City University of New York (CUNY)
CUNY Academic Works

Publications and Research | New York City College of Technology

Fall 2012

Who Defaults on their Home Mortgage?

Eric Doviak

Sean P. MacDonald
CUNY New York City College of Technology

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/ny_pubs

Part of the Finance Commons

Recommended Citation

Doviak, E., & MacDonald, S. (2012). Who Defaults on their Home Mortgage?. New York Economic Review, 43(1), 75-98. http://EconPapers.repec.org/RePEc:nye:nyervw:v:43:y:2012:i:1:p:75-98

This Article is brought to you for free and open access by the New York City College of Technology at CUNY Academic Works. It has been accepted for inclusion in Publications and Research by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.

Who Defaults on their Home Mortgage?

Eric Doviak[*] **Sean MacDonald**[†]

ABSTRACT

Since Feb. 13, 2010, detailed information on every home mortgage default and foreclosure in New York State must be filed with the New York State Department of Financial Services (DFS). The data come from pre-foreclosure filing (PFF) notices that mortgage servicers must send to both the borrower and the DFS 90 days prior to initiating the foreclosure process and when a foreclosure has commenced. Pairing the PFF data with data on originations from the Home Mortgage Disclosure Act (HMDA) reveals the race and ethnicity of borrowers who defaulted on their home mortgages. HMDA analyses consistently reveal strong racial and ethnic disparities in lending practices. Our analysis shows that the same disparities reappear in the default data (i.e the PFF data), which suggests that lending disparities contributed to the higher default rates that we observe among black and Latino borrowers. Our analysis also suggests that labor market recovery would do the most to reduce the rate of mortgage default.

1 INTRODUCTION

In 2006, borrowers' inability to repay subprime mortgages sounded the first warning bell that the nation's housing bubble was about to burst. Subprime lending – which was virtually non-existent at the peak of the previous real estate boom in 1989-90 – had increased from 5 percent of total mortgage originations in 1994 to almost 20 percent in 2005 (Doms et al., 2007). More disturbingly, at the beginning of the decade, the US Department of Housing and Urban Development (2000a) had already identified a pattern of racial and ethnic disparities in subprime lending and noted that the pattern transcended income level.

By the time markets tumbled in 2008, the racial and ethnic character of subprime lending ensured that minority borrowers would be particularly hard hit by the accelerating foreclosure crisis. To shed more light on the causes of the foreclosure crisis and its impact on minority borrowers, this article takes a closer look the factors affecting a homeowner's probability of default. Specifically, we look at defaults among owner-occupied, first-lien mortgages originated between 2004 and 2008 (the period when the most risky loans were originated).

To study the causes of default, we combine data on originations from the Home Mortgage Disclosure Act originations (HMDA) to the pre-foreclosure filing (PFF) data from the New York State Department of

[*]Brooklyn College, City University of New York – eric@doviak.net

[†]New York City College of Technology, City University of New York – smacdonald@citytech.cuny.edu

Financial Services (DFS), formerly known as the New York State Banking Department (NYSBD)[1,2], and trace loans from origination to default.

The HMDA data are particularly valuable because their geographic focus enables state bank regulators (like the DFS) to track institutions' lending neighborhood-by-neighborhood. When combined with other sources of information (e.g. reports from bank examinations), the HMDA data help bank regulators explore the question of whether local financial institutions are meeting the saving, borrowing and housing needs of low-to-moderate income communities and minority communities.

Academics frequently shun the HMDA data however because the data do not provide a detailed picture of each loan application. The HMDA data do contain borrower's income, loan amount and a small window on the interest rate, but critical details like the borrower's credit score and loan-to-value ratio are missing.

In defense of the HMDA data, we argue that they are a very important data source because they are the most comprehensive and it's the data source that the US Department of Housing and Urban Development had used in its (previously mentioned) research on racial and ethnic disparities in subprime lending. In his 2007 *Report to the Interagency Task Force on Subprime Mortgages*, NYS Banking Superintendent Neiman declared that "analysis of HMDA data is a priority" because the HMDA data is publicly available and because it is the data that regulators use to track lending neighborhood-by-neighborhood.

One year later, the task force was promoted to a governor's level task force and issued a follow-up report (Neiman, 2008) detailing its analysis of data from the Mortgage Bankers Association, the LoanPerformance Data, the HMDA data and the RealtyTrac data. The report noted a sharp increase in foreclosures since 2005, noted the racial disparities in lending practices, noted that subprime loans constituted almost half of serious delinquencies and noted that subprime loans with adjustable interest rates were seriously delinquent at rates far above the average for all loans (22 percent vs. 3 percent in New York State).

In the same report, Neiman also cited the State Foreclosure Prevention Working Group (2008) findings that seven out of ten seriously delinquent borrowers were not on track for any loss-mitigation outcome, that loss-mitigation departments were severely over-worked and – critically – that loss mitigation procedures (when followed) do increase the chances that homeowners will receive a loan modification.

Given these findings and upon Neiman's recommendation, the New York State legislature passed and Gov. David Paterson signed (on Dec. 15, 2009) the Mortgage Foreclosure Law which amended the *Real Property Actions and Proceedings* and inserted a new section (§ 1306) to require mortgage servicers to send borrowers a 90-day notice prior to commencing foreclosure proceedings on owner-occupied residential mortgages.

Additionally, the new law required mortgage servicers to electronically submit the pre-foreclosure filings (PFF) to the NYSBD (later 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." The language in § 1306 does not permit state bank regulators to sanction a lender or mortgage servicer for infractions discovered in the PFF filings. Enforcement of the law is left to the courts. Consequently, servicers have a strong incentive to submit honest and accurate filings.

When deciding what information about the loans to collect from the mortgage servicers, the NYSBD chose to collect information that would help it match the pre-foreclosure filings to the corresponding HMDA filings. Furthermore, in its two reports analyzing the PFF data (2010a; 2010b), the NYSBD 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.

Because the PFF data were designed to be matched to the HMDA data and given HMDA's historic and regulatory importance, this article continues the tradition of HMDA analysis by merging the PFF data into the HMDA data and asking what characteristics make a borrower more likely to default on his/her home mortgage.

We begin by discussing the literature on discrimination in mortgage lending in section 2. We then describe the PFF data in more detail and explain how we paired it with the HMDA data in section 3. Section 4 discusses the racial and ethnic disparities that we observe in the HMDA data (on originations) and the HMDA data (on defaults). The analysis there shows that blacks and Latinos tend to take high-cost loans at a higher rate than their white and non-Latino counterparts and those disparities in lending are reflected in the higher default rates among black and Latino borrowers.

Section 5 provides a basic regression analysis to further explore some of the questions that arise in the PFF to HMDA comparisons. Missing variables hinder efforts to properly understand what we observe in the HMDA data, but the estimated coefficients on race and ethnicity are striking. Our models also suggest that labor market conditions have the strongest effect on a borrower's probability of default.

2 REVIEW OF THE DISCRIMINATION LITERATURE

Although there is a link between lending disparities and the foreclosure crisis, we must separate the two events in time. To that end, this section will first review the literature on subprime lending in predominantly black and Latino communities before turning to its relationship with the subsequent foreclosure crisis. That having been said, several studies of the foreclosure crisis did look for its roots in the residential segregation seen in many of the nation's metropolitan areas (Rugh and Massey, 2010) and the lack of alternatives to subprime lenders in predominantly minority communities (US Department of Housing and Urban Development, 2000b). However, other factors also played a role. Doms et al. (2007) include changes in home prices and home price volatility among its causes, while Morgan et al. (2012) discuss the inability to exclude the mortgage on a primary residence from bankruptcy protection after passage of the 2005 Bankruptcy Reform Act.

Returning to the character of lending in black and Latino communities, we remind the reader that the HMDA data are limited. Analysis of the HMDA data strongly suggests that blacks and Latinos had difficulty obtaining loans on terms comparable to their white and non-Latino counterparts, but because the HMDA data omit important variables (such as the borrower's credit score and the loan-to-value ratio) one cannot prove a pattern of discrimination. In other words, it is easy to show that high-cost lending was most prevalent in predominantly minority communities, but it is difficult to take the next step and use the HMDA data to show that such lending is evidence of discrimination.

In an effort to overcome some of HMDA's limitations, Bocian et al. (2006) paired the 2004 HMDA data with a proprietary dataset of 177,000 subprime loans and found that after controlling for other factors (such as the borrower's FICO score and the loan-to-value ratio), blacks and Latinos received a disproportionate share of high-cost loans. The major limitation of their study however is that it does not sample from the universe of originations. It is a particular firm's sample of loans. The findings may suffer from selection bias and are certainly not generalizable to the broader market. More importantly, Bocian et al.'s work cannot be considered evidence of discrimination because it does not explain why borrowers took a subprime loan as opposed to a prime loan.

When employing the HMDA data to study the broader market researchers are generally confined to finding a correlation between racial segregation and the probability of receiving a high-cost loan. For example, 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 US. 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.

Other studies have also found a link between the racial composition of a neighborhood and the share of subprime lending in that neighborhood. For example, in a joint study conducted by several community organizations, Bromley et al. (2008) focused on subprime lending activity in 2006 across seven large US metropolitan areas. 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 a census tract's share of minority residents.

The US Department of Housing and Urban Development (2000b) also found a disproportionate concentration of subprime lending in predominantly minority – and particularly – African-American communities. In the study, which focuses primarily on subprime refinance lending, the number of subprime refinance loans originated in the New York metropolitan area between 1993 and 1998 increased by an estimated 350 percent. The study also found that subprime loans were three times more likely to be originated in lower-income neighborhoods in the New York metropolitan area than in higher-income neighborhoods, and more than four times more likely in predominantly black than in predominantly white neighborhoods.

It's particularly interesting to note that their study was published in 2000, which indicates that subprime lending expanded rapidly into minority communities long before the subprime mortgage meltdown began in 2006. According to Laderman (2001), one factor which contributed to the expansion of subprime mortgage lending in the early 1990s was the increasing frequency with which mortgages were securitized. Securitization reduced the risk associated with lending to subprime borrowers and it enabled large sums to be assembled for the purpose of subprime lending. Another factor that Laderman cites was deregulation. Prior to passage of the Depository Institutions Deregulation and Monetary Control Act in 1980, limits were imposed on the interest rates that lenders could charge. Once the caps were lifted, lenders could raise interest rates high enough to absorb the risk associated with lending to subprime borrowers.

In a separate but related report, the US Department of Housing and Urban Development (2000a) found that the pattern of originating subprime loans to minorities transcended income level and that this pattern established itself long before the subprime loan market reached its peak during the early 2000s.

Instead, borrowers in high-income black neighborhoods were twice as likely as those in low-income white neighborhoods to take out a subprime loan. Specifically, the study found that just six percent of borrowers in high-income white neighborhoods had subprime loans while 39 percent of borrowers in upper-income black neighborhoods had subprime loans. This figure was more than twice the 18 percent rate for borrowers in low-income white neighborhoods.

Such findings are disturbing. 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 hard to see how this could be applicable across neighborhoods. In other words, it is easy to imagine individual cases where a high-income black borrower's credit score is lower than a low-income white borrower's credit score; however it is difficult to see how the average credit score of a high-income black neighborhood could be lower than the average credit score of a low-income white neighborhood.

Given that blacks and Latinos took a disproportionately high share of subprime loans, one would also expect a disproportionately high rate of foreclosure in black and Latino communities. This is precisely what two other studies have found.

Rugh and Massey (2010) 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 obtained the total number of foreclosures between 2006 and 2008 from RealtyTrac's foreclosure database and computed the foreclosure rate as the number of filings per household unit. They then used the 2004-2006 HMDA data to compute the share of high-cost loans[3] in each MSA. To derive a measure of regulatory oversight, they also computed the share of loans within the MSAs that were originated by institutions covered under the Community Reinvestment Act (CRA). Rugh and Massey then 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 regression 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 frustrating omission in their published paper however is the lack of 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 they should have regressed the high-cost lending share on measures of segregation. If the coefficient were positive and statistically significant, then their claims of racial and ethnic targeting would have a firmer foundation.

Gerardi and Willen (2008) also examine the relationship between foreclosures and subprime lending in urban and minority communities. By matching the 1998-2006 HMDA data to deed registry data in the State of Massachusetts, they generate a dataset that contains the universe of mortgages, foreclosures, purchases and sales. In their analysis of the data, they find that a disproportionate share of subprime loans were originated to blacks and Latinos, but these loans proved unsustainable when home prices fell. The records of property sales in their dataset indicate that a "sudden and severe fall in the share of minority home ownership" began in 2005 due to a significant increase in foreclosures among minority homeowners.

The studies reviewed above show that blacks and Latinos took a disproportionately high share of high-cost and subprime loans, but the evidence that this trend reflects discrimination suffers from the limitations of the HMDA data. Nonetheless, the studies do help explain our finding that blacks and Latinos defaulted on their mortgages at a higher rate than their white and non-Latino counterparts.

3 THE NEW YORK STATE PRE-FORECLOSURE FILING DATA

As mentioned previously, our findings come from an analysis of the data that the NYSBD began collecting home mortgage defaults in Feb. 2010. (The DFS later assumed those responsibilities). When borrowers default on their primary residence, their mortgage servicers must send them a pre-foreclosure notice at least 90 days before commencing foreclosure proceedings and file the notice with the DFS.

The DFS collects an extraordinary level of detail on the loans. In addition to names and address, the DFS also collects 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 or not and whether an investor's approval is necessary to modify the loan. If the default progresses to a *lis pendens* filing (i.e. the first step in the foreclosure process – the filing of the complaint), then servicers are also required to follow up on their initial filing and provide information on the entity filing for foreclosure.

The New York State Banking Department (2010a,b) provided basic analysis of the PFF data. In another paper (Doviak and MacDonald, 2011), we compare the characteristics of loans that did and did not progress from default to a foreclosure filing. The analysis presented in this article uses our combined HMDA-PFF dataset to examine the loan characteristics which make a borrower more likely to default.

Prior to making such comparisons however, we first explain how we prepared the PFF dataset for statistical analysis in subsection 3.1. Then, in subsection 3.2, we explain how we matched the PFF data to the HMDA data. After providing those explanations, we discuss our comparisons in section 4 and we provide a very basic regression analysis in section 5. Section 6 concludes with a discussion of what we learned from the pre-foreclosure filing project.

3.1 PREPARING THE DATA FOR ANALYSIS

One of our first steps in preparing the dataset was to remove duplicate filings. Servicers who missed the three-business day deadline or submitted incorrect information would "re-file" the loan. Some servicers also submitted one filing for each borrower on the loan. The duplicates were fairly easy to identify however, because servicers almost always included their loan numbers with the filing, so the combination of the servicer's identity and the loan number enabled us to uniquely identify each loan[4]. In cases where a servicer submitted one filing for each borrower, we compared the borrower's first and last name to the names of other borrowers on the loan to see if there was a co-applicant or not.

Because servicers re-filed a loan to correct mistakes, we assumed that the filing which was submitted last contained the correct information. However if one of the duplicates contained information on a *lis pendens* filing, we retained that information. Using this method, we found a total of 214,705 unique loans and 33,859 duplicates in the PFF dataset. From there, we removed records that contained obvious errors (e.g. loans that were originated in the future) and records of 90-day letters that were not mailed in the year 2010. This reduced the PFF dataset to 211,962 clean records.

To ensure comparability across loans, we chose to focus on first-lien mortgages. This reduced the PFF dataset to 186,366 records, but it was a necessary step because a first-lien mortgage is very different from a home equity line of credit (HELOC). The former is frequently taken for the purpose of purchasing

Table 1: Distribution of Pre-Foreclosure Filings by Year of Origination

year	total	percent
1976-1989	2,502	1.3%
1990-1999	13,692	7.3%
2000	2,414	1.3%
2001	4,390	2.4%
2002	7,470	4.0%
2003	16,706	9.0%
2004	18,669	10.0%
2005	28,506	15.3%
2006	35,947	19.3%
2007	31,771	17.0%
2008	16,019	8.6%
2009	6,957	3.7%
2010	1,323	0.7%
total	186,366	100.0%
Data: Full PFF		

a home, while the latter is often used for home improvement.

Our analysis pays particular attention to the 130,912 first-lien mortgages that were originated in the years 2004-2008. Table 1 shows that these five years account for 70 percent of all PFF filings on first-lien mortgages. We chose to work with the years 2004-2008 because we wanted to compare 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 available in the pre-2004 HMDA data were quite limited. At the time of this writing, the 2009 HMDA data were available to us, but we chose not to work with it because lending practices changed dramatically after the subprime mortgage crisis crippled the world financial system in late 2008. Loans originated in 2009 were very different from loans originated in previous years, so – for this analysis – we wanted to focus on loans originated in the years leading up to and including the crisis. One avenue for future research is to compare lending patterns in the periods before and after the crisis to see how those differences affect the rate at which borrowers default.

3.2 MATCHING THE PRE-FORECLOSURE FILING DATA TO THE HMDA ORIGINATIONS DATA

The HMDA originations data contain the FIPS county code and census tract number of each property. This is a particularly valuable piece of information because census tracts have a small population (between 2,500 and 8,000 people) which is fairly homogeneous in terms of socio-economic characteristics and living conditions (US Census Bureau, 2000).

So our first step in matching the PFF data to the HMDA data was to identify the census tract of each property in the PFF dataset from the address. To identify the census tracts, we used Erle's (2005)

Table 2: Pre-Foreclosure Filings by Loan Amount

amount ($1000s)	no PFF	received PFF	overall
under 50	4.9%	2.8%	4.8%
50 to 99	16.5%	13.4%	16.3%
100 to 249	36.1%	27.7%	35.4%
250 to 399	25.8%	33.7%	26.4%
400 to 499	8.3%	12.7%	8.6%
500 and up	8.4%	9.7%	8.5%
total	1,544,118	130,722	1,674,840
Data: Combined HMDA-PFF			

Table 3: Pre-Foreclosure Filings by Applicant Income

income ($1000s)	no PFF	received PFF	overall
under 40	10.9%	9.9%	10.8%
40 to 59	18.0%	15.6%	17.8%
60 to 79	19.2%	18.3%	19.1%
80 to 99	15.8%	17.3%	15.9%
100 to 119	10.9%	12.9%	11.1%
120 to 159	11.9%	14.0%	12.0%
160 to 199	5.0%	5.4%	5.0%
200 and up	8.4%	6.6%	8.2%
total	1,465,078	123,878	1,588,956
Data: Combined HMDA-PFF			

"Geo-Coder-US-1.00" Perl module in conjunction with the US Census Bureau's (2007) TIGER/Line Files.

After using Erle's Perl module to create a database of New York State addresses from the TIGER/Line Files, we queried the database to obtain the latitudes and longitudes of the property addresses in the PFF dataset. Once we had the coordinates, we compared them to a database of census tract coordinates that we generated from the US Census Bureau's (2005) "Cartographic Boundary Files."

Using this method, we were able to identify the census tracts for 96 percent[5] of the addresses in the PFF database. To avoid losing the information that the other four percent contain, we identified each of the census tracts within the property's five-digit zip code and counted the number of times each census tract corresponded to that zip code. We then randomly assigned the property to one of those census tracts (using the number of occurrences as weights).

Once the Census Tracts of each property had been identified and we had purged the duplicates, matching the pre-foreclosure filing data to the HMDA originations data was fairly simple. We divided owner-occupied[6], first-lien mortgages in the HMDA data and first-lien mortgages in the PFF data into buckets by year of origination, census tract and co-applicant status. On average, there were 34 loans in each HMDA bucket and 3 loans in each PFF bucket, so to figure out which HMDA origination corresponded to the pre-foreclosure filing, we compared the loan amounts and chose the closest match.

4 WHO DEFAULTS ON THEIR HOME MORTGAGE?

Having identified the defaults in the HMDA data, we could quickly proceed to our most striking finding: that black and Latino borrowers defaulted at a higher rate than their white and non-Latino counterparts. But proceeding with such haste would be unjust. First, we must identify the financial characteristics that are correlated with default. Then we must compare the loan characteristics of minority and non-minority borrowers. Only after these first two steps have been conducted can we examine the default rates among black and Latino borrowers in an impartial manner.

4.1 FINANCIAL CHARACTERISTICS

Using the combined HMDA-PFF data, we find that one strong predictor of default is the amount borrowed. As table 2 shows, 56 percent of the borrowers who received a pre-foreclosure filing took loans in excess of $250,000, whereas only 43 percent of the borrowers who did not default borrowed more than $250,000.

It would be particularly insightful to compare the amounts that borrowers owe to the value of their homes. Unfortunately, HMDA does not provide the loan-to-value ratio or any information on the down payment, so we cannot make such a comparison. Nonetheless, if individuals who borrowed less have a larger equity stake in their homes, then these findings would illustrate the general principle that borrowers who have a larger equity stake in their home are less likely to default and enter the foreclosure process.

Repaying a mortgage also depends on the ability to pay, of course. But it's particularly striking to note that borrowers with income in the $80,000 to $199,999 range received pre-foreclosure filings at a higher rate than borrowers above and below that range (as shown in table 3). Why borrowers in the $80-200K income range default at a higher rate than lower-income borrowers is puzzling. The regression models discussed in section 5 suggest however that borrowers with higher incomes are less likely to receive a pre-foreclosure filing after controlling for other factors, such as: loan amount, predicted rate spread, changes in county-level employment and changes in the FHFA home price index.

Another good predictor of default is the interest cost of the loan. Table 4 shows that borrowers who took "high-cost" loans were more likely to receive a pre-foreclosure filing. When viewed in a risk-premium context, this finding should not be surprising. Borrowers who are more likely to default will have to compensate the lender for the additional risk by paying a higher interest rate.

However, there is also a risk that the additional cost of the loan will make the borrower more likely to default and go into foreclosure. In particular, a borrower's monthly payment is an increasing function of the interest rate, so a higher interest rate reduces a borrower's ability to repay the loan.

Lenders do not set interest rates exogenously however. Since a borrower's income and loan amount affect his/her probability of default, all else equal one would expect lenders to compensate for the additional risk by charging a higher interest rate to low-income borrowers and borrowers who take out a larger loan.

Table 4: Pre-Foreclosure Filings by Loan Cost

loan cost	no PFF	received PFF	total
non-high cost	92.8%	7.2%	1,364,557
high cost	89.4%	10.6%	310,283
overall	92.2%	7.8%	1,674,840
Data: Combined HMDA-PFF			

Table 5: High Cost Loans by Applicant Income

income ($1000s)	non-high cost	high cost	overall
under 40	10.1%	13.9%	10.8%
40 to 59	17.8%	18.0%	17.8%
60 to 79	19.0%	19.4%	19.1%
80 to 99	15.6%	16.9%	15.9%
100 to 119	10.8%	12.1%	11.1%
120 to 159	12.1%	11.9%	12.0%
160 to 199	5.3%	4.1%	5.0%
200 and up	9.3%	3.8%	8.2%
total	1,290,774	298,182	1,588,956
Data: Combined HMDA-PFF			

Table 6: High Cost Loans by Loan Amount

amount ($1000s)	non-high cost	high cost	overall
under 50	4.4%	6.6%	4.8%
50 to 99	15.9%	18.0%	16.3%
100 to 249	37.1%	28.1%	35.4%
250 to 399	25.8%	29.0%	26.4%
400 to 499	8.1%	10.8%	8.6%
500 and up	8.7%	7.6%	8.5%
total	1,364,557	310,283	1,674,840
Data: Combined HMDA-PFF			

Table 7: High Cost Loans by Additional Applicant

status	non-high cost	high cost	total
no co-applicant	77.8%	22.2%	952,877
co-applicant	86.3%	13.7%	721,963
overall	81.5%	18.5%	1,674,840

Data: Combined HMDA-PFF

Table 8: Pre-Foreclosure Filings by Additional Applicant

status	no PFF	received PFF	total
no co-applicant	91.1%	8.9%	952,877
co-applicant	93.6%	6.4%	721,963
overall	92.2%	7.8%	1,674,840

Data: Combined HMDA-PFF

In line with this reasoning, we find that low-income borrowers are more likely to receive a high-cost loan than borrowers with higher income. Table 5 shows 80 percent of high-cost loans were originated to borrowers with income below $120,000, whereas only 73 percent of loans that were not high-cost loans were originated to such borrowers.

Surprisingly however, there does not appear to be any systematic relationship between loan amount and the likelihood of the loan being a high-cost loan. Table 6 shows that loan amounts below $100,000 were more likely to be high-cost loans and loan amounts in the $250,000 to $499,999 range were also more likely to be high-cost loans.

It is difficult to understand why small loan amounts (i.e. those under $100,000) were more likely to be high-cost loans and why large loan amounts (i.e. those over $500,000) were less likely to be high-cost loans. Regression analysis (which controls for other factors like income) does not even help to explain this puzzle. As discussed in section 5, borrowers who took out larger loan amounts tended to receive lower interest rates on their mortgages after controlling for other factors even though the larger loan amounts made them more likely to default.

Another important factor in explaining interest rates is whether there is a co-borrower on the loan or not. As table 7 shows, 22 percent of loans without a co-applicant were high-cost loans, whereas only 14 percent of loans with a co-applicant were high-cost loans. This may be attributable to the fact that a second borrower is a (potential) second source of income, which helps to mitigate the risk that the loan will go into default. As table 8 shows, 9 percent of loans without a co-borrower received a pre-foreclosure filing, whereas only 6 percent of loans with a co-borrower received a pre-foreclosure filing.

Table 9: High Cost Loans by Applicant Race

race	non-high cost	high cost	total
Asian	89.7%	10.3%	89,998
Black/Afr. Am.	64.9%	35.1%	166,380
White	84.2%	15.8%	1,161,960
not provided	76.8%	23.2%	234,393
overall	81.5%	18.5%	1,674,840

Data: Combined HMDA-PFF

Table 10: High Cost Loans by Applicant Ethnicity

ethnicity	non-high cost	high cost	total
Hispanic/Latino	71.9%	28.1%	134,937
Not Hispanic/Latino	82.8%	17.2%	1,263,971
not provided	77.5%	22.5%	232,693
overall	81.5%	18.5%	1,674,840

Data: Combined HMDA-PFF

4.2 RACE AND ETHNICITY

In section 2, we reviewed evidence of racial and ethnic discrimination in lending practices. The HMDA data captures one form of such discrimination – the difference in the rate spread between loans originated to minorities and loans originated to whites. As tables 9 and 10 show, blacks and Latinos received a disproportionately high share of high-cost loans. Asians, by contrast, received a disproportionately low share. Tables 11 and 12 show that blacks and Latinos also received a disproportionately high share of pre-foreclosure filings, so one also has to wonder if racial and ethnic discrimination in lending practices contributed to the disproportionately high share of defaults among blacks and Latinos.

One way to address this question is to ask if fundamental differences between minorities and non-minorities justify the difference in rate spreads. If so, then the next question to ask is if those fundamental differences could have caused blacks and Latinos to default at disproportionately higher rates.

The first fundamental factor that we'll consider is income. If minority borrowers tended to have lower income than their non-minority counterparts, then one could justify the difference in rate spreads on the basis of income. Such a hypothesis only finds partial support in the data. Table 13 shows that 26 percent of Asian borrowers and 18 percent of white borrowers had income over $140,000, while only 11 percent of black borrowers did. The distribution of income by ethnicity shows a similar pattern. As table 14 shows, 18 percent of non-Latino borrowers had income over $140,000, while only 14 percent of Latinos did.

At first glance, the fact that there is more weight in the upper region of the income distribution among non-minority borrowers than there is among minority borrowers appears to lend support to the hypothesis

Table 11: Pre-Foreclosure Filings by Applicant Race

race	no PFF	received PFF	total
Asian	92.8%	7.2%	89,998
Black/Afr. Am.	88.0%	12.0%	166,380
White	92.8%	7.2%	1,161,960
not provided	91.7%	8.3%	234,393
overall	92.2%	7.8%	1,674,840

Data: Combined HMDA-PFF

Table 12: Pre-Foreclosure Filings by Applicant Ethnicity

ethnicity	no PFF	received PFF	total
Hispanic/Latino	89.0%	11.0%	134,937
Not Hispanic/Latino	92.4%	7.6%	1,263,971
not provided	92.0%	8.0%	232,693
overall	92.2%	7.8%	1,674,840

Data: Combined HMDA-PFF

that differences in income help explain why blacks and Latinos received a disproportionate share of high-cost loans. However, the lower region of the income distributions refutes the hypothesis. It appears to have been easier for low-income whites and non-Latinos to obtain a mortgage. Specifically, 31 percent of white borrowers had income below $60,000, while only 25 percent of black borrowers did. Similarly, 30 percent of non-Latinos had income below $60,000, while only 19 percent of Latinos did. Consequently, it would be hard to justify the disproportionate share of high-cost loans that blacks and Latinos received on the basis of income differentials.

Turning to default rates, the fact that a larger share of black and Latino borrowers fall into the $80-200K income range (than their white and non-Latino counterparts) provides some support for the hypothesis that income differences may help explain why blacks and Latinos were more likely to default, but the default rates among Asians casts doubt on the hypothesis. Specifically, table 13 shows that 50 percent of black borrowers fell in the $80-200K income range. That's higher than the 42 percent of white borrowers, but less than the 58 percent of Asian borrowers. Table 14 shows that 57 percent of Latino borrowers had income between $80,000 and $199,999 income, but only 43 percent of non-Latinos did.

Given the inability of income to explain the racial and ethnic disparities in loan cost and defaults, we now consider the amount of the original loan. Differences in loan amounts help explain why blacks and Latinos received a disproportionate share of pre-foreclosure filings, but they do not necessarily explain why they received a disproportionate share of high-cost loans.

Specifically, minorities tended to borrow much more than their non-minority counterparts. Table 15

Table 13: Applicant Income by Applicant Race

income ($1000s)	Asian	Black/Afr. Am.	White	not provided	overall
under 40	4.0%	8.0%	12.2%	8.7%	10.8%
40 to 59	11.7%	16.5%	18.9%	16.1%	17.8%
60 to 79	16.3%	23.0%	18.7%	19.4%	19.1%
80 to 99	17.3%	20.1%	15.1%	16.0%	15.9%
100 to 119	14.4%	13.6%	10.4%	11.0%	11.1%
120 to 159	17.6%	12.4%	11.5%	12.5%	12.0%
160 to 199	8.3%	3.7%	4.9%	5.5%	5.0%
200 and up	10.5%	2.7%	8.4%	10.8%	8.2%
total	85,965	156,030	1,105,913	220,741	1,588,956

Data: Combined HMDA-PFF

Table 14: Applicant Income by Applicant Ethnicity

income ($1000s)	Hispanic/Latino	Not Hispanic/Latino	not provided	overall
under 40	5.8%	11.6%	8.9%	10.8%
40 to 59	12.9%	18.5%	16.3%	17.8%
60 to 79	20.6%	18.9%	19.2%	19.1%
80 to 99	21.4%	15.4%	15.8%	15.9%
100 to 119	15.9%	10.6%	10.9%	11.1%
120 to 159	14.8%	11.7%	12.4%	12.0%
160 to 199	4.8%	5.0%	5.5%	5.0%
200 and up	3.8%	8.2%	11.0%	8.2%
total	125,440	1,203,686	219,669	1,588,956

Data: Combined HMDA-PFF

suggests that 62 percent of white borrowers borrowed less than $250,000 whereas only 40 percent of blacks did. Interestingly however, Asians appear to have borrowed even more than blacks (only 34 percent borrowed less than $250,000), but had the lowest rate of high-cost loans. Turning to ethnicity, table 16 shows that 59 percent of non-Latinos borrowed less than $250,000, whereas 36 percent of Latinos borrowed less than that amount.

The finding that blacks and Latinos tended to borrow more helps explain why they received a disproportionately high share of pre-foreclosure filings, but it does not explain why they took high-cost loans at a higher rate than their white, Asian and non-Latino counterparts. Asians also borrowed more, but took fewer high-cost loans. Moreover, as mentioned previously, the regression analysis in section 5 also refutes the hypothesis that borrowers who took out larger loan amounts would receive lower interest rates. The opposite is true. All else equal, the rate spreads on larger loans tend to be lower.

Table 15: Loan Amount by Applicant Race

amount ($1000s)	Asian	Black/Afr. Am.	White	not provided	overall
under 50	1.0%	3.2%	5.7%	2.7%	4.8%
50 to 99	6.3%	8.4%	19.1%	12.1%	16.3%
100 to 249	26.3%	28.3%	37.2%	35.2%	35.4%
250 to 399	33.3%	40.5%	23.0%	29.9%	26.4%
400 to 499	18.0%	12.8%	7.1%	9.4%	8.6%
500 and up	15.1%	6.8%	7.8%	10.7%	8.5%
total	89,998	166,380	1,161,960	234,393	1,674,840
Data: Combined HMDA-PFF					

Table 16: Loan Amount by Applicant Ethnicity

amount ($1000s)	Hispanic/Latino	Not Hispanic/Latino	not provided	overall
under 50	2.1%	5.4%	2.9%	4.8%
50 to 99	7.1%	17.8%	12.7%	16.3%
100 to 249	26.4%	36.1%	35.6%	35.4%
250 to 399	41.7%	24.4%	29.0%	26.4%
400 to 499	13.6%	8.1%	9.0%	8.6%
500 and up	9.2%	8.1%	10.7%	8.5%
total	134,937	1,263,971	232,693	1,674,840
Data: Combined HMDA-PFF				

In summary, neither income nor loan amount appear to justify the higher rate spreads on loans originated to blacks and Latinos. This finding is particularly disturbing because borrowers who took out high-cost loans were more likely to default, but the finding is not evidence of discrimination because the HMDA data does not contain critical information, such as the credit score and loan-to-value ratio.

5 ECONOMETRIC MODELS OF RATE SPREADS AND DEFAULTS

Section 4 describes several questions raised by the combined HMDA-PFF dataset. The most striking questions are why blacks and Latinos were more likely to take high-cost loans and why they are more likely to default on their mortgages. But there were other questions too. One is why there isn't a clear relationship between the amount of the original loan and the whether the loan was a high-cost loan. Another was why borrowers in the $80-200K income range default at a higher rate than borrowers with income both above and below that range.

In an attempt to answer some of these questions, this section presents a basic regression analysis. Although we have not developed a formal theory from microeconomic foundations, the analysis below presents an intuitive reduced form model. We acknowledge that our model may contain omitted-variable bias if race and ethnicity are correlated with credit score or loan-to-value ratio, but the bias is unlikely to be so severe that it invalidates all of the findings of the studies discussed in our literature review. We attempt to mitigate some of the omitted-variable bias in the model of default probability by including the log difference in the FHFA Home Price Index and the log difference in county-level employment between the year that the loan was originated and the year 2010.

Furthermore, the reader should notice that our Tobit model of a borrower's rate spread is a reduced form model of the price at which a lender and a borrower agree to originate a loan. While lenders charge a higher interest rate to borrowers at greater risk of default, the higher interest rate also makes the loan more difficult to repay. Therefore, our predicted rate spread from the first stage serves as an instrument in the second stage model of a borrower's probability of defaulting on his/her home mortgage.

The limitations of the HMDA data prevent us from testing more sophisticated models of the implicit supply and demand decisions and the estimates presented in this article suffer from omitted-variable bias. Nonetheless, if the signs of the regression coefficients are correct, then the model does provide us with insight into the causes of mortgage default.

5.1 ECONOMETRIC METHODS

One problem confronting any econometric analysis of the HMDA data is how to work with the rate spread. The HMDA data only provide a value of the rate spread when the difference between the interest rate on the mortgage and the yield on the comparable U.S. Treasury exceeds three percentage points[7]. Consequently, when addressing the question of why black and Latino borrowers were more likely to take out a high-cost loan, we have to find a way to work with the rate spread.

The simplest method is to reduce the rate spread to a binary variable (i.e. one if high-cost, zero otherwise) and employ a probit or logit model to estimate the probability that a borrower took a high-cost loan. The trouble with such a strategy is that is discards valuable information on the magnitude of the differences in rate spread among borrowers.

The alternative is to employ a Tobit model to obtain an estimate of the rate spread itself. The trouble with this strategy is that 81 percent of the loans in the combined HMDA-PFF dataset are not high-cost loans, so no value of the rate spread is reported for these loans. Therefore, instead of using the Tobit model to estimate the tail of the distribution, the Tobit model has to estimate 81 percent of the distribution.

We chose to use the Tobit model however because it provides an estimate of the rate spread which can be used as an instrument in a second-stage regression on the probability of defaulting on the home mortgage. One must use an instrument for the rate spread in the second-stage to overcome the endogeneity problem that arises when lenders charge higher interest rates to borrowers who are more likely to default.

To obtain efficient estimates of the parameters in the second-stage probability model, we used an algorithm that Adkins (2009) developed to implement Amemiya's Generalized Least Squares (AGLS). Adkins (2008) shows that the AGLS estimator yields consistent estimates of the parameters' standard errors and can be used to test the statistical significance of the parameters.

The AGLS algorithm requires estimates of the residuals from the first-stage regression, but – because the rate spread is censored at three percentage points – we could not use response residuals as we would if the first-stage regression were a standard OLS regression model. Therneau and Lumley's (2009) "survival" package for R (R Development Core Team, 2010) provides a viable alternative however. As its "survreg" function iteratively maximizes the log-likelihood function, it predicts the value of the dependent variable and calculates a correction term, called the "working residual" (Therneau, 1999), which we use in place of the response residual.

5.2 DISCUSSION OF THE REGRESSION RESULTS

As shown in table 17, the rate spreads on owner-occupied, first-lien mortgages originated to blacks and Latinos were higher than those originated to their white and non-Latino counterparts and the differences were statistically significant, even after controlling for other variables such as income, loan amount, whether there was a co-borrower on the loan, the purpose of the loan and region of the state and year of origination.

As emphasized repeatedly throughout this article, the estimated coefficients suffer from omitted-variable bias, but the racial and ethnic disparities in interest rates are too large to ignore. The coefficient estimates in model 1 suggest that the interest rate on a loan originated to a black borrower was 1.36 percentage points higher than a the interest rate originated to an equivalent white borrower. Model 2 suggests a slightly smaller difference: 0.86 percentage points. Turning to Latinos, the coefficient estimates in model 1 suggest that Latinos paid 0.92 percentage points more than an equivalent non-Latino borrower, while model 2 puts the gap at 0.64 percentage points. While this is deeply disturbing, the HMDA data omits many important variables (such as the borrower's credit score and the loan-to-value ratio), so we cannot conclude that this is evidence of discrimination.

With one exception, the signs of the other coefficients in the model are not surprising. The coefficient on loan amount is the exception. It seems odd to us that borrowers who took out larger loans would pay a lower interest rate. In the case of the HMDA data however, a large loan amount may be acting as a proxy for variables that we do not observe and thus indicate that the borrower is more creditworthy.

Before accepting our findings at face value however, one must note an important limitation of using the Tobit model to predict the rate spread: the estimates are far from perfect. By adding the average yield on a 30-year U.S. Treasury to the predicted rate spread, we can compare the Tobit models' predicted interest rates to the ones in the pre-foreclosure filing data. As tables 18, 19 and 20 show, the predicted interest rates do not have as much weight in the upper region as the interest rates in the PFF dataset. We believe that the predicted rate spread is correlated with the unobserved true values of the rate spread, but there is no way to check the validity of our model.

Turning to the second-stage model of the probability that a borrower will default, we find that the coefficient on the predicted rate spread is positive (suggesting that borrowers with higher rate spreads were more likely to default), but is only statistically significant at the 10 percent level in model 1 and is not statistically significant at all in model 2.

Both models suggest that black race and Latino ethnicity are positively correlated with the probability of default after controlling for other factors, such as income, loan amount and whether there is a co-applicant on the loan. We do not believe however that the melanin level in a person's skin affects his probability of

Table 17: Two-Stage: Tobit predicts Rate Spread, then Probit predicts PFF

	Model 1				Model 2			
	Tobit		probit		Tobit		probit	
Intercept	−0.0513	***	−2.1133	***	0.0037		−2.1071	***
	(0.0004)		(0.1183)		(0.0054)		(0.1715)	
Pred. Rate Spread			0.4093	.			0.3302	
			(0.2434)				(0.3173)	
ln(Loan Amount)	−0.0005	***	0.2511	***	−0.0005	***	0.2486	***
	(0.0001)		(0.0252)		(0.0001)		(0.0366)	
ln(App. Income)	−0.0014	***	−0.2067	***	−0.0009	***	−0.2054	***
	(0.0001)		(0.0251)		(0.0001)		(0.0365)	
Co-Applicant	−0.0053	***	−0.1044	***	−0.0049	***	−0.1059	**
	(0.0001)		(0.0243)		(0.0001)		(0.0352)	
Conv'l Loan	0.0156	***			0.0158	***		
	(0.0002)				(0.0002)			
Home Purchase	0.0114	***			0.0112	***		
	(0.0001)				(0.0001)			
Home Improve.	0.0075	***			0.0073	***		
	(0.0001)				(0.0001)			
Hispanic/Latino	0.0092	***	0.1705	***	0.0064	***	0.1702	**
	(0.0001)		(0.0424)		(0.0001)		(0.0616)	
Asian	−0.0017	***	−0.0447		−0.0034	***	−0.0456	
	(0.0002)		(0.0510)		(0.0002)		(0.0742)	
Black/Afr. Am.	0.0136	***	0.2381	***	0.0086	***	0.2396	***
	(0.0001)		(0.0395)		(0.0001)		(0.0575)	
Race not provided	0.0060	***	0.0662	*	0.0047	***	0.0640	
	(0.0001)		(0.0334)		(0.0001)		(0.0485)	
Female	0.0019	***	−0.0174		0.0018	***	−0.0180	
	(0.0001)		(0.0249)		(0.0001)		(0.0363)	
Δ ln(County Emp.)			−1.8524	**			−1.9836	*
			(0.5722)				(0.8206)	
Δ ln(House Price Idx.)			−0.3514	.			−0.3530	
			(0.1844)				(0.2678)	
Minority Pop. Pct.					0.0001	***		
					(0.0000)			
ln(HUD Median Family Income)					−0.0059	***		
					(0.0005)			
AIC	−561,338		827,003		−572,134		826,728	

*** $p < 0.001$, ** $p < 0.010$, * $p < 0.050$, . $p < 0.100$

Standard errors in parenthesis. Models also contain geographic, year and purchaser-type dummies.

Data: Combined HMDA-PFF

Table 18: Pre-Foreclosure Filings by Predicted Interest Rate (Tobit Model #1)

interest rate	no PFF	received PFF	overall
under 4.000	18.9%	13.2%	18.4%
4.000 to 5.999	49.2%	47.0%	49.0%
6.000 to 7.999	26.0%	30.4%	26.4%
8.000 to 9.999	5.8%	9.1%	6.0%
10.000 to 11.999	0.1%	0.3%	0.1%
total	1,435,566	122,402	1,557,968
Data: Combined HMDA-PFF			

Table 19: Pre-Foreclosure Filings by Predicted Interest Rate (Tobit Model #2)

interest rate	no PFF	received PFF	overall
under 4.000	19.4%	13.7%	19.0%
4.000 to 5.999	48.4%	45.6%	48.2%
6.000 to 7.999	26.0%	30.9%	26.4%
8.000 to 9.999	6.0%	9.6%	6.3%
10.000 to 11.999	0.1%	0.3%	0.2%
total	1,435,566	122,402	1,557,968
Data: Combined HMDA-PFF			

Table 20: Distribution of Interest Rates in Pre-Foreclosure Filing Data

interest rate	total	percent
under 4.000	11,133	6.0%
4.000 to 5.999	49,876	26.8%
6.000 to 7.999	94,870	50.9%
8.000 to 9.999	21,643	11.6%
10.000 to 11.999	7,060	3.8%
12.000 to 13.999	1,430	0.8%
14.000 and up	354	0.2%
total	186,366	100.0%
Data: Full PFF		

default. Instead we believe that black race and Latino ethnicity are acting as a proxy for some missing variable that does increase their probability of default, such as differences in socio-economic status, racial and ethnic disparities in the impact of the recent economic recession and/or forms of discrimination that we cannot measure with the HMDA data.

As one would expect, the coefficient on applicant income was negative and statistically significant in both models. We could have used a quadratic term in the regression model to reproduce the result in table 3 (where we found that borrowers in the $80-200K income range were more likely to default), but given the possibility that income is correlated with some of the other explanatory variables, we were reluctant to over-fit the model. Testing a quadratic term is left to future research.

It's interesting to note that the coefficient on the percentage change in the home price index is only statistically significant at the 10 percent level in model 1 and is not statistically significant at all in model 2. By contrast, the coefficient on the percentage change in county-level employment is statistically significant at the 5 percent level in both models. Importantly, the effect of changes in county-level employment is large. *Labor market recovery would sharply reduce the rate of mortgage default in New York State*.

6 CONCLUSION

After a year of collecting data, the NYSBD had collected enough data to support this analysis, to begin studying the causes of mortgage default and foreclosure and to begin studying the role that racial and ethnic disparities played in causing the foreclosure crisis.

Importantly, this analysis also suggests that employment growth may have the strongest effect on the home mortgage default rate. *In the absence of employment growth, even large principal balance reductions would only have a minimal effect on the rate of mortgage default.* The coefficient estimates in section 5 imply that, for a borrower with a 20 percent probability of default, having taken out a loan 10 percent smaller (i.e. the equivalent of a 10 percent principal balance reduction) would only reduce the default probability to 19.3 percent. Regardless of how downward-biased our coefficient estimate might be, it seems clear that the large losses that principal balance reductions would impose on lenders would not be outweighed by lower default rates. There are no easy solutions.

Because mortgage servicers found themselves too understaffed to handle the wave of defaults (State Foreclosure Prevention Working Group, 2008) and because the foreclosure crisis has had a disproportionate impact on minority communities, Neiman's (2008) report concluded that lenders, servicers, counselors and governments must work together to identify deliquent borrowers, assign to them counselors who will help them achieve the best possible outcome given the circumstances. That enforced cooperation took the form of the pre-foreclosure filing project, which supplied the data for this analysis.

In his 2011 "State of the State" address, newly-inaugurated New York Gov. Andrew Cuomo assigned blame for the foreclosure crisis on both "Washington" and "Albany." Cuomo then proposed a merger of the state banking and insurance departments, which was enacted and became effective on Oct. 3, 2011. Since Cuomo's speech, neither the NYSBD nor its successor, the DFS, have issued a single report on the pre-foreclosure filing project. Nor has the pre-foreclosure filing project generated any policy response.

7 ACKNOWLEDGEMENTS

We would like to thank the New York State Banking Department for making the Pre-Foreclosure Filing data available to us and for supporting our research. The views expressed in this paper are our own opinions and do not necessarily reflect the opinions of the New York State Banking Department or its successor, the New York State Department of Financial Services.

NOTES

[1]Both authors of this paper are former employees of the NYSBD. One of us wrote code for the PFF database, worked with the mortgage servicers who filed the notices, took calls from foreclosure lawyers and, on occasion, took calls from terrified homeowners. As heads of the NYSBD's research unit, we supported the department's regulatory oversight of the home mortgage market by providing regular analysis of the HMDA data to the department's executive team. Neither the NYSBD nor its successor, the DFS, have compensated us for conducting this analysis or for writing this article. We wanted to write it because we believe that it is important to understand the causes of the subprime mortgage crisis and how it has disproportionately affected minority communities.

[2]New York Gov. Andrew M. Cuomo's 2011 budget abolished the state banking and insurance departments and merged their functions into the Department of Financial Services on Oct. 3, 2011.

[3]Rugh and Massey use the term "subprime" to describe high-cost loans.

[4]In the rare cases where the servicer did not include a loan number, we used the property address instead of the loan number.

[5]238,830 of the 248,556 (non-unique) addresses

[6]Mortgage servicers only file pre-foreclosure filing notices when the property is a primary residence, so when matching the PFF data to the HMDA data, we focus on mortgages originated for owner-occupied properties.

[7]More precisely, the HMDA data provide a value for the rate spread of a first-lien mortgage when it exceeds three percentage points. For other lien statuses, the HMDA data provides a value for the rate spread when it exceeds five percentage points. Our analysis focuses exclusively on first-lien mortgages.

REFERENCES

L. C. Adkins. "Small Sample Performance of Instrumental Variables Probit Estimators: A Monte Carlo Investigation". *Joint Statistical Meetings Proceedings, Business and Economics Statistics Section*, 4 Aug. 2008.

L. C. Adkins. "An Instrumental Variables Probit Estimator using gretl". *Econometrics with gretl, Proceedings of the gretl Conference 2009*, pages 59–74, 28-29 May 2009.

D. G. Bocian, K. S. Ernst, and W. Li. "Unfair Lending: The Effect of Race and Ethnicity on the Price of Subprime Mortgages", 31 May 2006. URL http://www.responsiblelending.org/mortgage-lending/research-analysis/rr011-Unfair_Lending-0506.pdf.

C. Bromley, J. Campen, S. Nafici, A. Rust, G. Smith, K. Stein, and B. van Kerkhove. "Paying More for the American Dream: The Subprime Shakeout and Its Impact on Lower-Income and Minority Communities", Mar. 2008. URL http://www.policyarchive.org/handle/10207/19021.

A. Cuomo. "State of the State Address", 5 Jan. 2011. URL http://www.governor.ny.gov/sl2/stateofthestate2011transcript.

M. Doms, F. Furlong, and J. Krainer. "Subprime Mortgage Delinquency Rates". Working Paper 2007-33, Nov. 2007. URL http://www.frbsf.org/publications/economics/papers/2007/wp07-33bk.pdf.

E. Doviak and S. MacDonald. "Who Enters the Foreclosure Process?". *Proceedings of the New York State Economics Association*, 4:20–35, Sept. 2011. URL http://nysea.bizland.com/nysea/publications/proceed/2011/Proceed_2011_p020.pdf.

S. Erle. "Geo-Coder-US-1.00: Geocode (estimate latitude and longitude for) any US address". Comprehensive Perl Archive Network, 17 May 2005. URL http://search.cpan.org/~sderle/Geo-Coder-US-1.00/.

K. S. Gerardi and P. S. Willen. "Subprime Mortgages, Foreclosures, and Urban Neighborhoods". Public Policy Discussion Papers, no. 08-6, 22 Dec. 2008. URL http://www.bos.frb.org/economic/ppdp/2008/ppdp0806.pdf.

E. Laderman. "Subprime Mortgage Lending and the Capital Markets". *FRBSF Economic Letter*, 2001(38), 28 Dec. 2001. URL http://www.frbsf.org/publications/economics/letter/2001/el2001-38.pdf.

D. P. Morgan, B. Iverson, and M. Botsch. "Subprime Foreclosures and the 2005 Bankruptcy Reform". *FRBNY Economic Policy Review*, 18(1):47–57, Mar. 2012. URL http://www.newyorkfed.org/research/epr/12v18n1/1203morg.pdf.

R. Neiman. "Report to the Interagency Task Force on Subprime Mortgages". NYSBD document, 27 July 2007.

R. Neiman. "Report to the Governor's Interagency Task Force to Halt Abusive Lending Transactions (HALT)". NYSBD document, 30 Apr. 2008.

New York State Banking Department. "90-Day Pre-Foreclosure Notice Report", 10 June 2010a. URL http://www.dfs.ny.gov/reportpub/pff100610.pdf.

New York State Banking Department. "90-Day Pre-Foreclosure Notice Report", 7 Oct. 2010b. URL http://www.dfs.ny.gov/reportpub/pff101007.pdf.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL http://www.R-project.org.

J. S. Rugh and D. S. Massey. “Racial Segregation and the American Foreclosure Crisis”. *American Sociological Review*, 75(5):629–651, 2010.

G. D. Squires, D. S. Hyra, and R. N. Renner. “Segregation and the Subprime Lending Crisis”. EPI Briefing Paper, 4 Nov. 2009. URL http://www.epi.org/page/-/pdf/110409-briefingpaper244.pdf.

State Foreclosure Prevention Working Group. “Analysis of Subprime Mortgage Servicing Performance”. Data Report No. 2, Apr. 2008. URL http://www.csbs.org/regulatory/Documents/SFPWG/DataReportApr2008.pdf.

T. Therneau and T. Lumley. *survival: Survival analysis, including penalised likelihood.*, 2009. URL http://CRAN.R-project.org/package=survival. R package version 2.35-8.

T. M. Therneau. “A Package for Survival Analysis in S”, 27 Jan. 1999.

US Census Bureau. “Census Tracts and Block Numbering Areas”, 19 Apr. 2000. URL http://www.census.gov/geo/www/cen_tract.html.

US Census Bureau. “Cartographic Boundary Files”, 27 June 2005. URL http://www.census.gov/geo/www/cob/tr2000.html.

US Census Bureau. “2006 Second Edition TIGER/Line Files”, 5 Mar. 2007. URL http://www.census.gov/geo/www/tiger/tiger2006se/tgr2006se.html.

US Department of Housing and Urban Development. “Unequal Burden: Income and Racial Disparities in Subprime Lending”, May 2000a. URL http://archives.hud.gov/reports/subprime/subprime.cfm.

US Department of Housing and Urban Development. “Unequal Burden in New York: Income and Racial Disparities in Subprime Lending”, May 2000b. URL http://www.huduser.org/Publications/pdf/newyork.pdf.

.