

# **Neighborhood Effects in Mortgage Default Risk**

Prepared for:  
U.S. Department of Housing and Urban Development  
Office of Policy Development and Research

Prepared by:  
Robert F. Cotterman  
Unicon Research Corporation  
Santa Monica, CA

Under Contract: C-OPC-18484

March 2001

## **ACKNOWLEDGMENTS**

The author wishes to thank the staff of HUD Policy Development and Research, who have provided invaluable advice and assistance throughout the course of this project. Special thanks are due to Harold Bunce, Bill Reeder, and Sue Neal for extensive comments on an earlier draft and to Randy Scheessele for help in obtaining and understanding the FHA data.

The contents of this report are the view of the contractor and do not necessarily reflect the views or policies of the U.S. Department of Housing and Urban Development or the U.S. Government.

## Preface

This study examines the effect of neighborhood characteristics on the default of FHA mortgages. The analysis includes both neighborhood characteristics and characteristics of the individual loan and borrower, so that the effects of the neighborhood can be distinguished from those of the individual loan. In particular, the analysis seeks to distinguish the effects of neighborhood race, ethnicity, and income from the effects of the individual borrower's status. Research on the effects of neighborhood characteristics on default has been somewhat limited in the past, and this study's contribution to the literature is the inclusion of credit history data. The analysis finds that lower tract income and higher tract black composition are associated with higher rates of default, whereas individual borrower race or income are unrelated to default.

The study then goes on to examine possible causes for these findings, including whether higher defaults reflect more limited access to mortgage finance (as measured by refinance probability) or a response to previous defaults in the neighborhood. The findings regarding access to refinancing are not definitive. FHA refinancing probabilities seem to be higher in minority tracts and are at least equal to low-income tracts. Refinancing through other non-FHA sources of refinance funds, including conventional, is statistically less in predominantly Hispanic and lower-income tracts for holders of FHA mortgages. The effects of neighborhood race and income on default are reduced when lagged defaults and prepayments and neighborhood house price change are included in the analysis. Although the higher default rate in lower-income tracts remains significant, the higher default rate in minority tracts becomes insignificant, suggesting that lower house price appreciation is an important factor in the higher default rates observed in minority neighborhoods.

## TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
SUMMARY .....	iii
ACKNOWLEDGMENTS .....	ix
1 INTRODUCTION .....	1
1.1. Motivation and Background .....	1
1.2. Specific Aims .....	6
1.3. A Roadmap for the Remainder of the Paper .....	8
2 ESTIMATION OF NEIGHBORHOOD EFFECTS IN LOAN-LEVEL DATA ....	10
2.1. Data Sources and Key Definitions .....	10
2.2. A Preliminary Look at the Data .....	13
2.3. Statistical Models and Methods .....	16
2.3.1. A Default Model .....	16
2.3.1.1. Possible Advantages and Disadvantages of the Empirical Approach .....	17
2.3.1.2. A Motivation for the Structure of the Hazard Model .....	20
2.3.2. A Prepayment Model .....	23
2.4. Variable Definitions and Concepts .....	24
2.5. Statistical Estimates of Default and Prepayment Hazard Models .....	28
2.5.1. Estimates of Basic Hazard Models of Default and Prepayment .....	28
2.5.1.1. Default .....	28
2.5.1.2. Prepayment Other Than FHA Refinancing .....	30
2.5.1.3. FHA Refinancing .....	32
2.5.2. Adding Neighborhood Characteristics to Default and Prepayment Models .....	33
2.5.3. Adding Individual Characteristics to Default and Prepayment Models .....	35
2.6. Interpretation and Conclusions .....	38
3 ALTERNATIVE EXPLANATIONS FOR DIFFERENCES IN DEFAULT PROBABILITIES ACROSS TRACTS .....	42
3.1. Estimation of Tract-Level Models for House Prices in the Chicago MSA ....	42
3.1.1. Aggregate Models and Data .....	45
3.1.2. Estimation Results .....	49
3.2. Revisiting Hazard Models of Default for the Chicago MSA .....	53
4 CONCLUSIONS .....	58
REFERENCES .....	62
APPENDIX	
List of Tables .....	A-1

## SUMMARY

The primary purpose of this paper is to investigate the effect of neighborhood characteristics, especially mean income and racial or ethnic composition, on the default of individual FHA-insured loans. En route we also present estimated default effects of individual income and race or ethnicity. We attempt not only to estimate neighborhood default effects, but also to understand what might lie behind these effects. We first ask whether neighborhood default effects might be traceable to differential access to funds for refinancing mortgages. To this end, we examine effects of race, ethnicity, and income at both the individual and tract level on two kinds of individual prepayment activity: (a) prepayment for the purpose of obtaining FHA refinancing, and (b) prepayment for all other reasons, one of which may be to obtain conventional refinancing. Attention then turns to two other explanations that might underlie the effect of neighborhood characteristics on individual default behavior: neighborhood default activity that might lead to abandoned structures, or undesirable instability associated with turnover of homeowners in the neighborhood. The methods used and conclusions drawn from the study may be conveniently organized around four research questions that are used to guide the analysis.

**1. Once one controls for a variety of borrower- and loan-related factors, including time-varying characteristics, in an appropriate econometric model of default, do neighborhood effects seemingly related to income or race persist?**

We answer this question in the first portion of the study via a statistical analysis of individual default behavior in which both individual and tract (neighborhood) effects are permitted. The analysis is based on samples of FHA-insured loans<sup>1</sup> that were endorsed in 1992 or 1994, or for which applications were submitted in one of these years, and for which the subject property was contained within 22 selected MSAs. In contrast with the data used in many previous studies, the

---

<sup>1</sup> Streamline refinances are excluded because the applications for such loans lack critical information, such as loan-to-value ratios.

data files used here contain information on individual race, ethnicity, income, and a summary credit measure, the FICO score. Moreover, this study utilizes a hazard model that offers two principal advantages over the probit and logit models that have more typically been used in related studies: first, the ability of the hazard approach to take proper account of time-varying characteristics and, second, the ability of the hazard approach to accommodate the censoring that occurs when the observation window closes while a loan remains active.

We find that decreases in tract income and, less clearly, increases in representation of blacks within the tract, are associated with higher default probabilities of individual loans, and this relationship holds even when one controls for the race, ethnicity, and credit history of the borrower in an appropriate hazard model of default behavior. We also observe that although neighborhood impacts remain, their importance is dramatically affected by introducing critical controls for characteristics of individual loans, borrowers, and the economic environment.

**2. Do neighborhood characteristics, such as race and income, have effects on default that are separate and distinct from the effects of these same characteristics at the individual level?**

Estimates of the default hazard shows that while greater tract representation of blacks is probably associated with higher individual default probabilities, individual race effects --- black or Hispanic --- are absent. Hispanic representation at the tract level does not seem to matter in default behavior. Although tract income does seem to affect default behavior, individual income has no statistically significant impact.

**3. Is there evidence that differences in default probabilities reflect differences in the probability of refinancing?**

To assess this possibility, the study combines the evidence from the default hazard with a parallel set of estimated prepayment hazards utilizing the same set of individual FHA-insured mortgages. If differential access to funds for refinancing were the culprit in generating higher default rates in more heavily black neighborhoods, we might expect to find that either FHA refinancing probabilities or other prepayment probabilities would be lower for blacks. Instead the estimated

prepayment hazard models reveal that FHA refinancing probabilities, as well as other prepayment probabilities (which include conventional refinancing activities) tend to be higher, if anything, among loans in tracts with heavier minority (black or Hispanic) representation. Among these neighborhood effects, only the effect of tract-level Hispanic representation for other prepayment is statistically significant at conventional levels.

We do find that individual Hispanic ethnicity is associated with lower FHA refinancing probabilities, and individual blacks and Hispanics have lower probabilities of other types of prepayment, which may or may not be the result of discrimination in refinancing against individual borrowers on the basis of race or ethnicity. Yet there are no corresponding default effects traceable to the race of individuals.

Tract income is positively related to the probability of prepayment (other than for FHA refinancing) but has no significant effect on the FHA refinancing probability; as noted tract income is negatively related to default probabilities of individual loans.<sup>2</sup> The sign pattern is consistent with, but surely does not prove, that higher tract income supports lower default probabilities partially through greater access to conventional refinancing. Clearly, other explanations for this sign pattern are also possible.

We emphasize that this assessment of the role of differential access to funds is, by design, very narrow. It is restricted to effects arising among holders of FHA mortgages; conventional mortgage holders are excluded. The evidence is also entirely indirect. We do not account in any way for possible differences in the rate at which groups actually apply for refinancing from either FHA or conventional sources.

---

<sup>2</sup> It is less easy to assess the role of individual income within the context of default and refinancing, mainly because the estimation also uses income indirectly as part of the front-end ratio. Evaluating income effects under the assumption that the front-end ratio is to be held fixed, we find that individual income has no significant effect on default but has positive impacts on probabilities of FHA refinancing and other prepayment.

#### **4. Do neighborhood differences in default probabilities arise because of earlier defaults and, if so, what is the mechanism?**

Next we entertain the possibility that neighborhood race (or income) effects on individual default behavior are traceable to (a) past defaults within these same neighborhoods, which may lead to abandoned structures and declining neighborhood amenities, or (b) past turnover of homeownership within these same neighborhoods, which may be associated with undesirable neighborhood instability. Under either scenario, lagged neighborhood default activity and (under the second alternative) neighborhood prepayment activity, may directly affect individual default behavior, or effects may occur indirectly via changes in neighborhood house prices. To investigate these ways in which neighborhood default effects might arise, we turn to additional statistical analysis that is restricted to the Chicago MSA.

We first utilize quarterly data from tract aggregates --- dubbed “supertracts” --- within the Chicago MSA from 1994 to mid-1999 to explore the aggregate relationship between house prices, on the one hand, and previous defaults and prepayments, on the other. Regression analysis with these aggregate data generally appears to show that lagged defaults lead to lower neighborhood house price growth. It is unclear whether the reason is that defaults lead to abandoned structures and neighborhood deterioration, whether defaults are one component of undesirable instability in homeownership, or whether some other default-related impact is at work. The regression estimates of the effects of current and lagged prepayment rates tend to be more irregular, sometimes suggesting positive effects on house price growth --- findings which are inconsistent with the turnover-leads-to-default theory --- but sometimes suggesting negative impacts.

These results should be interpreted with caution for several reasons. First, the statistical results are fragile, with estimated impacts varying with whether fixed effects are permitted, with the number of origination years over which the default and prepayment rates are calculated, with the number of lagged values of house prices included as explanatory variables, and with the number of lags of default and prepayment rates that are included in the regression. In addition, other specifications sometimes show that one quarter leads in default rates and prepayment rates



are sometimes statistically significant, suggesting that such rates may proxy the effects of other influences on house prices.

To see whether lagged default and prepayment rates in the neighborhood might exert a direct effect on individual default probabilities, in addition to any indirect effects that occur via reductions in house prices, we reestimate default hazards for individual loans in the Chicago MSA. The estimation sample is composed of all 1994 Chicago MSA applications or originations for homes located in one of the supertracts that contain the requisite data. The new hazard specification augments the original specification by including a direct measure of supertract house price growth (measured in various ways), as well as lagged values of the supertract default and prepayment rates.

The results are mixed. The impact of supertract-level house price growth is often not significantly different from zero at conventional levels, though its effect is always of the anticipated sign. Lagged neighborhood defaults generally seem to have a positive impact on the probability of default of an individual loan, but lagged neighborhood prepayment rates are generally of mixed signs, casting substantial doubt on the role of prepayment activity (at least as measured here) as having a direct effect on default behavior.

It is noteworthy that introducing direct measures of neighborhood house price growth and lagged default and prepayment rates changes other estimated default effects. In particular, the estimated default impacts of supertract income and of supertract black representation are substantially reduced. While the effect of supertract income remains statistically significant, the neighborhood race effect is now far from statistical significance. Again caution is urged in interpreting these findings, in part because there is some evidence that lagged neighborhood default rates proxy other omitted variables that affect current individual default probabilities.

To conclude, the findings from the reestimated hazard models reinforce the idea that lagged neighborhood defaults may induce later individual defaults, and that the local price effects induced by lagged defaults (and other factors) may affect individual default probabilities as well. These default effects may arise because defaults result in vacant properties, leading to neighborhood decay, or because defaults are one component of undesirable turnover in neighborhood homeownership. In either case, a temporary rise in local default rates may tend to

persist. There is no support here for a role of lagged prepayments as a trigger that induces defaults directly, and there is an uncertain role for prepayments affecting house prices. Although inconsistencies, anomalies, and serious data limitations (including the narrowness and nonrandomness of the Chicago MSA data) make these conclusions highly tentative, it appears that this wider and more precise set of neighborhood controls may reduce the remaining estimated impact of neighborhood racial composition on individual default behavior.

# SECTION 1

## INTRODUCTION

### 1.1 Motivation and Background

This paper examines the existence and nature of the effect of neighborhood characteristics, particularly income and race or ethnicity, on the default of FHA loans. Such a study is motivated by several considerations. First, there is a continuing public policy interest in the fortunes of minority and low-income residents in general, and thus neighborhood effects tied to race and income are of special concern. Second, neighborhood effects related to race and income may be indicative of other problems, such as inability of low-income or minority borrowers to obtain funds for refinancing their homes; and such problems may (or may not) call for policy changes. Third, the existence of neighborhood effects implies that samples of loans drawn from specific, narrowly defined geographic areas will tend to covary in their default behavior, even after controlling for individual characteristics of loans and borrowers. The assessment of riskiness of loan portfolios should presumably recognize this covariance, which implies greater variance in outcomes for loan portfolios that are not distributed randomly with respect to geography.

Several existing studies deal in some way with the effects of neighborhood racial composition and neighborhood income on default. Given that these studies differ substantially in data sources, the types of information available, and the statistical methods that are employed, there is no unanimity in findings. For example, using data on loans purchased by Freddie Mac to estimate a proportional hazard model, but lacking data on borrower race, Van Order and Zorn (1995),<sup>1</sup> find that both borrower income and tract income help explain default behavior, but the effect of neighborhood income is stronger and more stable. Tract racial composition (share of households that are black) also appears to affect default probabilities. In contrast, in a logit estimation procedure using FHA data that include borrower race information, Schnare and

---

<sup>1</sup> Van Order, Robert, and Zorn, Peter. "Income, Location, and Default: Some Implications for Community Lending," July 1995.

Gabriel (1994)<sup>2</sup> find no consistent statistically significant impacts on default of either income or race at either the borrower or tract level. Berkovec, et al (1994),<sup>3</sup> also use FHA data, but they find that tract income is associated with reduced default probabilities, black representation within the tract is “not strongly and consistently” associated with default, and increasing Hispanic representation is associated with lower default probabilities. Although not focusing on neighborhood race and income *per se*, Calem and Wachter (1999)<sup>4</sup> examine the effect of neighborhood housing market conditions, among other factors, on long-term delinquency of home purchase loans that were originated in Philadelphia from 1988 through 1994 as part of an affordable home loan program. They find evidence that increases in housing market activity are associated with a lower probability of delinquency, and more expensive homes relative to the neighborhood tend to have higher delinquency rates.<sup>5</sup>

This paper attempts to improve upon and extend these existing studies. In contrast to the work of Van Order and Zorn (1995) and Van Order, Westin, and Zorn (1993),<sup>6</sup> which rely on race measured at the area (census tract or zip code) level, this study uses a data set that has information on borrower’s race as well. This detail will permit us to take a more refined look at the factors that may lie behind differences in default across neighborhoods. In particular and of

---

<sup>2</sup> Schnare, Ann, and Gabriel, Stuart A. “The Role of FHA in the Provision of Credit to Minorities,” ICF Incorporated, April 1994.

<sup>3</sup> Berkovec, James A., Canner, Glenn B., Gabriel, Stuart A., and Hannan, Timothy H. “Race, Redlining, and Residential Mortgage Loan Performance,” *Journal of Real Estate Finance and Economics*, 9: 263-294 (1994).

<sup>4</sup> Calem, Paul S., and Wachter, Susan M., “Community Reinvestment and Credit Risk: Evidence from an Affordable-Home-Loan-Program,” *Real Estate Economics*, V27, Number 1: 105-134 (1999).

<sup>5</sup> An NTIC study (“The Devil’s in the Details,” National Training and Information Center, October 1997) examines neighborhood effects of FHA loans, but the statistical analysis is generally informal. As required by Congress, the Federal Reserve Board is studying default of CRA loans in neighborhoods. See “The Performance and Profitability of CRA-Related Lending,” Report by the Board of Governors of the Federal Reserve System, submitted to the Congress pursuant to section 713 of the Gramm-Leach-Bliley Act of 1999, July 17, 2000.

<sup>6</sup> Van Order, Robert, Westin, Ann-Margret, and Zorn, Peter. “Effects of the Racial Composition of Neighborhoods on Default and Implications for Racial Discrimination in Mortgage Markets,” March 1993.

special importance, we will be better able to tell whether factors like race affect default at the neighborhood level (through, say, correlation with unobserved rates of house price growth) or only at the individual level (because, say, race is correlated with the unobserved occurrence of trigger events, like unemployment). In addition, our findings will be more directly applicable to FHA policy questions since we will be using data on FHA-insured loans rather than the conventional loans used in the latter set of studies.

In contrast to studies by Schnare and Gabriel (1994), Berkovec, et al (1996)<sup>7</sup>, and Berkovec, et al, (1994), all of which are based on FHA-insured loans, we will see whether interarea differences in default rates remain even after using more appropriate estimation techniques and controlling for events that occur after loan origination (*e.g.*, changes in principal balance). In our view, the empirical work in the latter studies offers unconvincing evidence of area effects because the estimation technique (a logit on whether a default has occurred since origination) cannot account properly for post-origination changes in the housing or economic environment, which in many cases may be substantial. Indeed, as revealed by the list of explanatory variables used in these studies,<sup>8</sup> there is no attempt to recognize this source of variation across areas, thus leaving open the possibility that interarea differences (or lack thereof) represent nothing more than spurious correlations of area-level measures with post-origination economic events.<sup>9</sup>

Finally, unlike almost all of these other studies, this study will be able to exploit a database containing credit history information on most borrowers.<sup>10</sup> This capability will enable us to

---

<sup>7</sup> Berkovec, James A., Canner, Glenn B., Gabriel, Stuart A., and Hannan, Timothy H. "Mortgage Discrimination and FHA Loan Performance," *Cityscape* Volume 2, Number 1, February 1996.

<sup>8</sup> Berkovec, et al (1994) justify considering only factors known at origination by arguing that they are looking for discrimination in the lending decision. Of course, some characteristics, such as the current, scheduled principal balance on a fixed-rate loan, are time-varying, but the time path is known in advance.

<sup>9</sup> Van Order, Westin, and Zorn (1993) recognize post-origination differences across loans by introducing a variable representing the probability of negative equity over an eight-year horizon, but the variable is not entered into the hazard model as a time-varying covariate. Again, there is no attempt to build in the appropriate dynamics.

<sup>10</sup> Calem and Wachter (1999) have credit scores on about 56 percent of the sample; they impute credit scores for the remainder.

isolate more clearly the effects of neighborhood by permitting us to control for a factor --- differences in past credit performance --- that is often alleged to lie behind area or individual differentials in default behavior. (See, for example, Van Order and Zorn (1995).)

The availability of additional data is in a sense a mixed blessing. Because we will have the luxury of more detailed data, we shall be forced to turn elsewhere for explanations in the event that area or individual differences related to race or income persist. That is, when one attempts to estimate area or individual effects of race or income on default using a database that lacks information on potentially important factors, one can reasonably wonder whether the estimated effects are attributable in total or in part to omitted variable bias. For example, in attempting to estimate the effects of race at the neighborhood level by using a database that lacks individual race information, one must contend with the possibility that findings are seriously distorted by omitted variable bias. Moreover, even if race at both the neighborhood and individual level is recognized in estimation, findings can be strongly affected by the omission of potentially important data on, for example, credit history.<sup>11</sup> Because this study includes credit history information in the default analysis, there is at least some hope that we have effectively dealt with this potentially important source of variation, and thus it will be more difficult to claim that remaining race or income effects at the individual or area level are due to correlations with missing credit history data. Although one could reasonably argue that there are other omitted variables, as well as a host of other potential econometric problems (some of which are discussed below), we shall turn elsewhere for explanations. We consider three possibilities, and we attempt to explore each in some detail.

One explanation revolves around the possibility that racial impacts on default are traceable to difficulty in refinancing loans when it would be advisable to do so, *i.e.*, when there is a sufficiently large reduction in mortgage rates. Such an explanation is suggested by the possibility of racial disparities in the probability of obtaining refinancing loans --- “redlining” when operating at the neighborhood level, or simple race discrimination when operating at the

---

<sup>11</sup> For this reason, Berkovec, et al (1994) attempt to assess the size of the bias in estimated race effects attributable to the omission of credit history data. Unfortunately, there appears to be an error in their derivation.

level of the individual borrower.<sup>12</sup> In this regard, note that Schnare and Gabriel (1994) find that individual blacks have lower loan acceptance probabilities than do whites in both the conventional and FHA sectors.<sup>13</sup> They do not, however, find a consistent pattern of effects of neighborhood racial composition on the probability of loan acceptance. Schill and Wachter (1993)<sup>14</sup> study the accept/reject decision for conventional loans in Philadelphia and Boston and find little evidence of redlining once controls for neighborhood risk are included in the analysis.<sup>15</sup> (They do report significant individual race effects, and they note that these could be traceable to the omission of individual borrower characteristics, such as credit scores and LTV.)

If redlining in obtaining refinancing is the culprit in generating higher default probabilities, one might expect to find that racial composition of the neighborhood enters the probability of default as well as the probability of refinancing. If discrimination in refinancing at the individual level generates higher default probabilities, one might expect to find that individual race influences both the probability of default and the probability of refinancing. Naturally, other less insidious explanations — *e.g.*, lack of information or unwillingness to pursue refinancing — could account for the same pattern.<sup>16</sup> Hence, we shall more realistically be able only to reinforce or cast doubt on these explanations, rather than deduce the single correct explanation.

A second explanation for interarea differences in default propensities (assuming they exist) is

---

<sup>12</sup> Noneconomic discrimination in obtaining home purchase loans is expected to result in lower default probabilities (on the margin) for affected individuals or groups. See, for example, Berkovec, et al (1994). Here we focus instead on possible discrimination in obtaining refinancing.

<sup>13</sup> It is unclear whether the paucity of relevant data on individual borrowers is playing an important role in generating these findings.

<sup>14</sup> Schill, Michael H., and Wachter, Susan M., "A Tale of Two Cities: Racial and Ethnic Geographic Disparities in Home Mortgage Lending in Boston and Philadelphia," *Journal of Housing Research*, Volume 4, Issue 2: 245-275 (1993).

<sup>15</sup> The evidence for conventional mortgage refinance applications reported in Schill and Wachter is more ambiguous: Boston, but not Philadelphia, shows differences in acceptance probabilities based on the racial composition of the tract even when neighborhood risk characteristics are included in the analysis.

<sup>16</sup> One might question the importance of refinancing difficulties in generating higher default probabilities. This question is considered at greater length below.

that higher levels of default within the neighborhood, set off by any reason, result in abandoned structures<sup>17</sup> that tend to make the neighborhood less desirable, which in turn propagates additional defaults in the future. The abandoned structures could act as a trigger for default, causing borrowers to default on homes that were ripe for default in any case. Alternatively, declines in house prices may act as a mediating factor. That is, abandoned structures may adversely affect house price appreciation in the area, thus increasing the probability of default in the area. Such an effect should be apparent in house price information at the neighborhood level. This explanation, of course, begs the question of what sets off the initial increases in defaults in certain neighborhoods, and why these are correlated with racial composition or neighborhood income. It does, however, provide a reason for defaults to continue at a high level for some time following a neighborhood shock in the default rate, perhaps even for a very long time.

A third possible explanation is that instability in the form of higher turnover of homeowners, as reflected in default or in simple changes of residence, makes minority or low income neighborhoods less desirable.<sup>18</sup> Again, higher turnover may act as a trigger event that brings on default directly. Alternatively, local house price changes may act as a mediating factor; that is, higher turnover may result in locally decreasing demand for housing which reduces local house prices, leading to default. This explanation again begs the question of what sets off the initial increases in turnover in certain neighborhoods, and why these are correlated with racial composition or neighborhood income.

## **1.2. Specific Aims**

The latter considerations suggest the following specific questions that the study will attempt to

---

<sup>17</sup> On this point, see the aforementioned study by NTIC (1997).

<sup>18</sup> Note in this regard that while high-income families may be more likely to make long-distance moves, low-income families are more likely to move at all. For example, the March 2000 Current Population Survey shows that about 28 percent of families with annual incomes less than \$10,000 have moved in the previous year. This percentage declines across income classes, reaching 11 percent for families with annual incomes exceeding \$90,000.



answer; the remainder of the paper is organized around these questions.

**Once one controls for a variety of borrower- and loan-related factors, including time-varying characteristics, in an appropriate econometric model of default, do neighborhood effects seemingly related to income or race persist?** There may, for example, be differences across neighborhoods in house price growth and in the rates at which loans amortize (due to differences in note rates and mortgage terms), and these differences may give rise to differences in default behavior even when controlling for conditions at loan origination, such as LTV. Spurious correlations between these factors and other neighborhood characteristics, such as race or income, may make it appear that differences in race and income, rather than these primary factors, are at work. If so, statistically correcting for differences in these primary factors may eliminate the appearance of neighborhood effects based on these spurious correlations.

Naturally, our ability to introduce a comprehensive set of relevant time-varying characteristics will be circumscribed by the available data; in particular, we lack data on house price growth at the neighborhood level,<sup>19</sup> and thus we rely on MSA-wide house prices for a major portion of the study.

**Do neighborhood characteristics, such as race and income, have effects on default that are separate and distinct from the effects of these same characteristics at the individual level?**

That is, are previous findings of neighborhood impacts truly traceable to the neighborhood, or are they instead proxying individual effects that cannot be measured in databases that are often used.

**Is there evidence that differences in default probabilities reflect differences in the**

---

<sup>19</sup> There is only a small literature on house price change at the neighborhood level in general and in lower-priced neighborhoods in particular. Exceptions are Pollakowski, et al, and Newburger, et al, who indicate that rates of house price appreciation may not be lower in the lower-valued portions of the market. (See Pollakowski, Henry O., Stegman, Michael A., and Rohe, William, "Rates of Return on Housing of Low- and Moderate-Income Owners," *AREUA Journal*, 19 (3), 417-425, and Newburger, Harriet, Pollakowski, Henry, Stegman, Michael A., and Rohe, William, "House price Appreciation in Neighborhoods with Modestly-Valued Properties," (undated).)

**probability of refinancing?** As noted, one possibility, prompted by earlier studies suggesting that blacks suffer discrimination in loan qualification, is that neighborhoods or individuals differ in default propensities because of differential access to refinancing alternatives. If so, we expect to find that default probabilities are positively related to factors related to (unrealized) refinancing probabilities and, second, that there are neighborhood or individual race effects in refinancing probabilities, conditional on the measurable incentives to prepay.

**Finally, do neighborhood differences in default probabilities arise because of earlier defaults and, if so, what is the mechanism?** That is, does an increase in the default rate at one point in time lead to higher default rates at later times? If so, are the default-inducing effects of earlier defaults transmitted through reductions in neighborhood house price growth or through some other mechanism?

### **1.3. A Roadmap for the Remainder of the Paper**

The exercise begins in Section 2 with a brief discussion of data sources and some tables that illustrate some of the interesting features of the data. Next is a discussion of statistical models of default and prepayment, as well as the formulation of variables that attempt to match the theoretical constructs. We then move to a presentation of statistical estimates that rely on both relatively standard default-related factors measured at the time of origination, but also on default-related factors that change over the course of loan duration. To provide a benchmark that may permit comparisons to existing studies, the analysis initially excludes information on individual race and credit history, as well as neighborhood characteristics. We then add, in sequence, neighborhood characteristics, individual race, and finally a summary credit score, assessing at each stage the remaining estimated impact of neighborhood characteristics. This work permits us to assess the roles of neighborhood and individual characteristics, as well as the possibility that difficulties in refinancing lie at the heart of neighborhood differences in default rates.

In Section 3, we begin a second strand of the analysis that asks more directly whether neighborhoods differ in identifiable ways that would logically lead to differences in default rates.

In particular, there we examine questions of whether default behavior seems to spawn subsequent default behavior, either by leaving vacant structures or because default is one component of undesirable turnover of ownership. The analysis turns first to an analysis of house price changes at the neighborhood level in the Chicago MSA and investigates the possibility that past default behavior and turnover of homeownership in general affect house price growth. We then apply the findings from this analysis to reestimate default models for the Chicago MSA.

Section 4 offers some tentative conclusions.