SECTION 1

INTRODUCTION

1.1. Purpose and Methodological Approach

The purpose of this paper is to answer two related questions. First, are FHA defaults geographically concentrated in a set of high-default neighborhoods? Second, are FHA defaults concentrated in a set of high-default lenders? To answer these questions in the most useful way, we must be more precise about what is meant by "concentrated." Even if loan activity were evenly distributed across neighborhoods and lenders, we would not realistically expect to find defaults distributed absolutely evenly in these same dimensions. Pure chance alone --- in the form of death or debilitating illness of the borrower, for example --- would likely cause some lenders and geographic areas to have more than the expected share of default activity. It is only when there are surprisingly large numbers of defaults, given the level of loan activity, that we might possibly want to delve further into possible causes. Presumably, we are surprised by the outcome, however, only when the observed number of defaults is highly unlikely to have arisen because of chance alone. That is, we may rephrase the initial question more usefully as follows:

> Is it highly unlikely that the geographic concentration of defaults could have arisen from chance alone?
>
> Similarly, is it highly unlikely that the concentration of defaults among lenders could have arise from chance alone?

Although the term "highly unlikely" must still be defined, rephrasing these questions in this way invites statistical analysis, for the purpose of one kind of statistical analysis is to assess how unlikely a particular event is to occur strictly as a result of chance. Hence, in this paper we use standard statistical methods to gauge how unlikely outcomes are to have occurred as a result of chance alone. Our approach to answering these questions is thus essentially statistical.

Despite the fact that the main line of inquiry is statistical, we first take a nonstatistical, descriptive look at the data on FHA originations in 1992 and 1994 in 22 MSAs,[5] characterizing tracts and lenders that seem to have high default rates relative to the MSA-wide average. We then provide a simple statistical analysis of these raw default rates, moving later to a more sophisticated statistical model that takes account of the presence of other measurable factors that could give rise to intertract and interlender differences in default probabilities.

This study contrasts with a recent, provocative study by the National Training and Information Center (see National Training and Information Center [1997]) that answers the same initial questions with an entirely nonstatistical methodology. Using FHA originations from 1991 through 1994 in a sample of 20 cities, the NTIC study identifies high-default tracts as those with a

---

[5] The individual MSAs are listed in Table 1 in Appendix A.

default rate that is at least 1.5 times the MSA-wide default rate. No consideration is given to the influence of sample size, much less systematic influences, on one's ability to draw conclusions from tract-level data on raw default rates. From this ambiguous evidence, the NTIC report concludes that defaults are too heavily concentrated geographically.

The NTIC report's attempt to find culprits for the observed geographical dispersion in default rates raises even more questions. The ten lenders in each city that have the largest number of defaulted loans are singled out as the "ten worst lenders" in each city.[6] Not only does this method ignore the role of chance, it fails to give any consideration to the sheer volume of loans in determining the number of defaults. Using the NTIC methodology, a high volume lender may show up in the "10 worst" list even though it has an exceptionally careful loan screening process, fair underwriting standards, a practice of offering numerous alternatives to foreclosure, and a low default rate.

Nowhere in the NTIC study is there any mention of variation in other factors, such as income or the prevalence of first-time homebuyers, that could lead to geographic and interlender dispersion in default rates in the absence of any wrongdoing on the part of lenders. That is, not all borrowers who are acceptable loan risks have the same propensity to default, and one should not be surprised to find that underlying default-related factors vary systematically, both geographically and across lenders. Below we see that there is indeed such variation.

1.2. Organization of the Report

The remainder of this paper proceeds through a statistical analysis that attempts to isolate the differences in default probabilities across tracts and lenders after allowing for the effects of chance and the influence of measurable systematic factors. The series of steps is as follows. In the next section, we lay the foundation for the subsequent work by explaining the nature of the FHA data that underlie our analysis and the conventions we adopt to measure defaults. In Section 3 we examine, in a largely nonstatistical and informal way, the distribution of default rates across tracts and lenders under three different measures of default and, within a measure, across origination years. We then go on to characterize tracts and lenders in a variety of dimensions, showing in particular how tracts and lenders with high default rates differ in key ways from other tracts and lenders --- ways that make intertract and interlender differences in default rates readily understandable.

Section 4 begins a more formal statistical analysis, first looking for evidence on whether there is any detectable overall association between tract and default activity. We next move into simple statistical tests at the basis of the individual tract, asking whether the default rate in each individual tract differs from the MSA rate beyond that which would be expected from random influences alone. Such tests are performed on data from each separate origination year and for both origination years together. The battery of statistical tests is then repeated for lenders.

One of the more important insights in this section is that random influences alone can cause

---

[6] It is unclear whether lenders' default volume is measured at the city level or within high-default census tracts alone. The same statistical flaws are inherent either way.

default rates for individual tracts or lenders to differ in a statistically significant way from that in the MSA. Thus, when conducting such tests on a large number of tracts or lenders, one should expect to find the appearance of significant differences even if nothing is awry. Such is the nature of statistical tests. The evidence here also suggests that some of the apparently significant differences in default rates across tracts or lenders vanish over a two-year period, suggesting that some perceived problems are only temporary.

Section 5 takes a deeper statistical look at interarea and intertract differentials through an examination of default at the level of the individual loan. The idea is that if interarea or interlender differentials exist, they should show up as determinants of default in microdata. We control for a few of the standard underwriting factors, as well as for a more extensive set of other factors that characterize the loan, the borrower, and the economic environment, and we ask how much influence of tract and lender remains after controlling for these other default-related factors. Although the estimates reveal that tract and lender influences remain even after controlling for other measurable factors, their influence declines considerably once such controls are introduced.

Unfortunately, the fact that FHA data do not contain many of the important determinants of default behavior, including possibly crucial elements of credit history, almost guarantees that tract and lender effects will remain, and other more subtle statistical factors tend to reinforce this tendency. For the same reasons, it proves impossible to tell whether lenders are following underwriting guidelines. The data do, however, permit a rather crude and cursory examination of other aspects of lender behavior. Section 5.2 briefly outlines some simple tests of whether high-default lenders intervene in delinquencies more quickly than do other lenders and whether they less frequently offer alternatives to foreclosure.

Section 6 closes with a summary that highlights the methodology. The "Summary of Findings" above provides a more complete summary of the analytical and policy findings.

The paper contains two appendices, each of which is devoted to tables. The text of the paper refers to 27 tables, most of which are composed of multiple panels. Appendix A presents a complete list of these tables, as well as all of the tables themselves. To help minimize searching through appendix material, the most critical tables are also copied and inserted at the appropriate place in the text, though they still appear in Appendix A as well. Appendix B contains MSA-specific results that either underlie some of the main tables or else repeat an aggregate analysis at the MSA-level. Because of its length, Appendix B has been omitted from the paper; Appendix B is available from HUD.

SECTION 2

DATA AND DEFINITIONS

2.1. Data Construction and Data Sources

The data to be used in this study come primarily from FHA data files on approximately 650,000 FHA-insured loans that were originated in calendar years 1992 and 1994[7] on properties in 22 MSAs.[8] Ten of these MSAs are also included in the NTIC study. The data contain the usual array of FHA data on the loan and the borrower, as well as information on the geographic area (state, county, census tract) in which the property is located and the identity of the lender.

These FHA data have been supplemented with tract-level data from two sources: 1990 Decennial Census data (which generally measure activity in 1989 or 1990) and mortgage data for 1992 through 1996 that have been generated under the Home Mortgage Disclosure Act (HMDA).

Following the NTIC report, we generally use the census tract as a neighborhood. The advantage of such a choice is that the Bureau of the Census attempts to choose tract boundaries so as to maintain the separate identity of neighborhoods. The disadvantage for the current study is that many census tracts have so few FHA loans in the period under consideration that statistical analysis at the tract level is all but impossible. Aggregating tracts into larger "supertracts" is one solution, but it is not obvious how to define an appropriate, feasible aggregation procedure, given the costs of the inevitable mistakes in aggregation. The cost of inappropriate aggregation is that dissimilar neighborhoods are put together, masking problems that would be exposed if tracts were kept separate, and this cost of aggregation must be traded against the benefits of subjecting additional areas to statistical analysis.

In this study we have opted for aggregation of many tracts with low loan volume. The aggregation method used here tries to minimize information loss in aggregation in two ways. First, we generally attempt to aggregate only those tracts with so few loans that they could not easily support statistical analysis in the absence of aggregation. Second, we demand that the tracts that are to be aggregated have similar tract incomes and minority representation in their populations and be in reasonable geographic proximity. The basic aggregation method is as follows. For each of the

[7] More precisely, the data set includes loans that originated in calendar year 1992 or 1994 or for which the amortization start date was in the interval February 1, 1992, through January 31, 1993, or the interval February 1, 1994, through January 31, 1995.

[8] The choice of these particular MSAs was based on the number of defaults over a broader period (1989 through 1994) for which we expected to be able to obtain suitable FHA data on the loan and the borrower. In the current application, only the data on 1992 and 1994 originations are available. In these FHA data, we assign a loan to a particular MSA based on the state and county in which the property is located. The alternative of using the coded value of MSA in the FHA data base leads to approximately the same set of loans, however. Using the coded values of tract and state to identify loans in the MSA leads to substantially fewer loans in the MSAs, especially in the 1992 origination data. Presumably, the tract coding in the 1992 FHA data fails to reflect consistently the then-current tract definitions and codes.

separate MSAs at issue, we use the Decennial Census data to calculate the deciles of tract-level median family income. For each census tract in these MSAs, we then record that tract's median family income decile. Next, we categorize the percentage minority into ten intervals of 10 percentage points each (0 to 10%, 10 to 20%, etc.) and use the Decennial Census data to record the bracket into which each tract's percentage minority falls. Finally, we make a grid of approximate rectangles that are roughly four miles on each side, adjusting the precise dimensions of the "rectangle" for each MSA so as to fit an integral number of gridlines between the extreme tract centroids in each MSA. We use the Census data to record the particular "rectangle" containing the centroid of each census tract.

These three aspects of each tract are used to form cells, each of which contains all tracts that are identical in these three dimensions, *i.e.*, have centroids in the same grid cell, are in the same median family income decile, and are in the same 10-percentage point minority representation bracket. Within each cell, all tracts having less than 30 loans[9] in the FHA data are aggregated together. If, within a cell, the aggregate thus formed still contains fewer than 30 loans, then the next smallest tract (in terms of loan volume) with more than 30 loans is included as well. Although this aggregation method is unlikely to match tracts for aggregation in an optimal way, it is guaranteed to aggregate only those tracts that are similar in minority composition, income, and geographic proximity. The effect of this aggregation procedure is to reduce the number of tracts having less than 30 loans by 15 percent (in the Fort Lauderdale, FL PMSA) to 59 percent (in the Chicago, IL PMSA), with over a 30 percent reduction in most MSAs.

2.2. Preliminary Matters: The Identification of Defaults and Observation Intervals

Analysis of these data requires the resolution of several immediate issues: the identification of defaults (*i.e.*, what constitutes a "problem" loan), the number of origination years to include in the analysis, and a resolution of two related timing issues: the length of loan duration over which default activity will be observed and the calendar date at which loan status will be recorded. As to the first of these items, default definitions can range from claims paid to simple 90-day delinquencies, with the choice hinging in part on the severity of the payment problem that is to be analyzed. As to the second item, the number of origination years, which is here limited to a maximum of two by virtue of the data available for this study, could be further limited in a particular analysis to focus on a more homogeneous lending environment; yet difficulties (*e.g.*, lax underwriting) that persist for only brief periods may not be worth treating, and limiting the range of origination years further reduces sample sizes.

2.2.1. Timing Issues

The third set of issues --- those related to timing --- is complex, in part because of possible

---

[9] The cutoff at 30 loans is arbitrary. Note that only tracts appearing in the Census data files are eligible for aggregation under this procedure. This fact precludes aggregation of those FHA loans for which tract identifiers do not match the Census files, including those loans with missing tract identifiers.

data problems. The FHA delinquency data in particular seem likely to contain substantial error and, moreover, our understanding is that these data have become more complete and more accurate with the passage of time. The latter consideration would suggest some advantage in measuring default status as late in calendar time as possible if delinquencies are to come under the definition of default. Clearly, however, measurement of loan status at a particular date implies measurement at a certain interval of loan duration (given the date of loan origination), and thus one must consider appropriate loan durations as well. The NTIC study argues that one should look long after loan origination to pick up the bulk of default activity. While it is true that the bulk of claims occur after the first few years of loan duration, the kinds of problems singled out by the NTIC study --- leaky roofs, absence of heating systems, and other structural problems --- could seemingly not go unnoticed for more than a couple of years if these problems were known or detectable at the point of sale. Indeed, one could argue underwriters and inspectors have done their jobs if the loan survives the first couple of years following origination, for the informational content of much of what underwriters and inspectors can observe deteriorates quickly with age. There is, for example, evidence that underwriting factors like credit history are much less predictive at longer loan durations. (See Holloway, MacDonald, and Straka [1993].)

The latter observations are not meant to deny that underwriting factors can affect default at longer loan durations; there is little doubt that they can. Lowering the maximum acceptable loan-to-value ratio to, say, 50 percent would generate a large cushion that would protect equity against substantial declines in local housing markets and would surely reduce the incidence of claims over a wide range of loan durations. Changes like these, however, involve policy changes at the national level and involve major tradeoffs in other dimensions. The search in this study is for indications of improper lender behavior, such as failure to follow existing underwriting guidelines, and these seem likely to be exposed rather rapidly after loan origination. It also seems reasonable to assume that other potential problem areas, such as improper loan servicing, will show up in loans at all durations.

2.2.2. Definitions of Default

For this paper we identify defaults in three ways, and we provide virtually all analyses for each of the three definitions. The three definitions are

1. Claims that have been made within two years of the amortization start date. We shall generally refer to this definition as "claims at two years."

2. Claims that have been made within two years of the amortization start date, plus all loans that are in delinquency (for 90 or more days) two years after the amortization start date and that are not observed to cure by May 1, 1997 (the time at which the data tapes were made). We will generally refer to this definition more simply as "uncured delinquencies at two years," though it is understood that this definition includes all claims included in the first definition as well.

3. Claims that have been made as of December 31, 1995, plus loans that are delinquent as of that date and are not observed to cure by May 1, 1997. We shall generally refer to this definition of default as "uncured delinquencies at 12/95," though again this definition additionally incorporates claims that have been made by December 31, 1995. This particular cutoff date is that utilized in the NTIC report. Note that this definition, unlike the first two, does not use a two-year window.

These choices represent compromises among the numerous ways in which a "problem" loan may be identified and compromises among the timing issues discussed above as well. Restrictive definitions, such as the first one above, capture those loans that are virtually guaranteed to impose costs on FHA as the loan insurer; looser definitions, such as all loans that are in delinquency status as of a particular date, surely capture numerous loans that will never end in a claim. Moreover, only loans that end in foreclosures (or related terminations, such as deeds-in-lieu) are likely to result in vacant properties that might lead to the deterioration of neighborhoods suggested by the NTIC study.

An important disadvantage of the most restrictive definition (claims at two years), however, is that it fails to include those ongoing delinquencies that will evolve into claims. For this reason, it makes sense to build in those delinquencies that do not show a cure by the end of the time over which we can observe such behavior. Notice that for loans that begin amortization in January 1995 --- the latest start among loans examined here --- the two-year window on observing claims or ongoing delinquencies expires in January of 1997, leaving little remaining time over which to observe a cure before the observation period ends.

There is another potential problem with the second of the two definitions above: the FHA data show marked increases in delinquencies as calendar time passes, and this feature suggests (but does not prove) that delinquency data have become more complete in recent years, as noted above. As a consequence, we may be missing delinquencies, particularly for loans that originate in early 1992, for which the two-year window ends in early 1994. We have attempted to minimize the problem of incomplete delinquency data by supplementing the FHA F42 delinquency data with A43 claims data, picking up some claims that are not reflected in the former data set and imputing a delinquency date for these claims. It seems likely that this imputation procedure has only a minimal effect in making the delinquency data more complete.

The third definition above may be less sensitive to data problems. This definition has the advantage of measuring ongoing delinquencies early enough (at December 31, 1995) that 16 months remain over which cures could occur and still be observed; it seems very likely that those delinquencies not observed to cure over this long an interval will end in a claim. In addition, this date is late enough that improvements in tracking delinquencies over time are likely to result in more complete coverage by this date, likely yielding a more comprehensive measure of delinquencies for the 1992 originations than is obtained at the two-year mark. Finally, this particular cutoff date is that used in the NTIC study, thus facilitating comparisons with that study. The major disadvantage of the third measure --- and the reason that it is not adopted as the primary measure here --- is that it is asymmetric in its treatment of loans that originate at different points in time, allowing too little time to elapse for loans that originate at the end of 1994 but (arguably) allowing too much time to elapse for loans that originate in early 1992 (about four years). The adoption of a two-year window for the first two definitions used here seems long enough to capture default behavior that arises out of poor underwriting, yet not so long that default behavior is completely overwhelmed by external factors that cannot possibly be reckoned with in the underwriting process.

The fact that these different definitions of default enjoy their own particular advantages means that the choice of any specific definition inevitably involves trade offs. In addition, we find that at times the alternative measures of default behave differently in a particular analysis, and sometime these differences may be instructive. At other times, the three measures behave similarly, and that is noteworthy. For these reasons, we make virtually all calculations in the paper for all three definitions of default, and our discussion sometimes covers all three measures as well. Nonetheless, because we believe that the second definition (uncured delinquencies at two years) offers, on the whole, slightly more important advantages than the other two definitions, and because the use of a single definition simplifies the exposition, we sometimes focus the discussion on only the second default definition, and we present tables in the text for only that definition. A complete set of all 27 tables referenced in the text, including those relying on the other two definitions, is contained in Appendix A.

Table 1 (in Appendix A) shows, for each of the 22 MSAs, the default rates calculated under all three definitions for each year of origination separately and for the two origination years together. We see, for example, that when default is defined as uncured delinquencies at two years following the amortization start date ("uncured delinquencies at two years"), the default rate in the Atlanta, GA, MSA is 1.4 percent in 1992 originations, 2.73 percent in 1994 originations, and 2.04 percent overall. The table shows that there is clearly variation from definition to definition within each MSA, as well as variation between origination years within each definition, even when, as in the first two definitions, the same number of years (two) elapse after loan origination.[10] For the second of these definitions (uncured delinquencies at two years), the 1994 originations display a higher rate than the 1992 originations in each MSA, perhaps in part because delinquencies are reported more completely as time passes.

[10] Notice also that similarities in default rates under the claims definition do not imply similarities in default rates under the other two definitions. See, for example, the data for Minneapolis and Philadelphia, which have very similar claim rates but very different default rates under either of the other two definitions.

SECTION 3

DISPARITIES IN DEFAULT RATES ACROSS AREAS AND LENDERS

3.1. The Geographic Concentration of Defaults

Using the two years (1992 and 1994) of loan originations together, the three panels of Table 2 show the percentage distribution of tracts[11] within each MSA across default rate categories. (Panel B of Table 2 is presented below; the remaining two panels are in Appendix A.) As is true for many of the tables in the remainder of this paper, only those tracts with more than 30 loans are classified into specific default rate categories, while all tracts with 30 or fewer loans are lumped into a single category. In the Atlanta MSA, for example, we see from Panel A (in Appendix A), which uses claims at two years as the measure of default, that among the 480 tracts in that MSA, 30.21 percent of the tracts have 30 or fewer loans in the two origination years together, and 42.29 percent of the tracts have more than 30 loans and a default rate that is between 0 and 0.5 percent. In Panel B below, which measures default as an uncured delinquency at two years, we see that in the Atlanta MSA, 20.83 percent of the tracts have more than 30 loans and a default rate that is in the range of 0 to 0.5 percent. Quite clearly, within each MSA there is substantial variation in raw default rates across census tracts under any of the three definitions entertained here. In addition, numerous tracts have 30 or fewer loans. (As we shall see, however, these low-volume tracts contain only a small fraction of loans in the aggregate.)

3.1.1. Tract Default Rates Relative to MSA Default Rates

A somewhat different look at the intertract variation in default rates is obtained by using the *relative* default rate of each tract, *i.e.*, by taking the ratio of the default rate of the tract to the default rate of the MSA in which the tract is located. Such a calculation presumes, of course, that the comparison of the tract rate to the MSA rate is not only appropriate, but is more meaningful than a comparison of the tract rate to, say, the national default rate. A major argument in favor of using the MSA rate as the basis for comparison, as is done throughout this paper, either implicitly or explicitly, is that there are strong idiosyncratic factors affecting local housing markets that should not be permitted to influence (in either direction) the labeling of a tract (or lender) as "high-default." Relying on MSA default rates as a benchmark controls for MSA-wide differences in the strength of the housing market and the local economy, as well as certain institutional features (such as the legal setting) that probably vary less within MSAs than across MSAs.[12] The alternative of making

[11] In the remainder of this paper, we use tracts that have been aggregated according to the procedures outlined in Section 2, and we generally exclude the approximately 10 percent of the loans for which tract identifiers are missing.

[12] Comparisons of tract rates with MSA rates precludes the possibility of picking out tracts (or lenders) that actually do perform poorly but whose poor performance is hidden by comparison with an MSA composed disproportionately of other poorly performing tracts (or lenders). That is, the assumption is that, absent local factors and differences across borrowers, tracts (lenders) would on average perform similarly in all MSAs. The more sophisticated analysis in Section 5 minimizes reliance on the latter assumptions by controlling for

TABLE 2

PERCENTAGE DISTRIUBTION OF TRACTS ACROSS DEFAULT RATE CLASSES, BY MSA
1992 AND 1994 ORIGINATIONS

PANEL B: UNCURED DELINQUENCIES AT TWO YEARS

MSA NAME	Total Tracts	< 31 Loans	Raw Default Rates (Percent) in Tracts with > 30 Loans									
			0 to 0.5 %	>0.5 to 1.0 %	>1.0 to 1.5 %	>1.5 to 2%	>2 to 3 %	>3 to 4 %	>4 to 7 %	>7 to 10 %	>10 to 15%	>15 %
ATLANTA, GA MSA	480	30.21%	20.83%	4.79%	7.71%	5.42%	12.71%	7.50%	7.92%	2.71%	0.21%	0.00%
BALTIMORE, MD PMSA	479	40.50	17.12	2.92	5.01	4.59	13.15	5.01	7.10	3.97	0.42	0.21
CHICAGO, IL PMSA	1142	46.23	18.30	1.58	3.50	5.87	8.32	4.29	7.88	2.80	0.70	0.53
DALLAS, TX PMSA	489	36.81	20.45	4.09	7.77	4.70	9.00	9.41	5.32	2.25	0.20	0.00
DENVER, CO PMSA	404	11.88	49.26	13.86	9.90	5.94	5.69	2.48	0.74	0.25	0.00	0.00
DETROIT, MI PMSA	1079	73.49	13.99	0.09	1.11	1.76	3.61	1.39	2.32	1.39	0.65	0.19
FORT LAUDERDALE, FL PMSA	147	34.01	16.33	2.04	2.72	10.88	10.88	8.16	12.93	1.36	0.00	0.68
FORT WORTH-ARLINGTON, TX PMSA	305	33.77	16.39	2.30	5.57	9.18	13.11	9.51	7.21	2.62	0.33	0.00
HOUSTON, TX PMSA	544	63.24	13.05	2.57	3.49	3.13	7.72	4.41	1.65	0.74	0.00	0.00
LOS ANGELES-LONG BEACH, CA PMSA	842	72.80	2.14	0.00	0.48	0.59	2.85	2.85	7.96	5.70	3.80	0.83
MEMPHIS, TN-AR-MS MSA	190	30.53	15.26	4.21	5.26	6.32	8.95	6.84	15.79	4.74	2.11	0.00
MIAMI, FL PMSA	225	52.89	8.44	2.22	3.56	3.56	7.11	5.78	12.44	2.67	0.89	0.44
MINNEAPOLIS-ST PAUL, MN-WI MSA	619	15.67	49.43	9.37	8.24	5.17	5.98	2.91	2.58	0.65	0.00	0.00
NORFOLK-VIRGINIA BEACH-NEWPORT	315	47.30	16.51	2.54	3.81	2.86	11.43	6.35	4.13	4.13	0.95	0.00
ORLANDO, FL MSA	222	44.14	12.61	1.35	6.31	5.41	9.01	10.81	5.86	2.25	1.80	0.45
PHILADELPHIA, PA-NJ PMSA	782	61.38	10.10	0.51	1.66	3.84	8.44	4.35	6.78	2.17	0.77	0.00
PHOENIX-MESA, AZ MSA	490	20.00	26.12	8.16	8.98	9.39	12.04	9.39	5.10	0.61	0.20	0.00
RIVERSIDE-SAN BERNARDINO, CA PMSA	328	20.73	2.44	0.61	2.13	3.05	6.71	8.84	32.93	17.99	3.35	1.22
SACRAMENTO, CA PMSA	273	56.04	14.29	0.37	2.93	4.40	7.33	4.40	8.06	1.83	0.37	0.00
ST. LOUIS, MO-IL MSA	367	36.78	22.89	7.36	5.45	4.90	7.90	7.08	5.99	1.36	0.27	0.00
TAMPA-ST PETERSBURG-CLEARWATER	359	52.65	16.16	0.56	3.62	6.41	6.13	4.46	8.36	1.39	0.28	0.00
WASHINGTON, DC-MD-VA-WV, PMSA	769	43.04	19.90	2.60	5.59	5.07	8.32	6.24	6.63	1.56	0.91	0.13

comparisons against a national default rate would ignore area-wide events that cause sharp divergence in default rates across areas.

Table 3 below summarizes relative default rates for all MSAs in a single table. (A breakdown for each individual MSA is presented in Appendix B, which is available from HUD.) Here we consider not only the three definitions of default discussed above, but three additional, related definitions as well: "delinquencies at two years," which includes all claims completed and (90-day) delinquencies in progress as of two years after the amortization start date, whether or not a cure is subsequently observed; "claims at 12/95," which includes only claims as of December 31, 1995; and "delinquencies at 12/95," which includes all claims completed or delinquencies in progress as of December 31, 1995 (the default measure used by NTIC). In this table, for each of the six default definitions, the raw default rate for each tract having more than 30 loans is divided by the raw default rate for the MSA as a whole, and the resulting relative default rate is used to categorize the tract into one of five categories; tracts having 30 or fewer loans are treated in a separate category. For each default rate class and each definition of default, the body of the table gives the number and percentage of tracts falling in the class, the number and percentage of all loans contained in those tracts, and the number and percentage of defaults that are contained in those tracts. Thus, for example, the first section of the table considers the first definition of default: claims occurring by two years after amortization start. The first row of the table shows that under this definition 5,053 tracts, or 46.71 percent of the 10,818 tracts, have 30 or fewer loans in the two origination years together. The second row shows that tracts containing 30 or fewer loans contain 56,862 loans, or only 9.54 percent of the total of 596,188 loans (for which tract is identified). Finally, the third row shows that tracts with 30 or fewer loans contain 471 defaults, which amounts to only 9.5 percent of the total of 4,956 defaults.

The next set of columns to the right (*i.e.*, the fourth and fifth columns of numbers from the left) provide the same information for those tracts having more than 30 loans and a relative default rate of 0 up to 0.5 (*i.e.*, zero up to one-half of the MSA average default rate). Identifying defaults as claims within two years (the first three rows), we see that there are 3,520 such tracts, which represent 32.54 percent of the total; that these tracts contain 267,057 loans, or 44.79 percent of the total number of loans; and that these tracts contain 67 defaults, or 1.35 percent of the total number of defaults.

It is worth emphasizing that a substantial number of tracts have 30 or fewer loans (5,053 tracts, or nearly 47 percent of the total[13]), but these low-volume tracts contain very few loans in the aggregate (56,862 loans, or about 9.54 percent of the total). Thus, while it may be unsettling to have a vast number of tracts that have so few loans that they drop out of some of our comparisons, these tracts are relatively unimportant in the sense that they contain relatively few loans. In addition, these low-volume tracts contain defaults in approximately the same proportion as they do loans --- around 9.5 percent of the defaults --- thus indicating that, at least in this dimension, these low-volume tracts are similar in the aggregate to tracts with higher loan volume.

---

observable differences across loans when estimating the influence of tracts or lenders on default probabilities.

[13] Note that these counts are made after the aggregation procedure has been used to collapse many of the tracts with 30 or fewer loans.

Table 3

Number and Percentage Distribution of Tracts, Loans, and Defaults Across Default Rate Classes
1992 and 1994 Originations

Default Definition	Type	Total	<=30 Loans		Default Rate of Tract Relative to MSA Rate (Tracts with >30 Loans)									
					0 to <0.5		0.5 to <1.0		1.0 to <1.5		1.5 to <3.0		3.0 and Above	
			Number	Percent of Total	Number	Percent of Total	Number	Percent of Total	Number	Percent of Total	Number	Percent of Total	Number	Percent of Total
1. Claims at Two Years	Tracts	10818	5053	46.71%	3520	32.54%	380	3.51%	410	3.79%	807	7.46%	648	5.99%
	Loans	596188	56862	9.54	267057	44.79	72643	12.18	64755	10.86	89234	14.97	45637	7.65
	Defaults	4956	471	9.5	67	1.35	596	12.03	896	18.08	1632	32.93	1294	26.11
2. Uncured Delinquencies at Two Years	Tracts	10818	5053	46.71	2295	21.21	1163	10.75	971	8.98	993	9.18	343	3.17
	Loans	596188	56862	9.54	182068	30.54	127208	21.34	109466	18.36	97099	16.29	23485	3.94
	Defaults	14378	1420	9.88	549	3.82	2579	17.94	3794	26.39	4372	30.41	1664	11.57
3. Delinquencies at Two Years	Tracts	10818	5053	46.71	2010	18.58	1428	13.2	999	9.23	1060	9.8	268	2.48
	Loans	596188	56862	9.54	160749	26.96	148366	24.89	113712	19.07	98905	16.59	17594	2.95
	Defaults	18138	1806	9.96	776	4.28	3541	19.52	4916	27.1	5487	30.25	1612	8.89
4. Claims at 12/95	Tracts	10818	5053	46.71	3158	29.19	479	4.43	534	4.94	1010	9.34	584	5.4
	Loans	596188	56862	9.54	262220	43.98	74196	12.45	68505	11.49	95920	16.09	38485	6.46
	Defaults	6797	697	10.25	204	3	844	12.42	1168	17.18	2305	33.91	1579	23.23
5. Uncured Delinquencies at 12/95	Tracts	10818	5053	46.71	2218	20.5	1174	10.85	964	8.91	1049	9.7	360	3.33
	Loans	596188	56862	9.54	185807	31.17	135062	22.65	98467	16.52	95492	16.02	24498	4.11
	Defaults	14659	1457	9.94	667	4.55	2748	18.75	3231	22.04	4737	32.31	1819	12.41
6. Delinquencies at 12/95	Tracts	10818	5053	46.71	1992	18.41	1412	13.05	973	8.99	1106	10.22	282	2.61
	Loans	596188	56862	9.54	164006	27.51	153858	25.81	105901	17.76	97484	16.35	18077	3.03
	Defaults	18383	1856	10.1	844	4.59	3723	20.25	4287	23.32	5886	32.02	1787	9.72

Another noteworthy feature of Table 3 is that claims tend to be more highly concentrated in the highest relative default rate class than do uncured delinquencies, and delinquencies as a whole are least concentrated in the uppermost relative rate class. More specifically, 23 to 26 percent of the tracts fall in the highest relative rate class (3 or more times the MSA rate), while about 12 percent of the uncured delinquencies and only about 9 to 10 percent of all (90-day) delinquencies are in the top relative rate category. We note that alongside this pattern, default rates are lowest for claims, higher for uncured delinquencies, and highest for all delinquencies.

3.1.2. Variation Across Definitions and Across Years

Before turning to a more systematic examination of the geographic concentration of defaults, it is of interest to point out two additional features of FHA data that promise to make it difficult for anyone to identify problem tracts easily and unambiguously. Here and in the remainder of this section we adopt, without endorsing, the NTIC practice of referring to a tract as a "high-default census tract," or more simply a "high-default tract," if it exhibits a raw default rate that is at least 1.5 times the MSA average. The first point to note is that while the tables above demonstrate that defaults exhibit geographic concentration under all definitions considered here, the identities of the particular tracts selected as high-default tracts depend on the default definition chosen. Table 4 (in Appendix A) illustrates this point. (Recall again that Appendix A contains all tables referenced in the text --- Tables 1 through 27. Only selected tables, or portions thereof, are copied and inserted into the text.) The first column of Table 4 shows, for each MSA, the number of tracts appearing in both origination years and having more than 30 loans in the two origination years together. The second column shows, among these tracts, the percentage of tracts having raw default rates that are at least 1.5 times the MSA average (are "high-default tracts") under the "claims at two years" definition of default only; the third column gives the percentage that are high-default tracts under the "uncured delinquency at two years" definition only; the fourth column gives the percentage that are high-default tracts under both of the latter two measures of default; and the fifth column shows the percentage that are not high-default tracts under either definition. We see, for example, that in the Atlanta MSA, there are 335 tracts having more than 30 loans and appearing in both origination years. Among these tracts, 13.13 percent of the tracts are identified as high-default under the claims definition only, another 8.66 percent of the tracts are high-default under the uncured delinquency definition only, and 16.72 percent are high-default tracts under both definitions. Looking over the table as a whole, it is clear that the particular tracts that constitute high-default tracts depend on the definition employed, and hence the choice of definition is likely to be crucial in singling out tracts that deserve further study.

The other feature of interest is that the particular tracts isolated as high-default tracts depend on the origination year examined. Each panel of Table 5 (in Appendix A) shows, for each MSA, the percentage of tracts (among those with more than 30 loans in each year) that are high-default tracts for 1992 originations only, high-default tracts for 1994 originations only, high-default tracts for both origination years, or high-default tracts in neither origination year. For example, the first line of Panel A indicates that, when defaults are defined as claims at two years, among the 281 Atlanta tracts with more than 30 loans in each origination year, 16.37 percent are high-default tracts for 1992 originations only, 16.01 percent are high-default tracts for 1994 originations only, 7.12 percent are high-default for both origination years, and 60.5 percent are not high-default tracts for either origination year. The table shows quite clearly that the identification of a tract as a "problem" tract depends on which origination year is used. The changing identity of high-default tracts suggests that this method picks up, in part, purely transitory factors, rather than long lasting effects that would suggest more serious problems.

- **To conclude, we find that when tracts are labeled as "high-default" by comparing their default rate to that of the MSA as a whole, the identity of such high-default tracts depends on which definition of default is adopted. In addition, within a definition of default, the identity of tracts labeled as "high-default" varies with the origination year selected, a result that suggests that some of whatever causes tracts to have high default rates is transitory, and thus presumably less serious.**

3.2. Characteristics of Tracts with High Relative Default Rates

Even if one believes that chance alone is unlikely to be responsible for the observed geographic differences in raw default rates, it by no means follows that lax underwriting practices, inappropriate underwriting standards, or any other particular cause is responsible. There are numerous reasons for variation in default rates across tracts, only some of which can be properly accounted for in underwriting even under ideal circumstances. In this section we group tracts into relative default rate categories and examine how numerous default-related factors vary across these relative default rate groups.

The three panels of Table 6 use the FHA data, Decennial Census data, and HMDA data to summarize conditions in all 22 MSAs for census tracts categorized by their default rate relative to (divided by) the MSA rate over the two origination years together. Corresponding tables for each individual MSA are included in Appendix B (available from HUD). The three panels of Table 6 consider the three measures of default; Panel B is presented below while the remaining two panels may be viewed in Appendix A. The discussion here considers all three default definitions because some of the differences across definitions are noteworthy.

Among the factors listed down the left-hand side are many that could plausibly affect loan success and might potentially explain differences in default rates across tracts. The five columns on the right give relative default rate classes into which tracts of over 30 loans are categorized; the leftmost column of numbers pertains to tracts of 30 or fewer loans regardless of default rate.

The first thirteen rows are calculated directly from the FHA data. The first three rows of each panel of Table 6 repeat the information from Table 3 above, and are included solely for ease of reference. The fourth row of a panel gives the overall raw default rate for tracts falling into each relative default rate category.

3.2.1. Attributes of Borrowers and Loans in High-default Tracts

Rows 4 through 13 of Table 6 consider characteristics of tracts that are often found empirically to be related to default rates. (See, for example, Neal (1989) and Quercia and Stegman (1992) for useful reviews of the literature on loan default.) The row labeled "FHA % Black" shows

TABLE 6

CHARACTERISTICS OF TRACTS IN VARIOUS RELATIVE DEFAULT RATE CLASSES
ALL MSAs

1992 AND 1994 ORIGINATIONS

PANEL B: UNCURED DELINQUENCIES AT TWO YEARS

		Default Rate of Tract Relative to MSA Rate (> 30 Loans Per Tract)				
Characteristic	Tracts With <31 Loans	0 to < 0.5	0.5 to < 1.0	1.0 to <1.5	1.5 to <3.0	3.0+
% of All Tracts in Default Rate Class	46.71%	21.21%	10.75%	8.98%	9.18%	3.17%
% of All Loans in Default Rate Class	9.54	30.54	21.34	18.36	16.29	3.94
% of All Defaults in Default Rate Class	9.88	3.82	17.94	26.39	30.41	11.57
Default Rate (%) in Class	2.50	0.30	2.03	3.47	4.50	7.09
FHA % Black	14.01	5.64	10.16	14.68	24.37	34.40
FHA % Hispanic	14.35	9.07	14.14	15.79	12.97	10.51
First Time (%)	52.80	41.63	43.58	44.63	47.50	51.91
% LTV .97 +	22.27	22.87	22.61	24.86	31.10	39.27
% Front end .29+	20.30	17.36	20.68	20.79	18.62	15.28
% Back end .41+	16.71	16.04	17.56	17.26	16.10	12.97
Income-MSA average	132.10	122.76	90.46	-9.74	-237.47	-657.26
Mortgage-MSA average	3452.04	2102.14	3616.56	988.76	-6694.90	-19306.42
Assets-MSA average	815.92	881.43	507.72	-253.44	-1496.62	-3632.94
FHA/Tot originations (%)	11.36	23.17	29.16	34.50	37.20	44.48
Black FHA/Blk originations (%)	30.72	39.75	43.64	47.45	52.39	56.56
Hispanic FHA/Hisp originations (%)	23.03	37.46	42.32	49.78	51.64	53.14
Conventional denials/applications (%)	15.55	11.00	13.42	15.58	17.57	22.42
Census % Hispanic	16.08	10.32	15.22	14.21	13.51	8.86
Census % Black	15.14	7.01	11.22	18.58	29.86	50.28
Census % Unemp Rate (%)	7.71	5.74	6.46	7.54	9.36	14.50
Census Income Ratio	1.02	1.12	1.06	1.03	0.93	0.79
Census Poverty Rate (%)	13.33	7.37	9.15	10.98	14.99	21.73
Census Home Ownership Rate (%)	57.49	66.76	65.16	63.11	60.34	54.13

the percentage of FHA loans that go to blacks within tracts in each relative default rate class.[14] Note that this percentage rises as one moves to higher default rate classes, particularly in Panels B and C, where the uncured delinquency definitions of default are used. This pattern is consistent with a common empirical finding (which will be reestablished in the loan-level analysis below) that blacks tend to have higher default rates. The percentage of FHA loans in these tracts going to Hispanics, given in the row "FHA % Hispanic," shows a brief increase and then a decline as one moves into the higher default rate categories. The "First Time (%)" row shows the percentage of FHA loans going to first-time home buyers within the tracts in each default rate class. The tendency of this percentage to rise across default rate classes, again especially in Panels B and C, is again consistent with, and may be causally related to, the pattern of defaults. The row "% LTV .97+" shows the percentage of loans in each default rate class with loan-to-value ratios exceeding 97 percent, a level that would generally be considered high. Again, the increase in the percentage of high-LTV loans as one moves into the higher default rate categories --- more impressive in Panel B --- is entirely consistent with, and perhaps causally related to, the attendant rise in default rates.

The rows "% Front end .29 +" and "% Back end .41+" give the percentage of FHA loans with front end ratios (monthly housing expenses divided by monthly income) exceeding 29 percent and the percentage of FHA loans with back end ratios (monthly housing expenses and other debt payments divided by monthly income) exceeding 41 percent. These percentages appear to rise and then decline as one moves across the various default rate categories. The statistical analysis later in this study shows that these ratios have mixed estimated effects on default probabilities at the level of the individual loan as well.

The next three rows again show a pattern consistent with, and possibly causally related, to the pattern of default rates. The row "Income-MSA average" shows average income in the relevant tracts, expressed as a deviation from MSA average income. The row "Mortgage-MSA average" shows the average mortgage amount in these tracts, expressed as a deviation from the average mortgage amount in the MSA. The row "Assets-MSA average" shows how assets after closing (again expressed as a deviation from the MSA average) vary, on average, across tracts in each default rate class. All three of these rows display the pattern one would expect: tracts with higher relative default rates generally display lower average incomes, smaller mortgage amounts, and smaller values of assets after closing relative to MSA averages. Trends again tend to appear somewhat stronger in Panels B and C.

3.2.2. FHA vs. Conventional Lending in High-Default Tracts

The next four rows of Table 6 are obtained from tract-level HMDA data (aggregated for

[14] Note that the calculations with the FHA data across all 22 MSAs implicitly weight each MSA in accordance with its size in the relevant dimension. That is, each tract is assigned a default rate category based on its default rate relative to the MSA average (if its total loan volume exceeds 30). Data on all loans within the relative default rate category are aggregated across all MSAs, and thus larger MSAs tend implicitly to be given more weight. The alternative of performing calculations for each MSA, and then taking, say, an arithmetic average across the MSAs, seems problematic since such a procedure would entirely ignore size differences. As noted, tables for individual MSAs are presented in Appendix B, which is available from HUD.

1992 through 1996), rather than from FHA data.[15] The rows "FHA/Tot Originations (%), " "Black FHA/ Blk originations (%)," and "Hispanic FHA/ Hisp originations (%)" show the fraction of total, black, and Hispanic originations, respectively, that are made by FHA within the tracts in each of the relative default rate categories. The patterns in Panels B and C clearly show increasing FHA presence overall, as well as within each of the two minority groups, as one moves into higher default rate tracts, but these patterns are muted in Panel A.

The latter observations admit of more than one interpretation. One is that these figures demonstrate the importance of FHA in providing funding, particularly to minorities, in areas that are plagued by high default rates, which, as indicated above (and as will be reinforced below), are areas that tend to be poorer. The less charitable interpretation is that because FHA does a larger share of the originations in the higher default areas, it is somehow responsible for the higher default rates among FHA loans. Notice, however, that the conventional sector originates the majority of the loans even in the highest relative default rate category. We see in Panel B, for example, that conventional loans make up 56 percent of originations in the highest relative default rate tracts.

- **We conclude that while FHA has a larger share of the market in tracts with higher default rates, conventional lending is not driven out of tracts with high default rates. Even within high-default tracts, conventional lenders apparently find borrowers who are acceptable risks.**

The row "Conventional denials/applications (%)" shows the percentage of conventional applications that are denied within the tracts in each relative default rate category. The general rise in denial rates as one moves into higher default rate categories, which is again especially clear in Panels B and C, may indicate a general increase in riskiness of the mortgage-seeking population. This increase in riskiness among mortgage seekers may in turn be reflected in the riskiness of FHA loans, especially if those who are denied conventional loans are often accepted for FHA loans.[16] Indeed, FHA is presumably not intended simply to displace conventional borrowing, but is instead intended to extend home ownership opportunities to many who would otherwise not be able to obtain conventional funding.

3.2.3. Other Attributes of High-default Tracts

The bottom six rows of Table 6 provide additional measures of the characteristics of the various default rate groups calculated from Census data. The percentage of residents who are Hispanic (Census %Hispanic) rises and then falls as one moves into higher default tracts, but the percentage black (Census %Black) increases as one moves to higher default rate tracts, and again the trend is especially strong in Panels B and C. The unemployment rate (Census Unemp Rate)

---

[15] Calculations requiring HMDA or Census data at the census tract level require us to attach the appropriate census tract to the FHA loan data. This matching procedure fails for the approximately 15 percent of loans with nonmissing tract identifiers. Loans with unmatchable tract identifiers are excluded from analyses when their inclusion was impossible or when it was deemed likely to lead to misleading results.

[16] We do not have information on the default rate on conventional loans in these tracts.

rises, tract median income divided by MSA median income (Census Income Ratio) declines, the poverty rate (Census Poverty Rate, the percentage below the poverty line) rises, and the home ownership rate (Census Home Ownership Rate) declines as tract-level relative default rates increase. As before, trends again appear to be stronger in Panels B and C. All indications point to lower economic status in tracts characterized by higher relative default rates.

Given the tendencies displayed in Table 6, the pattern of default rates across tracts should perhaps not come as a surprise. To see if tract characteristics can fully account for differences in default rates across tracts, we shall later turn to a statistical examination of defaults on individual loans.

Before concluding this portion of the investigation, it is worth considering one feature of the data that seems to occur throughout this section and elsewhere as well. Variation in default-related factors seems more closely related to default rates that include uncured delinquencies as well as claims (uncured delinquencies at two years and uncured delinquencies at 12/95) than to default rates based on claims alone. In terms of Table 6, patterns in Panels B and C appear more pronounced than in Panel A. One interpretation of this apparent difference is that the claims only definition contains disproportionately those defaults that occur very early in the life of the mortgage, and these defaults may tend more frequently to be due to factors such as divorce, serious illness, and death that are much more weakly correlated with the measured default-related factors. That is, the measured default-related factors may better explain those defaults that occur later in the life of the loan, and these defaults are weighted more heavily in the default measures that include uncured delinquencies.

- **To conclude, we find that a wide variety of default-related factors vary across tracts classified by their relative default rates. We find, for example, that first-time homebuyers, black borrowers, and high loan-to-value ratios are all more common in tracts with higher relative default rates. At least some of the observed pattern of default rates across tracts is likely explained by the characteristics of loans within these tracts.**

3.2.4. Default Rates in Low Income and Minority Tracts

The observations in the last section may seem to imply that all low income or heavily minority tracts suffer from high default rates. This section presents some data that should serve to dispel that overly pessimistic view.

Table 7 (in Appendix A) gives the first piece of evidence. Each set of three rows presents information for all tracts in a group defined by the median family income of the tract relative to the median family income of the MSA. Originations from both years are pooled. Each row in one of these relative income groups of tracts presents the distribution of tracts, loans, or defaults across the tracts in various relative default rate classes. For example, the first row in Panel B shows that (for the indicated definition of default), among tracts having median family incomes no higher than 80 percent of the MSA median, 12.11 percent of the tracts have more than 30 loans and a default rate that is from zero to one-half of the MSA default rate. This same relative default rate category contains 19.32 percent of the loans within this tract income group, as well as 1.17 percent of the defaults. It is quite clear from the numbers in this table that many low income tracts have relatively low default rates. A bit of arithmetic will show that among tracts with more than 30 loans, 40 percent of the low income tracts (*i.e.*, those with median incomes no more than 80 percent of the MSA median) are in the lowest two default rate categories, and thus have tract default rates that are less than the average across the MSA as a whole.

Table 8 (in Appendix A) tells a similar story, but tracts are subdivided according to minority representation in the population of the tract, rather than income. For example, the first three rows show the distribution of tracts, loans, and defaults across relative default rate classes of tracts for those tracts in which minorities (blacks and Hispanics) make up zero to under ten percent of the population. It is clear from these numbers that there are indeed many tracts with substantial minority representation and yet relatively low default rates. Some arithmetic will show that among tracts with more than 30 loans and minority representation of 30 to 50 percent, about 45 percent of tracts have a default rate that is below the MSA average.

- **The message from Tables 7 and 8 is clear. Although low incomes may be associated with high default rates, there are many tracts with low median incomes or substantial minority representation that have default rates that are below the MSA average.**

3.3. Are Defaults Concentrated in Particular Lenders?

We now take up the question of whether defaults in each MSA appear to be concentrated in particular lenders, following a similar, though somewhat abbreviated, methodology as that employed above in asking whether defaults appear to be concentrated geographically. As above, this initial look at the data is essentially descriptive and nonstatistical. The three panels of Table 9 are modeled after the corresponding panels of Table 2, except that here the focus is on lenders within each MSA, rather than on tracts. (Panel B is presented below; the remaining panels of Table 9 are in Appendix A.) Table 9 displays, for all three measures of default, the variation in raw default rates across lenders originating more than 30 loans[17] within an MSA. There is apparently substantial variation in raw default rates across lenders under all three definitions.

We emphasize that our primary focus is on a lender's default *rate* in an MSA. In contrast, as noted in Section 1, the NTIC study focuses on the *volume* of defaults by a lender in the MSA, a procedure that is questionable at best. Such a method has an obvious tendency to penalize high volume lenders even if such lenders are quite conservative in underwriting and strongly supportive of delinquent borrowers.

---

[17] Each lender is distinguished solely by the existence of a unique lender identification number in the FHA data file. Note that we do not attempt to aggregate low volume lenders. In contrast to the case for census tracts, where non-FHA data are available to help in identifying similar tracts, here we have no external sources of information that would aid us in aggregation. The small volume of loans that make aggregation of a lender desirable also make for imprecise measurement of characteristics that could be used to identify similar lenders that could properly be aggregated.

TABLE 9

PERCENTAGE DISTRIBUTION OF LENDERS ACROSS DEFAULT RATE CLASSES, BY MSA

1992 AND 1994 ORIGINATIONS

PANEL B: UNCURED DELINQUENCIES AT TWO YEARS

MSA NAME	Total Lenders	<31 Loans	Raw Default Rates (Percent) for Lenders with > 30 Loans									
			0 to 0.5 %	>0.5 to 1.0 %	>1.0 to 1.5 %	>1.5 to 2 %	>2 to 3%	>3 to 4%	>4 to 7%	>7 to 10%	>10 to 15%	>15 %
ATLANTA, GA MSA	303	50.83%	12.21%	3.96%	7.26%	6.27%	8.58%	5.61%	3.96%	0.66%	0.66%	0.00%
BALTIMORE, MD PMSA	226	56.19	9.73	3.54	4.87	5.75	7.08	6.64	5.31	0.44	0.44	0.00
CHICAGO, IL PMSA	310	53.87	9.03	5.48	5.48	6.13	9.68	3.87	4.19	0.97	0.97	0.32
DALLAS, TX PMSA	252	51.19	10.32	3.57	6.75	5.56	11.51	3.17	7.14	0.79	0.00	0.00
DENVER, CO PMSA	259	44.40	27.80	11.58	6.56	4.63	3.47	0.39	0.77	0.39	0.00	0.00
DETROIT, MI PMSA	164	50.61	11.59	8.54	5.49	7.93	6.10	3.66	3.66	1.22	0.61	0.61
FORT LAUDERDALE, FL PMSA	238	66.81	7.56	0.84	2.10	4.62	5.04	5.88	4.62	1.68	0.84	0.00
FORT WORTH-ARLINGTON, TX PMSA	228	61.40	7.02	3.95	6.14	5.26	4.39	4.82	4.39	1.32	1.32	0.00
HOUSTON, TX PMSA	188	56.38	10.11	4.79	7.45	5.85	9.04	4.26	2.13	0.00	0.00	0.00
LOS ANGELES-LONG BEACH, CA PMSA	438	68.04	0.91	0.00	0.00	2.05	1.83	3.20	9.13	10.05	3.42	1.37
MEMPHIS, TN-AR-MS MSA	121	57.02	7.44	4.13	2.48	1.65	9.92	8.26	3.31	4.13	1.65	0.00
MIAMI, FL PMSA	279	59.86	9.68	1.08	3.23	2.51	7.17	8.60	5.02	2.15	0.72	0.00
MINNEAPOLIS-ST PAUL, MN-WI MSA	200	53.00	21.00	9.50	6.50	5.00	1.00	3.00	1.00	0.00	0.00	0.00
NORFOLK-VIRGINIA BEACH-NEWPORT	150	52.67	8.67	2.67	7.33	6.67	10.00	6.00	5.33	0.67	0.00	0.00
ORLANDO, FL MSA	174	57.47	5.75	3.45	6.32	1.15	9.77	6.90	6.90	1.72	0.57	0.00
PHILADELPHIA, PA-NJ PMSA	189	61.38	6.88	2.12	3.17	5.29	11.64	4.76	4.23	0.00	0.53	0.00
PHOENIX-MESA, AZ MSA	190	37.37	12.63	7.37	12.63	11.05	14.21	3.16	1.58	0.00	0.00	0.00
RIVERSIDE-SAN BERNARDINO, CA PMSA	532	70.86	0.75	0.00	0.75	0.19	1.69	3.76	11.28	6.95	3.01	0.75
SACRAMENTO, CA PMSA	240	63.33	7.92	0.42	3.75	3.75	9.17	5.42	4.17	1.67	0.42	0.00
ST. LOUIS, MO-IL MSA	121	54.55	8.26	2.48	6.61	8.26	9.92	3.31	6.61	0.00	0.00	0.00
TAMPA-ST PETERSBURG-CLEARWATER	200	59.50	10.50	3.00	4.00	3.00	11.00	4.50	3.50	1.00	0.00	0.00
WASHINGTON, DC-MD-VA-WV, PMSA	282	53.90	9.57	5.32	7.09	6.38	9.93	3.90	2.84	0.71	0.00	0.35

3.4. Characteristics of High-default Lenders

As with tracts, we adopt for convenience the term "high-default lender" to refer to a lender with a default rate in the MSA that is at least 1.5 times the MSA average. Two questions that deserve attention at the outset are (a) whether such high-default lenders operate in all areas and (b), if so, whether they have higher default rates than other lenders in all areas in which they lend or only in some areas. These questions are of interest for two reasons. First, if there is no overlap of the geographic areas in which high-default and non-high-default lenders operate, it will be difficult to separate the effects of lenders from the effects of area. Second, if we find that within each kind of geographic area, default rates do not vary across lenders, then we may want to focus all of our attention on area dispersion in default rates.

3.4.1. Does Lender Performance Vary Across Areas?

Table 10 provides some insight, though its format may be somewhat confusing and thus requires explanation. The discussion centers around Panel B, presented below; the remaining two panels (A and C) are in Appendix A. In Table 10, each lender is classified according to its relative default rate, *i.e.,* its default rate in the MSA relative to (divided by) the average default rate in the MSA as a whole. The calculations pool both origination years. The first row gives the overall raw default rate of lenders in each of the relative default rate categories. Rows 2 through 6 show, for each lender relative default rate group, the fraction of their loans, as well as the default rates on these loans, in high-default tracts with greater than 30 loans, in non-high-default tracts with greater than 30 loans, and in tracts with less than 30 loans (low volume tracts). The next six rows show similar calculations for central cities and suburban areas.

Some examples may help to fix ideas. The first row of Panel B shows that the default rate (uncured delinquencies at two years) for lenders with less than 31 loans is 3 percent, while among lenders having a default rate that is from 0 to 0.5 of the MSA average, the default rate is 0.59 percent. Looking further down the second column of numbers, we see that among lenders having a default rate that is 0 to 0.5 of the MSA average, 16.28 percent of the loans are in high-default-rate tracts (row 2); within these high-default tracts, the loans from these lenders have a default rate of 1.45 percent (row 3). An additional 76.99 percent of loans from this same group of lenders are in non-high-default rate tracts (row 4), and the default rate for these loans is 0.40 percent (row 5). For these same lenders, 6.73 percent of the loans are in low volume tracts, and the default rate for these loans is 0.68 percent. Continuing down the table, these same lenders have 26.97 percent of their loans in city tracts, and the default rate on these loans is 0.66 percent; 57.45 percent of the loans by these lenders are in suburban tracts, and the default rate on these loans is 0.56 percent.

Looking at Table 10, we see that there appears to be some tendency for lenders in higher relative default rate classes to have a greater share of their business in high-default rate tracts and in central city areas. Within a lender default rate class, default rates in high-default rate tracts are always higher than in the non-high-default rate class, and higher in central cities than in suburban areas. Within high-default rate tracts and within non-high-default rate tracts, as well as within tracts classified by city/suburban status, default rates tend generally to rise as one moves along a row to the right, *i.e.,* into higher lender default rate classes.

TABLE 10

PERCENTAGE DISTRIBUTION AND OTHER CHARACTERISTICS OF LOANS IN EACH LENDER RELATIVE DEFAULT RATE CLASS ACROSS TRACTS CLASSIFIED BY RELATIVE DEFAULT RATE AND BY CENTRAL CITY/SUBURBAN STATUS
ALL MSAs

1992 AND 1994 ORIGINATIONS

PANEL B: UNCURED DELINQUENCIES AT TWO YEARS

	Lenders With	Default Rate of Lender Relative to MSA Rate (Lenders with > 30 Loans)				
	<31 Loans	0 to < 0.5	0.5 to < 1.0	1.0 to <1.5	1.5 to <3.0	3.0+
% Default Rate	3.00%	0.59%	1.86%	3.16%	4.27%	8.15%
% of Loans in High Default Tracts	30.66	16.28	20.79	26.30	24.93	33.61
% Default Rate in High Default Tracts	5.69	1.45	4.43	6.80	8.68	14.83
% of Loans in Non-High Default Tracts	62.41	76.99	73.87	68.79	70.32	62.01
% Default Rate in Non-High Default Tracts	1.69	0.40	4.02	1.78	2.68	4.34
% of Loans in Low Volume Tracts	6.93	6.73	5.34	4.91	4.75	4.38
% Default Rate in Low Volume Tracts	2.91	0.68	1.90	2.95	4.69	10.88
% of Loans in City Tracts	28.42	26.97	30.60	32.58	31.13	47.59
% Default Rate in City Tracts	3.58	0.66	2.14	3.40	4.84	10.04
% of Loans in Suburban Tracts	58.95	57.45	53.98	53.09	55.92	42.69
% Default Rate in Suburban Tracts	2.71	0.56	1.73	3.02	3.95	6.06
% of Loans in unknown City/Suburban Tracts	12.63	15.59	15.42	14.33	12.95	9.72
% Default Rate in unknown City/Suburban Tracts	3.07	0.58	1.74	3.14	4.30	8.08
% of Borrowers in High Default Tracts Who are Black	20.79	21.66	21.04	23.53	33.95	53.03
% of Borrowers in Non-High Default Tracts Who are Black	11.53	7.61	8.55	10.56	14.48	21.82
% of Borrowers in Low Volume Tracts Who are Black	11.49	9.11	11.20	13.03	18.48	36.18

These observations admit of more than one interpretation. One possibility, of course, is that high-default lenders fail to follow underwriting standards, or they pursue foreclosure too aggressively. For these reasons, their default rates are higher than those of other lenders, even after controlling crudely for the default rate of the tract.

A second possibility is that our split of tracts into high-default and non-high-default rate tracts is too coarse, and a more detailed breakdown would show that all lenders have identical default rates in properly defined, homogeneous tracts. That is, the existing two-way categorization of tracts surely leaves much variation among tracts within each category. If each of these two categories could be subdivided into more homogeneous categories, we might find that lenders have the same default rate within each of these more narrowly defined groups. Under this explanation, then, all lenders act essentially the same within a homogeneous area, and differences at higher levels of aggregation are traceable to underlying heterogeneity of areas.

A related interpretation of the findings in Table 10 is that lenders specialize in different types of borrowers. Perhaps lenders are located in different areas and there are informational efficiencies in tailoring lending practices to the kinds of borrowers most frequently encountered in their local markets. Alternatively, different lenders may specialize in different kinds of borrowers because market efficiencies dictate such a structure even if all lenders have identical access to all kinds of borrowers. Under this interpretation, even if we were to isolate more homogeneous areas, we might find that differences in default rates across lenders remain because of differences in their clientele within an area. By this explanation, the patterns in Table 10 may suggest that lenders with higher default rates tend to specialize in higher risk borrowers, though not necessarily those with unacceptably high risk.

The final three rows in each panel of Table 10 provide a bit of additional information along the latter lines. Each of these three rows gives, for the indicated kind of tract, the fraction of the lenders' borrowers who are black. Thus, for example, the second entry from the left in the third row from the bottom in Panel B states that, among lenders with relative default rates that are 0 to 0.5 times the MSA average, 21.66 percent of their borrowers in high-default rate tracts are black, but as shown in the row immediately below, 7.61 percent of their borrowers in non-high-default rate tracts are black. Quite clearly, within each of the tract default rate categories (*i.e.,* along a row), the fraction of black borrowers tends to rise as one moves into the higher default rate classes of lenders. The rise is especially dramatic in Panels B and C, again suggesting that systematic factors play a bigger role in determining uncured delinquencies than in determining claims at two years. It appears that even within an area classified by its relative default rate, lenders with high relative default rates have very different lending patterns than do non-high-default rate lenders. It is possible that such differences in clientele may account for their differing default rate experience. The material in the next section and the more detailed statistical analysis in Section 5 will provide some additional information on this issue.

3.4.2. Attributes of Borrowers, Loans, and the Population Served by High-default Lenders

To characterize more fully the type of borrowers serviced by lenders in the various relative default rate categories, we turn to Table 11, which provides a more detailed characterization of borrowers in each relative default rate group of lenders, like that provided in Table 6 for tracts in

TABLE 11

CHARACTERISTICS OF LENDERS IN VARIOUS RELATIVE DEFAULT RATE CLASSES
ALL MSAs

1992 AND 1994 ORIGINATIONS

PANEL B: UNCURED DELINQUENCIES AT TWO YEARS

Characteristic	Lenders With <31 Loans	Default Rate of Lender Relative to MSA Rate (> 30 Loans Per Lender) 0 to < 0.5	0.5 to < 1.0	1.0 to <1.5	1.5 to <3.0	3.0+
% of All Lenders in Default Rate Class	60.84%	11.81%	10.92%	8.33%	6.69%	1.41%
% of All Loans in Default Rate Class	4.32	17.63	37.64	26.78	12.33	1.30
% of All Defaults in Default Rate Class	5.38	4.31	29.00	35.08	21.84	4.40
Default Rate (%) in Class	3.00	0.59	1.86	3.16	4.27	8.15
FHA % Black	14.36	10.00	11.29	14.09	19.53	32.94
FHA % Hispanic	18.23	9.37	11.44	15.19	13.08	11.68
First Time (%)	39.98	39.69	43.35	48.30	50.88	59.81
% LTV .97 +	25.35	21.64	24.50	24.66	30.53	39.28
% Front end .29+	23.61	16.77	19.40	20.12	18.26	13.57
% Back end .41+	17.86	16.40	16.38	17.10	16.00	12.28
Income-MSA average	173.48	53.16	53.73	-18.56	-163.25	-450.65
Mortgage-MSA average	5083.77	-1383.87	1709.29	1162.72	-5244.91	-14422.89
Assets-MSA average	813.15	561.81	341.68	-144.99	-1090.71	-3286.62
FHA/Tot originations (%)	35.50	33.70	36.74	38.52	37.71	45.29
Black FHA/Blk originations (%)	39.64	37.68	40.12	41.84	41.44	48.28
Hispanic FHA/Hisp originations (%)	41.33	38.77	41.56	63.94	41.36	43.44
Conventional denials/applications (%)	17.93	14.97	15.47	16.40	17.19	19.62
Census % Black	13.60	11.06	11.87	13.16	16.95	25.85
Census % Hispanic	14.91	9.43	9.85	11.64	10.43	9.43
Census Unemp Rate (%)	7.89	6.71	7.05	7.41	7.44	10.83
Census Income Ratio	1.03	1.05	1.04	1.03	1.00	0.92
Census Poverty Rate (%)	9.67	7.90	8.30	9.01	9.20	12.13
Census Home Ownership Rate (%)	67.11	69.17	68.78	67.76	67.81	67.17

various relative default rate groups.[18] Panel B of Table 11 is presented below, and the remaining two panels (A and C) may be viewed in Appendix A. (MSA-level versions of Table 11 are presented in Appendix B, which is available from HUD.) Calculations underlying Table 11 are similar to those in Table 6, except in the case of calculations from Census or HMDA data. Whereas in Table 6 we are able to aggregate census tracts together to obtain tract-level measures from the HMDA or Census files, for the purposes of Table 11 we would need the equivalent measures at the lender level, rather than the census tract level. Because we do not have such lender-level data, we instead produce estimates by calculating weighted averages of the tract-level Census and HMDA data. The weights are the fraction of loans of each lender type that fall within each tract. Hence, in effect we produce measures that show "lender exposure" to the tract-level variables.

Notice that the patterns observed in Table 11 are qualitatively very similar to those seen in Table 6, though there are a few exceptions.[19] As might be expected, causal factors generally exhibit trends across lender groups classified by their relative default rates similar to the pattern across tract groups classified by tract relative default rates. We again reach the sensible conclusion that some of the pattern in default rates across lenders may arise because borrower, loan, and neighborhood characteristics lead to the observed differences in default rates. The more detailed statistical procedures in Section 5 will shed additional light on this issue.

- **To conclude, we find that characteristics of loans differ sharply across lenders classified by their relative default rates. It would not be surprising to find that these differences in loan composition account for at least some of the differences in default rates across lenders.**

3.5. Conclusions

The raw data show quite clearly that default rates vary substantially across tracts and lenders, both in raw terms and relative to the MSA average. As emphasized above, however, these simple comparisons cannot tell us whether defaults are so concentrated in tracts or lenders that randomness is unlikely to be the explanation. If systematic factors, rather than randomness, lie behind the variation in default rates across tracts and lenders, we have already provided suggestions about what those systematic factors might be. The summaries that report on the characteristics of borrowers and loans in high-default tracts and lenders show that a variety of default-related attributes --- first-time home ownership and high loan-to-value ratios, for example --- vary in a sensible way across tracts and lenders. If systematic factors are responsible for disparities in default rates across tracts and lenders, we may not have to look far for at least some of the candidates.

---

[18] Lenders with no more than 30 loans, while common, originate a very small percentage of loans (4.32 percent). Moreover, the fraction of all defaults accounted for by these low-volume lenders is similar to the fraction of loans, particularly for the claims-at-two-years and the uncured-delinquencies-at-12/95 definitions. In this sense, the small-volume lenders appear similar, on average, to the high-volume lenders.

[19] We see, for example, that there is little variation in home ownership rates (the bottom row of each panel) across lender default rate categories. In Panels A and C, we find little variation in the "exposure" to the black population (Census % Black) across categories.