studies lacking borrower data are confronted with the limitations inherent within the HMDA data, which lacks critical borrower information such as credit scores and loan-to-value measures, making a generalization applicable across neighborhoods difficult.

Thus in terms of our research, there would be a two-step process in documenting evidence of discriminatory lending. First is the necessity of identifying the correlation between some characteristics of census tracts and the distribution of pre-foreclosure notices and high cost loans. Second would be a more formal analysis that examines the probability of default within predominantly minority and non-minority census tracts, measured as a function of both census tract characteristics and borrower characteristics.

The focus of the current investigation, however, is limited to *the first of these goals* - to identify variables that could provide a more robust measure of the correlation between community characteristics, high cost loans and default rates. Abundant data document the increased reliance on debt financed household spending merely to 'keep up' with the rising cost of living. This trend is a strong indicator of rising income inequality. Thus measures of household income, poverty rates, employment to population ratios, and measures of loan to value are key in this inquiry. Did differences in median household income, poverty rates and employment levels within census tracts contribute to different outcomes in terms of the likelihood of default (as measured by the receipt of pre-foreclosure filing notices) or receipt of a high cost loan?

The data employed for this study include 2010 New York State pre-foreclosure (PFF) filings, Home Mortgage Disclosure Act data for loan originations from 2004 through 2008, and American Community Survey data from the U.S. Census Bureau. The detail obtained from the matched PFF-HMDA originations data has enabled an analysis of the characteristics of borrowers who were more likely to default, a comparison of loans that entered the foreclosure process to those that did not, and a comparison of defaulted loans across years in which they were originated. The American Community Survey (ACS) data facilitate development of a more comprehensive picture of the characteristics of the communities in which borrowers reside through the inclusion of data on poverty and employment rates, the ratio of applicant income to median household income, and the ratio of loan amount to median census tract home value, which serves as a proxy for a measure of loan to value.

The paper's methodology begins with the matched pre-foreclosure-HMDA originations dataset for the five boroughs of New York City and then identifies the census tract of each property in the PFF dataset. These data are then compared to a database of census tract coordinates available from the Census Bureau. Thus, with demographic information on borrowers from HMDA and loan characteristic data from the pre-foreclosure filings, the ACS data make it possible to extract aggregate demographic data at the census tract level to obtain measures of default rates and high cost loans based upon census tract characteristics.

We then conclude with a summary of the principal findings of our analysis. In brief, we find that the share of both high-cost loans and pre-foreclosure filing notices tended to be greater in census tracts characterized by higher poverty rates, lower employment-population ratios, and higher loan-to-value ratios.

### 3 REVIEW OF THE LITERATURE

A number of recent studies have looked at HMDA data in the context of the communities within which borrowers reside to discern lending patterns in minority communities. In many such studies, it is not

difficult to demonstrate that high-cost lending was most prevalent in predominantly minority communities, but it has always been difficult to take these analyses a step further to show conclusive evidence of discriminatory lending practices because of the lack of key variables – principally credit scores and loan-to-value data - in the HMDA data. In cases where studies have had access to proprietary or other loan datasets, there has been more of a foundation on which to link loan type to race (Bocian et al., 2006; Gerardi and Willen, 2008).

However, when employing just the HMDA data to study the broader market, researchers are generally confined to finding correlations between community characteristics such as race, ethnicity, and similar demographics and the probability of receiving a high-cost loan. This is still the case with studies that examine trends at the MSA level.

Rugh and Massey (2010) for instance attempt to link the correlation between high-cost lending and the patterns of residential segregation to the subprime foreclosure crisis. To find the link, they obtain the total number of foreclosures between 2006 and 2008 from RealtyTrac's foreclosure database and compute the foreclosure rate as the number of filings per household unit. They then use the 2004-2006 HMDA data to compute the share of high-cost loans<sup>2</sup> in each MSA. To derive a measure of regulatory oversight, they also compute the share of loans within the MSAs that were originated by institutions covered under the Community Reinvestment Act (CRA). Employing an OLS multiple regression model, the authors regress the number and rate of foreclosures in the nation's 100 largest MSAs on two measures of segregation: residential unevenness and spatial isolation. Their results suggest that residential segregation and the share of high-cost loans are both positively correlated with the number and rate of foreclosures across U.S. metropolitan areas.

One omission in their published paper however is a regression of the high-cost lending share on measures of racial and ethnic segregation. If segregation enabled lenders to target minorities for high-cost loans (as Rugh and Massey claim) then the next step should have been to regress the high-cost lending share on measures of segregation. If the coefficient were positive and statistically significant, then their claim of racial and ethnic targeting would have a firmer foundation.

Other studies have employed either census tract or MSA level data in conjunction with the HMDA data to obtain measures of the racial/ethnic composition of metropolitan areas to examine which communities had the highest concentration of high cost loans (Squires et al., 2009). Following is an overview of studies that have examined HMDA data in the context of where borrowers reside to discern patterns of lending in minority communities. Some of these studies examine data at the MSA level, while others examine census tract level data in investigating lending patterns.

Squires et al. (2009) use the 2000 Census data to construct a dissimilarity index to obtain a measure of the ten most segregated and the ten least segregated metropolitan areas in the U.S. They then compare the indices derived to the percentage of high-cost loans originated. Using 2006 HMDA data and the 2006 American Community Survey, they employ a multivariate OLS model (to control for several MSA-level variables) and find that racial segregation is a significant predictor of the percentage of high-cost loan originations in an MSA. Their results suggest that a 10 percent increase in black segregation was associated with a 1.4 percent increase in high-cost loans.

In another study using 2001 American Housing Survey data, Susin (2003) examines differentials in mortgage interest rates by race and ethnicity at the census tract level. Employing a sample of homeowners with mortgages, Susin examines interest rates in relation to a number of both loan and borrower characteristics in the context of the census tracts in which they reside. His study considers

the race/ethnicity of the borrower, home value and loan characteristics as well as neighborhood characteristics, including poverty rates and the percentage of Black and Latino residents in the census tracts in which the borrowers reside. His study finds that African American borrowers paid an average of 44 basis points more than white borrowers on their loans.

In an extensive 2006 study, Boehm and Schlottmann examine a pooled sample of MSA specific American Housing Survey data. Specifically, they select a large sample consisting of 5,000 households for each of 41 MSA's from 1998 through 2004, resulting in a rather large sample of 200,000 observations. The authors look at both conventional and FHA/VA mortgages to identify differences in mortgage interest payments made by Black, Latino and white borrowers of first lien mortgages. They find that across the full sample, Black borrowers paid higher interest rates on first mortgages in both the conventional and FHA/VA markets. Both non-white and white Latino borrowers paid significantly more in the conventional loan market than their white counterparts at 14.6 and 9.2 basis points, respectively. Blacks paid an additional 30.6 basis points.

Others have also found a link between the racial composition of a neighborhood and the share of subprime lending in that neighborhood at the MSA level. In a joint study conducted by several community organizations, Bromley et al. (2008) focused on subprime lending activity in 2006 across seven large metropolitan areas in the U.S. Data collected on the number of high-risk loans originated by a sample of 35 subprime lenders during that year indicated that these lenders accounted for an estimated 20 percent of the market share of subprime loans in predominantly minority neighborhoods within these metropolitan areas. Further, more than 40 percent of the loans made by high-risk lenders in these metropolitan areas were in neighborhoods where the share of minority residents was 80 percent or more. Subprime lenders' market share was also positively correlated with the percentage of minority residents within a given census tract.

While these and similar studies clearly demonstrate that Black and Latino borrowers obtained a disproportionately greater share of high-cost and subprime loans, the evidence that this trend reflects discrimination suffers from the limitations inherent within the HMDA data. The lack of information on credit scores in the HMDA data may explain some of the disparities in the rate spreads among individual borrowers, but it is difficult to see how this could be applicable across neighborhoods. In other words, it is certainly possible to imagine individual cases where a high-income black borrower's credit score is lower than a low-income white borrower's credit score; at the same time, however, it is difficult to see how the average credit score of borrowers in a high-income black neighborhood could be lower than the average credit score of borrowers in a low-income white neighborhood.

With the exception of Susin (2003), these studies have focused their analyses mostly at the MSA level in their examination of the relationship between the race and ethnicity of borrowers and the disproportionately greater share of high cost and/or subprime loans. While they have, at the same time, contributed extensively to the literature on the evidence of discriminatory lending practices, the focus at the MSA level and or the characteristics of MSAs poses limitations in getting at the widely varying characteristics of communities within those MSAs.

The present study seeks to expand upon the scope of these inquiries into high cost lending to focus on census tract level data that include several additional characteristics of the communities that are believed to have played a role in the percentage of high cost loans and the incidence of default. Studying census tract data for the New York MSA, this analysis examines the influence on the percentage of high cost loans of poverty rates, employment to population ratios, the ratio of loan amount to median home

value, and the ratio of applicant income to median household income. The study then examines the relationship between the percentage of high cost loans and rates of default across each of these variables by employing census tract data in conjunction with HMDA-PFF data. Using such a framework, this census tract level study offers the opportunity to move beyond some of the limitations posed by use of the HMDA data alone in the study of community characteristics by including data on default rates at the census tract level. The information such a study yields can then provide the foundation for future examinations of lending practices.

#### 4 DATA

Three datasets are employed for this analysis – the New York State Pre-foreclosure filing data for 2010, Home Mortgage Disclosure Act data for 2004 through 2008, and American Community Survey Data from the U.S. Census Bureau. The Pre-foreclosure filing data from the New York State Division of Financial Services became available in 2010 with the passage of legislation requiring lenders and loan servicers to provide delinquent borrowers with 90-day advance notice that their loans were in default to give these borrowers time to work with loan counselors to reach a modification agreement before formal foreclosure proceedings began.

Thus, since February 13, 2010, mortgage servicers have been required to file the notices with the New York State Division of Financial Services (formerly the NYS Banking Department), which has collected an extraordinary level of detail on the loans. Among the many data fields collected are: the property address, the names of the borrowers, the current monthly payment, the delinquent contractual payments, the interest rate, whether the loan is a fixed-rate or adjustable-rate mortgage, the date and the amount of the original loan, the lien type, the loan term, whether the loan has been modified and whether an investor's approval is necessary to modify the loan. If the loan progresses to a *lis pendens* filing (i.e. the first step in the foreclosure process or the filing of the complaint) then servicers are also required to follow up on their initial filing with information on the entity filing for foreclosure.

The detail captured in the PFF data makes three forms of analysis possible. First, the defaulted loans can be matched to publicly available data on originations from the Home Mortgage Disclosure Act (HMDA). By combining the HMDA and PFF data, it is possible to see which borrowers were more likely to default. Second, it is possible to compare the loans that entered the foreclosure process to those that did not, and finally, the "Full PFF" dataset can be used to compare defaulted loans across the years in which they were originated.

The first two datasets – which compare originations to defaults and compare defaults to *lis pendens* filings – effectively generate a quasi-longitudinal analysis, which makes it possible to track the universe of New York State home mortgages from origination to default to foreclosure. We use the term "quasi-longitudinal" however, because the PFF data only provide information on borrowers who defaulted in 2010. Data on the borrowers who did not default or defaulted in subsequent years are not available. While there is no way to perfectly match the PFF data to the HMDA data, the matching strategy employed generates a reasonably accurate result, although we cannot be certain of full verification of all matches.

These combined data have facilitated the identification of borrower characteristics and loan level characteristics that are more likely to be correlated with default. Principal among some of the findings of an earlier study, was that Black and Latino borrowers were more likely than other borrowers to have received high-cost loans, pre-foreclosure filing notices and to have obtained relatively larger loans than

the population overall (Doviak and MacDonald, 2012). While the evidence discussed in our earlier study pointed to the possibility of discriminatory lending practices, and several related studies bore out similar findings, the lack of critical data either from HMDA or PFF on loan to value and credit scores made it difficult to definitively “prove” such a finding.

With the incorporation into the present study of a third data set from the American Community Survey, it is our goal to obtain a more comprehensive picture of the relationship of default rates to the communities within which borrowers reside. Thus, the ability to include data on poverty rates, employment-to-population ratios, the ratio of applicant income to median household income, and the ratio of loan amount to median home value at the census tract level effectively adds the kind of information about borrowers and their communities that may provide more comprehensive information about the correlation between the receipt of a pre-foreclosure filing notice, the receipt of high-cost loans and community demographics.

## 5 METHODOLOGY

The starting point here is the pairing of a HMDA dataset on loan originations from 2004 – 2008 with pre-foreclosure filing notices issued in 2010 to borrowers within the five boroughs of New York City. The study examines first-lien mortgages to ensure comparability across loans. The resulting 130,912 first-lien mortgages that were originated in the years 2004-2008 account for 70 percent of all PFF filings on first-lien mortgages. The years 2004-2008 were selected to enable a comparison of the PFF data to the data on originations from the Home Mortgage Disclosure Act (HMDA). We chose 2004 as the first year, because the variables in the pre-2004 HMDA data were much less extensive.

The HMDA originations data contain the FIPS (Federal Information Processing Standards) county code and census tract number of each property. The FIPS code is a five-digit code assigned to each county, the first two digits representing the state and the last three digits, the specific county within the state. This is particularly valuable because census tracts have a small population (typically between 2,500 and 8,000 people) which is fairly homogeneous in terms of socio-economic characteristics and living conditions (U.S. Census Bureau, 2000).

The first step in matching the PFF data to the HMDA data is to use the address information to identify the census tract of each property in the PFF dataset. To identify the census tracts, Erle’s (2005) “Geo-Coder-US-1.00” Perl module is used in conjunction with the U.S. Census Bureau’s (2007) TIGER/Line Files. After using this module to create a database of New York State addresses from the TIGER/Line Files, the database was queried to obtain the latitudes and longitudes of the property addresses in the PFF dataset. Once the coordinates were generated, these were compared to a database of census tract coordinates that was generated from the U.S. Census Bureau’s (2005) “Cartographic Boundary Files.”

The HMDA data provide demographic data on borrowers such as race, ethnicity, and income, as well as information on loan originators. The 2010 New York State pre-foreclosure filings data provide information on loan type, loan amount, length of delinquency and other loan characteristics. The matching of these two datasets provides us with the characteristics of borrowers who defaulted on their loans. It also provides us with information on the type of loan where default was more likely to occur.

Next, data from the U.S. Census American Community Survey are introduced to extract aggregate demographic data at the census tract level to obtain measures of default rates based upon

census tract characteristics, as well as a measure of the tract-level variables that are most significantly linked to default. Supplementing the HMDA-PFF dataset with ACS data provides a rough measure of loan-to-value and provides a clearer picture of the relationship between poverty rates, employment and labor market conditions and the foreclosure crisis at the community level.

Census tracts within the five boroughs of New York City are the focus of this study of the relationship between borrower and loan level characteristics and of the economic characteristics of the census tract within which the loan was originated. Thus, several census tract level variables are employed to determine the extent to which these are correlated with borrower default rates.

We begin with variables that can provide some measure of the relationship between income, employment levels and poverty rates and the probability of receiving a pre-foreclosure filing notice at the census tract level. Specifically, we examine borrower income relative to median household income, the employment to population ratio for the population over 16 years, poverty rates, and borrower loan data relative to median home values at the tract level. Through this process, we seek to obtain a picture of the characteristics of borrower income, loan amount, receipt of a pre-foreclosure filing and loan type, relative to the characteristics of the census tract within which the mortgage was originated.

The pairing of total number of pre-foreclosure filings (PFFs) to HMDA originations at the census tract level will make possible an analysis of default rates which can then be correlated directly with census tract characteristics. With this information, it is the goal of this study to be able to obtain a clearer picture of the correlation between the characteristics of communities at the census tract level, loan type and rates of default.

## 6 ANALYSIS, FINDINGS, SUMMARY TABLES

The study finds that the distribution of high-cost loans and default rates (as measured by receipt of a pre-foreclosure filing notice) tended to be greater in census tracts characterized by higher poverty rates, lower employment-population ratios, and higher loan-to-value ratios. This trend is revealed in the new data set combining aggregate census tract data with the HMDA-PFF data.

This combined HMDA-PFF-ACS data indicate that census tract poverty rates are significantly correlated with the receipt of a high-cost loan. As illustrated in Table 1, where census tract poverty rates range from below 2.0 percent to less than 4.0 percent, the lowest percentage of high cost loans was originated. As the percentage of population below the poverty level income rises, the percentage of high-cost loans increases substantially.

**Table 1: Percentage of Families Below Poverty Income Level by Receipt of a High Cost Loan**

Percentage of population Below poverty income in tract	High-cost	Non high-cost	Total loans	Percent high-cost
Under 2.0%	44,660	305,077	349,737	12.8%
2.0 to 3.9%	41,470	228,513	269,983	15.4%
4.0 to 7.9%	65,395	288,586	353,981	18.5%
8.0 to 11.9%	41,204	146,060	187,262	22.0%
12.0 to 15.9%	23,659	80,097	103,756	22.8%
16.0 to 19.9%	15,684	45,938	61,622	25.5%
20.0% and over	34,026	82,453	116,479	29.2%
<b>Totals</b>	<b>266,098</b>	<b>1,176,724</b>	<b>1,442,822</b>	

Source: Combined HMDA-PFF-ACS

The absolute number of high-cost loans is greatest at poverty rates between 4.0 and 7.9 percent; however, even as the number of such loans declines as the percentage in poverty rises, the percent that are high-cost continues to increase. An interesting observation here is that as the share of population below poverty within census tract reaches 20 percent and above, both the number and percentage of high cost loans increase significantly.

Interestingly, however, the percentage of pre-foreclosure filings (notices of default) by percentage of families with census tract income below poverty level seems to reveal an opposite trend (as reflected in Table 2). The percentage of PFFs received is greatest where the percentage of families below poverty level is less than 2.0 percent and between 4.0 and 7.9 percent. In fact, the 4.0 to 7.9 percent range accounts for nearly a quarter of all PFFs, while at poverty rates of 7.9 percent and below, the percentage of PFFs accounts for nearly two-thirds of all such notices, suggesting that the PFF percent declines as poverty rate within the census tract rises.

**Table 2. Percent of Families Below Poverty Income by Receipt of a Pre-Foreclosure Filing Notice**

<b>Percentage below poverty in tract</b>	<b>Non-PFF</b>	<b>PFF</b>	<b>Total loans</b>	<b>PFF percent</b>	<b>Ratio high cost loans to defaults</b>
Under 2.0	327,169	23,552	350,721	20.7%	1.90
2.0 to 3.9	250,032	19,456	269,488	17.1%	2.13
4.0 to 7.9	327,169	27,762	354,931	24.4%	2.36
8.0 to 11.9	171,564	15,588	187,152	13.7%	2.64
12.0 to 15.9	94,427	9,216	103,643	8.1%	2.56
16.0 to 19.9	55,858	5,689	61,547	5.0%	2.75
20.0 and over	103,736	12,402	116,138	10.9%	2.74
<b>Total</b>	<b>1,329,955</b>	<b>113,665</b>	<b>1,443,620</b>		

Data: Combined HMDA-PFF-ACS

Such findings may seem counter intuitive. However, if we consider the likelihood that fewer loan originations would be made in census tracts characterized by very high poverty rates, the finding appears less surprising. One interesting finding here is that despite this trend, the ratio of high cost loans to defaults clearly rises as the share of families with incomes below poverty within census tract rises, suggesting that among those who did receive high cost loans, the default rate was significantly higher on those loans in census tracts characterized by higher rates of poverty.

The data also reveal that community employment rates are highly significant in relation to the likelihood of receiving a high-cost loan and/or a pre-foreclosure filing notice. High-cost loans appear to be much more likely in census tracts characterized by lower employment/population ratios, likely reflecting a greater instability of employment (Table 3a). Where the employment/population ratio is 55 percent or less of the tract population over the age of 16, the percent of high cost loans is greatest and accounts for more than 44 percent of such loans. One possible explanation is that in tracts where there was a greater concentration of residents with less employment and income stability, there was a disproportionately higher percentage of high cost loans to compensate for the increased risk.

**Table 3a: Employment to Population Ratio by Receipt of a High-Cost Loan**

<b>Employment/Population (pop over age 16)</b>	<b>Number non high-cost loans</b>	<b>High-Cost loans</b>	<b>Total loans</b>	<b>Percent high cost</b>
Under 50.0%	70,603.5	21,553.9	92,157.4	23.4%
50.0 to 54.9%	125,909.6	32,996.2	158,905.7	20.8%
55.0 to 59.9%	274,176.9	65,194.0	339,370.9	19.2%
60.0 to 64.9%	375,375.3	84,353.1	459,728.3	18.3%
65.0 to 69.9%	242,405.4	50,558.6	292,964.0	17.3%
70.0% and over	88,254.4	11,442.2	99,696.6	11.5%

Source: Combined HMDA-PFF-ACS

The percentage of pre-foreclosure filing notices follows a similar pattern, although it appears somewhat less pronounced in the data (Table 3b). The share of both high cost loans and pre-foreclosure filing notices is noticeably higher where the employment/population ratio is less than 55.0 percent. In the middle of the employment/population distribution – at ratios from 55.0 percent to 66.9 percent – there is little difference in the percent of high-cost vs. non-high cost loans received or in the percent of pre-foreclosure filings. As would be expected at high rates of employment relative to population - 70.0 percent or higher – the incidence of both high cost loans and pre-foreclosure filings are lowest.

Clearly, a high employment/population ratio suggests a comparatively more stable community, one in which employment stability is most likely established and where the probability of job loss is significantly lower. It is also more likely that job loss in such communities would be temporary and re-employment easier, reducing the probability of entering into default and the pre-foreclosure process. Further, residents in tracts characterized by higher rates of employment may have been comparatively less affected by the job losses associated with the financial crisis and thus, had a greater ability to continue making mortgage payments, resulting in a lower proportion of PFFs.

**Table 3b: Employment to Population Ratio by Receipt of a Pre-Foreclosure Filing Notice**

<b>Employment/Population (pop over age 16)</b>	<b>Number no PFF</b>	<b>Number PFF</b>	<b>Total loans</b>	<b>Percent PFF</b>	<b>Ratio high-cost loans to defaults</b>
Under 50.0%	83,787.2	8,078.3	91,865.5	8.8%	2.66
50.0 to 54.9%	144,965.2	13,653.5	158,618.7	8.6%	2.41
55.0 to 59.9%	312,539.7	26,956.6	339,505.3	7.9%	2.42
60.0 to 64.9%	424,256	36,409.3	460,665.2	7.9%	2.32
65.0 to 69.9%	271,311.0	22,300.7	293,611.7	7.6%	2.27
70.0% and over	93,069.9	6,257.8	99,354.8	6.3%	1.82

Data: Combined HMDA-PFF-ACS

Also consistent with the finding of higher percentages of both high cost loans and pre-foreclosure filings at lower employment to population ratios is the observation that the ratio of high-cost loans to defaults is also higher. Although the absolute number of high cost loans and defaults at rates of

employment below 50 percent are comparatively lower than at higher employment to population ratios, the likelihood of receiving a high cost loan and then defaulting is significantly greater.

As discussed in both the literature review and in our own earlier study, one of the shortcomings of drawing a strong conclusion regarding loan characteristics relative to home value is the lack of key loan-to-value data in HMDA. The employment of the combined HMDA-PFF-ACS dataset here, however, allows for the possibility of addressing this issue by obtaining a ratio of borrower loan amounts relative to the median home value by census tract.

Analysis of pre-foreclosure filings in relation to the ratio of loan amount to median home value in census tract (Table 4) clearly indicates that as this tract loan-to-value indicator rises, the percentage of PFFs does as well. Clearly, there is a correlation indicating that the greater the amount of the home purchase financed, the greater the incidence of default. Nearly a quarter or all loans were financed at rates of 90 percent and greater, and accounted for 21.8 percent of total defaults. Such a finding isn't surprising since a greater loan burden poses a considerably higher risk of default in a declining housing market.

**Table 4. Receipt of a Pre-foreclosure Filing by Ratio of Loan Amount to Median Home Value in Tract**

<b>Ratio of loan amount to median home value in tract</b>	<b>Number no PFF</b>	<b>Number PFF</b>	<b>Total loans</b>	<b>Percent PFF</b>	<b>Ratio – high cost loans to defaults</b>
Under 50.0%	363,078.0	17,863.3	380,941.3	4.7%	2.74
50.0% to 69.0%	339,138.8	25,372.7	364,511.5	7.0%	2.35
70.0% to 89.0%	319,189.4	33,337.2	352,526.7	9.5%	2.24
90.0% to 99.0%	101,076.7	12,970.8	114,047.5	11.4%	2.24
100.0% or greater	208,803.1	24,234.9	233,038.0	10.4%	2.21

Source: Combined HMDA-PFF-ACS

The relationship between tract loan-to-value and high-cost loans (Table 5) indicates that the percentage of high cost loans increases significantly as loan-to-value increases. Close to 50 percent of high cost loans were originated where tract loan-to-value was 90 percent or more. Interestingly, while just 30 percent of all high cost loans were at loan-to-values of 90 percent or more, the data indicated the significantly greater concentration of high-cost loans where 90 percent or more of the loan was financed. By contrast, just 12.9 percent of high cost loans were associated with tract loan-to-values of less than 50 percent.

**Table 5: Ratio of Loan Amount to Median Home Value in Census Tract**

<b>Ratio of loan amount To median home value in tract</b>	<b>Number non-high cost loan</b>	<b>Number high cost loan</b>	<b>Total loans</b>	<b>Percent high cost loan</b>
Under 50.0%	331,836.5	48,962.0	380,798.5	12.9%
50.0% to 69.0%	304,771.8	59,606.0	364,377.7	16.4%
70.0% to 89.0%	277,707.1	74,773.5	352,480.6	21.2%
90.0% to 99.0%	84,724.2	29,004.7	113,728.9	25.5%
100.0% or greater	178,862.2	53,485.7	232,347.9	23.0%

Source: Combined HMDA-PFF-ACS

One conclusion that might be drawn from this is that borrower ability to repay was likely a factor

not simply where a greater share of a home purchase was financed, but in the fact that high cost loans were more likely to be made to borrowers financing nearly all of a home's purchase price.

This finding would be expected given the increased risk associated with financing proportionately more of the value of the home, saddling the borrower with a greater liability, particularly in the event of home price depreciation or job loss, increasing the probability of default. As the housing and financial crisis revealed, this is precisely what happened to so many overleveraged borrowers.

Finally, the ratio of applicant income (from HMDA) relative to median household income at the census tract level reveals a comparatively weaker relationship to the distribution of high cost loans and pre-foreclosure filings (Tables 6 and 7). The significance of income as measured in this way may be somewhat obscured by the fact that many census tracts tend to be more homogeneous. This trend finds support in several studies. Iceland and Steinmetz (2003) for instance, characterize census tracts as being relatively homogeneous with respect to population in terms of economic status and living standards. Thus, there is likely not to be the kind of large observable disparities in income that would be seen at a broader level of analysis, such as at the MSA level. This characteristic of census tracts appears to be reflected in the distribution of both pre-foreclosure filings and high cost loans at nearly all ratios of applicant income to median census tract household income.

One interesting observation is that even at high ratios of applicant income to median tract income, there are still some high cost loans. At all ratios of applicant income to median tract income, the percentage of high cost loans shows little variance, ranging from 17.8 percent where the income ratio is less than 50 percent to 19.9 percent where applicant income is between 250 and 299 percent of median tract income.

**Table 6: Pre-foreclosure Filing by Ratio of Applicant Income to Median Household Income in Tract**

<b>Ratio Applicant Income to Median Household Income in Tract</b>	<b>Number Non PFF</b>	<b>Number PFF</b>	<b>Total loans</b>	<b>Received PFF</b>	<b>Ratio high cost loans to defaults</b>
Less than 50.0%	62,507.9	3,754.7	66,262.6	5.7%	3.14
50.0% to 99.0%	424,256.0	32,085.7	456,341.6	7.0%	2.47
100.0% to 149.0%	388,347.2	34,930.2	423,277.3	8.3%	2.29
150.0% to 199.0%	199,493.4	18,659.8	218,153.2	8.6%	2.25
200.0% to 249.0%	97,086.8	9,329.9	106,416.7	8.8%	2.24
250.0% to 299.0%	53,198.2	5,233.8	58,432	9.0%	2.22
300.0% and higher	103,736.6	9,898.8	113,635.3	8.7%	2.07

Source: Combined HMDA-PFF-ACS

**Table 7: High-cost loan by Ratio of Applicant Income to Median Household Income in Census Tract**

<b>Ratio Applicant Income to Median Household Income in Tract</b>	<b>Number non-high cost loan</b>	<b>Number high cost loan</b>	<b>Total loans</b>	<b>Percent high cost loan</b>
Less than 50.0%	53,306.1	11,974.4	67,280.5	17.8%
50.0% to 99.0%	377,728.7	79,297.2	457,025.9	17.4%
100.0% to 149.0%	342,427.0	79,829.4	422,256.4	18.9%
150.0% to 199.0%	175,332.0	41,777.4	217,109.4	19.2%
200.0% to 249.0%	85,900.9	21,021.7	106,922.7	19.7%
250.0% to 299.0%	47,069.0	11,708.3	58,777.3	19.9%
300.0% and higher	92,961.3	20,489.5	113,450.8	18.1%

Source: Combined HMDA-PFF-ACS

Quite possibly, at higher income levels, there may have been a higher proportion of "jumbo" loans, which typically carry higher interest rates or that at these higher income ratios, borrowers may

have been placed into higher risk categories by lenders if they were employed in high paying, yet cyclical industries. This may also offer at least a partial explanation for the slightly higher rates of default as the ratios of applicant to median tract income rise. Higher loan amounts at higher levels of income, in addition to carrying higher rates of interest, may be at higher risk of default in a declining and/or weak housing market coupled with rising unemployment.

## 7 CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

The findings of this analysis incorporating census tract data into the matched HMDA-PFF data file provide a considerably more comprehensive picture of the relationship between community characteristics and the distribution of both high-cost loans and pre-foreclosure filings at the height of the housing boom. The additional information drawn from census tract characteristics clearly indicates that the share of both high-cost loans and pre-foreclosure filing notices tended to be greater in census tracts characterized by higher poverty rates, lower employment-population ratios, and where the loan-to-value ratio was very high. These data lend fairly robust support to the argument that there is a strong correlation between loan characteristics and the demographic characteristics of the communities within which borrowers live.

The findings also raise the question of the potential significance of other demographic characteristics that could be explored in a future study. For instance, it is likely that the level of educational attainment is closely related to employment to population ratios, which would offer further insight into incidence of both high cost loans and default rates. The occupational make-up of communities, along with the dominant industries at the census tract level could reveal much about median incomes of communities. Household size may also be significant as loan characteristics and default probability might be expected to vary based upon the number of workers, and thus income, in a household.

The present study, lacking key demographic information on borrowers *within* the census tracts, suggests that any correlations found in the data are not sufficient to draw conclusions regarding discriminatory lending practices. However, the findings do raise some important questions that should be the focus of future research. Why for instance were the distribution of high cost loans and pre-foreclosure filings greater in census tracts characterized by higher poverty rates? Why were loan-to-value ratios greater in higher poverty census tracts?

To investigate the question of possible discriminatory lending, further inquiry would clearly need to look more closely at the relationship of variables such as the racial and ethnic composition of census tracts, a measure of the ratio of applicant income to median census tract income, and how a community's average FICO scores are related to the probability of people in minority vs. non-minority dominated census tracts receiving high cost loans. While the finding of a clear correlation between lower employment-population ratios and a higher distribution of pre-foreclosure filings and high cost loans is not surprising here, expanded research which includes an examination of employment-population ratios within census tracts based upon their predominant racial and ethnic composition might provide some added insight into lending patterns. Thus, an examination of the likelihood of default (as measured by the probability of receiving a pre-foreclosure filing) could be examined as a function of the percent in census tract below poverty level income, the percent employed within census tract (as measured by the employment/population ratio), loan to median tract home value, applicant income to median tract income,

and dummy variables to identify applicants' race and ethnicity. The outcome of such an inquiry could provide a more comprehensive picture of the relationship between race, ethnicity and the probability of receiving a high-cost loan and/or a pre-foreclosure filing. Based upon one of the key findings of our previous research (Doviak and MacDonald, 2012), - that 35.1 percent of Black borrowers and 28.1 percent of Latino borrowers received high-cost loans - suggests and that establishing these connections between the characteristics of census tracts and their demographic make-up can provide some more definitive answers.

## ENDNOTES

1. The pre-foreclosure filing data had its origins in December 2009, when New York State's Mortgage Foreclosure Law amended the Real Property Actions and Proceedings, inserting a provision (1306) requiring mortgage servicers to send borrowers a 90-day notice prior to commencing foreclosure proceedings on owner-occupied residential mortgages. The new law also required mortgage servicers to electronically submit the pre-foreclosure filings (PFF) to the NYSBD (now the Division of Financial Services or DFS) for the purpose of putting borrowers in touch with non-profit mortgage counselors and "to perform an analysis of loan types which were the subject of a pre-foreclosure notice." Since enforcement of the law is left to the courts, servicers have had a strong incentive to submit honest and accurate filings. When deciding what information about the loans to collect from the mortgage servicers, the DFS chose to collect information that would facilitate a matching of the pre-foreclosure filings to the corresponding HMDA loan originations. Furthermore, in its two reports analyzing the PFF data, the DFS compared the PFF data to the HMDA data to estimate the mortgage default rate by county and to compare mortgage default rates by loan amount.
2. Rugh and Massey (2010) use the term "subprime" to describe high-cost loans.

## REFERENCES

- Bocian, D.G., K.S. Ernst, and W. Li. 2006. "Unfair Lending: The Effect of Race and Ethnicity on the Price of Subprime Mortgages." 31 May, *Center for Responsible Lending*. <http://www.responsiblelending.org/mortgage-lend-ing/research-analysis/unfair-lending-the-effect-of-race-and-ethnicity-on-the-price-of-subprime-mortgages.html>
- Boehm, T.P. and A.M. Schlottmann. 2006. "Mortgage Pricing Differentials Across Hispanic, Black and White Households: Evidence from the American Housing Survey." *U.S. Department of Housing and Urban Development, Office of Policy Development and Research*. [http://www.huduser.org/Publications/PDF/hisp\\_homeown5.pdf](http://www.huduser.org/Publications/PDF/hisp_homeown5.pdf)
- Bromley, C., J. Campen, S. Nafici, A. Rust, G. Smith, K. Stein, and B. van Kerkhove. Mar 2008. "Paying More for the American Dream: The Subprime Shakeout and Its Impact on Lower-Income and Minority Communities." *PA Policy Archive*. <http://www.policyarchive.org/handle/10207/19021>.
- Doviak, E. and S. MacDonald. "Who Defaults on their Home Mortgage?" 2012. *New York Economic Review*, 43, 75-98. <http://nysea.bizland.com/>
- Erle, S. "Geo-Coder-US-1.00: Geocode (estimate latitude and longitude for) any US address." 17 May 2005. *Comprehensive Perl Archive Network*. <http://search.cpan.org/~sderle/Geo-Coder-US-1.00/>.
- Gerardi, K.S. and P.S. Willen. 2008. "Subprime Mortgages, Foreclosures, and Urban Neighborhoods. 2008." *Federal Reserve Bank of Boston Public Policy Discussion Papers*, no. 08-6, 22 December. <http://www.bos.frb.org/economic/ppdp/2008/ppdp0806.pdf>.
- Iceland, J. and E. Steinmetz. 2003. "The Effects of Using Census Block Groups Instead of Census Tracts When Examining Residential Housing Patterns." *U.S. Census Bureau*. <http://www.census.gov>.

- [gov/housing/patterns/publications/unit\\_of\\_analysis.pdf](http://www.gov/housing/patterns/publications/unit_of_analysis.pdf)
- New York State Division of Financial Services. 2009. "Facing Foreclosure: Know Your Rights." <http://www.dfs.ny.gov/consumer/hetprights.htm>
- Rugh, J.S. and D.S. Massey. 2010. "Racial Segregation and the American Foreclosure Crisis." *American Sociological Review*, 75(5): 629–651.
- Squires, G.D., D.S. Hyra, and R.N. Renner. 2009. "Segregation and the Subprime Lending Crisis." *EPI Briefing Paper*, 4 Nov. <http://www.epi.org/page/-/pdf/110409-briefingpaper244.pdf>.
- Susin, S. 2003. "Mortgage Interest Rates and Refinancing: Racial and Ethnic Patterns." unpublished manuscript, U.S. Census Bureau. 2003. [http://www.huduser.org/Publications/PDF/hisp\\_homeown5.pdf](http://www.huduser.org/Publications/PDF/hisp_homeown5.pdf)
- U.S. Census Bureau. 19 Apr. 2000. *Census Tracts and Block Numbering Areas*. [http://www.census.gov/geo/www/cen/sdo5\(tract\).html](http://www.census.gov/geo/www/cen/sdo5(tract).html).
- U.S. Census Bureau. 27 June 2005. *Cartographic Boundary Files*. <http://www.census.gov/geo/www/cob/tr2000.html>.
- U.S. Census Bureau. 5 Mar 2007. *2006 Second Edition TIGER/Line Files*. <http://www.census.gov/geo/www/tiger/tiger2006se/tgr2006se.html>