FEDERAL RESERVE BANK OF MINNEAPOLIS COMMUNITY AFFAIRS REPORT

Report No. 2007-1

Targeting Foreclosure Interventions: An Analysis of Neighborhood Characteristics Associated with High Foreclosure Rates in Two Minnesota Counties

Michael Grover
Associate Economist ▪ Community Affairs ▪ Federal Reserve Bank of Minneapolis

Laura Smith
Assistant Professor ▪ Department of Geography ▪ Macalester College

Richard M. Todd
Vice President ▪ Federal Reserve Bank of Minneapolis

Revised June 2007*

A report series from the Community Affairs Department of the Federal Reserve Bank of Minneapolis.

**This revised version reflects comments made at the 2007 Federal Reserve System Community Affairs Research Conference.*

Targeting Foreclosure Interventions: An Analysis of Neighborhood Characteristics Associated with High Foreclosure Rates in Two Minnesota Counties

By Michael Grover, Laura Smith, and Richard M. Todd

Abstract: This study examines the statistical association of foreclosure sales with social, economic and housing variables measured at the census tract level for two purposes of interest to foreclosure mitigation practitioners—to assess whether it is feasible to identify in advance neighborhoods likely to have high rates of foreclosure, and to explore the socioeconomic traits of high-foreclosure neighborhoods so as to design appropriate mitigation programs. We collected data on foreclosure sales in 2002 from the sheriff's departments of Hennepin and Ramsey counties, the two core counties that comprise the Minneapolis-St. Paul MSA. We find that several factors commonly associated with high-foreclosure sale rates could have correctly identified, in advance, most neighborhoods with high rates of mortgage foreclosure. To guide the design of foreclosure mitigation programs, we also present evidence that foreclosure risks in our two counties were highest in neighborhoods with elevated credit risk indicators and a high proportion of homeowners who are recent minority buyers or young. We show that an accurate credit risk variable is among the best predictors of foreclosure and also critically affects our multivariate analysis of factors associated with foreclosure. To limit social losses associated with foreclosures, we conclude that consideration should be given to enhancing public access to data on mortgages, foreclosures, and foreclosure risk factors, especially the neighborhood distribution of credit scores.

I. Introduction

Public policy promotes homeownership generally and recently has also sought to close racial and ethnic homeownership gaps. Many such policies focus on barriers to becoming a homeowner, but research shows that the ability to sustain homeownership is also a policy concern. It has been shown that disparate rates of exit from homeownership account for a significant share of racial homeownership gaps and that exits via foreclosure involve externalities that could lead to an inefficient allocation of resources.[1]

In a number of cities, campaigns backed by public and private resources have been mounted to mitigate the personal and public impact of foreclosures generally and spatially concentrated foreclosures in particular. Although many of these efforts seem to be achieving good results already, [2] our working assumption is that analysis of foreclosure data can enhance their effectiveness and efficiency in two ways. First, by predicting the neighborhoods where future foreclosure rates will be high, analysis can help

[1] See Haurin and Rosenthal (2004); Reid (2004); and Apgar and Duda and Nawrocki Gorey (2005).

practitioners target their resources to the neediest neighborhoods. Second, by providing reliable information on the characteristics of neighborhoods and borrowers that are most closely associated with high foreclosure risks, analysis can help practitioners shape the content of their interventions to more effectively address the needs of borrowers in foreclosure.

Our analysis illustrates that useful information can be easily generated if the necessary data on foreclosures and key related factors are readily available. Using data from Hennepin and Ramsey counties,[3] the core of the Twin Cities metropolitan area, we show that indicators of the low-income, minority population, or credit risk could have been used to identify neighborhoods with high foreclosure rates in advance, thereby allowing foreclosure counselors and others to focus their prevention and mitigation efforts on these areas.[4]

Although numerous studies have discussed factors correlated with foreclosure, we believe we are the first to explicitly analyze how accurately high foreclosure neighborhoods can be identified in advance. We are also one of the few studies of neighborhood foreclosure patterns that measures neighborhood credit risk levels with data based on actual credit scores. This variable is not only our best single predictor of future foreclosure rates, but it is also critical to our multivariate analysis of the neighborhood traits associated with foreclosures. For our study area, a high percentage of adults with very low credit scores and an upward trend in the percentage of minority homeowners were the two factors most strongly and robustly associated with high foreclosure rates in 2002. When we substitute a well-known but less accurate credit risk proxy, based on mortgage denial rates, our analysis of related factors is qualitatively different. An association between foreclosures and the minority share of the general or home owning population also becomes significant, apparently because the proxy systematically underestimates credit risks in neighborhoods with high minority populations. The proxy's tendency to focus attention directly

[2] See Apgar and Duda (2005b); Collins and Nawrocki Gorey (2004); Baker (2004); Quercia, Cowan, and Moreno (2004); Home Ownership Center (2004); and Neighborhood Housing Services of Chicago (2005).
[3] Hennepin and Ramsey counties include the cities of Minneapolis and St. Paul respectively; both contain numerous suburbs.
[4] A complication to our analysis is that some well-regarded foreclosure counseling organizations were operating in our study area during our study period. So our data on foreclosures should be thought of as foreclosure sales rates net of foreclosure sales prevented by existing foreclosure counseling efforts.

on minority presence, rather than on the credit history issues that seem to underlie the minority share and foreclosure correlation, could mislead mitigation program design.

We also discuss how our multivariate results square with the literature on foreclosures and neighborhood transition, arguing that existing interpretations derived from experience in cities with falling housing prices may not provide a good explanation for foreclosures in Minneapolis and St. Paul. As an alternative, we suggest a simpler interpretation relating foreclosures to increased home buying by households who may have been economically vulnerable or financially inexperienced.

Our results suggest that there may be social benefits to making some of the data we use more routinely available to the public. Public foreclosure and mortgage records, which are now often unavailable to the public in electronic formats that facilitate analysis, can help identify existing foreclosure hotspots and vulnerable populations for targeted interventions. They can also help identify the types of loans and the individual lenders most commonly involved in foreclosures. Although further analysis of the costs and privacy issues associated with greater availability is needed, we conclude that there is at least a potential public interest in making these data more readily available to foreclosure mitigation efforts, and we make some specific suggestions as to the kinds of changes that should be considered. In addition, our most effective predictor and the key variable in our analysis of related factors—the prevalence of low credit scores in each neighborhood—is normally not available to public and nonprofit foreclosure mitigation efforts. We suggest that consideration also be given to how that data might be made more available to foreclosure mitigation practitioners.

II. Factors Affecting Mortgage Foreclosures

In the 1990s, researchers developed models of the factors associated with foreclosures, in order to more accurately price mortgage-backed securities. The research has focused on two distinct causes.[5] According to the *options theory* of foreclosures, foreclosures occur when the value of the property falls sufficiently below the outstanding loan balance, so that the borrower finds it profitable to exercise the option to put the collateral to the lender in lieu of continuing a stream of payments whose present value exceeds the value of the property. In empirical work, these so-called ruthless foreclosures are often associated with homes whose market value has stagnated or declined and with loans with a high initial loan-to-value ratio. According to the *trigger event theory*, foreclosures occur when the borrower experiences financial setbacks, such as job loss, divorce, or medical expenses, that make it difficult to continue making payments. In empirical work, these foreclosures are often associated with low credit scores or high unemployment rates. The two types of causes can merge, as when a borrower with limited equity loses a job, misses payments, and incurs penalties that push the amount owed above the value of the home. We draw on both theories below.

III. Data

The data for this study come from several sources. Researchers from the University of Minnesota and Federal Reserve Bank of Minneapolis jointly collected public data on 1,199 foreclosure sales for Hennepin and Ramsey counties in 2002.[6] The foreclosure sale documents contained information about the property, original mortgage, mortgagor and mortgagee names, mortgage assignments, and the sheriff's sale. In total, we obtained usable data on 1,178 residential properties from the sheriff's sale documents.[7]

Collecting the 2002 data through public sources proved problematic from the start. Neither county had the sales records available in electronic format, so all 1,199 records were obtained as hard copies. Since the foreclosure sale documents only listed legal descriptions, we needed to match the legal descriptions of all foreclosed properties to unique street addresses. After this step, the 1,178 usable street addresses were geocoded and matched to a census tract. All of the other information on the sale document had to be manually entered into the database before it could also be analyzed. Overall, this process took several months to complete.

---

[5] Ambrose and Capone (1998).

[6] Our team collected and entered the 2002 foreclosure sale data for Ramsey County. We thank Jeff Crump and his associates at the University of Minnesota for collecting and entering the data for Hennepin County.

The address of the foreclosed property was the most important piece of information for this research, but not the only one considered. The researchers also sought to identify other patterns in the foreclosures, such as the lenders involved, interest rates, mortgage riders or conditions, and the dates of the mortgage originations. Additional time-consuming data collection at the property records departments of both counties was required in order to complete, revise, or add to the foreclosure sales records.

Two pieces of data proved especially laborious to confirm. First, it was nearly impossible to distinguish refinance from home purchase loans, even after reviewing mortgage documents filed at the counties' property records offices. Second, for some documents it was difficult to determine the lender of record. For example, a small portion (roughly 5 percent in 2002) of the sale documents listed the Mortgage Electronic Registry System (MERS) as the lender.[8] While additional reviews of county property records did allow for the resolution of most unconfirmed lending transactions in 2002, unless this step is taken it is not possible to know who the lender was. The additional steps we took to collect and refine the foreclosure sale information limited the amount of data that we could reasonably obtain and forced us to focus on a single year and the two core counties.

After this initial phase, we made several additional steps to improve the quality of the data. These efforts resulted in the creation of three datasets used in our analysis. Two datasets use individual property records as the unit of analysis. The first one refines and enhances the 1,178 individual foreclosure sale records. In order to standardize the list of mortgagees, we used information from the Federal Reserve System's National Information Center (NIC) to place all foreclosed mortgages under a current lender. We used the NIC database to identify and categorize lending institutions as either a bank, bank affiliate, or

---

[7] The total number of residential foreclosure sales recorded by both counties was actually 1,199 in 2002. Data on 21 of the foreclosure sales, however, were excluded because each of these sales could not be accurately matched to a street address.

[8] MERS-related loan volume increased substantially in just three years, accounting for more than one-third of the foreclosure sales in Hennepin County alone in 2005. A Virginia-based company, MERS acts as nominee in the county land records for the lender and the loan servicer and eliminates the need for these two entities to file assignments when trading loans. This service can make the process of trading mortgages on the secondary market more efficient. MERS can appear as either the record mortgage holder (commencing the foreclosure) or the purchaser of the foreclosed property, or both. MERS is not the only entity that obscures information about a lender. Other financial entities, such as banks, often act as trustees in property transactions. See Apgar and Duda (2005a), 26–27.

other lending institution. We also identified lenders as being either prime or subprime using the 2002 Department of Housing and Urban Development's list of Subprime and Manufactured Lenders.[9]

We took several steps to augment the data contained on the sheriff's sale document. We used public property records databases in both counties to collect data on the interest rate and mortgage riders for the mortgages under foreclosure. Of the 1,178 sets of documents reviewed, 62 percent contained interest rate and mortgage rider information. To see if foreclosed mortgages tend to have above-average interest rates, we matched the original interest rate on each foreclosed mortgage to the average monthly rate on 30-year conventional mortgages originated in the same month as the foreclosed mortgage, using data posted by the Board of Governors of the Federal Reserve System from the Federal Home Loan Mortgage Corporation.

We created a second dataset of properties that were not under foreclosure sale in 2002. We took this additional step in order to compare pricing and mortgagee information for these properties with those that were under foreclosure. To do this, we selected a residential area north of downtown Minneapolis, which exhibited a high rate of foreclosure sales in 2002, and selected a 20 percent random sample (198 properties identified from a county parcel data set) from a list of total properties sold in the area between 1998 and 2002. This five-year time period roughly corresponds to the origination dates for most foreclosed mortgages in the area. Using a public-property records database at Hennepin County, we gathered information on these other mortgages, including the mortgagee, interest rate, and rider information. We also matched the interest rate on each of these mortgages to the average rate on 30-year conventional mortgages originated in the same month. We use these two datasets to provide an overview of foreclosure sale patterns in both counties.

Finally, we organized a third dataset of economic and demographic information for the census tracts of Hennepin and Ramsey counties. We merged some census tracts to facilitate analysis, specifically to allow for the linking of census data from 1990 and 2000 due to tract boundary changes and to the fact

[9] See the following HUD Web site for more information on how subprime lenders were identified, *www.huduser.org/datasets/manu.html*.

that some tracts had too few mortgaged properties to allow computation of a meaningful foreclosure sale rate.[10] We base most of our analysis of the neighborhood factors associated with foreclosures on 327 "merged tracts" for which we have complete data. With the exception of two essentially nonresidential tracts that were omitted, our 327 merged tracts cover all of Hennepin and Ramsey counties. From this point on, we will use the terms "tract" and "merged tract" synonymously, unless otherwise noted. As is conventional, these tracts are what we mean by a "neighborhood." The location information available for each foreclosure sale permits mapping the precise location of each foreclosure as well as assigning each one to a distinct tract.

A key variable from Census 2000 is each tract's number of owner-occupied housing units with a mortgage outstanding. Following Apgar and Duda (2004) and others, we use this variable as the denominator in calculating the 2002 foreclosure sale rate for each tract. The approximate two-year gap between our measure of mortgaged units and our measure of foreclosure sales, although imperfect, nearly aligns with the median time between mortgage origination and foreclosure in our sample. In that sense, we regard the Census 2000 data on mortgaged units to be an acceptable denominator for computing an annual foreclosure sale rate.

We augmented the census data with data on credit scores and mortgage market activity. From mortgage data collected under the Home Mortgage Disclosure Act (HMDA), we selected home purchase origination and denial rates from 1996–1999. We also used 1999 credit worthiness data obtained by the Federal Reserve Bank of Minneapolis through the Board of Governors of the Federal Reserve System from PCI Corporation. Using the PCI data for each tract on the total number of adults and the number of adults whose credit score in 1999 was in the "very high risk" category (i.e., who had very low credit scores), we created a variable for the percentage of adults in each tract who had a very risky credit score in 1999.[11] Before using our data to see how well foreclosure hotspots could have been predicted, we

[10] In all cases, tracts with low counts of mortgaged properties were merged with adjacent tracts.
[11] Credit risk categories ranged from very low to very high credit risk and also included the number of adults in the population with no credit information. In our analysis, we experimented with additional categories such as "high risk" and "no credit score" but found that they added little. Credit risk variables for merged tracts were computed using population-based credit risk data obtained from PCI Corporation and CRA Whiz through a contract with the Board of Governors of the Federal Reserve System.

provide a general overview of Hennepin and Ramsey County foreclosure sales and foreclosure sale rate patterns.

IV. Overview of Hennepin and Ramsey Counties Foreclosure Sales

National foreclosure rates began to rise in the late 1990s and hit new peak levels (at least by modern standards) in the early 2000s before declining again.[12] Rising foreclosure rates coincided with the rapid expansion of subprime lending and widespread allegations of “predatory” lending, often in the same neighborhoods that were experiencing high rates of foreclosure. This combination focused attention on the public policy issues raised by foreclosures, especially in cities where foreclosures tended to concentrate their public and private burdens in poor neighborhoods. As a result, numerous studies of metropolitan real estate markets have documented some of the borrower, lender, and neighborhood factors that correlate with high rates of foreclosure.[13] Since average housing prices were rising in most cities during the late 1990s and early 2000s, most of these factors relate to the trigger event theory of foreclosures, not options theory. Our data generally support previous research findings.

When standardized by the number of mortgaged units in a neighborhood, the foreclosure sale rate was similar in our two counties in 2002 (Table 1). [Tables begin on Page 36 of this report; Figures are listed immediately after the tables section.] However, the central cities of Minneapolis (Hennepin) and St. Paul (Ramsey) have much higher foreclosure sale rates than the suburban municipalities around them, as also shown in Figure 1. This figure displays Hennepin (to the west) and Ramsey counties, their major bodies of water and highways, and the boundaries of the cities of Minneapolis and St. Paul. Each dot in the figure represents the location of one of the 1,178 residential foreclosure sales from 2002 that we analyze. Foreclosures occurred throughout the two-county area but varied considerably in density, with the highest density areas to the northwest and southeast of downtown Minneapolis and in a belt around

---

[12] These conclusions are based on quarterly data from the Mortgage Bankers Association of America on the percentage of outstanding mortgages for which the legal process of foreclosure has been initiated but not yet completed.
[13] See Apgar and Duda (2004); Apgar and Duda (2005a); Garcia (2003); Gruenstein and Herbert (2000); Immergluck and Smith (2004); Lauria and Baxter (2000); Newburger (2006); and The Reinvestment Fund (2005).

downtown St. Paul. Converting to neighborhood foreclosure sale rates—2002 foreclosures per 100 owner-occupied, mortgaged units in Census 2000—largely preserves this spatial pattern of concentration (Figure 2).

As in other studies, our foreclosed mortgages are disproportionately of recent origin and carry higher rates of interest. In our 2002 data, the median duration before a foreclosure sale in Hennepin and Ramsey counties was 2.6 years. Also, as shown in Figure 3, most (86 percent) of the foreclosure sales with interest rate information had a rate above the simultaneous prevailing rate on 30-year fixed-rate prime mortgages. Most of the foreclosure sales with adjustable-rate mortgages had an interest rate significantly higher than the simultaneous prevailing rate; by contrast, the majority of foreclosure sales with fixed-rate mortgages had an interest rate within one percentage point of the prevailing rate. In short, most Hennepin and Ramsey County foreclosure sales in 2002 occurred within three years of origination of the mortgage and about half involved loans with high rates of interest. Figures 4 and 5 further illustrate these patterns.

Given the high interest rates on many foreclosed mortgages, and the prevalence of nonbank subprime lenders in the high-risk mortgage market, it is not surprising that most of the mortgages foreclosed in 2002 were originated by lenders other than federally regulated banks or their affiliates (Figure 6). Examining individual lenders reveals that the largest share of foreclosed sales (135 out of 1,178, or 13 percent) had mortgages originated by Ameriquest or two of its affiliates (Long Beach Mortgage Company and Town and Country Credit). This is significant because allegations of serious lending abuses in the late 1990s and early 2000s have been made regarding Ameriquest offices in the Twin Cities.[14] It is possible, therefore, that a sizable portion of the 2002 foreclosure sales in our data may reflect what many would deem abusive or predatory lending.

---

[14] See Ameriquest Press Release, "Ameriquest Announces Agreement with States," and *Los Angeles Times*, "Workers Say Lender Ran 'Boiler Rooms.'"

The high interest rates and other high-risk attributes typical of mortgages foreclosed in 2002 are not necessarily typical of all mortgages in their neighborhood, even in neighborhoods with relatively low incomes and high foreclosure sale rates. For a high-foreclosure area north of downtown Minneapolis, Table 2 shows some statistically significant differences between our foreclosed mortgages and a random sample of other mortgages originated in the same area between 1998 and 2002, when the majority of the area's foreclosed mortgages were originated. Compared to typical mortgages in their neighborhood, the foreclosed mortgages were smaller, had higher interest rates, and were more likely to have been originated by a nonbank or subprime lender. Foreclosed properties in this area were also more likely to have another mortgage on the property as well.

Overall, we found that 80 percent of borrowers in foreclosure owed more on their mortgages than the original principal amount. The median difference between the principal and the amount outstanding at the time of the foreclosure sale was about $4,500 and was likely due to missed payments, fees, and unpaid taxes. Still, in nine out of 10 foreclosure sales, the sheriff sale amount was more than the amount outstanding on the mortgage, with a median difference of approximately $3,400. The fact that most foreclosed properties were sold for more than the outstanding amount due on the mortgage is likely a reflection of the strong and appreciating housing market in the region that began in the mid-1990s. This finding suggests that many borrowers in foreclosure in the Twin Cities in 2002 might have paid their outstanding debt and departed with some equity. This finding further supports the view that timely, well-targeted mitigation efforts can reduce the overall costs of foreclosures to borrowers, lenders, and the general public.

Switching our perspective from the individual loans to the Hennepin and Ramsey county neighborhoods that had high foreclosure sale rates in 2002, we find that they display many of the characteristics that earlier studies have associated with elevated foreclosure sale rates, as shown in Table 3. Consistent with the trigger event theory, foreclosure sale rates in our 327 merged tracts have strong (more than 0.65) positive correlations with three indicators of credit risk and three indicators of high or increasing minority presence. Figure 7 further illustrates the tendency for foreclosure sale rates to rise with the minority share of a neighborhood's population. Other credit risk and minority variables, as well as the neighborhood's rate of unemployment, are fairly strongly positively correlated with foreclosure sale rates as well. A high share of young homeowners, who often have less wealth and less stable employment than older borrowers, is somewhat positively correlated. By contrast, levels of education and income correlate negatively with foreclosure sale rates to a moderate or strong degree. Figure 8 shows that both high-income and some low-income neighborhoods have low foreclosure sale rates, while high foreclosure sale rates are concentrated in other low-income neighborhoods. The positive correlation between FHA lending and foreclosure sale rates provides some support for the options theory of foreclosures too, since FHA loans can indicate a tendency toward higher loan-to-value ratios.

V. An Assessment of Variables That Can Predict Foreclosure 'Hotspots'

Having described some of the factors associated with high foreclosure sale rates in our study area, we now assess how well individual variables, that would have been known in advance and that foreclosure theory and studies say are correlated with foreclosure rates, would have performed as predictors of high foreclosure risk neighborhoods in 2002. We find that most of the variables we tried would have correctly identified in advance about two-thirds of the top quintile of foreclosure rate neighborhoods in 2002. Overall, our 1999 neighborhood credit score variable is the most accurate in identifying high foreclosure-sale rate tracts, but other variables available before 2002, like average 1990 household income or the 1990 non-white population share, also do well. We present maps further illustrating how several variables could have been used to direct foreclosure mitigation resources to most Twin Cities areas of highest need in 2002, but the credit score variable seems at least marginally superior, due to a lower tendency to falsely identify neighborhoods as high risk.

Table 4 shows how seven variables available before 2002 compare in their ability to correctly rank tracts according to their foreclosure sale rates in 2002. None of the variables correctly ranks even half of the 20 highest foreclosure-sale rate tracts in its top 20,[15] but the percent of tracts correctly identified improves when the comparison expands to the top 50 and top 65 (first quintile) foreclosure-sale rate tracts. The ranking based on credit score correctly identifies 72 percent of the top 50 foreclosure-sale rate tracts, and rankings based on the percent of the population living in poverty in 1990, (the negative of) average 1990 household income, and a high incidence of subprime refinance mortgages originations in 1999 were nearly as good, identifying 70 percent of the top 50 foreclosure-sale rate tracts. The credit score variable has a run of success as the number of tracts ranked expands to the top quintile of foreclosure-sale rate tracts, at which point it correctly identifies 77 percent, compared to 55 percent to 68 percent for all of the other variables in Table 4. Expanding the rankings further, to the top 100 foreclosure-sale rate tracts, narrows most of these differences, with the credit score correctly identifying 72 percent compared to 71 percent for the mortgage denial variable and 65percent to 69 percent for most of the rest (except that the percent of non-white homeowners in 1990 gets only 52 percent correct).

The figures on the number of top foreclosure-sales rate tracts correctly identified were chosen in part because they focus where we think practitioners would want to focus—only on the top end of the distribution of foreclosure incidence. To assess the ability of our variables to rank the entire distribution of foreclosure rates, Table 5 shows Spearman rank correlation coefficients between each variable's ranking of all 327 tracts and the actual ordering of foreclosure rates for all tracts. Again, the 1999 credit score variable performs best, with a correlation of 0.640. Its closest competitors are the 1996-99 mortgage denial rate, the incidence of subprime refinancing, and (the negative of) average 1990 household income, whose correlations are lower by 7 or 8 hundredths. The other 1990 Census variables have correlations that are also lower than the 1999 credit score variable.

[15] We discovered, accidentally, that a tract's 1990 public assistance income per capita ranks 11 of the top 20 foreclosure sales rate tracts in its top 20, but we have not focused on this result because per capita public assistance income is not a variable typically considered in analyses of foreclosure incidence and hence would probably not have been considered, in real time, for predicting areas of high foreclosure rates in 2002. However, it also performed well at identifying tracts in the top 50 (35 correct), top 65 (44 correct), and top 100 (68 correct), so it may be useful to evaluate its predictive potential with other data sets.

Although the credit score consistently leads the pack, its edge is often slim and only sometimes significant by conventional standards, especially when we focus just on identifying the top 100 or fewer foreclosure-sale rate tracts. We compute statistical significance with a series of pairwise modified permutation tests, each based on the null hypothesis that rankings based on the credit score variable and one of the other variables are equivalent with respect to predicting or correlating with the ranking of foreclosure-sale rate tracts. (In all cases, rankings are listed in an unchanging order, from the ranking of Tract no. 1 in row no. 1 to the ranking of tract no. 327 in row no. 327.) For example, consider the null that rankings based on the credit score variable and the percent poor variable are equivalent. If so, then randomly swapping subsets of data between the two variables should not lead to a systematic difference in their performance. To exploit that idea, we randomly select 1,000 327x1 vectors of ones and zeros. For each of these vectors, we then identify all the rows with ones and ignore those with zeros. For the rows that have a one, we exchange the rank in the credit score variable with the corresponding rank from the other variable, and then recompute the rankings for each modified variable. We also compute our statistics of interest for each of the 1,000 pairs of randomly modified variables as well as the difference between the credit score variable's statistics and the other variable's statistics. The result is a nonparametric estimate, which imposes the null hypothesis of equivalence but no other assumptions, of the distribution of the difference in the statistics.

We use this estimated distribution to judge whether the actual difference between our credit score variable's statistics and the other variables' statistics is consistent with the null hypothesis of no difference. For the Spearman rank correlations, which put equal weight on all portions of the entire ranking of the 327 tracts, the fourth column of Table 5 shows that we consistently reject the null that the other variables are equivalent to the credit score variable. When we focus only on the top 100 or fewer foreclosure-sale rate tracts, the results are much less decisive and depend on exactly how many top tracts we include. Among the top 20 foreclosure-sale rate tracts, the credit score variable is clearly only on par with the three other variables that also correctly identify nine of the 20. For the top 50 tracts, the credit score variable has a numeric edge, but the only variable with a statistically significant performance lag, judging by the p-values in parentheses in Table 4, is the percentage of non-white homeowners in 1990. With the credit score variable performing especially well as the focus expands to the top 65 tracts, its advantage over all the other variables becomes significant at that point, using a 10 percent significance level and making no allowance for the fact that multiple tests are being conducted. However, it gives back some of the significance of its edge if we further expand to the top 100 foreclosure-sale rate tracts. At that point, the p-values are 0.037 for percent poor in 1990, 0.034 for percent non-white in 1990, 0.000 for the percent of non-white homeowners in 1990, 0.122 for (the negative of) average 1990 household income, 0.165 for subprime mortgage refinancing incidence, and 0.469 for mortgage denial rates.[16,17]

Additional insights into how these variables might have guided resource allocation can be gained from the maps in Figures 9 to 15. The map corresponding to each variable codes each tract as a "hit" (correctly identified as in the top quintile of foreclosure sale rates in 2002, shown in red), a "miss" (incorrectly identified as out of the top quintile when actually in the top quintile, shown in yellow), a "false positive" (identified as in the top quintile when actually out of that quintile, shown in blue), or as "other" (correctly identified as not in the top quintile, shown as white). Figure 9, for example, shows the hits, misses, and false positives based on ranking tracts by the percentage of adults with a very risky credit score. The red areas corresponding to hits cluster in four high foreclosure sale rate areas—central St. Paul, south-central Minneapolis, northwest (known as "North") Minneapolis, and a suburban area just north of Minneapolis. The credit score variable would have successfully guided resources to these four high-

[16] We tried other variables and simple combinations of variables with similar results, and we suspect that additional variables, like the percent of the population with a high school degree, might also work well. In recognition of Apgar and Duda's (2004, 2005) analysis of foreclosure patterns based on a combination of income and race, we also created a variable that flagged a tract as high risk if the tract was in the riskiest quintile for both 1990 household income and 1990 minority population share. The performance of this indicator for the top 50 or 65 was in line with the other good performers shown in Table 4.

[17] Because data from Census 2000 would not have been available to guide resource allocation in 2002, the results presented use Census variables from just the 1990 Census. This gives an advantage to variables that are more frequently updated like the credit score or the HMDA-derived variables, whose values through 1999 are used above. To assess whether more up-to-date estimates of the Census variables would have been useful, we also evaluated Census 2000 versions of all the Census variables. The results were similar to the 1990-based results presented, except that the performance of the percentage of non-white homeowners' variable improved markedly. Had it been available in advance, it would have correctly identified 11 of the top 20 tracts, 36 of the top 50, and 47 of the top 65. Its Spearman rank correlation is 0.044 lower than that of the credit score variable, but this difference is not significant (p=0.125). The fact that the percent of non-white homeowners changes from a weak competitor to the credit score variable using 1990 data to a strong competitor using 2000 data is related to one of the findings of the next section's multivariate analysis—that change in the percentage of non-white homeowners is significantly correlated to foreclosure rates even after controlling for the neighborhood distribution of credit scores. However, it is not clear that this relationship could have been exploited in real time to better anticipate foreclosure rates in 2002.

foreclosure hotspots. Its misses, in yellow, are often adjacent to the red areas correctly identified, so, in practice, resources allocated to the red areas also would have been easily accessed by residents in many of the missed tracts as well. Similarly, most of its false positives, in blue, were near to actual high risk tracts, so directing resources there would not have been especially expensive (on the margin) or wasteful. However, this variable's failure to identify high foreclosure risks in suburban areas along Minneapolis' western border might be problematic, to the extent that foreclosure mitigation resources sometimes do not cross such political boundaries easily. Its remaining misses are widely scattered and include some high-income neighborhoods; it is unclear whether public and nonprofit foreclosure mitigation resources could or would have been especially targeted there even if these areas had been identified in advance.

Figure 15 shows that using the rate of prime mortgage denials to guide resource allocation would have given fairly similar results. However, this variable has more misses and its false positives are more extensive and not as near to actual high foreclosure-sale rate tracts. So, using this variable to allocate resources could have led to increases in cost and potential waste. Figures 10 to 14 map the results for the other variables in Table 4 and show large areas of false positives in central and south Minneapolis, and sometimes in suburban areas too. Compared to Figure 9, they also show more misses, especially in the northwest portion of Minneapolis and its near northern suburbs.

Overall, we think that the credit risk variable would have led to the most effective allocation of foreclosure mitigation resources in Hennepin and Ramsey counties in 2002. It correctly picks out the major areas where high foreclosure-sale rate tracts cluster, and its misses and false positives are mostly near the same areas. It has the lowest rate of false positives for the top quintile of foreclosure risk and would have led to only limited waste. Because it is designed to identify households whose debt repayment ability is weak, we also have some faith that its success was not just accidental and might generalize to other years or places. It has the potential advantage that it is kept up-to-date continuously, unlike Census and other official data.[18,19] We are less confident that the reasonably good performance of some of the

[18] We also mapped many of the same Census variables using Census 2000 data, even though it would not have been available to practitioners before 2002, to assess whether more up-to-date values would improve the somewhat inferior performance of the

other variables in 2002 in our study area would hold up over time and space, as conditions in society and the mortgage business change. For now, however, the Twin Cities' experience in 2002 suggests that variables like prime mortgage denial rate, household income, minority population share, and the prevalence of subprime refinancing may be effective supplements to or substitutes for credit risk information, which is often not readily available to nonprofit foreclosure mitigation organizations. Of course, additional research is needed to assess whether these findings generalize to other times and places.

VI. A Multivariate Analysis of Factors Associated with High Rates of Foreclosure Sales

In this section, we are no longer analyzing forecasts but instead conduct a multivariate after-the-fact analysis of factors associated with high foreclosure sale rates. The hope is that understanding those factors, even after the fact, can help improve the design of future foreclosure prevention programs.

A. Overview

The many neighborhood factors that correlate with high foreclosure sale rates are generally moderately to strongly correlated with each other as well. Pairs of variables from Table 3, for example, mostly exhibit correlations of 0.50 or higher, and correlations of 0.70 or higher are also present. The general pattern of multicollinearity among the factors commonly associated with foreclosures makes it difficult to understand the nature of the relationships between these factors and foreclosures. Although we think a full, causal understanding of these relationships cannot be extracted from our cross-sectional data, for the purpose of guiding the design of foreclosure mitigation programs we would like to shed some light on the relative independence or redundancy of the information conveyed about foreclosure sales rates by the various factors. We find that high 2002 foreclosure sales rates in our study area seem most clearly associated with two dimensions of our neighborhood data—credit risk and rising minority homeownership. This type of finding may be of use in developing effective intervention strategies now

---

Census variables. The credit risk variable remained superior, in our judgment, even when compared to Census 2000 data for income and minority shares, using the same criteria as above.

[19] Of course, fully exploiting its advantages would require making neighborhood credit score data more readily available to government agencies and nonprofit organizations engaged in foreclosure mitigation. We discuss this later in the paper.

and could grow in importance as a resource-targeting tool as additional years of foreclosure and related data accumulate. We also comment on whether these results validate existing causal models of foreclosure sale rates.

B. Methodology

The basic variable of interest is the 2002 foreclosure sale rate, fr_i, defined for each merged tract as the number of foreclosure sales in the tract in 2002 divided by the number of owner-occupied, mortgaged units reported in the tract in Census 2000. A foreclosure sale in a given year is an unlikely event—the highest rate among tracts in our 2002 sample was under five foreclosure sales per 100 mortgaged units. For that reason, a large population of mortgaged units is needed to accurately measure a neighborhood's probability of foreclosure. However, many of our tracts, even after merging those with the lowest numbers, had only 100 to 300 mortgaged units. Others had far more. It is thus likely that our foreclosure sale rate variable measures underlying foreclosure probabilities with error, and that the variance of that error varies systematically between tracts with high and low numbers of mortgaged units. This is a common problem for analysts of rate or proportions data.

To adjust for the resulting heteroscedasticity of the error term in our regression results, we use the logit version of the minimum Chi-squared estimator (MCSE) for proportions data that is described in Greene (2003).[20] To avoid taking logarithms of zero in tracts that had no foreclosures, we define the dependent variable in our MCSE regressions as $lfr_i = \log [(fr_i^*+eps)/(1-fr_i^* +eps)]$, where eps is a small number.[21] For comparison, and because weighted regression statistics cannot be used to compare the fit of alternative specifications, we present the ordinary least-squares (OLS) results for lfr_i produced in the first stage of the MCSE procedure.

[20] See Greene (2003), Chapter 21, pp. 686–689.
[21] In the results we present, the small number is one-half of the smallest positive value for fr_i, or about 0.00019. We experimented with values that ranged between twice as large and five times as small, with no important qualitative change in our conclusions. Of course, the specific values of coefficients and goodness of fit statistics do vary with the choice of the replacement for zero values. All of our regressions were implemented by code we wrote for Estima's RATS for Windows 6.10 software package.

C. Results

For our study area in 2002, the lfr_i regressions show strong direct associations among high foreclosure sale rates and three categories of variables—credit risk variables, minority homeownership transition variables, and demographic factors, chiefly age. When measured by actual credit score information, the credit risk variable provides the strongest ability of any single variable we tested to account for the neighborhood pattern of foreclosure sales. The combination of this credit risk variable and a measure of the change in minority homeownership share from 1990–2000 is nearly as effective in explaining the pattern as all other sets of variables we tried. However, other variables can modestly improve the fit of our regressions or are significant in our minimum Chi-squared estimates. Some of the other variables seem to augment the model's credit risk information. They improve the overall fit and somewhat diminish the coefficient on the credit risk variable. Other variables, such as those pertaining to age, may add a new dimension to the model, beyond credit risk and minority transition.

We also find that the year 2000 level of (as opposed to the 1990–2000 change in) the share of minorities (in the population or among homeowners) is highly correlated with credit risk variables. This creates a strong indirect association between foreclosure sale rates and static minority share variables that is of independent interest. Static minority share variables can also appear to have a direct relationship to foreclosure rates if high-quality measures of credit risk are omitted from the analysis.

Table 6 summarizes our multivariate results. Model 1 attempts to explain the neighborhood pattern of foreclosure sale rates using only the variable VHCREDIT2, the percentage of neighborhood adults whose credit score is in the very risky (i.e., very low) category. Based on the percentage of variance explained in OLS regressions that include a constant and one explanatory variable at a time, we find that this variable has the strongest individual relationship with the neighborhood pattern of foreclosures.

Models 2 and 3 add two measures of racial/ethnic characteristics.[22] OPCT_NW00 equals the share of minority homeowners in the neighborhood in 2000. Specifically, it equals

[(minority homeowners in 2000)/(all homeowners in 2000)],

where "minority" is defined as any household head not coded as a non-Hispanic white person only. For example, minority in 2000 includes multirace individuals. OPCT_NW90 is the analogous variable for 1990, and CHG_NWOWNPCT equals OPCT_NW00 - OPCT_NW90. Thus, OPCT_NW00 is a static variable, measuring the share of minority homeowners at a point in time (2000), whereas CHG_NWOWNPCT is a dynamic, or neighborhood transition, variable, measuring the change in minority homeownership share during the 1990s. Both variables reflect aspects of the neighborhoods' racial and ethnic characteristics, and yet they are not redundant. Their bivariate correlation is not strong, 0.39, and their relationship to the neighborhood foreclosure pattern is also different, as shown in the Model 3 results for lfr_i, where the dynamic variable is highly significant while the static variable is insignificant. We explore this fact further below, where we note that the static measure of minority share is highly correlated with VHCREDIT2. We regard VHCREDIT2 as a relatively precise indicator of share of the high-credit-risk population in the neighborhood. As long as it is present, static measures of minority presence in the neighborhood add little to the fit of our models.[23]

By contrast, CHG_NWOWNPCT is very important to the fit of our model, as indicated by a comparison of Models 1 and 2. With VHCREDIT2 already present, no other variable we tried has so large an impact on the standard error of estimate, the log likelihood, or the R-bar-squared as this measure of increased minority homeownership. Adding further variables, as in Models 4 and 5, improves the fit of the model, but not to the same extent as adding CHG_NWOWNPCT to Model 1. In that sense, Model 2 provides a relatively good, two-dimensional account of the neighborhood pattern of foreclosures in Hennepin and Ramsey counties in 2002. The two key dimensions appear to be credit risk and change in

[22] We have also computed similar regressions with variables analogous to OPCT_NW00 and CHG_NWOWNPCT but specific to African-Americans, Asian-Hawaiian-Pacific Islanders, and Hispanics. In all cases examined, F-tests come nowhere near rejecting the hypothesis that the coefficients of the race and ethnicity variables are identical. (The similarity of Asian and other minorities here is likely linked to the large share of relatively low-income Southeast Asian former-refugee families in the Twin Cities in 2000 and might not hold elsewhere).

the racial and ethnic homeownership shares. We defer our discussion of how to interpret these two dimensions.

We could perhaps conclude our multivariate analysis with Model 2, as it is simple and fits nearly as well as any model we found. In fact, we did extensive robustness checking of Model 2 by testing whether additional variables could greatly improve the fit or overturn the significance of its two key variables. Because many variables in our dataset are highly correlated, adding more variables to the model tends to reduce the significance of any individual variable, including our two key variables. However, both key variables remained significant in almost all specifications. We present Models 4 and 5 as examples of the best-fitting and most informative alternative models.

In keeping with the options theory of foreclosures, Model 4 adds a variable that measures the presence of high loan-to-value mortgages in the neighborhood. FHA99PMORT is the number of FHA home purchase mortgages originated in 1999 divided by the number of mortgaged units in 2000. Since FHA mortgages are disproportionately used by borrowers who have difficulty qualifying for conventional credit and often feature low down payments, they are relatively at risk for foreclosure. This variable provides more neighborhood detail on the extent of risky mortgage lending and possible ruthless foreclosures, beyond the basic credit risk information in VHCREDIT2. It slightly improves the model's fit, as shown in Table 6. The new variable also partly substitutes for VHCREDIT2, as shown by the large drop in the coefficient on VHCREDIT2 between Model 3 and 4. This is not surprising since FHA99PMORT has a bivariate correlation of 0.59 with VHCREDIT2. It reinforces and adds extra detail to the model's credit risk dimension, without adding a truly new dimension to the model.[24]

According to Model 5, three other credit risk indicators are also significantly related to foreclosure sale rates but mainly substitute for already included variables in their effects on model fit. Young homeowners, measured in OWN45P as the percent of homeowners under 45, tend to be at higher risk of foreclosure for a number of reasons—low income, low personal wealth, and higher risk of

[23] In results not shown, we have verified this for other static measures of race and ethnicity, including population share.
[24] We also tried a credit risk variable based on the prevalence of subprime refinancing, but it was not significant.

unemployment, and relatively new mortgages with little equity to protect or draw on. VACANT00 reflects the percent of area housing units listed as vacant in Census 2000. LOAN2INC is the ratio average loan amount to median borrower income for the tract, both from HMDA data for 2000. VACANT00 is a broad indicator of trigger event prevalence, and LOAN2INC reflects financial vulnerability.

Adding OWN45P helps clarify that the effects of the key minority transition variable, CHG_NWOWNPCT, are not merely due to age effects. The minority population in the Twin Cities is relatively young by national standards because of substantial domestic and foreign in-migration of young adults. It is conceivable, therefore, that our minority transition variable mostly reflects age, not race or ethnicity. The addition of OWN45P suggests that this is not the case. OWN45P does take away some of the explanatory power of CHG_NWOWNPCT, as seen in its lowered coefficient in Model 5. However, CHG_NWOWNPCT remains highly significant, even controlling for age effects. In the absence of OWN45P, the importance of CHG_NWOWNPCT may be slightly exaggerated due to its ability to capture some age effects. However, CHG_NWOWNPCT's primary role in the model is as a measure of racial and ethnic transition into homeownership. In that role, it has few substitutes. When we drop variables one at a time from Model 5 (or models with additional variables added to explain lfr_i), the dropping of CHG_NWOWNPCT causes the largest decline in model fit. In results not shown, we find that OWN45P and VACANT00 are mainly responsible for the reduction in the coefficient on VHCREDIT2 between Model 4 and Model 5, while LOAN2INC mainly reduces the coefficient on FHA99PMORT.

Overall, our regressions suggest a strong role for credit risk, which can be measured by several only slightly independent variables, as well as a possibly independent role for minority homeownership transition. Applying factor analysis to an expanded set of variables yields somewhat similar results. One common factor is dominant (eigenvalue 10.8) and has strong positive weights on credit risk variables like VHCREDIT2, strong negative weights on financial stability variables like education level, and a moderately high weight on CHG_NWOWNPCT. Two additional factors may be marginally important (eigenvalues 1.13 and 0.72); these factors place relatively high weights on the new variables in Models 4 and 5 and a measure of housing appreciation 1990-2000.[25] However, those variables and CHG_NWOWNPCT are the variables whose variances are the least well explained by the three most important factors.[26]

D. The Key Role of Accurate Credit Risk Data

Although information on credit histories and credit scores has been widely recognized as relevant to foreclosure risks, data on the neighborhood distribution of credit scores often has not been available to public officials, foreclosure counselors, or foreclosure researchers. As a result, researchers and others have constructed proxies for credit risk from more widely available data. One proxy that has been used recently is the rate of denial on prime home-purchase mortgage applications.[27] This rate can be computed annually at the census tract level from data publicly available under HMDA.

We have computed a version of the HMDA denial proxy for our neighborhoods. However, when we substitute it for VHCREDIT2, a measure of risk based on actual credit scores, our findings change significantly. In particular, static race/ethnicity variables—variables representing the level of minority presence in the neighborhood—become significantly associated with the neighborhood pattern of foreclosures, even in the presence of our other control variables (including change in minority homeownership). This is true whether minority presence is measured as share of homeowners (the most clearly relevant group for foreclosure risks) or as share of population.

[25] The housing appreciation variable was not significant in our regressions, however. Other variables suggested by foreclosure theory that we found to be insignificant when added to our models include the percentage of individuals who lack a credit score, the percentages of household heads with a high school or college education, the tract unemployment rate, the percentage of renters, the prevalence of subprime refinancing, and mean tract income.

[26] Principal components gave broadly similar results.

[27] Regarding the quality of credit risk measures, in our data the use of denial rates on prime mortgage applications as a proxy for credit score information would suggest a strong direct relationship between minority shares and foreclosure sale rates, even though this relationship disappears (for lfr_i) when credit score information is included. Denial rate proxies have been used as credit risk proxies in recent foreclosure studies when credit score information was not available. Our result raises the possibility that other studies might have found a direct association between foreclosure sale rates and minority population (or homeownership) shares when only a more indirect relationship might have appeared if actual credit score measures had been available. See Apgar, Calder, and Fauth (2004); Apgar and Duda (2005a); Apgar and Duda (2005b); and Galster, Hayes, and Johnson (2005). Additional research, preferably involving other cities, would be necessary to evaluate this possibility.

Further analysis suggests the change may result because the HMDA denial proxy underestimates credit-score-based measures of credit risk in minority neighborhoods. The simple correlation between the HMDA proxy and VHCREDIT2, our credit-score-based measure of credit risk, is quite strong, 0.84, and regressing VHCREDIT2 on the HMDA proxy explains 70 percent of the variance of VHCREDIT2. This clearly suggests that the HMDA proxy can be useful in identifying neighborhoods with high foreclosure risks when actual credit score data are not available. However, the Year 2000 minority shares of population and of homeowners are also strongly correlated with credit risk in our sample, with correlation coefficients of 0.91 and 0.83, respectively. When either of these variables is added to the regression of VHCREDIT2 on the HMDA proxy, the minority presence variables have positive, highly significant coefficients, and the amount of the VHCREDIT2 variance explained rises noticeably (to 76 percent with minority share of homeownership and 83 percent with minority share of population). Our findings suggest that for research or analytical purposes, the HMDA denial rate proxy may be an imperfect substitute for accurate credit score data. We find that its use may exaggerate the direct linkages between neighborhood racial/ethnic composition and foreclosure risks.

E. Implications for program design and causality

We do not expect that our data can pin down a single causal explanation for the pattern of foreclosure sale rates observed in Hennepin and Ramsey counties in 2002. However, we think our multivariate results can help guide some of the responses to foreclosures and may challenge some of the existing causal models.

Our finding that rising minority homeownership seemed to have a strong association with foreclosure sale rates, independent of credit scores, and other factors may have implications for programs that seek to boost minority homeownership rates. Whatever the cause of this association, homeownership preservation or foreclosure mitigation programs in areas with rising minority homeownership rates may find it useful to conduct targeted outreach to new minority homeowners. This would ensure that they are aware of the program's services and will feel comfortable accessing them if they are needed. This conclusion is reinforced by other evidence that minority households have a significantly higher rate of exit from homeownership than white households.[28]

Our data, although extensive, are basically a single cross-section. Furthermore, although we also collected a limited sample of data on nonforeclosed mortgages in a portion of Minneapolis (detailed above), we have little information about the broader population of mortgages from which the 2002 foreclosures emerged. For these reasons, the possibilities for using our data to identify the fundamental underlying causes of foreclosures are limited. Our results may support the causal model of Baxter and Lauria (2000) and Lauria (1998) that foreclosures cause neighborhood racial transition. Lauria (1998) provides the following summary of the causal linkage: “Middle-income professional whites employed in businesses impacted by recession who had recently bought housing with high loan-to-value ratios were forced to sell or have their houses foreclosed upon. The depressed market, in turn, made such housing affordable to middle-class blacks interested in homeownership.”[29]

Although the Baxter and Lauria model has been cited as potentially relevant to foreclosure dynamics in other economically distressed cities,[30] we are not ready to judge it the best interpretation of our findings. Some of our facts match their interpretation, such as the association between racial transition and foreclosure sales, and we also measure foreclosures in the wake of a recession and during a period of elevated unemployment. However, unlike in Baxter and Lauria’s study, our 1990–2000 measure of racial transition predates the year of our foreclosures, 2002. This raises at least some doubts about whether foreclosures were the cause of the increase in minority homeownership that was common in our high foreclosure neighborhoods, although we would need data on foreclosures in earlier years to more fully address this point.

[28] Grover and Todd (2005); Haurin and Rosenthal (2004); and Reid (2004).
[29] Lauria (1998, p. 9).
[30] For example, see Garcia (2003).

As a competing hypothesis, we suggest a simpler possibility.[31] Expanded credit access and a strong economy allowed many economically vulnerable minority households to become homeowners in the Twin Cities in the late 1990s. When the economy turned down in 2000, many of these new minority homeowners experienced income declines and had too little home equity or other household wealth to maintain their mortgage payments. This timeline is also bolstered by the median age of foreclosed mortgages (2.6 years) in our analysis. With Ameriquest and its affiliates being by far the most common lenders among our foreclosures and in light of their recent agreement to settle (without admission of wrongdoing) claims of abusive lending within our study area, we also note the potential role of subprime and possibly abusive refinancing aimed at inexperienced minority homebuyers experiencing financial distress. Under either our simple or augmented hypothesis, increased minority homeownership is a signal of vulnerability to foreclosure, due in part to the low wealth-to-income ratios that are still common among minority households. Furthermore, it is a signal to enhance homeownership preservation and foreclosure mitigation programs in neighborhoods with increasing minority homeownership.

VII. Recommendations for Better Data on Foreclosures and Related Factors

We gained access to sufficient data for the purpose of analyzing the spatial distribution of foreclosures in Hennepin and Ramsey counties. Through the complications we experienced along the way, which are detailed in the first section of the paper, we also gained insights into how foreclosure data can be made more accessible and useful. Based on those insights, we have developed the following recommendations. While we suggest serious consideration of these recommendations and believe their implementation could make local foreclosure prevention efforts more efficient and effective, we recognize that no analysis of the public- and private-sector cost of implementing these proposals has been completed. Furthermore, some of these proposals may require changes to policy and/or practice at several political or organizational levels, depending how local, county and state laws work together.

[31] A version of this hypothesis also seems to be among the several hypotheses that Lauria and Baxter (2000) discuss, but it is not

Consider making more data available in an electronic format. Although private firms collect foreclosure information to sell to investors and others with a business interest in foreclosed properties, most government agencies and nonprofit foreclosure mitigation practitioners have no convenient access to these data, and there has been little effort to make foreclosure records readily available to the public. During the early part of our research, we spent considerable time entering information from paper foreclosure-sale documents into an electronic spreadsheet that was linked to geographic information system and statistical software programs. In order to speed up this process, we recommend an assessment of whether to make foreclosure filing and sale information available in an electronic format on a periodic basis, preferably on an easily accessible Web page. The results in this paper suggest that there would be benefits. These need to be quantified and compared to the costs.

Consider updating or adding key information to foreclosure notices and sales documents to make them complete. Addresses are not the only useful pieces of information in the foreclosure filings and sales documents. The growing number of transactions contracted through mortgage intermediaries, such as MERS, increases the need for accurate information about the lenders involved. We recommend that consideration be given to having foreclosure filing and sale records provide more comprehensive, up-to-date data to the public, including information on the originating lender or the type of loan. Other data, such as information on interest rates and riders, would also be valuable to foreclosure mitigation efforts. Much of this information already exists at the county property records departments and could be linked to foreclosure documents using the legal description or tax identification number of the property.

Consider maintaining a database repository of foreclosure notices and sales that contains historical and current records. At present, only sporadic research and data collection efforts on foreclosure patterns take place in most metropolitan areas. Therefore, we recommend the creation of a foreclosure database that would collect and maintain the information, both historic and current, that is included on the public filing and sale documents, augmented by additional information from the mortgage

---

the one that seems to make their model appealing to other foreclosure analysts.

documents found at county property records departments. Access to the database would be granted to public and nonprofit groups, especially those groups engaged in foreclosure prevention activities. One possible location for this database could be through the counties themselves, since they already maintain property-record files and, by law, collect foreclosure information.

Consider making neighborhood credit-score-based measures available to foreclosure mitigation efforts. Lastly, our final proposal relates to the key finding in our multivariate analysis, namely that the credit-score-based measure of credit risk was the factor most strongly associated with the geographic concentrations of foreclosures. As noted previously, the neighborhood distribution of credit scores is not readily available or free of cost to public officials, foreclosure counselors, or foreclosure researchers. Updated credit score data for the population are available, but only at a cost through vendors. Because this information has an important public use, we, therefore, recommend that the provision and public availability of these data at the census tract level on an annual basis be investigated.

VIII. Concluding Remarks

As public policy continues to promote homeownership, especially as it relates to closing racial and ethnic homeownership gaps, foreclosure mitigation efforts will be crucial to preserving the gains made through these efforts. In order to monitor foreclosure trends and deliver effective services, these efforts need access to accurate, timely, and inexpensive data on foreclosures and mortgages as well as to variables that can help them predict where foreclosures will concentrate and understand what factors are most closely related to foreclosures.

REFERENCES

Ambrose, Brent W., and Charles A. Capone. (1998). “Modeling the Conditional Probability of Foreclosure in the Context of Single-Family Mortgage Default Resolutions.” *Real Estate Economics* (3): 391-429.

Ameriquest Press Release, “Ameriquest Announces Agreement with States,” January 23, 2006.

Apgar, William C., Allegra Calder, and Gary Fauth. (2004). *Credit, Capital and Communities: The Implication of the Changing Mortgage Banking Industry for Community Based Organizations.* Harvard University, Joint Center for Housing Studies Working Paper CCC04-1 (March).

Apgar, William C., and Mark Duda (2004), *Neighborhood Level Mortgage Foreclosures in Los Angeles: An Assessment of Patterns and Examination of Policy Issues.* Los Angeles Neighborhood Housing Services (May), Unpublished Draft Report.

Apgar, William C., and Mark Duda. (2005a). *Mortgage Foreclosures in Atlanta: Patterns and Policy Issues,* NeighborWorks (December).

Apgar, William C., and Mark Duda. (2005b). *Collateral Damage: The Municipal Impact of Today’s Mortgage Foreclosure Boom.* Report prepared for the Homeownership Preservation Foundation (May 11).

Apgar, William C., Mark Duda, and Rochelle Nawrocki Gorey. (2005). *The Municipal Cost of Foreclosures: A Chicago Case Study*. Homeownership Preservation Foundation Housing Finance Policy Research Paper Number 2005-1 (February).

Baker, Christi. (2004). *Report 3: Essential Components of Post-Purchase Program Models.* Fannie Mae Foundation (July). (Based on work by Lucy S. Gorham, Roberto Quercia and William Rohe of the Center for Urban and Regional Studies at the University of North Carolina at Chapel Hill.)

Collins, J. Michael, and Rochelle Nawrocki Gorey. (2004). *Analyzing Elements of Leading Default-Intervention Programs.* White Paper, Policy Lab Consulting Group.

Cutts, Amy Crews, and Richard K. Green. (2004). *Innovative Servicing Technology: Smart Enough to Keep People in Their Houses?* Harvard University, Joint Center for Housing Studies Working Paper BABC 04-19 (February).

Galster, George, Chris Hayes, and Jennifer Johnson. (2005). “Identifying Robust, Parsimonious Neighborhood Indicators.” *Journal of Planning Education and Research* (24), 265–280.

Garcia, Ramon. (2003). *Residential Foreclosures in the City of Buffalo, 1990–2000.* Buffalo Branch of the Federal Reserve Bank of New York (June).

Greene, William H. (2003). *Econometric Analysis* (Fifth Edition). Upper Saddle River, New Jersey: Prentice Hall.

Grover, Michael, and Richard M. Todd. (2005). *A Case for Post-Purchase Support Programs as Part of Minnesota’s Emerging Markets Homeownership Initiative.* Federal Reserve Bank of Minneapolis Community Affairs Report 2005-1 (October).

Gruenstein, Debbie, and Christopher Herbert. (2000). *Analyzing Trends in Subprime Originations and Foreclosures: A Case Study of the Atlanta Metro Area.* ABT Associates (February).

Haurin, Donald, and Stuart Rosenthal. (2004). *The Sustainability of Homeownership: Factors Affecting the Duration of Homeownership and Rental Spells.* U.S. Department of Housing and Urban Development, Office of Policy Development and Research (December).

Hirad, Abdighani, and Peter M. Zorn. (2001). *A Little Knowledge Is a Good Thing: Empirical Evidence of the Effectiveness of Pre-Purchase Homeownership Counseling.* Harvard University, Joint Center for Housing Studies Low-Income Homeownership Working Paper 01.4 (August).

Home Ownership Center. (2004). *Minneapolis and St. Paul Mortgage Foreclosure Prevention Program Annual Report Ending December 31, 2003.* Unpublished.

Immergluck, Dan, and Geoff Smith. (2004). *Risky Business: An Econometric Analysis of the Relationship between Subprime Lending and Neighborhood Foreclosures.* Woodstock Institute (March).

Immergluck, Dan, and Geoff Smith. (2005). *There Goes the Neighborhood: The Effect of single-Family Mortgage Foreclosures on Property Values.* Woodstock Institute (June).

Lauria, Mickey. (1998). "A New Model of Neighborhood Change: Reconsidering the Role of White Flight," *Housing Policy Debate* (volume 9, issue 2), 395-424.

Lauria, Mickey, and Vern Baxter. (2000). "Residential Mortgage Foreclosure and Racial Transition in New Orleans," *Housing Policy Debate* (volume 11, issue 3), 675-699.

Los Angeles Times, "Workers Say Lender Ran 'Boiler Rooms,'" February 4, 2005.

Neighborhood Housing Services of Chicago. (2005). *The Chicago Home Ownership Preservation Initiative: A Learning Laboratory.* Unpublished report (June 15) available at *www.nhschicago.org/downloads/1219hopi_report_06_05_2005.pdf*.

Newburger, Harriet. (2006). *Foreclosure Filings and Sheriff's Sales Experienced by Low-Income, First-Time Homebuyers.* Unpublished Manuscript (April).

Quercia, Roberto G., Spencer M. Cowan, and Ana Moreno. (2004). *The Cost-Effectiveness of Community-Based Foreclosure Prevention.* Harvard University, Joint Center for Housing Studies Working Paper BABC 04-18 (February).

Reid, Carolina Katz. (2004). *Achieving the American Dream? A Longitudinal Analysis of the Homeownership Experiences of Low-Income Households.* University of Washington, Center for Studies in Demography and Ecology, Working Paper 04-04 (April).

The Reinvestment Fund. (2005). *Mortgage Foreclosure filings in Pennsylvania: A Study by the Reinvestment Fund for the Pennsylvania Department of Banking, 2005.* Unpublished report available at *www.trfund.com/resource/downloads/policypubs/TRF-Housing-Opp-Heinz.pdf*.

Todd, Richard M. (2005). "Homeownership gap calls for broad, balanced response." Federal Reserve Bank of Minneapolis, *Community Dividend*, Issue no. 2.

Toledano, Kathleen, and Michael V. Berry. (2004). "Neighborhood Housing Services of Chicago Home Ownership Preservation Initiative." Federal Reserve Bank of Chicago, *Profitwise: News and Views, special edition* (May).

This page is intentionally left blank.

Table 1: Foreclosure Sales Characteristics by Political Subdivisions				
County	Number	Percent	Rate Per 1,000 Mortgaged Units	Mortgaged Units
Hennepin	831	70.5%	3.95	210,533
Ramsey	347	29.5%	4.12	84,202
Total	1,178	100.0%	4.00	294,735
Central cities				
Minneapolis	414	35.1%	7.61	54,379
St. Paul	298	25.3%	7.36	40,494
All other cities	466	39.6%	2.33	199,862
Total	1,178	100.0%	4.00	294,735

Table 2: Neighborhood Foreclosure and Sold Property Comparison		
Characteristic	Foreclosed Properties (N=75)	Non-foreclosed Purchased Properties (N=198)
Mean interest rate*	10.1 %	7.9%
Mean mortgage principle*	$71,675	$95,492
Percent nonbank lender**	77.8 %	62.2%
Percent subprime lender*	47%	11%
Percent holding other mortgage*	35%	8%

* Significant at 0.001 level.
** Significant at 0.01 level.

Table 3: Correlations Between 2002 Neighborhood Foreclosure Rates and Other Neighborhood Characteristics (Hennepin and Ramsey Counties)	
Neighborhood Characteristic	**Correlation**
Credit risk variables	
1999 share of households with very risky (very low) credit scores	0.76
1996–99 rate of denial on prime home purchase mortgage applications (HMDAs)	0.69
1999 subprime refinancing originations per Census 2000 mortgaged unit	0.69
1999 FHA originations per Census 2000 mortgaged units	0.46
Race and ethnicity variables	
2000 minority* share of population	0.77
2000 minority share of homeowners	0.73
1990–2000 change in minority share of population	0.50
1990–2000 change in minority share of homeowners	0.67
Other	
2000 percent of population over 25 with a college degree	-0.56
2000 percent of population over 18 with a high school diploma or equivalent	-0.73
Logarithm of 1999 average household income	-0.54
2000 rate of unemployment	0.64
2000 percent of home-owning households with head younger than 45 years old	0.30
* For Minneapolis-St. Paul, minority means other than non-Hispanic white alone	

Table 4: Comparing 7 Variables' Ability to Identify Tracts with Highest 2002 Foreclosure Sale Rates*						
	Highest 20		Highest 50		Highest Quintile (65)	
	# Correct	% Correct	# Correct	% Correct	# Correct	% Correct
Risky credit scores in 1999	9	45	36	72	50	77
Percent poor in 1990	9	55	34	70	44 (0.071)	68
Percent non-white in 1990	9	45	34	68	43 (0.034)	66
Percent non-white owners in 1990	8	40	29 (0.039)	58	36 (0.000)	55
Low income in 1990	9	45	35	70	43 (0.001)	66
High subprime refi in 1999	9	45	35	70	43 (0.033)	66
High prime denial rate 1996-99	8	40	32	64	40 (0.004)	62

*Figures in parentheses are p values for the null that the variable is equivalent to the credit score variable in its ability to correctly identify top 50 or top quintile tracts, based on 1,000 random rank swaps. Values not shown are greater than 0.10.

Table 5: Spearman Rank Correlation Coefficients with Foreclosure Sale Rate			
	Coefficient	Difference from Credit Score Coefficient	Occurrence of Equal or Larger Differences in 1,000 Random Ranks Swaps
Risky credit scores in 1999	0.640	NA	NA
Percent poor in 1990	0.511	0.129	0
Percent non-white in 1990	0.480	0.160	0
Percent non-white owners in 1990	0.387	0.253	0
Low income in 1990	0.564	0.076	2
High subprime refi in 1999	0.559	0.081	14
High prime denial rate 1996-99	0.571	0.069	2

Table 6: Multivariate Regression Results					
	Model 1	Model 2	Model 3	Model 4	Model 5
Panel A: Dependent Variable lfr_i; Ordinary Least Squares Estimation					
CONSTANT	-7.78***	-7.62***	-7.57***	-7.55***	-8.84***
VHCREDIT2	11.41***	8.87***	8.40***	6.86***	5.98***
CHG NWOWNPCT		4.52***	4.44***	3.95***	4.02***
OPCT NW00			0.34	0.86	0.95
FHA99PMORT				7.61	6.58**
OWN45P					0.42
VACANT00					19.92*
LOAN2INC					0.49**
Std. error of est.	1.176	1.152	1.153	1.151	1.152
Log likelihood	-516.09	-508.71	-508.63	-507.53	-507.17
R-bar squared	0.347	0.373	0.372	0.374	0.373
Panel B: Dependent Variable lfr_i; Minimum Chi-Square Estimation					
CONSTANT	-7.51***	-7.32***	-7.36***	-7.38***	-8.94***
VHCREDIT2	10.79***	8.28***	8.68***	6.88***	5.79***
CHG NWOWNPCT		3.58***	3.60***	3.21***	2.98***
OPCT NW00			-0.22	0.42	0.64
FHA99PMORT				9.81***	8.24**
OWN45P					1.38***
VACANT00					12.93***
LOAN2INC					0.46***
*, **, or ***: Significant at 0.10, 0.05, or 0.01 level, respectively.					

This page is intentionally left blank.