Analysis of FHA Single-Family Default and Loss Rates

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

Contract Number DU100C000018484

March 2004

Acknowledgments

I am grateful to members of the HUD staff who have graciously provided information and guidance on numerous aspects of this project. I am especially indebted to Harold Bunce, William Reeder, Sue Neal, Randy Scheessele, and Daryll Getter for many helpful discussions that clarified the issues and to Judy May and Dominick Stasulli for useful insights on the nature of the foreclosure process, FHA loan loss data, and accounting procedures. I take responsibility for any errors.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Previous studies of mortgage risk in both the conventional and FHA sectors have focused almost exclusively on default behavior and on the factors that lead to default. This is the approach taken in numerous articles in the professional economics and finance literature, as well as in nonacademic studies produced by practitioners within the industry. In addition, recent extensions on FHA mortgage scoring have followed the main lines of previous research in focusing solely on the default probability as a metric for risk. In virtually all of this extensive research virtually no attention is given to other dimensions of loss and to the dollar value of losses in particular; thus, little is known about dollar loss and its determinants.

This focus on default in the mortgage scoring context means that observable factors affecting the likelihood of default assume a primary role. Because minorities tend to have less attractive distributions of factors leading to default, mortgage scoring systems tend to give minorities less favorable scores than nonminorities, justifying such patterns with well-reasoned arguments of business necessity. Some observers, understandably concerned by this racial discrepancy in scoring outcomes, have suggested that minorities generate smaller dollar losses on average when they default, and thus a mortgage scoring system relying on dollar losses rather than default alone might improve minorities' lot. In addition, a mortgage scoring system that recognizes both the probability of default and the dollar losses attendant upon default would provide a more complete, and thus superior, measure of risk that could be used for policy decisions as well as for underwriting.

The purpose of this paper is to use data on FHA-insured loans from 1992, 1994, and 1996 to examine the factors that influence both default probabilities and dollar loss rates, as well as the avenue by which impacts arise. En route we pay special attention to the possibility that minorities would fare better with a scorecard based in part on dollar losses. The analysis ranges from simple statistical summaries and descriptive regressions to more complete and sophisticated statistical analysis.

The simple statistical summaries examined first indicate that dollar loss rates are lower for loans that take longer to default, partly because of declines in the gap between unpaid principal balance and the price received in property disposition. Similar analysis shows that loss rates tend to rise with the amount of time spent in foreclosure and property disposition. Comparisons of means show that loss rates are higher for blacks than for whites, both overall and within each of several components of the loss rate.

More sophisticated statistical analysis of three-year defaults suggests that increases in the front-end ratio, LTV, the note rate, and borrower incomes are associated with increases in loss rates. Increases in FICO scores, mortgage payments held in reserve, loan amounts, house price growth, relative house prices, and tract incomes are associated with lower loss rates. Blacks, Hispanics, and those in judicial foreclosure states and in underserved areas have higher loss rates, other things the same.

The analysis shows that although there is a link between factors affecting default probabilities and those affecting loss rates, there are differences in the relative importance of factors affecting each. For example, by a couple of different measures, FICO scores appear to be more important in determining default behavior; house price growth and relative house price appear to be more important in affecting loss rates.

More sophisticated statistical analysis reinforces the findings from simple data summaries that differences in the timing of default-related events—the time from origination until default, the time spent in foreclosure processing, and the time spent in property disposition—are critical in determining loss rates. Moreover, the evidence as a whole suggests that the importance of at least some explanatory factors, such as FICO scores, appears to stem primarily from their effects on durations. Higher FICO scores are associated with lengthening the time prior to default and reducing time spent in foreclosure and property disposition. Similarly, blacks have higher loss rates that may be traceable mainly to behavioral differences in timing, but perhaps also to differences in some loss components, given timing.

A comparison of applicants in 1992 with those in 1996 suggests that default probabilities rose, loss rates per default fell, and loss rates per loan rose across these cohorts. The estimated models suggest that declines in FICO scores were the major contributor to the increase in default rates, while increases in house price growth were the major contributor to declines in loss rates among defaults.

Finally, the evidence, though highly tentative, suggests that basing underwriting on the expected loss rate per loan, rather than on default probabilities, will not substantially improve the lot of black applicants. Ranking risks according to the expected loss rate per loan results in even lower representation of blacks in the low risk category, and higher representation of blacks in the high-risk category, when compared with ranking risk by estimated default probabilities alone.

SECTION 1

INTRODUCTION

1.1. Background

In recent years there have been extensive research efforts into the nature of FHA mortgage defaults. Two studies by Cotterman, for example, have examined the extent to which the likelihood of default of FHA-insured loans is traceable to the separate influence of locational factors and borrower characteristics, including past credit performance.[1] Such studies may be usefully applied to the evaluation of underwriting guidelines, the assessment and management of risk, the calculation of insurance premiums, and the formulation of loss mitigation strategies.

HUD realizes, however, that the occurrence of default is only one dimension of loss and that a complete picture demands that the severity of loss be considered as well. In particular, defaults resulting in minimal losses are a less serious problem than defaults resulting in large losses, and policies and programs that take expected default rates into account should properly account for the differential losses among these defaults. Thus, risk management, loss mitigation strategies, and underwriting guidelines should recognize loss severity, not simply expected default probabilities. The factors underlying loss severity among FHA loans have, however, received much less systematic attention than have the factors underlying the occurrence of default.[2]

The purpose of this paper is to begin to redress this imbalance. The empirical work in this paper will provide estimates of the effect of various factors on loss severity, and in particular, the separate impact of locational factors and borrower characteristics. Three specific aspects of this research are of special interest. The first is the possibility that factors affect the probability of default in a very different way than they affect loss severity. In particular, locational and other factors beyond the borrower's control may be much more important in determining loss severity than are borrower characteristics. Such findings would suggest that portfolio risk assessments that focus solely on the influence of default-related factors could be substantially improved by bringing in locational characteristics.

The second and related aspect of special interest is the possibility that loss severity is associated with race in a very different way than is the probability of default. In particular, default rates are often found to be higher for blacks than for whites, and underwriting systems based on default behavior tend to accept lower fractions of black applicants, with the justification resting on business necessity. It has been suggested, however, that loan loss rates may be lower for blacks. If so, underwriting systems giving appropriate weight to loss severity would be more favorable for blacks than are default-based underwriting systems. This paper will attempt to assess these possibilities.

---

[1] See Robert F. Cotterman, *Neighborhood Effects in Mortgage Default Risk*, Report Prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research (March 2001), and Robert F. Cotterman, *Assessing Problems of Default in Local Mortgage Markets*, Report Prepared for the U.S. Department of Housing and Urban Development, Office of Policy Development and Research (March 2001).

[2] Loss rate regressions have occasionally appeared as part of a larger study of FHA lending but have rarely been the focus. See, for example, James A. Berkovec, Glenn B. Canner, Stuart A. Gabriel, and Timothy H Hannan, "Discrimination, Competition, and Loan Performance in FHA Mortgage Lending, *Review of Economics and Statistics 80,* 241-250.

A third area of interest is in identifying the way in which various elements influence loss rates, for such knowledge may be helpful in guiding policy. If, for example, the evidence suggests that the major contributor to loss is time spent in property disposition, one might search for methods to streamline that process. We offer several pieces of indirect evidence along these lines.

1.2. A Roadmap for the Remainder of the Paper

The analyses in this paper are designed to isolate the factors that influence loss rates and, if possible, identify the route by which their influence is felt. To that end, we use data on a sample of FHA-insured loans from three application years, first examining summary statistics and then turning to a more sophisticated statistical analysis. The plan for the remainder of the paper is as follows.

Section 2.1 begins with a brief clarification of the distinction between conditional and unconditional loss rates. Section 2.2 then moves to an overview and brief statistical summary of the default data and the loan loss data that underlie the statistical presentation. Included here is information on default rates at three, five, and seven years by racial or ethnic group; the distribution of defaults by type of claim; average loss rates by timing of default, by timing of property disposition, and by loan size; and a decomposition of the loss rate into its components. Various breakdowns of loss rates by race/ethnicity and for low-income borrowers are presented as well.

Section 3 discusses a variety of borrower attributes and characteristics of the loan, the housing market, and the geographic area that might be expected to affect default behavior and loss rates. Possible links between factors affecting default and those affecting loan losses are discussed as well.

Section 4 presents and discusses empirical estimates of the effects of a series of explanatory factors on the probability of default and on conditional and unconditional loss rates. Three methods are used to illustrate the relative importance of the factors. First, we calculate the changes in default probabilities and loss rates that would result from a variety of hypothetical changes to the explanatory variables. Second, we order our sample according to estimated risk and calculate the mean value for each explanatory variable within high-risk and low-risk segments of the borrower population. Third, we use the observed changes in explanatory factors between 1992 and 1996 to calculate the implied changes in default probabilities and loss rates over this interval; we then compare these to the actual changes over this same period.

Section 5 closes with some conclusions.

An appendix provides the major statistical estimates that are summarized in the main body of the paper, offers some graphical displays of the effect of various factors, and gives additional details on the estimation methodology.

SECTION 2

PRELIMINARIES

2.1. A Brief Overview of the Conceptual Framework

To begin, we clarify some terminology. First, we define the “loss rate” for a defaulted loan as the dollar loss divided by the initial loan amount. The loss rate thus records the number of dollars of lost per dollar lent. This traditional way of quoting loss rates is a particularly convenient metric for comparisons with mortgage insurance rates, which are also expressed in terms of premium dollars per dollar borrowed. We return to this point below.

Two kinds of loss rates are of interest—“conditional” loss rates and “unconditional” loss rates. Conditional loss rates are calculated over only those loans that default and in this way “condition” on the occurrence of default. The traditional method of calculating loss rates within the set of loans that have defaulted thus yields a conditional loss rate. In what follows, we shall often refer to this traditional calculation of a conditional loss rate as simply “the loss rate.” In contrast, the “unconditional” loss rate is calculated over all loans, both those that default and those that do not default, where nondefaulting loans are assumed to exhibit losses of zero. The calculation of the unconditional loss rate is the same as the calculation of the conditional loss rate except that nondefaulting loans, with a loss rate of zero, are included in the unconditional loss rate calculation. The conditional loss rate (CLR) and unconditional loss rate (ULR) are thus related by

$$ULR = CLR * DR, \tag{1}$$

where DR is the default rate.[3] If, for example, the default rate were 5 percent and the loss rate among those defaulting (*i.e.,* the conditional loss rate) were 60 percent, then the unconditional loss rate would be 3 percent (0.05 x 0.6 = 0.03). That is, even though losses are, on average, sixty cents on the dollar among defaulting loans, losses are on average only 3 cents on the dollar among all loans.

The unconditional loss rate is clearly a useful measure for many purposes. In valuing a portfolio of loans, the expected unconditional loss rate would be a useful piece of information. In deciding whether to underwrite a loan, one might compare the expected unconditional loss rate with the insurance rate.[4] Moreover, in both of these contexts, there is clearly an advantage in knowing the expected unconditional loss rate rather than simply the expected default rate.

---

[3] The latter holds for samples of any size. For purposes of statistical estimation, however, it is useful to think in terms of expected values in the population rather than sample averages. In particular, the expected loss rate E(L) for a loan selected at random from the population of all loans may be expressed as the product of two terms: (1) the probability that the loan will default Pr(D = 1), and (2) the expected loss rate L given that the loan defaults E(L|D = 1), or

$$E(L) = Pr(D = 1)\ E(L|D = 1)$$

where D is a default indicator. The purpose of the empirical work is to estimate the influence of various factors upon these expected values.

[4] If not all insurance is paid up front, there is uncertainty over future payments, and the appropriate comparison is to expected insurance rates. A more complete comparison would consider all expected gains and losses, including those arising probabilistically from prepayment.

In principle, losses from default may occur over the full term of a loan. In practice, the data used here permit us to look at only a relatively narrow time band following loan origination. As explained in somewhat more detail below, given the tradeoffs between the desire to use data on relatively recent loans and the conflicting desire to observe each loan for a long period of time, our main focus is on defaults occurring within the first three years of loan duration. Because default is much more heavily concentrated in the early loan years, however, a focus on the early years of loan duration may not be too misleading. That is, the hope is that default activity within three years of loan origination will be a good guide to what would be found over a longer post-origination horizon.

Before embarking on a systematic empirical exercise to estimate the effects of factors underlying loss rates, we examine the data that will form the basis for the empirical work. Simple statistical summaries will provide useful background for the more detailed and sophisticated work to follow.

2.2. Overview of the Data

The samples of FHA-insured loans used as the basis for estimation in this study were selected originally to support projects that examined the occurrence of default but not on the magnitude of the loss. The nature of these earlier studies motivated various aspects of the sampling scheme. Here we summarize the salient features of the sample.

Sample loans have been drawn from three application years: 1992, 1994, and 1996. In each case, the universe from which the samples were drawn consisted of endorsed loans that were not part of the FHA streamline refinance program.[5] For application years 1992 and 1994, samples were stratified according to claim status in May 1997; claims were oversampled. For the current project we have developed weights that account for these differential sampling probabilities, as well as for the different rates at which observations have been eliminated because various data elements are missing.[6] All statistical analyses presented in this paper use these weights.

The status of the sample loans has been tracked through March 2001. Because defaults resulting in a claim are not recorded as such until the claim is paid, and because there is a long and variable lag from the time of default until the payment of a claim, only those defaults that occurred well before the data files were made will appear as defaults in the data. In addition, because explorations have shown that the elapsed time from default to payment of the claim seems to vary with some of the same factors that influence default probabilities, there is some concern that including more recent defaults, which are sure to be more incomplete, will bias estimates of substantive parameters. We have therefore chosen to identify as defaults only those defaults that occurred on or before June 1999 (and which led to a claim).

The choice of a default horizon for the detailed empirical work in subsequent sections is dictated by the nature of the sample. There is a tradeoff in choosing the length of the default

---

[5] The automated records for streamline refinances are missing a crucial piece of information: the loan-to-value ratio.

[6] Weights have been calculated so that the sample cross tabulation of race (white, black, Hispanic, other) by initial claim status (claim or not as of May 1997) by final loan status (claim, prepaid, or active as of March 2001) matches that of the universe for each application year. For application year 1996, the cross classification by initial claim status is left out of the weight calculation because it was omitted from the sampling scheme.

horizon: using more recent application data so as to give the analysis greater currency implies using shorter default horizons. In particular, if we wish to use the most recent (1996) application data and to allow, say, 21 months for defaults to complete the claim process, we may obtain sufficient sample sizes for a default window of three years. Thus, the default model here examines those defaults occurring within three years of loan origination. Given that defaults must occur by June 1999, and given that we require that a loan have a full three years of exposure to the possibility of default, only those 1996 applications that originate by June 1996 are eligible for inclusion in the estimation sample used for the detailed empirical work in Section 4. For the 1992 and 1994 applications, the three-year restrictions on timing of default and length of exposure have virtually no impact on the set of applications included in the analysis, but they do affect which loans are classified as defaults in the detailed statistical work.

Although the empirical work below utilizes data from a variety of sources, most variables have been drawn from FHA files. FHA's A43 and CHUMS systems have supplied information on the nature of the loan and the borrower, as well as information on the evolving default status of the loan. More detailed information on the dollar losses arising from claims have been obtained from the SAMS and A43C systems. Finally, in the course of an earlier effort, FICO scores were obtained from Equifax and Trans Union and were attached to files from which all individual identifying information had been removed.[7] These scores are used for the current project as well. We now describe selected characteristics of the FHA data, with an emphasis on the loan loss data that are the focus of this study.

2.3. Descriptive Statistics

We first provide a brief summary of the default status of a sample of 240,901 loans[8] that are used for the descriptive work in this section. Table 1 shows, for each application year and for each racial/ethnic group, as well as for low-income borrowers,[9] the default rates for three years, five years, and seven years following loan origination. For each calculation, we include only those loans that originate early enough that there is a full three, five, or seven years, respectively, of elapsed time after origination and prior to July 1999. That is, to calculate the default rate at N years, we require that each loan originate early enough that it has potential exposure to default of N years by the time the default observation window effectively closes at the end of June 1999.

[7] The FHA data and the FICO scores are maintained separately and in anonymous form; at estimation they are linked via a randomly assigned case number.

[8] This sample is composed of 78,888 loans from 1992 applications, 71,115 loans from 1994 applications, and 90,898 loans from 1996 applications.

[9] Low-income borrowers are defined as those whose incomes are less than 80 percent of the county/MSA median.

Table 1

Default Rates (%) at 3 Years, 5 Years, and 7 Years,
by Race/Ethnicity, Low Income Status, and Application Year

Application Year	Race	3-Year	5-Year	7-Year
1992	White	2.03	3.54	4.27
	Black	4.91	8.32	10.81
	Hispanic	4.41	9.62	13.18
	Other	3.03	5.79	6.29
	Low Income	2.82	4.89	5.86
	Overall	2.53	4.57	5.47
1994	White	3.07	4.10	
	Black	6.77	9.14	
	Hispanic	6.72	9.47	
	Other	4.54	5.80	
	Low Income	4.35	6.01	
	Overall	4.03	5.26	
1996	White	3.34		
	Black	6.93		
	Hispanic	6.99		
	Other	5.40		
	Low Income	5.12		
	Overall	4.39		

Table 1 shows the anticipated increase in the default rate as the observation window lengthens: 7-year rates exceed the corresponding 5-year rates, which in turn exceed the corresponding 3-year rates. Looking across application years, we see a tendency for rates at a given horizon to increase for more recent applications. That is, the rates for 1996 applicants tend to exceed the corresponding rates for 1994 applicants, which in turn exceed those for the 1992 applicants. Default rates for Hispanics and blacks tend to be higher than for the other groups. Reasons for these patterns are, of course, unclear, though the especially high rates for Hispanics, particularly in the 1992 applications, are presumably traceable in part to the downturn in the California housing market in the 1990s.

2.3.1. Loss Rates

We now turn to the loan loss data, which are a major focus of the current study. Although we will eventually restrict attention to losses generated by defaults within three years of origination, the statistical summaries in the remainder of Section 2 generally use the full range of defaults available in the sample of 1992, 1994, and 1996 applications, regardless of the timing of default.[10] Loan losses in principle arise from a variety of types of claim. Table 2 shows the percentage distribution of claim types across categories for each calendar year of default from 1993 through 2000. The vast majority of claims are conveyance claims, presented in the first column. The second column shows the claims arising from loans that went into the assignment program. This program received approximately one-quarter of the claims prior to 1996, but nothing in recent

[10] The samples used to analyze loss rates in Section 2 generally consist of about 5,765 loans (defaults) from 1992 applications, 3,969 loans from 1994 applications, and 2,164 loans from 1996 applications.

years following the termination of the program. Finally, nonconveyance claims --- which include preforeclosure sales and the like --- tend to exhibit increasing use, in part a reflection of recent FHA efforts at loss mitigation.

Table 2

Percentage Distribution of Type of Claim by Year of Default

	Type of Claim			
Year of Default	Conveyance	Assignment	Nonconveyance	Total
1993	72.49	25.16	2.35	100.00
1994	72.26	24.99	2.76	100.00
1995	67.17	27.44	5.39	100.00
1996	86.98	4.77	8.25	100.00
1997	93.48	0.00	6.52	100.00
1998	92.15	0.00	7.85	100.00
1999	93.24	0.00	6.76	100.00
2000	89.36	0.00	10.64	100.00

Although loan losses could in principle be attached to any of these kinds of claims, the loss data utilized here focus on only one kind: conveyance claims. Loan loss data are present in the files for a very few assigned loans, and the infrequency with which these were observed led to questions about their representativeness. Thus, these few loan losses were excluded from the loan loss sample used in the analytic work to follow. Loan loss data on the remaining (nonconveyance) claims are unavailable, and these claims were thus eliminated from consideration.

Note that the focus on conveyance claims limits the generality of the empirical work and introduces a potential bias. Assuming that nonconveyance claims tend, on average, to yield lower loss rates, statistical models based on conveyance claims alone, such as those developed in this paper, will tend to overstate expected loss rates. Moreover, if nonconveyance claims as a share of all claims continue to rise as they have in the recent past, the overstatement bias will worsen over time.

The loss rates examined in this section are conditional loss rates, as defined above: the loan loss divided by the initial loan amount.[11] Loan losses in turn are defined as in the FHA data files: the undiscounted sum of all cash flows attributable to the default, including the costs of paying the claim, foreclosure expenses, costs of holding and selling the property, and the return from selling the property.[12]

Table 3 shows the mean value of the loss rate (and the raw count of sample loans upon which the mean is calculated) by application year and by timing of default. Separate columns record corresponding results for blacks, Hispanics, and low-income borrowers. We notice that loss rates tend to be higher for blacks and for low-income borrowers within an application year and

[11] The initial loan amount is defined to exclude up-front MIP.

[12] Rarely, (about 0.6 percent of the losses in the 1996 data, 0.9 percent in the 1994 data, and 0.7 percent in the 1992 data) losses are negative; that is, the default results in a gain.

time-to-default category. Within each application year and borrower group, we see a tendency for loss rates to decline when the time to default increases. In the 1992 loans for all borrowers (the "Total" column), for example, we see that the mean loss rate of 48.22 percent for defaults occurring within three years of origination declines to 46.36 percent for defaults occurring from three to five years of origination and then to 43.46 percent for defaults occurring from five to seven years after origination. Presumably, one factor in this pattern is the reduction in unpaid principal balance for loans that remain active for longer periods of time before defaulting. Notice finally that there also appears to be some tendency for loss rates to be smaller in more recent application years, given the timing of default.

Table 3

Mean Loss Rate (%) and Raw Counts, by Elapsed Time from Loan Origination to Default,

by Application Year, Race/Ethnicity, and Low Income Status

	Yrs to Def	Total		Black		Hispanic		Low Income	
		rate	count	rate	count	rate	count	rate	count
92	<=3	48.22	3793	55.81	525	52.44	547	49.50	1423
	>3 & <=5	46.36	1629	49.78	193	48.23	332	47.39	579
	>5 & <=7	43.46	343	49.71	41	42.36	60	48.21	117
	Total	46.84	5765	52.92	759	48.97	939	48.52	2119
94	<=3	44.15	3336	50.42	559	44.20	601	47.49	1420
	>3 & <=5	36.80	633	41.79	99	33.78	111	41.28	261
	Total	42.21	3969	48.48	658	41.34	712	45.84	1681
96	<=3	42.44	2164	48.71	349	38.65	413	47.37	1046
	Total	42.44	2164	48.71	349	38.65	413	47.37	1046

Table 4 presents an alternative breakdown by showing, for a given number of years from loan origination to default, how loss rates vary with the number of years from origination to property disposition. We see, for example, that loss rates among loans that defaulted within the first year following origination and for which property disposition occurred during the second year following loan origination were 38.04 percent on average; this average was calculated over a sample of 956 loans. Looking across each row, we notice a tendency for average loss rates to rise as property disposition occurs later and later. That is, holding constant the number of years from loan origination to default, we see that delaying property disposition seems to result in larger loss rates on average.[13] This delay could result from delays in foreclosure, which might be reflected in additional costs of foreclosure and additional interest due the lender, as well as additional time spent in the property disposition process, which could mean additional holding costs. Below we present additional information along these lines.

[13] This phenomenon appears to hold for a given application year as well.

Table 4

Mean Loss Rate (%) and Raw Counts (shaded), by Years to Property Disposition and Years to Default

	Years to Property Disposition								
Years to Default	1	2	3	4	5	6	7	8	9
1	28.31	38.04	46.65	59.45	68.79	81.31		82.45	
	13	956	1200	300	47	12		1	
2		31.87	36.55	47.52	59.10	77.37	66.70	114.38	
		124	2300	1607	295	53	6	4	
3			29.80	36.96	44.92	56.63	70.81	58.55	
			113	1722	1051	187	22	7	
4				26.61	37.06	44.49	62.05	62.79	
				58	952	459	82	17	
5					31.36	36.91	46.02	57.68	55.80
					25	311	215	48	1
6						34.12	37.80	49.73	66.82
						5	142	127	2
7							110.66	40.76	17.92
							5	30	2

Again using Table 4, looking along a diagonal down and to the right holds constant (approximately) the number of years from default to property disposition as the number of years from origination to default increases. Thus, for example, if we start with default occurring within the first year after origination and property disposition within the second year after origination, and then move down one cell and to the right one cell, we arrive at the figure for default within the second year after origination and property disposition within the third year after origination. In both cells, property disposition occurs within (approximately) one year of default. Looking down the diagonals in this fashion reveals a rather mixed picture. In particular, the declines in average loss rates for defaults that occur at higher durations do not appear as dramatic, if indeed they appear at all, when compared with what was seen in Table 3. A major source of this difference is that the figures in Table 4 are aggregated across borrower cohorts, and more recent borrower cohorts have lower loss rates. Because more recent cohorts are more heavily represented in the lower years-to-default categories, they introduce distortion. Declines in loss rates associated with later times to default would more clearly emerge in Table 4 if we were to separate borrowers by year of application.[14]

[14] Another source of contrast between Tables 4 and 3 is apparently that average loan losses in the more recent application years are somewhat understated in our data because defaults having lengthy times from default to property

Table 5 provides a more detailed look at the rate with which defaults complete the property disposition process. Table 5 records, for each racial/ethnic group, for low-income borrowers, and overall, the fraction of 3-year defaults that have completed the property disposition process by a given time interval following the date of default. Perhaps the most notable feature of this table is the relatively low fraction of black defaults that have completed property disposition at 18 and 24 months following default. We see, for example, that at 24 months after default, 74 percent of whites and only 56 percent of blacks have completed the property disposition process. Longer times for property disposition may contribute to relatively larger loss rates for blacks. In addition, this observation suggests that, for reasons outlined in footnote 14 in the last paragraph, loss rates are likely to be understated for blacks relative to others, particularly in the more recent application year (1996).[15] That is, our loss data are probably missing a greater proportion of the black defaults, and these defaults likely have a differentially higher loss rate.

Table 5

Cumulative Percentage of 3-Year Defaults Completing Property Disposition Process in Each 6-month Interval after Default, by Race/Ethnicity and Low Income Status

Number of Months After Default	Whites	Blacks	Hispanics	Other	Low Income	Overall
6	0.47	0.14	0.03	0.35	0.30	0.31
12	14.31	8.67	10.62	8.41	12.05	12.26
18	47.67	33.59	47.50	48.27	43.89	44.94
24	74.08	56.17	73.01	71.88	69.93	70.32
30	87.16	73.54	86.86	88.32	84.36	84.51
36	93.88	85.20	93.79	95.15	92.02	92.24
42	97.07	92.29	96.64	97.69	96.17	96.09
48	98.67	96.77	98.11	97.89	98.41	98.16
54	99.35	97.83	99.23	99.36	99.30	99.03
60	99.67	98.75	99.71	100.00	99.63	99.51
66	99.94	99.69	100.00	100.00	99.88	99.90
84	100.00	100.00	100.00	100.00	100.00	100.00

Finally, Table 6 carries the analysis in a new direction by looking at how loss rates vary with the initial loan amount. In particular, we first calculate the quartiles of the distribution of

disposition, and thus larger loan losses, are truncated. That is, we are unable to observe loan losses when property disposition occurs after March 2001. Loans that default in, say, June 1999 have only 21 months from the time of default to complete the foreclosure and property disposition processes. This limit precludes lengthy times spent in foreclosure and in property disposition, thus effectively excluding cases that are likely to exhibit especially large loan losses. This kind of truncation is most severe on the most recent defaults; thus, more recent application years and lengthier times to default will be more seriously affected. Hence, figures in Table 3 for more recent cohorts and, within cohort, for longer times to default, will be more seriously distorted.

[15] Because Table 5 uses only loans that have completed the disposition process, it necessarily excludes loans that take an extraordinarily long time to complete property disposition. Differences would probably be more dramatic if all defaults in the data could be followed until property disposition were complete.

initial loan amounts in each application year; we then group the loans in each year into the appropriate quartiles, and we calculate the average loss rates within each quartile. The quartile loan amounts used to group the loans in each year are presented in the leftmost column of Table 6. As with other tables, separate columns present the loss rates for each racial/ethnic group and for low income borrowers, as well as overall.

Table 6

Mean Loss Rate (%) within Quartiles of Loan Amount, by Application Year, Race/Ethnicity, and Low Income Status

1992 Applications

Quartile Loan Amount	White	Black	Hispanic	Other	Low Income	Overall
<=$51,542	48.29	60.46	42.10	62.40	52.14	51.19
>$51,542, <=$71,550	41.53	49.66	48.83	41.33	44.78	43.94
>$71,550, <=$98,650	42.24	52.58	54.96	56.47	48.73	46.15
>$96,850	47.09	59.93	54.19	54.41	51.24	51.19

1994 Applications

Quartile Loan Amount	White	Black	Hispanic	Other	Low Income	Overall
<=$53,762	51.68	57.86	45.85	41.61	52.39	52.47
>$53,762, <=$74,279	38.92	45.03	44.20	37.62	42.46	40.85
>$74,279, <=$99,144	37.42	39.67	44.51	37.52	41.26	39.44
>$99,144	37.84	55.46	43.56	44.03	47.93	43.11

1996 Applications

Quartile Loan Amount	White	Black	Hispanic	Other	Low Income	Overall
<=$55,797	52.14	52.54	49.74	47.94	50.44	51.70
>$55,797, <=$74,396	41.57	52.22	41.88	40.62	45.34	43.77
>$74,396, <=$100,000	34.58	40.04	42.02	33.80	38.61	37.04
>$100,000	31.04	47.88	30.77	33.23	38.62	33.28

The question of interest here is whether loss rates decline markedly with loan size and whether this decline leads to observed racial difference in loss rates. If, for example, dollar losses are composed of a fixed component unrelated to loan amount plus a second component that varies in direct proportion with loan size, we should see loss rates decline as loan amounts increase. This feature could imply higher average loss rates for blacks if their loan amounts are smaller. The only year in which this pattern appears to hold in the overall sample is 1996. In the remaining two years, the overall sample indicates a U-shaped relationship in which rates initially decline but ultimately rise with the loan amount.[16] Notice that within each loan size category, loss rates for blacks tend to be higher than those for whites; comparisons between whites and Hispanics are less consistent. This crude, preliminary evidence suggests that simple differences in the distribution of

[16] Such a pattern might have occurred for 1996 loans as well if we had used a finer categorization of loan amounts in the top loan size category.

loan amounts cannot account for differences in loss rates between blacks and whites. The more sophisticated analysis below will examine racial differentials while holding constant loan size.

2.3.2. Components of Loss

We now take a different perspective on loss rates by subdividing the total loan loss into its components. For this purpose we express the dollar loss as the following sum:

Loss = (unpaid principal balance – sales price received in property disposition) +
interest foregone + FAC + sales expense + holding cost ,

where FAC represents the sum of foreclosure, acquisition, and conveyance costs. Dividing each term in the latter equation by the initial loan amount yields an equation for the loss rate:

Loss/ initial loan amount =
((unpaid principal balance – sales price received in property disposition)/initial loan amount)
+ (interest foregone/initial loan amount) + (FAC/initial loan amount)
+ (sales expense/initial loan amount) + (holding cost/initial loan amount). (2)

The latter equation expresses the total loss rate as the sum of component loss rates, where each component loss rate is a loss component divided by the initial loan amount. Table 7 shows, by application year and number of years to default, the average total loss rate and the average loss rate contributed by each component. (Columns referring to loss components are labeled with the numerator of the component.) Within each row, the first column is the sum of the remaining columns, aside from rounding error.[17] Note that the rows are mutually exclusive. Thus, for example, the first row in each of the three panels shows the figures for all defaults occurring up to one year after loan origination; the second year gives the figures for all defaults occurring at more than one year and up to two years after origination.

Table 7 shows that unpaid principal balance (UPB) minus price is generally the largest portion of the loss rate; interest foregone, FAC, and sales expense are roughly equal contributors; and holding costs tend to account for the smallest share. Table 7 also illustrates again the substantial decline in loss rates as years to default increase. Looking at the breakdown of loss rates as time-to-default rises within each loan cohort, we see substantial declines in the component rate arising from the difference between UPB and the amount received from selling the home. Indeed, this factor appears to account for most of the decline in the total loss rate as time-to-default increases. With the exception of 1996 loans, the component arising from interest foregone shows little consistent pattern across time-to-default intervals. The FAC rate component shows a declining pattern across default times in 1994 loans and, more weakly, in the 1996 cohort, but it exhibits little discernable pattern in 1992 loans. The sales expense component declines consistently with the timing of default in 1994 and at the higher default times in 1992. Holding cost rates seem to vary little with the timing of default in any of the application years.

[17] Choosing to deduct sales price of the home from UPB is a convenient but somewhat arbitrary way to force loan loss to be the sum of a series of (typically) positive components.

Table 7

Mean Loss Rate (%) and Component Rates (%),
by Years to Default and Application Year

1992 Applications

Years to Default	Loss Rate	Component Rates				
		UPB-Price	Interest Foregone	FAC	Sales Expense	Holding Cost
1	48.93	17.86	10.86	6.98	8.04	5.17
2	48.72	17.18	9.67	7.94	8.48	5.39
3	47.51	15.66	9.57	8.64	8.66	4.90
4	46.80	15.48	8.43	9.29	8.24	5.35
5	45.85	15.10	10.13	7.39	7.68	5.56
6	44.13	14.06	9.80	6.81	7.47	5.99
7	38.32	10.81	9.20	7.05	6.47	4.78

1994 Applications

Years to Default	Loss Rate	Component Rates				
		UPB-Price	Interest Foregone	FAC	Sales Expense	Holding Cost
1	46.66	14.00	9.33	8.89	9.06	5.37
2	45.89	14.30	9.23	8.50	8.42	5.44
3	41.40	10.17	9.89	7.75	8.05	5.53
4	37.38	7.63	9.83	6.51	7.54	5.82
5	33.99	6.02	8.80	6.17	7.25	5.70

1996 Applications

Years to Default	Loss Rate	Component Rates				
		UPB-Price	Interest Foregone	FAC	Sales Expense	Holding Cost
1	46.48	17.27	9.73	6.80	7.35	5.33
2	41.28	12.48	9.09	6.37	7.36	5.94
3	37.82	10.58	7.93	6.15	7.22	5.91

Next consider a similar decomposition of loss rates by race/ethnic group and for low-income borrowers for each application year. The first column of Table 8 presents, for each group and each application year, the mean loss rate. In the next five columns are the component loss rates; within each row the sum of the component rates is the loss rate presented in the first column, aside from rounding error. The final three columns are explained below.

We see that loss rates tend to be 7 to 8 percentage points higher for blacks than for whites in all years; the orderings for Hispanics and Others relative to whites vary over the years. The highest loss rate in the table is recorded for Others in 1992, however. The component-specific rates in the table enable us to see the proximate source of these differences. We see, for example, that the component for UPB minus sales price is very large for Others and for Hispanics in 1992, perhaps a consequence of the substantial downturn in the California housing market in the 1990's. Comparing component rates for blacks and whites, we see that black-white differences in the first component (UPB minus sales price) is on the order of 3.5 percentage points, or nearly half of the overall black-white difference in the total loss rates. Black-white differences in interest foregone account for approximately 1.5 percentage points. Although black-white differences in FAC, running as high as 2 percentage points in 1992, account for most of the remaining difference, there are black-white differences in sales expenses and holding costs as well.

The final three columns show average elapsed times between various milestones in the default process. We hinted above and we argue below that differences in these elapsed times are important for generating differences in loss rates across groups. The first of the elapsed times shows the average time (in months) from the start of loan amortization until default occurs. As with the other figures in this table, the calculations are performed over loans for which loan loss data are present and which, therefore, have necessarily completed the foreclosure process. Average elapsed times until default are shortest for blacks. The next column shows average elapsed times in months from the time of default to the completion of foreclosure and property acquisition; these times are likely to respond in part to declarations of bankruptcy and to the presence of judicial foreclosure proceedings, both of which tend to lengthen the time required to process a foreclosure. Average times tend to be larger for blacks than for the other groups in all three application years. Finally, the last column shows the average number of months from property acquisition to disposition. We see little variation in these figures across the racial/ethnic groups, though the mean for blacks is largest in two of the three years.

Table 8

Mean Loss Rate (%), Component Rates (%) , and Elapsed Times (months), by Race/Ethnicity, Low Income Status, and Application Year

1992		Component Rates					Elapsed Times		
Race	Loss Rate	UPB-Price	Interest Foregone	FAC	Sales Expense	Holding Cost	Amortization Start to Default	Default to Property Acquisition	Property Acquisition to Disposition
White	44.08	13.15	9.52	7.93	8.18	5.28	38.03	15.80	5.15
Black	52.92	16.80	10.84	9.93	8.91	6.34	35.74	19.68	6.08
Hispanic	48.97	21.46	8.37	6.83	7.44	4.86	41.24	14.78	6.26
Other	55.09	25.02	10.77	7.42	7.27	4.61	41.01	16.72	5.65
Low Income	48.52	15.01	9.72	8.97	8.73	6.06	37.21	16.66	5.48
Overall	46.84	15.70	9.58	8.04	8.13	5.35	38.34	16.29	5.52

1994		Component Rates					Elapsed Times		
Race	Loss Rate	UPB-Price	Interest Foregone	FAC	Sales Expense	Holding Cost	Amortization Start to Default	Default to Property Acquisition	Property Acquisition to Disposition
White	40.49	9.99	9.01	7.77	8.11	5.60	27.67	15.62	5.12
Black	48.48	13.40	11.21	9.13	8.56	6.16	25.35	18.66	5.94
Hispanic	41.34	12.11	9.49	6.86	7.88	4.98	28.17	15.07	5.91
Other	38.96	10.78	9.37	5.86	8.16	4.78	28.33	13.93	5.55
Low Income	45.84	13.21	9.39	8.81	8.32	6.09	26.97	16.09	5.74
Overall	42.21	11.14	9.56	7.79	8.15	5.55	27.33	16.05	5.46

1996		Component Rates					Elapsed Times		
Race	Loss Rate	UPB-Price	Interest Foregone	FAC	Sales Expense	Holding Cost	Amortization Start to Default	Default to Property Acquisition	Property Acquisition to Disposition
White	41.89	13.17	8.90	6.54	7.41	5.86	17.43	14.49	5.54
Black	48.71	17.12	10.12	7.79	7.56	6.07	15.35	16.82	6.56
Hispanic	38.65	11.90	8.63	5.49	7.29	5.33	16.61	13.69	6.24
Other	43.88	16.65	10.00	5.89	6.03	5.23	15.95	15.61	5.83
Low Income	47.37	17.79	8.97	7.05	7.52	6.03	17.29	14.69	6.01
Overall	42.43	13.75	9.11	6.48	7.33	5.73	16.78	14.78	5.91

Before proceeding, we digress briefly to discuss the dates that were used to calculate the time intervals discussed in the last paragraph; these dates assume an important indirect role in the analytic work to follow. Four dates, all of which are measured at a monthly level, are especially critical: the amortization start date, the date of default, the date the claim is paid, and the date of property disposition. The amortization start date is used herein as the time of loan origination or the time that the loan begins; this date determines which loans are in the estimation samples. The default date is treated as an important behavioral milestone: the time that the borrower chooses irrevocably to stop repaying the loan. A particularly important use of the default date is in determining whether a loan counts as a default in estimation. The date that the claim is paid serves an important analytic role as the time that FHA pays the remaining principal balance, foregone interest, and other expenses. Less importantly, we also treat this date as the date that the property is acquired and the date at which foreclosure is completed. Thus, we somewhat loosely refer to the elapsed time from the default date to the claim payment date as the "time spent in foreclosure" or the "time spent in property acquisition" or the "time spent in claim processing." The date of property disposition serves an important analytic role as the time at which FHA is assumed to receive the proceeds from selling the foreclosed property. We sometimes refer to the elapsed time from the claim payment date to the property disposition date as the "time spent in property disposition." The most important analytic uses of the claim payment date and the property disposition date are in discounting loan losses, as described below. These dates are used for a variety of other essentially descriptive analyses, and in this context there is less concern about the precise characterization of each date and the related time interval.

SECTION 3

FACTORS UNDERLYING DIFFERENCES IN DEFAULT PROBABILITES AND LOSS RATES

In this section we discuss a set of explanatory factors that may be used in a statistical analysis to explain default probabilities and loss rates. The purpose of the statistical analysis is not only to permit us to understand what underlies default and loan loss, but also to enable us to apply what is learned to underwriting practice. For this reason, we select characteristics of the loan or the borrower that are, or could be, known at the time of loan application. Characteristics of the market or the geographic area may be included, but their purpose is to prevent spurious correlation that might otherwise bias statistical estimates of borrower or loan characteristics.

3.1. The Specification of a Default Relationship

As revealed by the fundamental Eq. (1), the default rate is a necessary component of the unconditional loss rate. As a consequence, any attempts to discover factors underlying unconditional loss rates must consider the specification and estimation of the probability of default as well. Moreover, as explained below, the default model is tied statistically to the specification of the conditional loss rate. The default model thus assumes an important role in the specification and understanding of both conditional and unconditional loss rates, and we estimate a model of default alongside a model of conditional loss rates.

The explanatory factors employed in the default model are similar to those used in other default studies. The list of basic variables and their definitions is provided in Table 9,[18] though Table 9 also includes a couple of variables that are not used in the default specification but do appear in the loss rate analysis below.[19] The explanatory factors in Table 9 are classified into five groups, though the particular classification is in some cases rather arbitrary. We briefly consider the five groups and the variables within each group.

The first group consists of credit characteristics and variables proxying a borrower's capacity to pay. The variables FRONT (front-end ratio) and BACK (back-end ratio) are traditional payment-to-income measures; increases are expected to raise the probability of default. The variable RSVpmts measures the number of monthly mortgage payments the borrower has in reserve after closing; the related variable NO RSVpmt indicates that there are no assets left after closing. A larger number of reserve payments is anticipated to decrease the probability of default. Finally, the FICO variable is based upon the well-know FICO scores for borrower and coborrower (if any).[20] FICO scores are scaled so that increases in FICO scores represent better credit histories and thus better credit risks.

[18] A complete variable list is provided in the appendix. The list in Table 9 excludes additional variables that serve to effect splines and secondary variables that account for missing values that occur for certain basic variables.

[19] Numerous other candidates were considered for inclusion but were eliminated when they failed to pass significance tests at standard levels. We also assume throughout that effects of variables are the same across application years. Some work on permitting differences in impact across years indicated that some effects do vary from year to year.

[20] FICO scores (up to two for each individual) were first averaged, and then the resulting averages for both individuals were averaged.

Table 9
List of Explanatory Variables

Factor	Description
Credit and Ability-to-Pay Characteristics of Borrowers	
FRONT	Front-end ratio (percent)
BACK	Back-end ratio (debt-to-income) (percent)
RSVpmts	Assets after closing divided by mortgage payment
NO RSV pmt	Indicates that assets after closing divided by mortgage payment is zero
FICO	Average of credit scores for borrower and coborrower
Characteristics of the Loan	
LTV	Loan-to-value ratio
LOANamt	Mortgage amount excluding MIP, divided by 10,000
ARM	Indicates ARM
TERM15	Indicates that the term of loan is 15 years or less
NOTErt	Note rate (%)
Characteristics of the Area Housing Market and of the Home Relative to the Area Market	
HPcMSA36	Proportional change in house price at 36 months after loan origination calculated from metro. data
HPcST36	Proportional change in house price at 36 months after loan origination calculated from state data
HPrelPW	House price relative to PriceWaterhouse median price
HPrelLL	House price relative to area median price calculated as (FHA loan limit/0.95)
Race- and Income-Related Characteristics of the Individual and Neighborhood	
BLACK	Indicates African-American
HISPANIC	Indicates Hispanic
INCOME	Monthly income divided by 1000
PCTBLK	Percentage of Census tract population that is African-American
PCTHSP	Percentage of Census tract population that is Hispanic
TRCT_MSA	Census tract income divided by MSA income
UNDERSERVED	Indicator for underserved area in 1996
Other Characteristics	
JUDforecl	Indicator for judicial foreclosure state
SUBURBAN	Indicator for Census tract in suburban area
INTchg36	Change in the 30-year fixed mortgage rate at 36 months after origination
CA	Indicator for California
UNEMPC	Change in state-level unemployment rate (in percentage points) at 36 months after origination
CONDO	Indicator for condominium

The second group of variables characterizes the loan. The LTV is included as well-justified standard fare in default models; higher values of LTV are expected to lead to higher default probabilities. The loan amount (LOANamt) is included to see if larger loans default at a different frequency, other things the same. An indicator for ARMs allows ARMS to differ from fixed-rate mortgages in their default probabilities. TERM15 (an indicator for loans with terms of 15 years or less) is included to incorporate the possibility that more rapid amortization of shorter-term loans, or the self-sorting of borrowers who choose shorter-term loans, results in lower default probabilities. Finally, NOTErt (the mortgage note rate) may proxy the absolute payment burden, other things the same, or it may pick up higher-risk borrowers who pay higher note rates; in either case, NOTErt is anticipated to be positively related to the default probability.

The third group of variables measures characteristics of the housing market and the position of the home within the local market. A pair of variables is used to measure house price growth, a key component in default behavior. The variable HPcMSA36 measures house price growth from origination to 36 months after origination for those homes located in MSAs for which we have MSA house price indices.[21] When we lack MSA house price data, we use HPcST36, which calculates the house price growth at the state level instead.[22]

A second set of variables is used to pick up the position of the house within the area house price distribution; homes in the lower tail of the price distribution may be harder to sell and thus more likely to default. The variable HPrelPW measures the price of the house relative to the median area house price as calculated by PriceWaterhouseCoopers. For areas for which the latter measure is unavailable, we use the variable HPrelLL, which measures the price of the home relative to the area median implied by the FHA loan limit.[23]

The fourth group of variables measures racial/ethnic characteristics of the borrower or the neighborhood, where the latter is taken to be the Census tract. Indicators for black borrowers and Hispanic borrowers are included to allow for possible differences in default probabilities. The variable INCOME permits default probabilities to vary with the borrower's monthly income. Corresponding tract-level variables attempt to capture possible differences in the likelihood of default based on neighborhood values for these same characteristics. In particular, the variables PCTBLK and PCTHISP measure the percentage of the tract population that is black or Hispanic, respectively. TRCT_MSA measures median tract income relative to median MSA income. Finally, the variable UNDERSERVED is an indicator for underserved areas (measured as of 1996). As such, it is a function of area income and racial composition.

The fifth group is composed of miscellaneous measures. (Some of these did not survive in the default analysis but do appear in the loss rate models and are mentioned here for convenience.) In particular, INTchg36 measures the change in market rates for 30-year fixed rate loans from the time of loan origination to 36 months after. One effect of rising mortgage rates may be to dissuade existing borrowers from refinancing, thereby increasing the number of loans remaining active and

---

21 House price growth is calculated from Freddie Mac house price indices. An alternative way to build in the effects of house price growth is through a variable that measures the probability of negative equity. The disadvantage of such a variable in the current context is that it incorporates other effects, such as LTV and the mortgage term. The latter effects are considered separately in the development here, partly because for our purposes it is desirable to separate factors that are subject to great uncertainty and are completely beyond the borrower's control (house price growth) from factors that can in principle be chosen (such as the mortgage term and LTV). Some experimentation suggests that the current formulation performs at least as well as one that incorporates the probability of negative equity.

22 An indicator is also set when the state-level variable is used. See the appendix.

23 The FHA loan limit is set to match 95 percent of the median area house price, subject to a maximum and a minimum value. Indicators are set when the FHA loan limit is at the continental maximum or minimum. See the appendix.

thus exposed to possible default. Another effect, however, is to reduce the discounted value of remaining payments on fixed rate mortgages, thereby reducing the real payment burden and decreasing the likelihood of default. An indicator for California loans is included to allow for the particularly poor California housing market in the early-to-mid 1990s. The variable SUBURBAN allows loans in suburban areas to have different default proclivities (for unexplained reasons) than loans in central cities and rural areas.[24] The change in unemployment rates (UNEMPC) from the time of loan origination to 36 months afterwards attempts to control crudely for a changing economic picture that could affect default behavior. CONDO allows loans for condominiums to default at a different rate. Finally, JUDforecl identifies judicial foreclosure states; this variable is used only in the loan loss analysis.

Having considered the factors that we use to explain default activity, we turn to factors that might be expected to influence loss rates.

3.2. The Ingredients of the Conditional Loss Rate Relationship

3.2.1. Default Threshold Factors and Default Timing Factors

An assessment of factors underlying loss rates raises interesting and important issues, particularly with regard to likely links between default behavior and loan losses. We start by going through a few scenarios that illustrate important points.

To begin, we focus for simplicity on dollar losses, rather than loss rates. Suppose that defaults occur when borrowers find that their unpaid principal balance exceeds the value of their home by, say, \$100, and only then. Assume that losses to FHA are simply UPB minus the price received in property disposition, and that there is no systematic difference between the latter price and the value of the home as viewed by the borrower. In this case, we would expect to find loan losses to be mostly positive and bunched around \$100. A few negative losses (*i.e.,* gains) might arise if house prices grow enough after default but prior to property disposition. We would expect to find no relationship between losses and time of default (or any other variable) because borrower default behavior fixes realized losses at about \$100, aside from random post-default variation in home values. That is, even though there might be a relationship between UPB and timing of default within the class of defaulting loans, this relationship would be offset by a similar relationship between the house price received in property disposition and time of default among these same defaulting loans.

Next we expand upon the last scenario. Now suppose that borrowers default when (and only when) the value of the home falls short of UPB by at least \$100, but the exact shortfall at which default occurs is determined by a borrower's desire to keep a clean credit record. Suppose that we can measure the intensity of such a desire by the borrower's FICO score. In this case, borrowers with higher FICOs would let their house value fall more than \$100 short of UPB before choosing to default. Hence, we would find that among defaulted loans, losses would be positively related to FICO scores; yet among all loans, the probability of default would be negatively related to FICO scores. An analogous argument can be made for other factors that either promote or retard a borrower's willingness to default as measured by the equity loss he/she is willing to tolerate.

[24] We found no evidence of differences in default or loss rates between homes in central cities and homes in rural areas.

Such factors affecting the required equity threshold can be expected to enter the probability of default, as well as the size of loan loss conditional on default, via their link to the difference between UPB and house value. This argument suggests that many of the default factors listed in Table 9 be included as factors that might affect conditional loss rates.

The argument in the last paragraph was based upon the assumption that borrowers default when home value minus UPB (*i.e.,* equity) is negative by a large enough amount to overcome their implicit costs of default, and that nothing else matters. In practice, factors beyond equity may clearly be involved in the default decision. The market value of the mortgage rather than the book value may be relevant; "trigger" events such as job loss and disability may come into play; and so forth. To the extent that factors unrelated to the difference between UPB and house value matter in the default decision, the close link between realized losses and default threshold factors will be weakened. Instead, the individual factors underlying realized UPB and the individual factors underlying realized house values would matter in determining loan losses and conditional loss rates.

Suppose, for example, that equity is completely irrelevant in the default decision. Suppose instead that borrowers default when financially distressed, and that those with higher FICO scores tend on average to take much longer after loan origination to become financially distressed. In this case, FICO scores will tend to be negatively related to loan losses (among defaults) because those with higher FICO scores tend to survive longer, thereby reducing UPB and generating smaller losses, other things the same.

In the real world in which a complex variety of motivations generate default behavior, one might expect that factors affecting the timing of default, as well as factors affecting equity thresholds that trigger default, will influence loss rates. The net results may be unpredictable on prior grounds. To continue with the above example, on the one hand, when FICO scores affect realized losses by increasing the negative equity threshold that must be crossed to generate a default, they will be positively related to losses. On the other hand, when equity is not the only factor in the default decision, FICO scores may be negatively related to losses since they tend to lengthen the survival time, even among defaulting loans, thus reducing losses via their effect on UPB. The upshot of this is that all factors influencing default can be expected to influence loss rates, even if the direction of effect cannot be anticipated at the outset.

While realizing that a complex set of interactions may be involved, it is worth discussing briefly some of the factors that might be expected to influence loss rates, particularly through their influence on timing. This discussion largely ignores possible threshold effects discussed above but considers a more complete set of loss components.

3.2.2. Components of Loss Revisited

The components of the loss rate, as written in Equation (2), provide a useful organizational device. As will be seen, the factors underlying some of the loss components seem fairly clear; in other cases underlying factors are not at all clear. To help clarify the ingredients of the component loss rates, we experimented with a set of exploratory regressions in which each component loss rate was regressed on a variety of explanatory variables.[25] These regressions were not meant to be definitive but were instead meant to be descriptive, to provide insight as to why factors affect loss rates the way they do, and to suggest structure. In particular, the explanatory variables included various aspects of timing, such as the elapsed time from origination to default. Such variables are likely to be statistically endogenous and are replaced by their determinants for the empirical implementation in Section 4. The discussion here identifies factors that seem likely to underlie components of the loss rate, and thus the total loss rate; where appropriate, we highlight some of the more interesting findings from the exploratory regressions.

To begin, we separate the first (equity) term in Eq.(2) into two parts: (UPB/initial loan amount) and (sales price received in property disposition/ initial loan amount), and we discuss them separately.

UPB/Initial Loan Amount. This term is simply the principal balance at default relative to the initial principal balance. As such, we expect its determinants to include the timing of default, as well as those factors affecting the rate of amortization. (An exploratory regression supported our expectations.) Presumably, factors affecting the timing of default are those affecting the probability of default, discussed above. Factors affecting the rate of amortization include the term of the loan (TERM15) and the note rate (NOTErt).

Price Received in Property Disposition/Initial Loan Amount. If the value of the home at property disposition were the same as at origination, this term would simply be the inverse of LTV at origination. Hence, this term should be related to initial LTV but would be expected to differ because of post-origination changes in house price. Growth in house prices 36 months after origination can be measured by the variables HPcMSA36 and HPcST36 in Table 9. Because these growth estimates cover large areas, price growth for a particular home within the area may well deviate from the area average. In addition, it is possible that the price received in property disposition is related to the amount of time that the home spends in the foreclosure and property disposition processes. An auxiliary exploratory regression did indeed suggest that increases in time spent in foreclosure processing and in property disposition are associated with lower prices received in property disposition.[26] Determinants of these elapsed times are considered below.

Interest Foregone/Initial Loan Amount. This component can be expressed as the product of three terms: UPB/Initial loan amount (discussed above), the rate at which interest accrues,[27] and the number of months from the time of default to the time that the claim is paid (discussed below).[28]

FAC/Initial Loan Amount. The dollar value of FAC is probably related to institutional features, such as the legal environment surrounding foreclosure proceedings, including the presence of

---

[25] The sample was restricted to loans defaulting during the first three years following origination. These regressions were generally specified using logarithmic values of dependent and many of the independent variables.

[26] The direction of causation is unclear, particularly with regard to time spent in property disposition. It may be that homes languishing in foreclosure and property disposition depreciate in the process. Instead, it may be that it takes longer to sell homes that are worth less than might otherwise be expected.

[27] We use the note rate to proxy the more appropriate debenture rate at which interest is reimbursed.

[28] An auxiliary regression unexpectedly showed that the unpaid balance and the interest rate had the wrong signs, and this was true even when the debenture rate was used rather than the note rate.

judicial foreclosure, and perhaps is increasing in the length of time spent in foreclosure processing. FAC is probably only weakly related to other available explanatory variables. The presence of judicial foreclosure is represented by an indicator variable (JUDforecl in Table 9 above). Variables underlying the length of time spent in foreclosure processing are examined below. Notice that if FAC has a fixed component, we would expect FAC/initial loan amount to decline with increases in the initial loan amount. An auxiliary descriptive regression confirms all of these expectations.

Sales Expenses/Initial Loan Amount. This term seems likely to vary directly with the ratio (price of the home received in property disposition/initial loan amount), discussed above, and with the amount of time spent in property disposition. It might also perhaps differ by type of structure (*e.g.,* for condos). An auxiliary exploratory regression confirms that these expectations do in fact hold.

Holding Costs/Initial Loan Amount. This component seems likely to vary directly with the ratio (price of the home received at disposition/initial loan amount), discussed above, and with the amount of time spent in property disposition. An auxiliary regression confirms these expectations. Factors underlying the time spent in property disposition are considered below.

The discussion above noted, and auxiliary descriptive regressions confirm, that several time intervals are important: the elapsed time from origination to default, the time spent in foreclosure processing, and the time spent in property disposition. As discussed above, the elapsed time from origination to default is expected to be related to all of the factors included in the default logit. We now consider factors underlying the time spent in foreclosure processing and the time spent in property disposition.

Time in Foreclosure Processing. One might expect time spent in foreclosure processing to be greater in judicial foreclosure states, a finding that is confirmed by an auxiliary regression. This same regression suggests that foreclosure processing time is less for those with higher FICO scores. One possibility is that those with higher FICO scores are more reluctant to declare bankruptcy. Similar reasoning might apply to those who have more expensive homes relative to the area median (HPrelPw and HPrelLL), who are found in the exploratory regressions to have shorter times spent in foreclosure processing.

Time in Property Disposition. An exploratory regression confirms that time spent in property disposition is shorter for homes that are (or were, at the time of origination) more valuable relative to the local area median (HPrelPW and HPrelLL). For reasons that are unclear, time spent in property disposition appears to be lower for those with higher FICOs. Aside from the possibility of spurious correlations, this effect might be traceable to better pre-default upkeep provided by those with higher FICOs, which might lead to less time required to make the home suitable to sell.

The upshot of the above discussion is that, aside from an indicator for judicial foreclosure, the factors affecting the components of conditional loss rates are included within the set of variables affecting the probability of default. Section 4 reports the empirical findings.

SECTION 4

ESTIMATION OF MODELS OF DEFAULT AND LOSS RATES

4.1. Empirical Estimates of Impacts in a Model of 3-Year Defaults

To estimate the impact of a variety of factors on the probability of a default within three years after origination, we employ logit estimation, a standard statistical procedure for estimating models with dichotomous dependent variables. A full set of estimation results is presented in the appendix.[29] Here we instead illustrate the importance of each factor by showing how much the probability of default is estimated to change, according to our statistical estimates, in response to a given hypothetical change in each underlying factor, holding constant all other factors.

Unfortunately, it is unclear how large a hypothetical change in each factor would best serve the dual purpose of (a) illustrating its effect and (b) permitting comparisons across different factors. For current purposes we have chosen two different kinds of hypothetical changes: one kind of change for indicator variables and the other for conceptually continuous variables. For indicator variables, we consider a hypothetical change in each variable from zero to one, *i.e.,* the hypothetical change from the absence to the presence of the factor in question. In making this change, we hold all other factors at their sample averages. For the ARM variable, for example, we start with a hypothetical fixed rate loan and estimate the probability of default. We then estimate the default probability of an otherwise identical ARM, and we difference the two estimated default probabilities to arrive at the estimated impact of an ARM on default.[30]

There is one important exception to the latter framework for indicator variables. For UNDERSERVED, we consider a hypothetical in which we begin with UNDERSERVED set to zero and with tract income and racial composition set to averages calculated over all non-underserved areas. That is, TRACT_MSA, PCTHSP, and PCTBLK are set to their means in non-underserved areas. We then estimate the change in the probability of default associated with changing the value of UNDERSERVED to one and changing tract income and racial composition to their means calculated over underserved areas. The idea is to estimate the change in the probability of default when moving from a typical non-underserved area (with income and racial composition as in a typical non-underserved area) to a typical underserved area (with income and racial composition as in a typical underserved area).

For continuous variables, we consider the estimated change in default probabilities associated with a change from the first quartile value to the third quartile value of the variable in question. That is, we calculate the first quartile and the third quartile value for each variable in the default estimation sample. We form a hypothetical loan with sample means for all variables other than the focus variable, which is set to the first quartile value. We estimate the default probability and compare it to the estimated default probability for the same hypothetical loan with the focus variable reset to the third quartile value.

---

[29] The sample for this logit is composed of 69,286 loans from 1992 applications, 63,155 loans from 1994 applications, and 35,841 loans from 1996 applications.

[30] When two indicators are used to distinguish among three or more categories, we set up the hypothetical change so that we start with the default probability for the omitted group and then change to the indicated group.

Table 10 may provide some context for both kinds of hypothetical changes. There we present the first, second, and third quartile values, as well as the mean and standard deviation, for all variables in the sample used to estimate the default probability model. The first and third quartile values in this sample are used as the basis for the hypothetical changes described above. (An analogous set of statistics is presented for the smaller sample of defaulted loans that is used to estimate the loss rate model below.)

Table 10
Summary Statistics on the Explanatory Variables

	Full Sample					Loss Sample				
Variable	25th percentile	Median	75th percentile	Mean	St. Dev.	25th percentile	Median	75th percentile	Mean	St. Dev.
FRONT	18.00	22.53	27.15	22.55	6.356	19.71	24.37	28.57	24.17	6.345
BACK	30.79	36.10	39.99	34.96	6.691	32.07	37.16	40.65	35.97	6.382
RSVpmts	0.000	0.829	4.163	5.334	16.536	0.000	0.000	1.772	2.911	10.694
NO RSV pmt	0.000	0.000	1.000	0.376	0.484	0.000	1.000	1.000	0.516	0.500
FICO	643	686	728	684	57.908	604	639	676	641	55.263
LTV	0.9308	0.9587	0.9708	0.9387	0.060	0.9397	0.9625	0.9733	0.9476	0.047
LOANamt	5.57	7.25	9.23	7.59	2.759	5.37	7.37	9.86	7.78	3.050
ARM	0.000	0.000	0.000	0.242	0.428	0.000	0.000	1.000	0.283	0.451
TERM15	0.000	0.000	0.000	0.028	0.166	0.000	0.000	0.000	0.009	0.092
NOTErt	7.00	8.00	8.50	7.74	1.060	7.00	8.00	8.50	7.85	1.101
HPcMSA36	0.068	0.122	0.167	0.118	0.089	0.033	0.097	0.146	0.082	0.090
HPcST36	0.091	0.136	0.161	0.130	0.066	0.077	0.123	0.151	0.111	0.062
HPrelPW	0.653	0.864	1.088	0.872	0.340	0.542	0.714	0.919	0.727	0.298
HPrelLL	0.496	0.630	0.766	0.623	0.210	0.449	0.577	0.718	0.571	0.214
BLACK	0.000	0.000	0.000	0.108	0.311	0.000	0.000	0.000	0.193	0.395
HISPANIC	0.000	0.000	0.000	0.113	0.317	0.000	0.000	0.000	0.200	0.400
INCOME	2.408	3.125	4.009	3.337	1.345	2.235	2.939	3.813	3.145	1.297
PCTBLK	0.804	2.809	9.422	10.052	18.412	1.412	5.210	16.457	15.349	23.469
PCTHSP	0.989	2.452	8.417	8.343	14.490	1.272	4.763	17.875	13.379	18.973
TRCT_MSA	0.859	1.007	1.169	1.027	0.263	0.793	0.950	1.106	0.958	0.257
UNDERSERVED	0.000	0.000	1.000	0.396	0.490	0.000	1.000	1.000	0.548	0.498
JUDforecl	0.000	0.000	1.000	0.429	0.495	0.000	0.000	1.000	0.404	0.491
SUBURBAN	0.000	0.000	1.000	0.495	0.500	0.000	0.000	1.000	0.484	0.500
INTchg36	-1.150	-0.810	-0.450	-0.817	0.510	-1.170	-0.890	-0.460	-0.922	0.531
CA	0.000	0.000	0.000	0.094	0.292	0.000	0.000	0.000	0.249	0.433
UNEMPC	-27.660	-22.078	-15.909	-20.882	10.589	-26.667	-22.059	-16.071	-20.488	9.922
CONDO	0.000	0.000	0.000	0.025	0.156	0.000	0.000	0.000	0.034	0.182

Tables 11 and 12 present the results of the calculations based on hypothetical changes. The first set of calculations (Table 11) pertains to a "basic model" that excludes the variables measuring race and income characteristics of the individual and area. The second set of calculations (Table 12) is for an "expanded model" that includes these additional variables. In both tables the first two columns give the hypothetical “before” and “after” values—zero and one for indicator variables and (approximate) first and third quartile values for conceptually continuous variables. The next three columns show the estimated default probabilities for the “before” and “after” states of the world and the difference (default probability “after” minus the default probability “before”).[31] The remaining six columns are considered later. Looking down the difference column in Table 11 and focusing first on the continuously measured variables, we note from the sign pattern that each factor works in the direction that might have been anticipated. In particular, increases in the front-end ratio, the back-end ratio, LTV, and the note rate are estimated to increase the default probability, while increases in reserve payments, FICO scores, loan amounts, post-origination house price growth, relative house price, and post-origination changes in mortgage rates result in reductions in default probabilities. The FICO impact seems to dominate the rest of the continuously measured variables, with the hypothetical change from the first to the third quartile value associated with a decrease in the default probability of 2.24 percentage points. Increasing reserve payments from 0 to 4 has a sizable effect as well. At the other end of the spectrum, increasing the loan amount from $55,000 to $92,000 reduces the estimated default probability by only 0.05 percentage points.[32]

The estimated impacts of indicator variables in Table 11 again conform to expectations as to direction. In particular, ARMs and California loans have higher estimated default probabilities, and shorter-term mortgages have lower estimated default probabilities. Among these, the larger adverse estimated effect of California residence is noteworthy.

Next turn to the expanded model in Table 12. Estimated default impacts of the variables in common to the two specifications are generally very similar, and our discussion is thus limited to the new entries.[33] Black borrowers have higher estimated default probabilities than otherwise identical non-blacks. Increases in income tend to lower estimated default probabilities. Higher representation of blacks and Hispanics and lower tract incomes are associated with higher estimated default probabilities. Given these findings, it is not surprising that borrowers in the “average” underserved area (in the sense described above) have higher estimated default probabilities than otherwise identical borrowers in “average” non-underserved areas.[34] None of these effects seems particularly large, and all are dwarfed by the effect of California residence or the hypothetical change in the FICO.

[31] A missing set of entries indicates that the variable was not used in the default logit.

[32] The impact of loan amount depends heavily on where it is evaluated. Our estimates imply that marginal increases in loan amounts below $65,000 reduce default probabilities, but the marginal effect on default probabilities becomes positive for loan amounts in excess of $65,000.

[33] Note that the indicator for Hispanic ethnicity is omitted from the expanded model because its estimated effect was not significantly different from zero at standard significance levels.

[34] The indicator for underserved areas was eliminated from the default logit for lack of statistical significance. There remains an entry for UNDERSERVED in the table, however, because, as discussed above, the conceptual experiment involved changes in area income and minority representation as well. There is apparently no additional default effect associated with the underserved label.

Table 11

Effects of Hypothetical Changes in Explanatory Variables on Default Probabilities and Conditional and Unconditional Loss Rates

Basic Model

Variable	Hypothetical Values		Default Probability (%)			Conditional Loss Rate (%)			Unconditional Loss Rate (%)		
	Before	After	Before	After	Difference	Before	After	Difference	Before	After	Difference
Credit and Ability-to-Pay Characteristics of Borrowers											
FRONT	18	27	1.90	2.38	0.48	31.3	32.3	1.0	0.594	0.769	0.175
BACK	31	40	2.07	2.21	0.14	--	--	--	0.658	0.702	0.044
RSVpmts	0	4	2.72	1.59	-1.12	32.8	30.6	-2.1	0.889	0.487	-0.402
FICO	642	728	3.49	1.25	-2.24	32.8	30.8	-2.0	1.145	0.385	-0.761
Characteristics of the Loan											
LTV	0.93	0.97	2.00	2.41	0.42	31.0	32.8	1.8	0.619	0.792	0.173
LOANamt	5.5	9.2	2.13	2.08	-0.05	31.3	30.4	-0.9	0.668	0.635	-0.033
ARM	0	1	1.99	2.61	0.62	31.2	33.8	2.6	0.622	0.884	0.262
TERM15	0	1	2.17	1.03	-1.15	--	--	--	0.692	0.327	-0.365
NOTErt	7	8.5	1.94	2.34	0.39	30.9	32.8	1.9	0.600	0.766	0.166
Characteristics of Area Housing Market and of the Home Relative to the Area Market											
HPcMSA36	0.07	0.17	2.49	1.79	-0.70	35.0	27.3	-7.7	0.871	0.488	-0.384
HPcST36	0.09	0.16	2.50	1.90	-0.60	35.8	32.2	-3.6	0.896	0.612	-0.284
HPrelPW	0.67	1.1	2.48	1.87	-0.61	35.0	30.2	-4.8	0.868	0.565	-0.303
HPrelLL	0.51	0.77	2.38	1.86	-0.52	32.5	27.4	-5.0	0.773	0.510	-0.263
Other Characteristics											
JUDforecl	0	1	--	--	--	28.1	36.7	8.6	0.599	0.782	0.183
SUBURBAN	0	1	--	--	--	33.8	30.7	-3.2	0.720	0.653	-0.067
INTchg36	-1.15	-0.45	2.24	2.01	-0.23	32.6	31.0	-1.5	0.730	0.624	-0.106
CA	0	1	1.99	4.04	2.05	--	--	--	0.633	1.287	0.653
UNEMPC	-28	-16	--	--	--	31.5	32.1	0.6	0.670	0.682	0.012
CONDO	0	1	--	--	--	31.9	29.7	-2.2	0.679	0.631	-0.047

Table12

Effects of Hypothetical Changes in Explanatory Variables on Default Probabilities and Conditional and Unconditional Loss Rates

Expanded Model

Variable	Hypothetical Values		Default Probability (%)			Conditional Loss Rate (%)			Unconditional Loss Rate (%)		
	Before	After	Before	After	Difference	Before	After	Difference	Before	After	Difference
Credit and Ability-to-Pay Characteristics of Borrowers											
FRONT	18	27	1.96	2.26	0.30	30.1	32.5	2.4	0.590	0.734	0.144
BACK	31	40	2.04	2.20	0.16	--	--	--	0.638	0.688	0.050
RSVpmts	0	4	2.65	1.59	-1.06	32.0	30.4	-1.6	0.851	0.485	-0.366
FICO	642	728	3.41	1.26	-2.15	32.1	30.5	-1.6	1.092	0.383	-0.709
Characteristics of the Loan											
LTV	0.93	0.97	1.98	2.38	0.40	30.6	32.3	1.7	0.606	0.769	0.163
LOANamt	5.5	9.2	2.02	2.16	0.14	31.8	29.5	-2.3	0.640	0.636	-0.004
ARM	0	1	1.98	2.57	0.59	30.9	32.5	1.6	0.612	0.836	0.224
TERM15	0	1	2.15	1.06	-1.09	--	--	--	0.673	0.331	-0.342
NOTErt	7	8.5	1.92	2.31	0.38	30.8	31.8	1.0	0.593	0.734	0.141
Characteristics of Area Housing Market and of the Home Relative to the Area Market											
HPcMSA36	0.07	0.17	2.42	1.74	-0.68	34.2	26.1	-8.1	0.829	0.455	-0.374
HPcST36	0.09	0.16	2.62	2.00	-0.62	37.3	33.0	-4.3	0.979	0.662	-0.317
HPrelPW	0.67	1.1	2.37	1.90	-0.47	33.6	30.3	-3.3	0.795	0.576	-0.219
HPrelLL	0.51	0.77	2.31	1.92	-0.39	31.8	28.3	-3.5	0.736	0.544	-0.192
Race- and Income-Related Characteristics of the Individual and Area											
BLACK	0	1	2.08	2.36	0.29	31.0	32.4	1.4	0.645	0.766	0.121
HISPANIC	0	1	--	--	--	31.0	32.2	1.2	0.645	0.669	0.025
INCOME	2.4	4	2.21	2.04	-0.18	30.0	32.2	2.2	0.664	0.656	-0.008
PCTBLK	0.8	9.4	2.01	2.07	0.07	31.3	32.0	0.7	0.628	0.663	0.036
PCTHSP	1	8.4	2.02	2.08	0.06	32.4	31.9	-0.5	0.655	0.663	0.007
TRCT_MSA	0.86	1.17	2.27	2.06	-0.21	33.1	29.4	-3.7	0.752	0.605	-0.147
UNDERSERVED	0	1	1.97	2.40	0.43	29.3	35.9	6.7	0.578	0.864	0.285
Other Characteristics											
JUDforecl	0	1	--	--	--	28.0	35.7	7.7	0.590	0.753	0.163
SUBURBAN	0	1	--	--	--	32.7	31.0	-1.8	0.678	0.641	-0.037
INTchg36	-1.15	-0.45	2.22	1.99	-0.22	32.2	30.4	-1.8	0.713	0.605	-0.108
CA	0	1	1.97	3.94	1.97	--	--	--	0.618	1.235	0.617

4.2. Empirical Estimates of Impacts in a Model of Conditional Loss Rates at Three Years after Origination

4.2.1. Refining the Measure of Loan Loss: The Issue of Discounting

The development to this point has measured the loss rate as dollar losses divided by the initial loan amount. This calculation implicitly assumes that dollars are equally valuable no matter when they are received or paid. The existence of positive interest rates, however, implies that the government should not be indifferent to the timing of costs or receipts. Because interest foregone is a real cost to FHA, cash flows should be discounted to some single point in time, thus putting all cash flows on a comparable (discounted) basis. In principle, on a typical conveyance, one could separate the timing of numerous payment events, including the payment of the claim, foreclosure costs, taxes, upkeep on the property, and sales expenses, as well as the receipt of rent (if any) and sales proceeds upon disposing of the property. In practice, however, the data at our disposal do not permit all such distinctions, nor is it clear that the finer distinctions would even matter. Hence, the detailed statistical estimation here recognizes the timing of two events: payment of the claim and disposal of the property. We assume that at the time that the claim is paid, FHA pays the full amount of acquisition costs, including the remaining unpaid principal balance (UPB), interest owed the lender, and all foreclosure, acquisition, and conveyance expenses (FAC). All remaining elements of the profit loss calculation --- the net of holding costs, sales expenses, and sales proceeds --- are assumed paid at the time of property disposition. Cash flows are discounted back to the month of loan origination at the 10-year Treasury rate.[35] Thus, we calculate discounted loan losses empirically as

$$\text{(UPB + interest foregone + FAC)} / (1 + r)^{tc} + \text{(holding costs + sales expenses – sales proceeds)} / (1 + r)^{td} \qquad (3)$$

where tc is the number of months from loan origination until the claim is paid, td is the number of months from loan origination until property disposition, and r is the discount rate.

The assumption in the last paragraph is that discounting matters, for otherwise the issues of whether to discount and, if so, the appropriate discount rate to use, become immaterial. To help gauge the importance of discounting, we calculated loan losses three different ways.[36] One way used the formulation given in Equation (3), thus drawing some distinctions in timing, which in turn affect discounted values. A second alternative used no discounting at all, simply adding the dollar value of all costs regardless of when they occurred; these undiscounted costs were the basis for the results in Sections 2 and 3 above. A third method discounted cash flows to loan origination, but with somewhat less loan-specific detail than that used in the first method. For this purpose we used the loan-specific time from origination to loan default, but then added sample average values of time elapsed from default to the payment of the claim to estimate tc, and added sample average values of the elapsed time from default to property disposition to estimate td. Comparing the three alternative calculations of loan losses with each other revealed a very high degree of correlation; all three pairwise correlations were above 0.96. This rather limited examination suggests that precise discounting does not appear to matter greatly in the current context though, of course, these comparisons surely hinge on the discount rate that is adopted; a sufficiently high discount rate would presumably overturn this verdict. To give some indication of the effect of discounting on loss rates in the aggregate, we note that discounting reduces conditional loss rates for 3-year defaults in 1992 applications from 48.22 to 40.52 percent, and in 1996 applications from 42.44 to 37.76 percent. For the remainder of this paper we measure loss rates using discounted cash flows.

[35] More specifically, we use a discount rate of 6.24 percent per year, the average of the 10-year Treasury rates over the period January 1992 though January 2001. Clearly, one could argue for other discount rates.

[36] These experiments were conducted with an earlier version of the loan loss data.

4.2.2. Empirical Estimates

We now turn to estimates of regression models of the (discounted) conditional loss rate estimated over samples of loans that defaulted within the first three years following origination.[37] As with the default model, the estimated parameters are presented in detail in the Appendix.[38] Here we instead illustrate the importance of various factors by considering the same series of hypothetical changes as were used to illustrate the importance of default factors. The estimated impacts on loss rates are presented in the remaining columns of Tables 11 and 12. In both of these tables, the three columns in the middle of the table show the conditional loss rates in the "before" and "after" states of the world, as well as the difference in the conditional loss rates (calculated as the "after" rate minus the "before" rate). The three rightmost columns use the estimated default probabilities and the estimated conditional loss rates to calculate the estimated unconditional loss rates. Again, the "before" and "after" unconditional loss rates are differenced to arrive at the "difference" column.

Looking first at the impact of the continuously measured variables on conditional loss rates in Table 11, we see that increases in the front-end ratio, LTV, and the note rate are associated with increases in the loss rate. Increases in reserve payments, FICO scores, loan amounts,[39] house price growth, relative house prices, and post-origination changes in interest rates work to reduce the conditional loss rate. We also note that the largest impacts arise from the house price growth and relative house price variables. Among indicator variables, we see that borrowers in judicial foreclosure states and non-suburban areas have higher conditional loss rates than their otherwise equivalent counterparts in other areas. The presence of judicial foreclosure appears to have a particularly large effect in increasing conditional loss rates.

As discussed above, the effects of many default-related factors on conditional loss rates are unclear on prior grounds, and thus empirical estimates like those in Tables 11 and 12 are necessary to resolve ambiguity. The estimated effects of the FICO scores are particularly interesting in this context, for they nicely illustrate the complexity of interpreting these conditional loss regressions. We take a brief detour to explain. The pure "threshold effect" discussed earlier (see pages 20-21) would lead higher FICO scores to be associated with higher loss rates (arising from higher negative equity thresholds to default) conditional on default. Pure timing effects suggest, however, that because higher FICOs are associated with longer times to default,[40] higher FICO scores will be associated with lower conditional loss rates by allowing additional amortization of the principal balance. In addition, as noted earlier, exploratory regressions reveal that times in foreclosure processing and in property disposition tend to be shorter for those with higher FICO scores, and these effects tend to reduce loss rate components arising from FAC, holding costs, interest foregone, and sales expenses. The effects of FICOs on time-to-default and time spent in

---

[37] Loans used in these exercises are a subset of the 3-year defaults appearing in the default logits, differing from the latter only because not all of the 3-year defaults had information on dollar losses and because we excluded a very few anomalous cases (such as the few assigned loans with loan loss data). The sample includes 3,796 loans from 1992 applications, 3,355 loans from 1994 applications, and 1,197 loans from 1996 applications.

[38] The appendix also briefly discusses heteroskedasticity adjustments made to the regression.

[39] The estimated effect of the loan amount on loss rates is negative throughout, but the marginal effect is smaller in absolute value at larger loan amounts.

[40] We also confirmed in a separate exploratory regression of defaulting loans that higher FICO scores are associated with increased time until default.

foreclosure and property disposition also imply effects on loss rates via their effects on discounting. In particular, longer times to default reduce the present value (at origination) of each dollar paid in claim. Shorter times from default to property disposition reduce the wait until the proceeds from sale are received, thus increasing their discounted value and reducing discounted loss rates. Finally, those with higher FICO scores might behave differently in other ways that affect loss rates.

According to the estimates presented in Tables 11 and 12, those with higher FICOs tend to have smaller conditional loss rates. By the reasoning in the last paragraph, the pure threshold effect of higher FICO scores, which would lead to higher conditional loss rates, is apparently dominated by the other effects leading to lower conditional loss rates.

We may also draw some tentative conclusions on the route by which FICO scores affect loss rates. As discussed in the last paragraph, higher FICO scores are associated with reduced times spent in foreclosure processing, reduced times spent in property disposition, and longer duration until default. Each of these timing impacts serves to reduce the loss rate. In exploratory regressions for each of the loss rate components in which we controlled directly for these durations, the FICO score was significantly different from zero only in the regression for interest foregone. The evidence from these exploratory regressions as a whole thus suggests that the principal influence of the FICO score may be through its effects on time-to-default and on the times required for foreclosure processing and property disposition.

Another set of regressions offers confirming evidence. We reran the conditional loss rate regressions reported in Tables 11 and 12 using the original explanatory variables plus a set of controls for each of the three durations (time to default, time in foreclosure processing, and time in property disposition). We found that in both regressions the durations had the expected signs and were statistically significant. The effect of including these durations, moreover, was to reduce the estimated FICO effect by about two-thirds, and in the expanded model, to render its coefficient insignificant by the usual standards. Again, the suggestion is that a major impact of the FICO is on timing: lengthening the time to default and reducing times spent in foreclosure processing and property disposition.

Turning to the conditional loss rate columns for the expanded model in Table 12, we see that estimated impacts are identical in sign to those in Table 11, though there are occasionally some fairly large differences in magnitudes. The new entries in the expanded model—race- and income-related characteristics—are of particular interest. We see that blacks and Hispanics are associated with higher conditional loss rates than whites and others, while higher individual incomes also appear to be related to higher loss rates, perhaps a result of a strong threshold effect. Furthermore, conditional loss rates are higher with increases in the percentage of area population that is black, reductions in the percentage Hispanic, and reductions in area relative income. Finally, the borrowers in areas that are underserved (and have income and minority representation like that of the average underserved area) have higher estimated conditional loss rates than similar borrowers in non-underserved areas (with characteristics like that of the average non-underserved area).

Evidence from other regressions suggests that the relationship between borrower race and loan loss is traceable to differences in several dimensions. Exploratory regressions like those in Section 3.2.2 show that blacks tend to have shorter times to default and longer times for foreclosure processing and property disposition than do whites. Holding constant these elapsed times, blacks tend to have higher values for FAC and holding costs, and a larger gap between UPB and price received in property disposition, relative to similarly situated whites. The underlying reasons for the Hispanic effect, however, is less clear in these exploratory regressions. Hispanics tend to have longer time in property disposition relative to whites but no significant difference in other aspects of timing. Holding constant elapsed times, Hispanics show a larger gap between UPB and price received in property disposition and larger values for interest foregone, but smaller values for holding cost and sales expense.

Other evidence suggests that association between race and elapsed times is primary. Rerunning the conditional loss rate regression in Table 12 using the variables from the expanded model plus controls for each of the three elapsed times (time to default, time in foreclosure processing, and time in property disposition) reduces the estimated black coefficient by over three-fourths and renders it statistically insignificant. The Hispanic coefficient falls by about one-third and also becomes statistically insignificant.

The unconditional loss rates provide a "bottom line" by giving the expected loss per loan arising from the combination of default propensities and conditional loss rates. With few exceptions, factors are estimated to have the same qualitative effect on default rates as on conditional loss rates, and thus on unconditional loss rates. Magnitudes of effects are of interest as well. Among the continuously measured variables in Table 11, the estimated effects of FICO scores stand out and, to a lesser extent, reserve payments, house price growth, and relative house prices. Among indicator variables, the estimated impact of California residence is noteworthy. These statements seem to hold for the unconditional loss rates in the expanded model in Table 12 as well. We also notice from Table 12 that black borrowers and borrowers in "average" underserved areas have higher estimated unconditional loss rates than otherwise identical borrowers without these characteristics.

To give additional perspective on the relative size of the effects of FICO scores and race, we made some additional comparisons using the estimated default and loss rate models. We first calculated the estimated default probability, conditional loss rate, and unconditional loss rate for a hypothetical individual with sample averages for all variables other than race, which was assumed to be black. We then calculated the FICO scores that a hypothetical white borrower with the same characteristics (other than FICO and race) would need in order to have the same values for the three risk measures. The calculated FICO value clearly depends on the size of the race effect as well as the impact of FICO scores, and for this reason the calculated FICOs vary strongly with the risk measure. To have the same estimated default probability as the hypothetical black borrower with the average FICO in the sample (a score of 683.7), the hypothetical white borrower would need to have a FICO score 11 points lower (a FICO score of about 673). To have the same conditional loss rate, the white borrower would need to have a FICO score of 609; to have the same unconditional loss rate, the white borrower would need to have a FICO score of 670.

We then made a similar set of calculations of compensating FICO scores with somewhat different hypothetical borrowers. The hypothetical black borrower in this second scenario was assumed to have values for all variables set at the sample averages for blacks only, including the FICO score (a value of 651). The hypothetical white borrower was assumed to have values for all variables other than the FICO set to sample averages for whites. We then calculated the FICO scores for the white borrower that would yield equivalent values for the three risk measures. In this case we found that a FICO score of 610 would equate the estimated default probabilities, a FICO score of 170 (outside of the feasible range of FICO scores) would equate the estimated conditional loss rates, and a FICO score of 591 would equate unconditional loss rates. Not surprisingly, considerably lower FICOs are required to compensate for riskier characteristics, on average, of black borrowers.

4.3. Graphical Comparisons of the Importance of Selected Factors

Figures 1, 2, and 3 attempt to provide a more complete comparison of the effects of selected continuously measured variables: LTV, the front-end ratio (FRONT), the back-end ratio (BACK), the initial loan amount (LOANamt), reserve payments (RSVpmts), house price change within MSAs (HPcMSA36), house prices relative to the area median (HPrelPW), and FICO scores.[41] For each of these variables, we plot the effect on each risk measure (default probability, conditional loss rate, unconditional loss rate) of gradually increasing the variable value from its 1st to its 99th percentile. Because exactly the same procedure was followed for each variable and each risk measure, we can illustrate the details by considering only the effect of LTV on the default probability. First, we used the full sample to calculate each percentile value of LTV from the 1st to the 99th percentile. We then used the enhanced model to estimate the series of default probabilities that would result if LTV were set at each of these percentile values while all other variables were set at the sample means. Next, to facilitate comparisons, we used as a benchmark the default probability we obtained for LTV set at its first percentile. That is, from each of the default probabilities, we deducted this benchmark default probability, thus yielding the series of increases (or decreases) in default probabilities relative to the benchmark. We then plotted the resulting differences in default probabilities (relative to the benchmark) at each percentile of LTV. Repeating this procedure for each risk measure and each explanatory variable yields Figures 1 through 3. Each diagram thus shows how a particular risk measure responds as each explanatory variable moves from its 1st up to its 99th percentile.[42] Horizontal lines are provided to facilitate comparisons between variables that increase, and variables that decrease, each measure of risk.

---

41 The appendix plots estimated default probabilities, conditional loss rates, and unconditional loss rates for a range of values of each of these explanatory variables.

42 We arbitrarily exclude values below the 1st and above the 99th percentiles with the idea that sample values in these regions are more likely errors, and in any case are so extreme as to be of limited interest for our purposes.

Figure 1
Change in Probability of Default by Percentile of Explanatory Variable

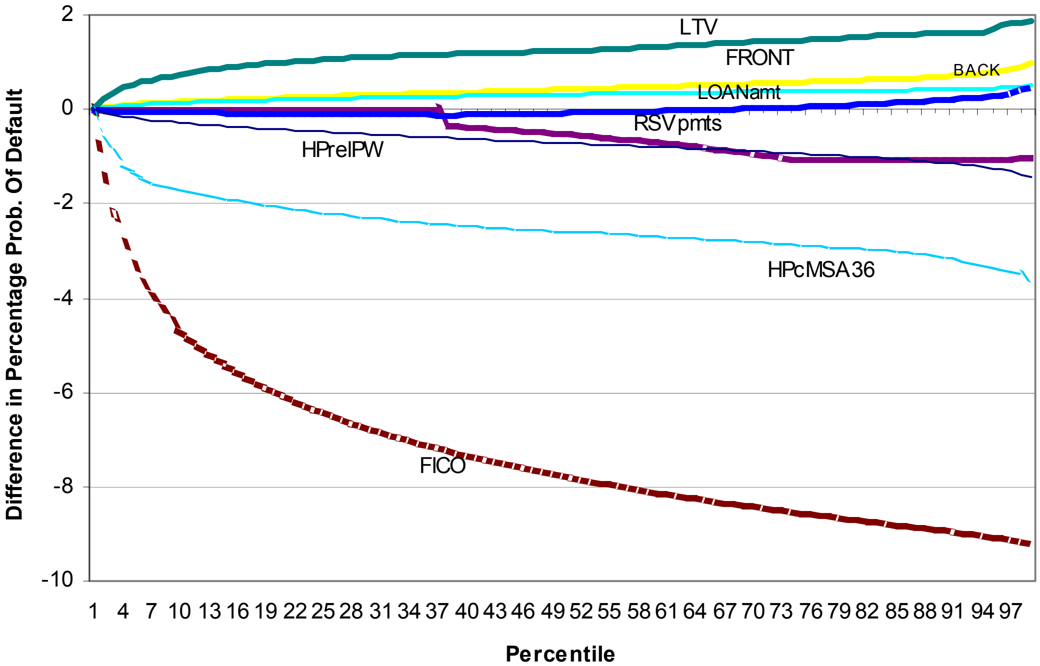


Figure 2
Change in Conditional Loss Rate by Percentile of Explanatory Variable

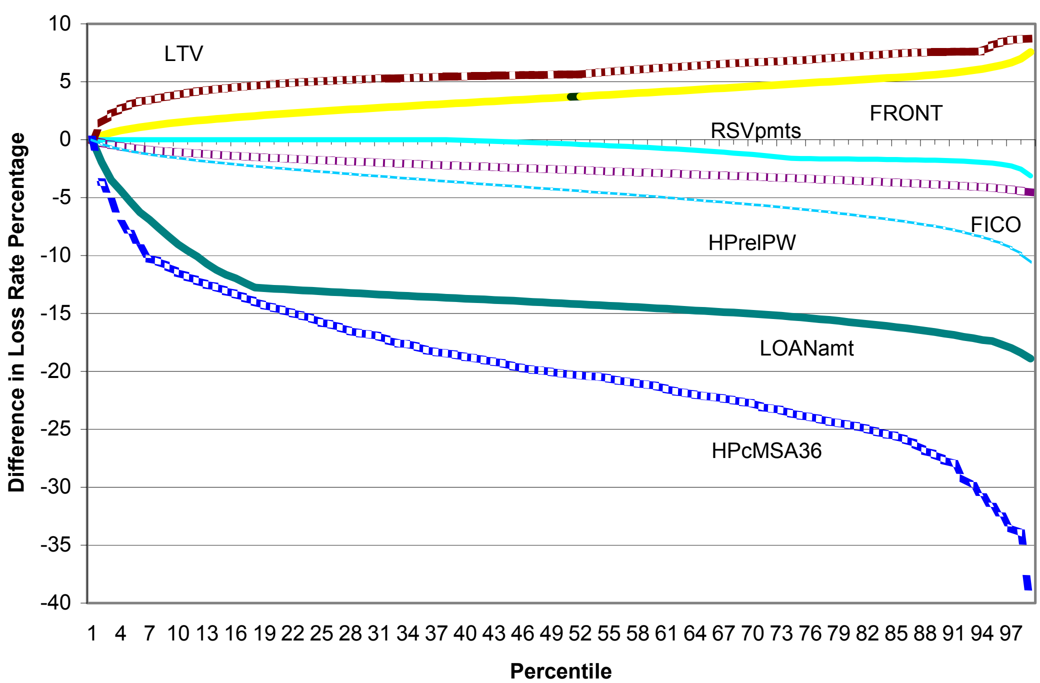
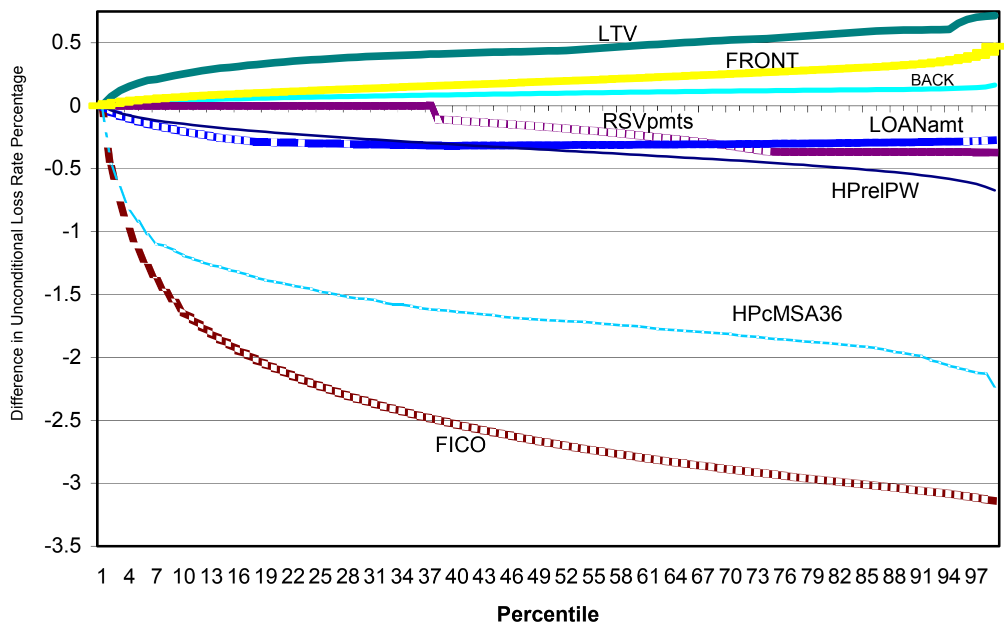


Figure 3

Change in Unconditional Loss Rate by Percentile of Explanatory Variable



Using this particular way of standardizing effects of changes in each variable, the importance of FICO scores in default and unconditional loss rates stands out, as does the importance of house price growth in all risk measures. Notice, moreover, that the major impact of some factors occurs at the extremes of their distribution. We see, for example, that the major effects of the initial loan amount on the loss rate occur at values of the loan amount below the 20th percentile. This way of evaluating the relative importance of various factors clearly depends on the range of data observed in our sample; alternative samples may well give different results.[43]

4.4. Assessments of the Importance of Explanatory Factors in an Underwriting Context

Tables 11 and 12 provide some useful estimates of the impact on default and on loss rates of a variety of factors and give helpful comparisons of relative importance. The estimates imply, for example, that estimated conditional and unconditional loss rates are higher for blacks than for whites and higher for borrowers in underserved areas than for other borrowers. These results suggest that incorporating loss rates into manual underwriting guidelines or automated scoring systems might not improve referral rates for minorities relative to whites. In this section we use the estimated models in a somewhat different way to provide a bit more direct evidence on this point.

In this exercise we use the estimated models to calculate the three risk measures (default probability, conditional loss rate, and unconditional loss rate) for each loan in the sample used in the default estimation. The resulting risk measures are, of course, based on the values of the explanatory variables for these sample loans; the variables consist of FHA data and credit scores that are in principle known at loan origination, plus a variety of data from external sources. The risk measures can be thought of as "scores" that would be yielded by an underwriting process that

[43] In view of the differences between Tables 11 and 12, relative importance also depends on model specification.

uses the parameter estimates obtained in this paper, though with one rather important *caveat*. The models estimated here use external data (*e.g.,* on house price growth) that would not be used in underwriting. In addition, the expanded model includes data on race/ethnicity that would surely be excluded from any underwriting system. We include the expanded model in this exercise for purposes of comparison.

The estimated default probabilities and conditional and unconditional loss rates can each be used to rank the loans according to that particular measure of risk, and as noted, aside from the proviso in the last paragraph, these risk rankings would result from faithfully applying underwriting guidelines built upon our models. For each of the three risk rankings, we have taken the highest (most risky) and the lowest (least risky) 10 percent of the sample loans and have calculated the mean values for a variety of explanatory factors within each risk class. Findings for the basic model are presented in Table 13; findings for the expanded model are in Table 14. In both tables, one set of four columns is devoted to each measure of risk. Within each measure of risk, we record mean values for all explanatory variables for the most risky 10 percent and the least risky 10 percent of the loans, as well as the difference between the means.

Notice that each explanatory variable is included in both tables even if that variable does not appear in the models. In particular, the set of race- and income-related characteristics are excluded from the statistical models underlying Table 13, yet we include those variables in the table so that one can see the mean values for these variables in the high- and low-risk loan categories. Because these variables are not included in the models underlying Table 13, differences in their means are produced by correlations between these variables and other variables that are included in the models.

The figures in Tables 13 and 14 may be used in the first instance to see if a factor is "more important" in separating high risk from low risk loans when the conditional or unconditional loss rate is used to measure risk than when the more standard, default-based risk measure is used. Thus, if we see that the difference in the means between high- and low-risk loans is larger for the loss-rate-based risk measures than for the default-based risk measure, we can say in this limited sense that that variable's "importance" has increased. Naturally, this measure of importance depends not only on the estimates from our statistical models, but also the particular risk classes we have chosen (highest and lowest 10 percent) and the distributions of explanatory variables within this sample. In particular, data from alternative years or alternative samples of borrowers may show very different patterns. These *caveats* aside, we see that many variables exhibit smaller differences in the means for the loss-rate-based risk measures than for the default-based risk measure, and in this sense these variables are less important for the loss-rate-based measures. There are some noteworthy exceptions, however. Most of the house price growth and relative house price variables are considerably more important in the loss-rate-based measures, as is the indicator for underserved areas. Notice also that the black indicator exhibits a slightly higher difference in the means when the unconditional loss rate is used to measure risk than when the default probability alone is used. This finding again suggests that incorporation of loss rates into underwriting criteria—automated or otherwise—is unlikely to provide substantial relief to black applicants.

It is also of interest to compare the figures in Table 13 to the corresponding figures in Table 14. Differences between these two tables seem surprisingly small, even among some of the race-and income-related characteristics, which are (where significant) included in the models in Table 14 but not in the models in Table 13. The black indicator, the percentage black in the area (PCTBLK), and the indicator for underserved areas do appear to exhibit considerably larger differences in means in Table 14 than in Table 13. Even in Table 13, however, the differences in means for these variables seem fairly large for most risk measures. As noted, differences in means for these race- and income-related variables in Table 13 arise solely because these variables are correlated with other variables that are included in the underlying models.

Table 13

Differences in Mean Values of Explanatory Variable in High Risk Vs. Low Risk Loans

Basic Model

Risk Measured by Expected Default Probabilities				Risk Measured by Expected Conditional Loss Rates				Risk Measured by Expected Unconditional Loss Rates			
Credit and Ability-to-Pay Characteristics of Borrowers											
	Low Risk	High Risk	Difference		Low Risk	High Risk	Difference		Low Risk	High Risk	Difference
FRONT	20.03	25.88	5.85	FRONT	22.02	22.20	0.18	FRONT	20.65	25.47	4.82
BACK	31.19	36.53	5.34	BACK	34.69	34.29	-0.40	BACK	32.29	36.25	3.97
RSVpmts	13.975	1.627	-12.348	RSVpmts	8.562	3.487	-5.075	RSVpmts	12.625	1.631	-10.994
NO RSV pmt	0.129	0.620	0.491	NO RSV pmt	0.288	0.466	0.178	NO RSV pmt	0.169	0.613	0.443
FICO	755	607	-147	FICO	708	662	-46	FICO	744	613	-131
Characteristics of the Loan											
	Low Risk	High Risk	Difference		Low Risk	High Risk	Difference		Low Risk	High Risk	Difference
LTV	0.906	0.951	0.045	LTV	0.930	0.943	0.013	LTV	0.914	0.951	0.038
LOANamt	7.15	8.41	1.26	LOANamt	8.61	5.89	-2.73	LOANamt	7.69	8.03	0.34
ARM	0.164	0.346	0.182	ARM	0.262	0.181	-0.081	ARM	0.199	0.328	0.130
TERM15	0.138	0.002	-0.136	TERM15	0.025	0.054	0.029	TERM15	0.100	0.003	-0.097
NOTErt	7.60	7.92	0.32	NOTErt	7.49	8.13	0.65	NOTErt	7.54	7.98	0.44
Characteristics of the Area Housing Market and of the Home Relative to the Area Market											
	Low Risk	High Risk	Difference		Low Risk	High Risk	Difference		Low Risk	High Risk	Difference
HPcMSA36	0.1391	0.0448	-0.0942	HPcMSA36	0.2129	0.0229	-0.1900	HPcMSA36	0.1739	0.0335	-0.1405
HPcST36	0.0404	0.0146	-0.0258	HPcST36	0.0165	0.0367	0.0202	HPcST36	0.0317	0.0167	-0.0150
HPrelPW	0.4291	0.3611	-0.0680	HPrelPW	0.6413	0.2077	-0.4336	HPrelPW	0.5111	0.3309	-0.1802
HPrelLL	0.3708	0.1803	-0.1905	HPrelLL	0.3412	0.1443	-0.1969	HPrelLL	0.3783	0.1647	-0.2136
Race-and Income-Related Characteristics of the Individual and Area											
	Low Risk	High Risk	Difference		Low Risk	High Risk	Difference		Low Risk	High Risk	Difference
BLACK	0.0350	0.2111	0.1761	BLACK	0.0402	0.1762	0.1360	BLACK	0.0296	0.2172	0.1876
HISPANIC	0.0450	0.2773	0.2323	HISPANIC	0.0496	0.1593	0.1097	HISPANIC	0.0407	0.2667	0.2260
INCOME	3.70	3.14	-0.56	INCOME	3.83	2.73	-1.10	INCOME	3.78	3.07	-0.71
PCTBLK	5.34	14.51	9.17	PCTBLK	3.45	15.52	12.07	PCTBLK	4.26	15.50	11.24
PCTHSP	4.98	15.50	10.52	PCTHSP	4.64	10.20	5.56	PCTHSP	4.67	15.03	10.36
NoCensus	0.130	0.092	-0.038	NoCensus	0.196	0.053	-0.142	NoCensus	0.169	0.085	-0.084
TRCT_MSA	0.935	0.855	-0.079	TRCT_MSA	0.871	0.834	-0.037	TRCT_MSA	0.902	0.847	-0.055
UNDERSERVED	0.244	0.531	0.287	UNDERSERVED	0.193	0.569	0.376	UNDERSERVED	0.212	0.553	0.341
Other Characteristics											
	Low Risk	High Risk	Difference		Low Risk	High Risk	Difference		Low Risk	High Risk	Difference
JUDforecl	0.428	0.313	-0.115	JUDforecl	0.114	0.717	0.603	JUDforecl	0.295	0.389	0.094
suburban	0.380	0.456	0.075	suburban	0.559	0.274	-0.285	suburban	0.434	0.417	-0.017
INTchg36	-0.627	-0.996	-0.369	INTchg36	-0.627	-0.882	-0.255	INTchg36	-0.600	-0.981	-0.382
CA	0.002	0.418	0.416	CA	0.027	0.154	0.127	CA	0.003	0.387	0.384
UNEMPC	-21.969	-20.375	1.594	UNEMPC	-24.053	-18.782	5.271	UNEMPC	-22.926	-19.924	3.002
CONDO	0.016	0.042	0.026	CONDO	0.023	0.022	-0.001	CONDO	0.017	0.040	0.024

Table 14

Differences in Mean Values of Explanatory Variable in High Risk Vs. Low Risk Loans

Expanded Model

Risk Measured by Expected Default Probabilities

	Low Risk	High Risk	Difference
Credit and Ability-to-Pay Characteristics of Borrowers			
FRONT	19.99	25.80	5.80
BACK	31.09	36.53	5.43
RSVpmts	13.951	1.621	-12.329
NO RSV pmt	0.134	0.620	0.487
FICO	754	609	-145
Characteristics of the Loan			
LTV	0.905	0.951	0.045
LOANamt	7.15	8.40	1.25
ARM	0.164	0.343	0.179
TERM15	0.138	0.002	-0.136
NOTErt	7.61	7.92	0.31
Characteristics of the Area Housing Market and of the Home Relative to the Area Market			
HPcMSA36	0.1426	0.0461	-0.0965
HPcST36	0.0367	0.0137	-0.0230
HPrelPW	0.4341	0.3606	-0.0736
HPrelLL	0.3609	0.1810	-0.1799
Race-and Income-Related Characteristics of the Individual and Area			
BLACK	0.0212	0.2752	0.2540
HISPANIC	0.0408	0.2843	0.2434
INCOME	3.76	3.13	-0.63
PCTBLK	4.11	19.45	15.34
PCTHSP	4.30	17.09	12.79
NoCensus	0.139	0.077	-0.061
TRCT_MSA	0.962	0.832	-0.130
UNDERSERVED	0.196	0.610	0.413
Other Characteristics			
JUDforecl	0.430	0.313	-0.117
suburban	0.391	0.439	0.048
INTchg36	-0.624	-0.991	-0.367
CA	0.002	0.415	0.413

Risk Measured by Expected Conditional Loss Rates

	Low Risk	High Risk	Difference
Credit and Ability-to-Pay Characteristics of Borrowers			
FRONT	22.47	22.29	-0.18
BACK	34.95	34.27	-0.69
RSVpmts	7.938	3.507	-4.431
NO RSV pmt	0.304	0.476	0.173
FICO	705	660	-45
Characteristics of the Loan			
LTV	0.928	0.944	0.015
LOANamt	8.53	6.12	-2.41
ARM	0.272	0.171	-0.101
TERM15	0.025	0.052	0.027
NOTErt	7.48	8.13	0.66
Characteristics of the Area Housing Market and of the Home Relative to the Area Market			
HPcMSA36	0.2153	0.0321	-0.1832
HPcST36	0.0134	0.0292	0.0158
HPrelPW	0.5988	0.2539	-0.3448
HPrelLL	0.3515	0.1468	-0.2047
Race-and Income-Related Characteristics of the Individual and Area			
BLACK	0.0266	0.3131	0.2865
HISPANIC	0.0516	0.1515	0.0999
INCOME	3.65	2.88	-0.77
PCTBLK	2.40	29.61	27.21
PCTHSP	4.93	9.61	4.68
NoCensus	0.151	0.025	-0.126
TRCT_MSA	1.051	0.758	-0.293
UNDERSERVED	0.089	0.789	0.700
Other Characteristics			
JUDforecl	0.149	0.682	0.532
suburban	0.539	0.299	-0.240
INTchg36	-0.633	-0.866	-0.233
CA	0.042	0.138	0.096

Risk Measured by Expected Unconditional Loss Rates

	Low Risk	High Risk	Difference
Credit and Ability-to-Pay Characteristics of Borrowers			
FRONT	20.87	25.40	4.53
BACK	32.40	36.22	3.82
RSVpmts	12.135	1.652	-10.483
NO RSV pmt	0.183	0.609	0.426
FICO	741	614	-127
Characteristics of the Loan			
LTV	0.913	0.951	0.038
LOANamt	7.68	8.01	0.33
ARM	0.201	0.317	0.117
TERM15	0.096	0.003	-0.093
NOTErt	7.55	7.98	0.43
Characteristics of the Area Housing Market and of the Home Relative to the Area Market			
HPcMSA36	0.1809	0.0366	-0.1443
HPcST36	0.0267	0.0151	-0.0116
HPrelPW	0.5135	0.3406	-0.1729
HPrelLL	0.3698	0.1645	-0.2053
Race-and Income-Related Characteristics of the Individual and Area			
BLACK	0.0168	0.3017	0.2848
HISPANIC	0.0396	0.2714	0.2317
INCOME	3.74	3.06	-0.68
PCTBLK	2.89	22.62	19.73
PCTHSP	4.35	16.28	11.93
NoCensus	0.160	0.068	-0.092
TRCT_MSA	0.993	0.802	-0.190
UNDERSERVED	0.130	0.668	0.538
Other Characteristics			
JUDforecl	0.307	0.387	0.080
suburban	0.443	0.406	-0.037
INTchg36	-0.597	-0.972	-0.375
CA	0.005	0.373	0.368

4.5. Implications of the Estimates: Accounting for Changes in Risk Between 1992 and 1996

In this section we use the models developed in this paper to account for the changes in each of the three risk measures between 1992 and 1996. That is, we decompose each change into the amount that is associated with changes in each of the explanatory variables. In addition to providing a useful decomposition of changing risk between loan cohorts at these two points in time, this analysis will provide another measure of importance for each of the risk factors, though one which may be heavily tied to the particular years we have chosen to compare. A different choice of years might yield a very different set of primary risk factors.

The method we use to account for changes in default rates is as follows. We calculate the default probability for a representative 1992 borrower who has all variables set to the 1992 mean values. These means are calculated within the sample of 1992 loans used to estimate the default probability model. We then change the value of one explanatory variable to its 1996 mean and recalculate the default probability for this individual. The difference in the default probabilities is identified as the change in the default probability between the1992 and 1996 cohorts that is accounted for by that explanatory factor. This process is repeated for each explanatory factor.[44]

We use a similar method to account for changes between 1992 and 1996 in conditional loss rates. That is, we calculate the conditional loss rate for a representative defaulted 1992 borrower who has all variables set to 1992 means (in this case, calculated within the loan loss sample that is used to estimate the conditional loss rate model). We then change one explanatory variable to its 1996 mean (in the loan loss sample) and recalculate the conditional loss rate. The difference between the loss rates so calculated is identified as the change in the conditional loss rate accounted for by that explanatory variable. Again, the process is repeated for each explanatory variable.

The accounting for changes in unconditional loss rates relies on the latter two sets of calculations. In particular, we use the default probability calculated with 1992 means in the full sample, together with the conditional loss rate calculated with 1992 means in the loan loss sample, to form the 1992 benchmark unconditional loss rate. We then use the default probability calculated with one factor set to its 1996 mean in the full sample, together with the conditional loss rate calculated with the same factor set at its 1996 mean in the loan loss sample, to calculate an unconditional loss rate that reflects one of the changes for 1996. The difference between the latter and the 1992 benchmark is identified as the change in the unconditional loss rate associated with that factor.

It should be emphasized that this method of accounting for change has its deficiencies even if all the statistical models used herein are correct. One problem is that logit default probabilities are nonlinear in the explanatory variables. Changing one variable at a time to its 1996 mean fails to recognize the implicit interactions with other variables that arise out of nonlinearity of the logit function. In addition, the average default probability depends on not simply the mean of each characteristic, but also in a complex way on other aspects of the distribution of each factor. The use of mean values fails to reckon with heterogeneity. Even for the conditional loss rate calculations, which do not share the latter problems, there is no guarantee that our method of variable-by-variable accounting for change from 1992 to 1996 will come close to the observed change.

With these drawbacks in mind, consider Tables 15 and 16, which utilize the basic model and the extended model, respectively, to carry out the accounting. In each table we present the 1992 and 1996 means for each variable in the full (default rate) sample and in the loan loss

---

[44] The choice of 1992 means for all other variables, rather than 1996 means, for the benchmark representative individual is completely arbitrary.

sample,[45] and we provide the implied changes in the default probability and the conditional and unconditional loss rates. We also calculate subtotals for changes within each category of variable and, at the bottom of the table, we give the total of all changes in a row labeled "explained, " as well as the actual change from 1992 to 1996 in the corresponding rates. Notice in this regard that because we are now using discounted losses in calculating loss rates, the loss rates for 3-year defaults in 1992 and 1996 are no longer those reported in Table 3. As indicated above in Section 4.2.1, the discounted loss rate for 3-year defaults in 1992 applications is 40.52 percent and in 1996 applications is 37.76 percent.

Looking first at the changes in default probabilities in Table 15, we see that only 0.69 percentage points of the actual 1.859 percentage point rise in default probabilities is accounted for by our method. Of this, most is traceable to the decline in FICO scores from 1992 to 1996; no other factors come close. For conditional loss rates, our accounting substantially overpredicts the decline in loss rates. Here the major factor in the explanation—again by far—is changes in house prices. In the final column, we see that only 0.185 percentage points of the actual increase of 0.632 percentage points in the unconditional loss rate is accounted for by our calculations. Changes in FICO scores again appear to be a major contributor. A similar picture is presented in Table 16.

[45] We do not provide means for the house price growth and relative house price variables because each of these is measured by two or more variables that are considered together. For example, we isolate the effect of changes in house prices by changing HPcMSA36 and HPcST36 together from their 1992 to their 1996 mean values.

Table 15

Accounting for Changes in Default Rates and Conditional and Unconditional Loss Rates Between 1992 and 1996

Basic Model

	Variable	Mean in Full Sample		Mean in Loan Loss Sample		Implied Percentage Point Change 1992 to 1996		
		1992	1996	1992	1996	Def. Prob	Cond. Loss	Uncond. Loss
Credit and Ability-to-Pay Characteristics of Borrowers	FRONT	22.01	23.10	23.70	24.51	0.046	0.092	0.020
	BACK	34.44	35.45	35.11	36.74	0.012	0.000	0.005
	RSVpmts	5.69	5.00	3.64	2.27	0.040	0.190	0.019
	FI CO	693	671	654	625	0.503	0.688	0.213
Subtotal						0.601	0.970	0.256
Characteristics of the Loan	LTV	0.9336	0.9415	0.9407	0.9515	0.042	0.187	0.020
	LOANamt	7.15	8.20	7.49	7.92	-0.006	-0.458	-0.010
	ARM	0.200	0.244	0.221	0.271	0.020	0.127	0.010
	TERM15	0.037	0.023	0.012	0.005	0.017	0.000	0.007
	NOTErt	7.87	7.33	7.94	7.45	-0.105	-0.625	-0.051
Subtotal						-0.031	-0.769	-0.024
Characteristics of the Area Housing Market & of the Home Relative to the Area Market	Changes in home prices	--	--	--	--	-0.181	-7.567	-0.181
	Relative house prices	--	--	--	--	0.107	-0.071	0.041
Subtotal						-0.074	-7.638	-0.141
Other Characteristics	JUDforecl	0.443	0.408	0.416	0.407	0.000	-0.074	-0.001
	SUBURBAN	0.412	0.492	0.416	0.465	0.000	-0.156	-0.003
	INTchg36	-0.506	-0.995	-0.499	-0.994	0.129	1.096	0.070
	CA	0.070	0.125	0.256	0.232	0.066	0.000	0.026
	UNEMPC	-24.40	-21.42	-23.08	-21.74	0.000	0.063	0.001
	CONDO	0.024	0.022	0.042	0.024	0.000	0.042	0.001
Subtotal						0.196	0.972	0.094
Summary	*Explained*					0.690	-6.465	0.185
	Actual					1.859	-2.763	0.632

Table 16

Accounting for Changes in Default Rates and Conditional and Unconditional Loss Rates Between 1992 and 1996

Expanded Model

	Variable	Mean in Full Sample		Mean in Loan Loss Sample		Implied Percentage Point Change 1992 to 1996		
		1992	1996	1992	1996	Def. Prob	Cond. Loss	Uncond. Loss
Credit and Ability-to-Pay Characteristics of Borrowers	FRONT	22.01	23.10	23.70	24.51	0.028	0.210	0.015
	BACK	34.44	35.45	35.11	36.74	0.014	0.000	0.005
	RSVpmts	5.69	5.00	3.64	2.27	0.038	0.151	0.017
	FICO	693	671	654	625	0.479	0.540	0.202
Subtotal						0.559	0.902	0.240
Characteristics of the Loan	LTV	0.9336	0.9415	0.9407	0.9515	0.040	0.220	0.020
	LOANamt	7.15	8.20	7.49	7.92	0.034	-0.562	0.004
	ARM	0.200	0.244	0.221	0.271	0.019	0.077	0.009
	TERM15	0.037	0.023	0.012	0.005	0.016	0.000	0.006
	NOTErt	7.87	7.33	7.94	7.45	-0.103	-0.318	-0.046
Subtotal						0.007	-0.583	-0.006
Characteristics of the Area Housing Market & of the Home Relative to the Area Market	Changes in home prices	- -	- -	- -	- -	-0.180	-7.936	-0.186
	Relative house prices	- -	- -	- -	- -	0.086	0.001	0.034
Subtotal						-0.094	-7.935	-0.151
Race- & Income-Related Characteristics of the Individual and Area	BLACK	0.093	0.110	0.182	0.180	0.004	-0.003	0.001
	HISPANIC	0.084	0.154	0.149	0.245	0.000	0.114	0.002
	INCOME	3.277	3.422	3.125	3.078	-0.012	-0.065	-0.006
	UNDERSERVED, tract race and income	- -	- -	- -	- -	0.035	0.386	0.020
Subtotal						0.027	0.432	0.018
Other Characteristics	JUDforecl	0.443	0.408	0.416	0.407	0.000	-0.066	-0.001
	SUBURBAN	0.412	0.492	0.416	0.465	0.000	-0.087	-0.001
	INTchg36	-0.506	-0.995	-0.499	-0.994	0.126	1.259	0.072
	CA	0.070	0.125	0.256	0.232	0.064	0.000	0.025
Subtotal						0.190	1.106	0.095
Summary	*Explained*					0.688	-6.078	0.195
	Actual					1.859	-2.763	0.632

SECTION 5

CONCLUSIONS

The evidence herein suggests that a variety of characteristics of the borrower, the lender, and the market affect loss rates. Increases in the front-end ratio, LTV, the note rate, and borrower incomes are associated with increases in conditional loss rates. Increases in FICO scores, reserve payments, loan amounts, house price growth, relative house prices, and tract incomes are associated with lower conditional loss rates. Blacks, Hispanics, and those in judicial foreclosure states and underserved areas have higher conditional loss rates, other things the same.

Although there is a link between default and loan loss, there are differences in the importance of factors affecting these two dimensions of loss. Some factors affecting the probability of default—such as back-end ratios and short-term mortgages—seem to have little discernible impact on loss rates. Using the difference between predicted values at the first and third quartiles of a continuous explanatory variable as a standard for comparison, there are marked differences in the relative importance of FICO scores, on the one hand, and house price growth and relative house price, on the other hand. FICO scores appear to be more important in determining default; house price growth and relative house price appear to be more important in determining loss rates. These findings seem to hold when looking over a wider range of values as well.

Simple statistical procedures and more sophisticated statistical analysis suggest that differences in the timing of default-related events are critical in determining loss rates. Simple comparisons of means illustrate that loss rates tend to decline with increases in time-to-default and to rise with the time spent to complete foreclosure and property disposition. These findings are reinforced when these durations are used in regressions to explain conditional loss rates, and descriptive regressions confirm the importance of these durations in affecting components of loss.

The importance of at least some explanatory factors appears to be traceable primarily to their effects on durations. Evidence from overall loss rates and from a set of descriptive component-by-component analyses suggests that FICOs affect loss rates partly if not mainly through impacts on duration. Higher FICO scores are associated with lengthening the time prior to default and reducing time spent in foreclosure and property disposition. Blacks have higher loss rates that may be traceable mainly to behavioral differences in timing, but perhaps also to differences in some loss components, given timing.

Using data on applicants in 1992 and 1996, we find that default probabilities rose while conditional loss rates fell, but the former effect dominated the latter so that unconditional loss rates rose between these cohorts. The models developed here suggest that declines in FICO scores were the major contributor to the increase in default rates, while increases in house price growth were the major contributor to declines in conditional loss rates.

Finally, the evidence here suggests that it is unlikely that basing underwriting on expected unconditional loss rates, rather than default probabilities, will substantially improve the lot of black applicants. Ranking risks according to unconditional loss rates results in even lower representation of blacks in the low risk category, and higher representation of blacks in the high-risk category, when compared with ranking risk by estimated default probabilities alone. Results from this limited comparison should be viewed as only suggestive, however.

APPENDIX

ADDITIONAL DETAILS ON ESTIMATION

Here we provide technical details and a more complete set of empirical results. Included in this appendix are a complete list of variables and their summary statistics, as well as parameter estimates from the default logits and the conditional loan loss regressions.

Table A-1 provides a complete list of the variables used in any of the statistical models, together with a brief description of each variable. Several comments on details of variable construction are in order here. First, the variables LOANamt and RSVpmts are splined to permit changes in marginal impacts over the range of the explanatory variable. LOANamt is splined at $65,000 (via LOANamt65) when used in the default logit and at $50,000 (via LOANamt5) when used in the conditional loss rate regression. RSVpmts is splined at a value of 4 (RSVpmt4).

Second, the variable HPcMSA36 is used whenever the appropriate MSA-level data are available. When those data are not available, then HPcMSA36 is set to zero, and the state-level variable HPcST36 is used instead, and the indicator STdata36 is set to one. (When HPcMSA36 is used, HPcST36 and STdata36 are set to zero).

Third, the variable HPrelPW is used whenever the PriceWaterhouseCoopers median house price is available. When the latter is not available, HPrelPW is set to zero and HPrelLL is used as long as the area FHA loan limit is not at the continental maximum or minimum. (HPrelLL is set to zero when HPrelPW is used.) If HPrelPW cannot be used and the area FHA loan limit is at one of its continental limits, HPrelLL is set to zero and either LLmax or LLmin is set to one.

The variable NO_TRCT is an indicator variable that is one when (and only when) tract income is unavailable. In that event TRCT_MSA is set to zero.

Finally, NoCensus is an indicator that is set to one when (and only when) Census information on tract racial composition and area type (SUBURBAN) is unavailable. When this occurs, the variables measuring racial composition and area type are set to zero.

Table A-2 gives the means and standard deviations of all variables in the sample used to estimate the default logits and in the sample used to estimate the conditional loss rate regressions. Note that not all variables are used in all specifications. Weighting is used in these calculations to account for unequal probabilities of selection from various strata.

Tables A-3 and A-4 give the parameter estimates for the default logits. Both of these logits are weighted to account for unequal selection probabilities, most of which result from purposeful oversampling of defaults. The choice-based nature of the selection process implies that weighting is necessary to secure statistical estimates with desirable large sample properties (*e.g.,* consistency).

Tables A-5 and A-6 give the parameter estimates for the conditional loss rate regressions that have been corrected for heteroskedasticity (unequal variances in the disturbances). Heteroskedasticity in the disturbances might be anticipated to arise from two sources: from possible variation with loan size and from the censoring process by which losses are observed only for defaulted loans. To detect possible heteroskedasticity, we implemented the Breusch-Pagan test on the ordinary least squares (OLS) residuals using all of the independent variables to explain the standard deviation of the disturbances. Having detected heteroskedasticity, we used the absolute values of the unweighted OLS residuals in auxiliary regression to obtain consistent estimates of the linear relationship between a subset of the explanatory variables and the standard deviation of the disturbance. Finally, we used the estimated standard deviation for each observation, as well as the inverse sampling probabilities, in a final weighted least squares procedure (asymptotically equivalent to GLS) to estimate the conditional loss rate regression. To help give additional assurance against misleading standard errors, we report White's heteroskedasticity-consistent (robust) standard errors.

Figure A-1 plots the estimated default rates, conditional loss rates, and unconditional loss rates for a range of values of selected explanatory variables. In these calculations, all variables except the focus variable are set at the sample means.

Table A-1
Complete List of Explanatory Variables

Variable Name	Variable Description	Comments (Note: Indicator variables are denoted by (I).)
Credit and Ability-to-Pay Characteristics of Borrowers		
FRONT	Front-end ratio (percent)	
BACK	Back-end ratio (debt-to-income) (percent)	
RSVpmts	Assets after closing divided by mortgage payment	Any amount
NO RSV pmt	Assets after closing divided by mortgage payment is zero	(I)
RSVpmt4	Assets after closing divided by mortgage payment (spline)	Amount by which exceeds 4
FICO	Average of borrower credit scores and coborrower credit scores	
Characteristics of the Loan		
LTV	Loan-to-value ratio	Any amount
LTV96	LTV spline	Amount by which exceeds 96
LOANamt	Mortgage amount excluding MIP, divided by 10,000	
LOANamt5	LOANamt spline	Amount by which exceeds 5
LOANamt65	LOANamt spline	Amount by which exceeds 6.5
ARM	Indicates ARM	(I)
TERM15	Term of loan is 15 years or less	(I)
NOTErt	Note rate (%)	
Characteristics of the Area Housing Market and of the Home Relative to the Area Market		
HPcMSA36	Proportional change in house price at 36 months after loan origination	Proportional change calculated from metro. data
HPcST36	Proportional change in house price at 36 months after loan origination	Proportional change calculated from state data
STdata36	Metro. data unavailable	(I)
HPrelPW	House price relative to area median price	Price/PriceWaterhouse median price
HPrelLL	House price relative to area median price	Price/ (loan limit/0.95)
LLmax	Loan limit at continental maximum	(I)
LLmin	Loan limit at continental minimum	(I)
Race- and Income-Related Characteristics of the Individual and Neighborhood		
BLACK	Indicates African-American	(I)
HISPANIC	Indicates Hispanic	(I)
INCOME	Monthly income divided by 1000	
PCTBLK	Percentage of Census tract population that is African-American	
PCTHSP	Percentage of Census tract population that is Hispanic	
NoCensus	Indicates no Census info available for PCTBLK, PCTHSP, or suburban	(I)
TRCT_MSA	Census tract income divided by MSA income	
NO_TRCT	Indicates that info on TRCT_MSA not available	(I)
UNDERSERVED	Underserved area in 1996	(I)
Other Characteristics		
JUDforecl	Judicial foreclosure state	(I)
SUBURBAN	Census tract in suburban area	(I)
INTchg36	Change in the 30-year fixed mortgage rate at 36 months after origination	
CA	Indicator for California	(I)
UNEMPC	Change in state-level unemployment rate in percentage points.	
CONDO	Indicator for condominium	(I)

Table A-2

Sample Means and Standard Deviations of Explanatory Variables

	Full Sample		Loss Sample	
Variable	Mean	St. Dev.	Mean	St. Dev.
FRONT	22.55	6.36	24.17	6.35
BACK	34.96	6.69	35.97	6.38
RSVpmts	5.33	16.54	2.91	10.69
RSVpmt0	0.376	0.484	0.516	0.500
RSVpmt4	3.74	15.89	1.87	10.05
FICO	684	57.91	641	55.26
LTV	0.9387	0.0599	0.9476	0.0468
LTV96	0.0058	0.0081	0.0073	0.0096
LOANamt	7.593	2.759	7.782	3.050
LOANamt5	2.762	2.540	2.984	2.799
LOANamt65	1.676	2.166	1.926	2.394
ARM	0.2416	0.4281	0.2831	0.4505
TERM15	0.0284	0.1661	0.0086	0.0924
NOTErt	7.74	1.06	7.85	1.10
HPcMSA36	0.0926	0.0922	0.0656	0.0869
HPcST36	0.0278	0.0615	0.0219	0.0520
STdata36	0.2137	0.4099	0.1962	0.3972
HPrelPW	0.3940	0.4904	0.3591	0.4198
HPrelLL	0.2903	0.3422	0.2431	0.3150
LLmax	0.0738	0.2615	0.0897	0.2857
LLmin	0.0221	0.1471	0.0103	0.1011
BLACK	0.1083	0.3107	0.1933	0.3949
HISPANIC	0.1132	0.3169	0.2000	0.4000
INCOME	3.337	1.345	3.145	1.297
PCTBLK	9.021	17.706	14.018	22.840
PCTHSP	7.487	13.958	12.218	18.519
NoCensus	0.103	0.303	0.087	0.281
TRCT_MSA	0.914	0.406	0.866	0.373
NO_TRCT	0.110	0.313	0.096	0.294
UNDERSERVED	0.344	0.476	0.486	0.500
JUDforecl	0.429	0.495	0.404	0.491
SUBURBAN	0.444	0.497	0.442	0.497
INTchg36	-0.817	0.510	-0.922	0.531
CA	0.094	0.292	0.249	0.433
UNEMPC	-20.88	10.59	-20.49	9.92
CONDO	0.0248	0.1557	0.0344	0.1824

Table A-3

Logit Model of 3-Year Defaults

Basic Model

Number of obs = 168282
Wald chi2(22) = 8005.96
Prob > chi2 = 0.0000
Pseudo R2 = 0.1176

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
LTV	3.236321	0.3173174	10.199	0	2.614391	3.858252
ARM	0.277804	0.0407832	6.812	0	0.1978701	0.3577371
BACK	0.007318	0.0021544	3.397	0.001	0.0030957	0.011541
FRONT	0.025608	0.0022412	11.426	0	0.0212153	0.0300006
RSVpmts	-0.0993	0.0124558	-7.972	0	-0.1237085	-0.0748824
RSVpmt4	0.099712	0.0127765	7.804	0	0.0746704	0.1247535
NO RSV pmt	0.148678	0.0367033	4.051	0	0.0767411	0.2206156
HPrelPW	-0.67222	0.0636219	-10.566	0	-0.7969167	-0.5475233
HPrelLL	-0.9684	0.0826123	-11.722	0	-1.130318	-0.8064834
LLmax	-0.58618	0.0730039	-8.029	0	-0.729261	-0.4430909
LLmin	-0.82005	0.0973166	-8.427	0	-1.010783	-0.6293094
CA	0.730067	0.0447593	16.311	0	0.6423406	0.8177939
TERM15	-0.7622	0.1208428	-6.307	0	-0.9990448	-0.5253496
HPcMSA36	-3.37867	0.1836813	-18.394	0	-3.738677	-3.018659
HPcST36	-4.02493	0.3973448	-10.13	0	-4.803715	-3.246152
STdata36	0.131735	0.0546244	2.412	0.016	0.024673	0.2387965
FICO	-0.01222	0.0002142	-57.033	0	-0.0126379	-0.0117982
LOANamt	-0.06266	0.0171009	-3.664	0	-0.0961809	-0.0291465
LOANamt65	0.077552	0.0195797	3.961	0	0.0391767	0.1159278
INTchg36	-0.15785	0.0262221	-6.02	0	-0.2092454	-0.1064565
NOTErt	0.125905	0.0170982	7.364	0	0.0923927	0.1594162
LTV96	6.441958	1.795315	3.588	0	2.923205	9.960711
_cons	0.848419	0.3632741	2.335	0.02	0.1364145	1.560423

Table A-4

Logit Model of 3-Year Defaults

Expanded Model

Number of obs = 168282
Wald chi2(29) = 8212.55
Prob > chi2 = 0.0000
Pseudo R2 = 0.1206

def36	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
LTV	3.107511	0.3203605	9.7	0	2.479616	3.735406
ARM	0.2691131	0.0407934	6.597	0	0.1891595	0.3490666
BACK	0.0085137	0.0021593	3.943	0	0.0042815	0.0127459
FRONT	0.0162069	0.0042907	3.777	0	0.0077974	0.0246165
RSVpmts	-0.0959263	0.0124833	-7.684	0	-0.1203931	-0.0714596
RSVpmt4	0.0964519	0.0128008	7.535	0	0.0713628	0.121541
NO RSV pmt	0.1371593	0.0367478	3.732	0	0.065135	0.2091836
HPrelPW	-0.5222322	0.0657528	-7.942	0	-0.6511053	-0.3933591
HPrelLL	-0.7356343	0.0859876	-8.555	0	-0.9041669	-0.5671017
LLmax	-0.4828189	0.0738243	-6.54	0	-0.6275118	-0.3381261
LLmin	-0.6883082	0.0990187	-6.951	0	-0.8823813	-0.494235
CA	0.7127316	0.0477681	14.921	0	0.6191079	0.8063553
TERM15	-0.721664	0.1221828	-5.906	0	-0.9611378	-0.4821902
HPcMSA36	-3.379191	0.1856346	-18.203	0	-3.743028	-3.015354
HPcST36	-3.929978	0.3926168	-10.01	0	-4.699493	-3.160463
STdata36	0.1979439	0.0542792	3.647	0	0.0915586	0.3043292
FICO	-0.0118437	0.0002202	-53.78	0	-0.0122753	-0.011412
LOANamt	-0.0135349	0.021555	-0.628	0.53	-0.0557819	0.0287122
LOANamt65	0.0440513	0.0200027	2.202	0.028	0.0048468	0.0832558
INTchg36	-0.1558547	0.0262984	-5.926	0	-0.2073985	-0.1043108
NOTErt	0.1241719	0.0181286	6.85	0	0.0886405	0.1597033
LTV96	6.26444	1.804773	3.471	0.001	2.727149	9.80173
BLACK	0.131945	0.0421768	3.128	0.002	0.0492801	0.2146099
INCOME	-0.0528804	0.027189	-1.945	0.052	-0.1061698	0.0004089
PCTBLK	0.0038586	0.0007423	5.198	0	0.0024037	0.0053135
PCTHSP	0.0037738	0.0008996	4.195	0	0.0020106	0.0055369
NoCensus	0.1739241	0.1392455	1.249	0.212	-0.098992	0.4468403
TRCT_MSA	-0.3267019	0.0604521	-5.404	0	-0.445186	-0.2082179
NO_TRCT	-0.5405627	0.1436732	-3.762	0	-0.8221569	-0.2589685
_cons	0.8592189	0.3783733	2.271	0.023	0.1176208	1.600817

Table A-5
Conditional Loan Loss Rate Regression
3-Year Defaults

Basic Model

Number of obs = 8348
F(22, 8325) = 90.23
Prob > F = 0.0000
R-squared = 0.2448
Root MSE = .16159

	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
FRONT	0.0011491	0.0003622	3.173	0.002	0.0004392	0.001859
HPrelPW	-0.1108163	0.0110325	-10.045	0.000	-0.1324427	-0.0891899
HPrelLL	-0.1940276	0.0145283	-13.355	0.000	-0.2225066	-0.1655486
LLmax	-0.0358957	0.0129421	-2.774	0.006	-0.0612653	-0.010526
Llmin	-0.1342088	0.017636	-7.610	0.000	-0.1687797	-0.0996378
FICO	-0.0002352	0.000041	-5.730	0.000	-0.0003157	-0.0001548
JUDforecl	0.0860921	0.0052648	16.352	0.000	0.0757718	0.0964124
LTV	0.1735152	0.0708211	2.450	0.014	0.0346882	0.3123423
LTV96	1.118457	0.3234724	3.458	0.001	0.4843701	1.752543
LOANamt	-0.0613092	0.0063203	-9.700	0.000	-0.0736986	-0.0489197
LOANamt5	0.0588867	0.0066782	8.818	0.000	0.0457957	0.0719777
NOTErt	0.0126532	0.002736	4.625	0.000	0.0072899	0.0180164
HPcMSA36	-0.7739209	0.0304101	-25.450	0.000	-0.8335321	-0.7143096
HPcST36	-0.513805	0.0539322	-9.527	0.000	-0.6195255	-0.4080845
Rchg36	-0.022126	0.0049127	-4.504	0.000	-0.0317561	-0.0124958
ARM	0.0259811	0.0069595	3.733	0.000	0.0123387	0.0396234
NoCensus	-0.0573808	0.007456	-7.696	0.000	-0.0719963	-0.0427652
SUBURBAN	-0.0315879	0.0051143	-6.176	0.000	-0.0416131	-0.0215626
RSVpmts	-0.0053179	0.0015142	-3.512	0.000	-0.0082861	-0.0023498
RSVpmt4	0.0051131	0.001582	3.232	0.001	0.0020119	0.0082143
CONDO	-0.0221246	0.0121992	-1.814	0.070	-0.0460381	0.001789
UNEMPC	0.0004711	0.0002351	2.004	0.045	0.0000103	0.000932
_cons	0.658482	0.080263	8.204	0.000	0.5011465	0.8158176

Table A-6

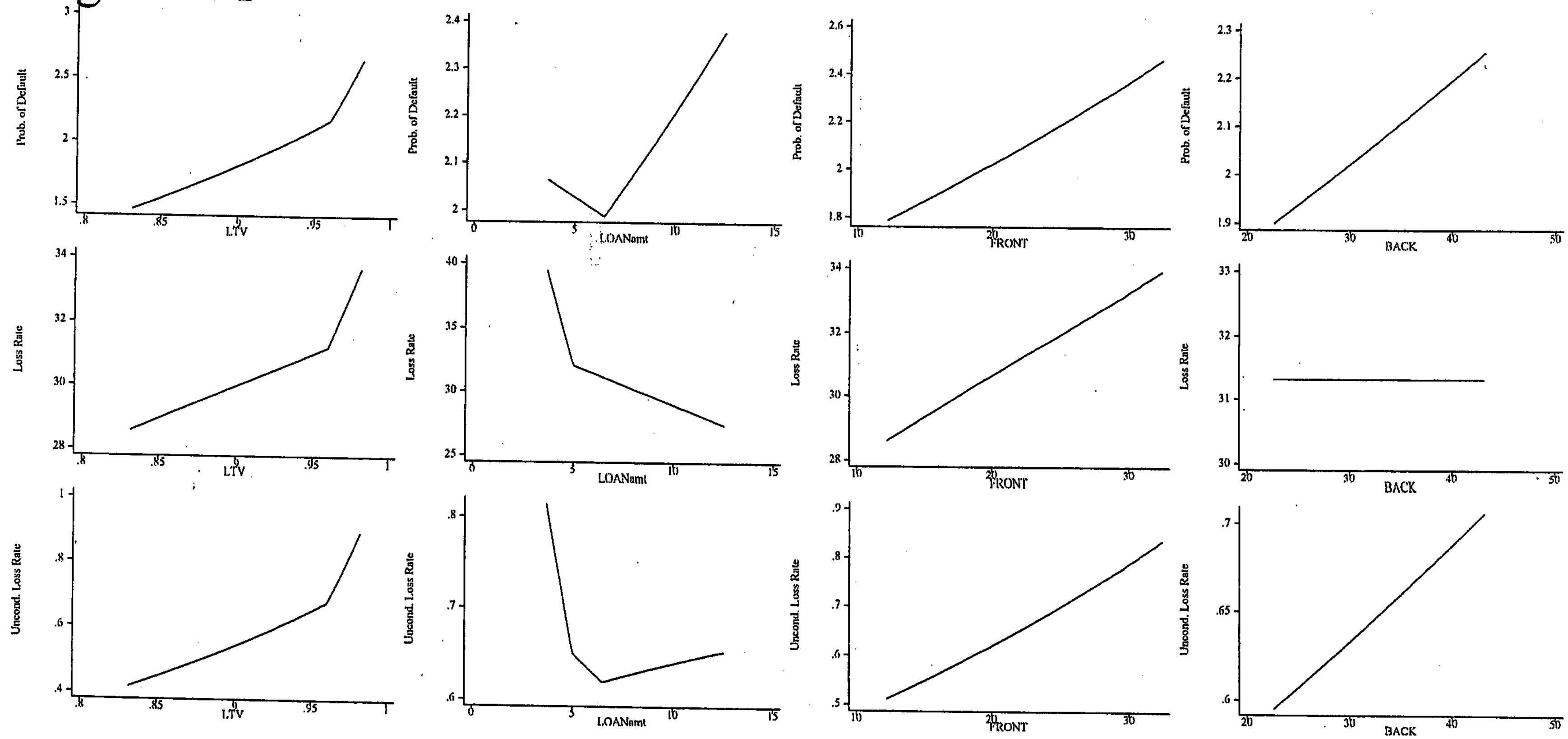
Conditional Loan Loss Rate Regression
3-Year Defaults

Expanded Model

Number of obs = 8348
F(29, 8318) = 86.42
Prob > F = 0.0000
R-squared = 0.2953
Root MSE = .15355

	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
FRONT	0.0026259	0.0006281	4.181	0.000	0.0013946	0.0038571
HPrelPW	-0.076299	0.0107286	-7.112	0.000	-0.0973297	-0.0552683
HPrelLL	-0.1339945	0.0137022	-9.779	0.000	-0.1608542	-0.1071348
LLmax	-0.0348229	0.012486	-2.789	0.005	-0.0592986	-0.0103472
LLmin	-0.0925337	0.0178436	-5.186	0.000	-0.1275116	-0.0575558
FICO	-0.0001847	0.0000382	-4.834	0.000	-0.0002596	-0.0001098
JUDforecl	0.0773477	0.0049242	15.708	0.000	0.067695	0.0870004
LTV	0.2042835	0.0658211	3.104	0.002	0.0752577	0.3333093
LTV96	0.9207909	0.3164902	2.909	0.004	0.3003914	1.54119
LOANamt	-0.0556332	0.006275	-8.866	0.000	-0.0679337	-0.0433327
LOANamt5	0.0494759	0.0061688	8.020	0.000	0.0373834	0.0615683
NOTErt	0.0064256	0.0026722	2.405	0.016	0.0011874	0.0116638
HPcMSA36	-0.8084711	0.0288184	-28.054	0.000	-0.8649623	-0.7519798
HPcST36	-0.6177413	0.0710467	-8.695	0.000	-0.7570105	-0.478472
Rchg36	-0.0254269	0.0047752	-5.325	0.000	-0.0347875	-0.0160664
ARM	0.0155926	0.006386	2.442	0.015	0.0030746	0.0281107
NoCensus	-0.063356	0.023979	-2.642	0.008	-0.1103607	-0.0163513
SUBURBAN	-0.0176955	0.0046731	-3.787	0.000	-0.026856	-0.008535
RSVpmts	-0.0040804	0.001437	-2.840	0.005	-0.0068972	-0.0012635
RSVpmt4	0.0038706	0.0015078	2.567	0.010	0.0009148	0.0068263
TRCT_MSA	-0.1194946	0.0114918	-10.398	0.000	-0.1420214	-0.0969678
NO_TRCT	-0.1047326	0.0253262	-4.135	0.000	-0.1543782	-0.055087
STdata36	0.0301756	0.0095052	3.175	0.002	0.011543	0.0488082
BLACK	0.0137241	0.0069174	1.984	0.047	0.0001643	0.0272839
HISPANIC	0.0118134	0.0066237	1.784	0.075	-0.0011706	0.0247974
INCOME	0.0138533	0.0037193	3.725	0.000	0.0065626	0.0211441
PCTBLK	0.0008418	0.0001507	5.587	0.000	0.0005464	0.0011372
PCTHSP	-0.0006976	0.0001615	-4.319	0.000	-0.0010142	-0.000381
UNDERSERVED	0.0207167	0.0059934	3.457	0.001	0.0089682	0.0324652
_cons	0.6043332	0.0741897	8.146	0.000	0.4589028	0.7497635

Figure A-1



Prob. of Default
Loss Rate
Uncond. Loss Rate
LTV
LOANamt
FRONT
BACK

Figure A-1 cont.