ACTUARIAL RESEARCH CLEARING HOUSE
1988 VOL. 2

SAMPLING INVESTORS AND OTHER DELIGHTS

Thomas N. Herzog

Actuarial Branch
U. S. Department of Housing
and Urban Development

October 1987

SAMPLING INVESTORS AND OTHER DELIGHTS

ABSTRACT

This work describes a study undertaken to determine whether the Federal Housing Administration (FHA) should modify or discontinue its single-family home mortgage insurance program for investor (i.e., non-occupant) loans. Three probability samples were drawn--one for each of endorsement years 1979, 1981, and 1983. The sampled data were analyzed using both Bayesian and sample reuse procedures.

The results indicate an adverse selection problem with investors. In 1979, when house prices and interest rates were both rising, the number of investor loans and their claim rates were relatively low. However, in 1981 and 1983 when housing conditions deteriorated in many parts of the United States, the number of FHA investor loans increased as did their claim rates in comparison to those of owner-occupants.

1. Introduction

1.1 Background

The Federal Housing Administration (FHA) was created in 1934 to encourage improvements in housing standards and conditions, to provide an adequate home financing system, and to exert a stabilizing influence on the mortgage market in the United States. In general, FHA does not make loans or build houses, but instead operates various insurance programs under the National Housing Act. One such program, Section 203(b), provides insurance for private lenders against losses on mortgages financing single-family homes (i.e., one- to four-family dwellings). Thus, under Section 203(b), FHA insures such mortgages against the risk of foreclosure, which arises from the borrower's failure to continue to make his monthly mortgage payments.

When a lender causes an FHA-insured home to be foreclosed, and the home is not worth the amount still owed, the lender has the right to convey the property to FHA in exchange for insurance benefits equal to the sum of the outstanding balance on the mortgage at the time of foreclosure and expenses relating to the foreclosure and claim processes. Such a lender is said to have filed a claim (for insurance benefits) against FHA. A claim can also arise if the lender assigns the mortgage to FHA.

1.2 Purpose

This work was motivated by an earlier work (Bak, Herzog, and Middaugh(1984)) in which we examined the claim rates of Section 203(b) single-family home mortgages as a function of the loan-to-value ratio (i.e., the proportion of the purchase price that is financed). In that work, we found an unusually high claim rate on loans whose loan-to-value ratio was between 80.1% and 85.0%. Since this group of loans includes investor (i.e., non-occupant borrower) loans with a minimum down-payment, we felt a study devoted to such loans was warranted. Such a study could help to determine whether FHA should modify its underwriting standards for investor loans on single-family homes and, if so, to what extent.

1.3 The Loans Examined

We restricted our attention to fully amortizing, level-payment loans having a term to maturity of 30 years, as in our earlier work. Such loans include about 80% of FHA's single-family activity and probably an even larger percentage of its investor loans. Because we were not able to identify individual investor loans on our automated database, we needed to examine individual casebinders* to do so. As a result, we constructed a proxy definition for investor loans and restricted our attention to mortgages which satisifed this proxy definition. The proxy

* A casebinder is a file containing documents on the borrower's creditworthiness, the valuation of the insured property, and the endorsement of the insurance.

definition was suggested by earlier analyses of the data, i.e., Bak, Herzog, and Middaugh (1984), Herzog and Fogel (1987), and Herzog and Stasulli (1987). Finally, because it was expensive to go through each casebinder manually, we sampled about 6,000 casebinders on mortgages endorsed in 1979, 1981, and 1983. We believe this is the first published paper which describes a study of individual loans identified as investor loans.

1.4 An Overview of Sections 2-5

We first present the proxy definition, the sample design, and the procedure used to determine if the mortgagor is an investor. Next, we examine the claim rates and their distribution. Finally, we discuss some additional assumptions and limitations of our analysis.

2. The Data

2.1 The Proxy Definition of Investor Loans

The maximum loan-to-value ratio* for FHA single-family home mortgages held by investors (i.e., those who are not owner-occupants) and insured prior to 1984 is 85 percent of the maximum loan-to-value ratio permitted for owner-occupants. The following table displays some typical maximum loan-to-value ratios:

	Maximum Loan-to-Value Ratio	
Purchase Price	Owner-Occupant	Investor
$25,000	97%	82.45%
$35,000	96.43%	81.97%
$50,000	96%	81.60%
$60,000	95.83%	81.46%

As a result, our proxy definition of investor loan encompasses the following loan-to-value ratio and mortgage amount combinations:

Loan-to-Value Ratio	Mortgage Amount
80.1-83.0%	< $35,000
80.1-82.0%	> $35,000

----

* The loan-to-value ratio is (roughly) the proportion of the purchase price that is financed.

2.2 The Sampling Frames

Three separate list frames--one for each of the endorsement years 1979, 1981, and 1983--were employed. These consisted of 7,946, 10,473, and 30,747 Section 203(b) 30-year term level-payment single-family home mortgages, respectively; all such mortgages satisfied our "proxy" definition of an investor loan. Within each of the three frames, the mortgages were sorted in ascending order of their FHA case numbers.

2.3 Selecting the Mortgages to be Sampled

Systematic probability samples were selected from each frame in the following fashion. For the 1979 endorsements, the 4th, 8th, ..., and 7,944th mortgages were selected, yielding a total sample of 1,986 mortgages. For the 1981 endorsements, the 5th, 10th, ..., and 10,470th mortgages were selected, yielding a total sample of 2,094 mortgages. Finally, for the 1983 endorsements, the 15th, 30th, ..., and 30,735th mortgages were selected, giving us a total sample of 2,049 mortgages.

2.4 The Casebinders

We requested that the casebinders on each of the selected mortgages be sent to HUD headquarters from the Federal Records Center in Suitland, Maryland. Jan Fogel of the Actuarial Branch then examined the HUD FORM 92900.1 in each casebinder to determine whether the mortgagor was an owner-occupant or an investor. The assumption was made that the mortgagor was an owner-occupant if the "occupant" box was checked in item 9B. The mortgagor was assumed to be an investor if one of the following boxes was checked: Landlord, Builder, Operative Builder, or Escrow Commitment. No box was checked on about 20 of the casebinders selected. These were then examined by members of the Underwriting Branch of FHA's Office of Single-Family to see if a firm decision could be made on the type of mortgagor. Using other information in the casebinders, the Underwriting Branch was able to classify all but two of the mortgagors whose casebinders were examined. Of the 6,124 cases in our three samples, 194 cases were "missing" according to the staff of the Federal Records Center. Of the missing 1981 endorsements, 11 were from Fresno and 12 were from Camden, N.J. Because such a large proportion of the missing cases were from two HUD area offices, we decided to contact the HUD staff in these locations to obtain information on these 23 cases. As a result, we were informed that all 11 of the Fresno cases were investor loans, that 8 of the Camden cases were owner-occupants, and that the four other Camden casebinders were missing. The remaining 175 "missing" cases, plus the 2 mortgages we were unable to classify, were distributed by endorsement year as follows:

Endorsement Year	Number of Casebinders Missing
1979	39
1981	64
1983	74

3. The Results

The results of the study are summarized in Table 1. For endorsement years 1981 and 1983, investor loans constitute 54.4% and 53.7% of the casebinders examined, respectively. Thus, we estimate that 5,701 and 16,517 mortgages endorsed in 1981 and 1983, respectively, were actual investor loans with a loan-to-value ratio in excess of 80%. In both instances the claim rate on investor loans is higher than the corresponding claim rate on owner-occupants. Moreover, the results are consistent with our previous analysis (see Table 2) in that the observed claim rate on actual investor loans is in both instances higher than the corresponding claim rate obtained via our "proxy" definition as well as the corresponding claim rate on all Section 203(b) 30-year term level payment loans.

For the 1979 endorsements, only about 18.1% of the sampled cases were determined to be investor loans. Thus, we estimate that only 1,437 of the 1979 endorsements are actual investor loans with a loan-to-value ratio above 80%. Thus, compared to the 1981 and 1983 results, there were few FHA investor loans endorsed in 1979.

The 3.1% claim rate on investor loans endorsed in 1979 is less than both the 4.8% rate of the owner-occupants who satisfied our "proxy" definition and the 6.2% rate on all Section 203(b) 30-year term level payment loans. Why did investors do better than owner-occupants on 1979 endorsements, but worse on 1981 and 1983 endorsements? We have no definite answers, only some hypotheses and/or partial answers. First, there were relatively few investor loans endorsed in 1979. In most parts of the country, single-family houses bought in 1979 experienced some appreciation during their first few policy years. In addition, the assumability of FHA mortgages increased their value substantially as interest rates rose sharply during the early 1980's. These factors probably helped investors more than they did owner-occupants, particularly owner-occupants in older, inner-city houses in Regions 2* and 5. Some of the houses bought in 1981 and 1983 may have been purchased because of what, in retrospect, was unfounded optimism about local housing markets. For example, the number of investor claims in Regions 6 and 9 increased from 13 and 58, respectively, on 1979 endorsements** to 375 and 764 on 1981 endorsements, while the number of investor loans increased by less than 75% in both regions. These regions included such over-built markets as Houston and Las Vegas. Finally, in declining housing markets, investors are more likely to make rational economic decisions to default on their mortgages since they usually have no psychological attachment to the house

---

* The Regions mentioned here consist of the following states:
Region 2 = New York, New Jersey, and Puerto Rico
Region 5 = Illinois, Ohio, Indiana, Michigan, Wisconsin, and Minnesota
Region 6 = Arkansas, Louisiana, Oklahoma, Texas, and New Mexico
Region 9 = Arizona, California, Hawaii, and Nevada
Region 10 = Alaska, Idaho, Oregon, and Washington.

**The claim experience through June 30, 1987, of 1979, 1981, and 1983 endorsements for all 10 HUD Regions and 41 HUD field offices is shown in Table 6 of Herzog and Stasulli[1987].

and may not have invested much money on decorating the house. Moreover, some unscrupulous investors in bad markets may resort to equity skimming to recoup some of their losses.

In summary, the results indicate an adverse selection problem with investors. In 1979, when house prices and interest rates were both rising, the number of investor loans and their claim rates were relatively low. However, in 1981 and 1983 when housing conditions deteriorated in many parts of the United States, the number of FHA investor loans increased as did their claim rates in comparison to those of owner-occupants.

4. Regional Data

In Tables 3 and 4, we summarize our 1981 and 1983 sampled data by HUD Region. In other words, we examine the experience within the areas covered by each of HUD's 10 Regional Offices. Regions 1-3, covering the East Coast from Maine to Virginia, had relatively few claims and investor loans in both endorsement years. For the 1981 endorsements, Regions 5, 6, 9, and 10 all had investor claim rates in excess of 25%. In each instance, the claim rate on investor loans was substantially higher than the corresponding claim rate on owner-occupant loans in our sample. For 1983 endorsements, investor loans seem to be doing particularly poorly in Regions 5 and 10.

5. Estimating the Dispersion of the Claim Rate

We next attempt to estimate the dispersion of the claim rates using two distinct approaches. The first, and the one we prefer, is based on Bayes' Theorem. The other is a frequentist approach* based on the "jackknife" statistic (see, for example, Mosteller and Tukey[1977, pages 133-163]). Sample reuse methods, such as the "jackknife", have recently been made popular by Efron[1982]. Nevertheless, we feel that more insight into the problem at hand is gained by using the Bayesian approach of calculating the (posterior) distribution conditional on the observed data. Specifically, (1) it is instructive to think about the entire distribution, (2) we believe that the observed data are all we have to base our inferences on (in addition to our subjective prior opinions which may be quite diffuse) since we feel it does not make sense to draw repeated sub-samples of our original sample as is done applying the bootstrap, (3) the Bayesian approach forces us to make explicit all of the assumptions used in our model, and (4) we can use the posterior distribution to perform a type of hypothesis testing which makes sense. The last is in contrast to the frequentist type of significance tests (i.e., based on the Neyman-Pearson Lemma) which Deming[1986, page 272] says "have no application here or anywhere".

---

* Hogg and Craig[1978, page 2] calls this the "relative frequency approach".

5.1 The Bayesian Approach

The usual assumption is that the data are realizations of a binomial distribution. Since the Beta distribution is the conjugate prior of the binomial distribution*, we have assumed that the prior distribution is a member of the Beta family of distributions:

$$f(x;a,b) = \begin{cases} \dfrac{\Gamma(a+b+2)}{\Gamma(a+1)\Gamma(b+1)} x^{a}(1-x)^{b} & 0<x<1 \\ = 0 & \text{elsewhere} \end{cases}$$

where a > -1 and b > -1. The non-informative prior of this family is obtained by setting a = -1 and b = -1. An alternate diffuse Beta prior distribution is obtained by letting a be the observed cumulative claim rate of the mortgages satisfying the "proxy" definition of investor loan and b = 1-a. So, for example, for the 1983 endorsements we have a = 0.072 and b = 0.928. Since there is little difference in the results when the non-informative prior is used in place of the above alternate prior, we restrict our attention to the non-informative prior. The results are summarized in Tables 5 and 6 where we present the mode, mean, and standard deviation of the posterior (Beta) distribution for investor and occupant loans, respectively. The mathematical expressions for the last three characteristics are:

posterior mode: $(k-1)/(n-1)$

---

* For a general discussion of this, see Lindley[1970, pages 141-153]. For a discussion of this in an actuarial context, see Herzog[1985].

posterior mean: k/n

posterior standard deviation: $(k)(n-k)/(n^2)(n+1)$

where k is the number of claims observed from a sample of n casebinders. In our opinion, the mode represents the best point estimate of the claim rate, although in most instances of interest here the values of the mode and mean are nearly equal.

We now illustrate how the above results could be used to do "hypothesis testing" within a Bayesian framework. To test the null hypothesis, H_0, that the claim rate on 1981 investor loans, I, is greater than that on 1981 owner-occupant loans, O, versus the alternative hypothesis, H_1, that 1981 owner-occupants have higher claim rates than 1981 investors, we determine the probability that

$$I-O > 0$$

where the Beta density function of I has mean 25.4% and standard deviation 1.31% and that of O has mean 15.0% and standard deviation 1.18%. Assuming I and O are stochastically independent and that both have approximately normal distributions (i.e., that the Central Limit Theorem applies), we find the desired probability to be almost 1.

Instead of using the normal approximation described here, we could alternatively use the approximation based on the F-statistic described in Chapter 7 of Lindley[1970]. A third method is to carry out a stochastic simulation of the difference of two Beta random variables. Herzog[1984] provides a discussion of stochastic simulation in an actuarial environment. Finally, a fourth method is discussed in Novick and Grizzle[1965].

5.2 The Jackknife Approach

The second method of estimating the standard deviations of the claim rates of interest is based on the "jackknife" statistic. The formulas for these statistics are on page 135 of Mosteller and Tukey[1977]. We calculated seven sets of estimates, using 7, 8,..., 13 (independent) replicates. For both investors and owner-occupants, the estimated mean claim rates were virtually identical to the values estimated under the Beta-binomial model of Section 4.1. The standard deviations estimated using 7, 8,..., 13 replicates are shown in Table 7. For both owner-occupants and investors, the jackknife estimates of the standard deviation show wide variations among themselves. For investors, the estimates range from 0.51% to 1.63%; while the owner-occupant estimates go from 0.78% to 1.31%. We were somewhat surprised with this wide range of values. We thought that, given the wide use of the jackknife, it produced estimated standard deviations which were more stable--i.e., less dependent on the number of replicates used--than the results shown in Table 7.

6. Other Assumptions and Limitations of the Analysis

(1) We have calculated the estimated claim rates for the various groups based only upon those casebinders which we have

thus far obtained. Although the number of currently missing casebinders is low, such casebinders represent a disproportionately large number of claims:

Endorsement Year	Number of Claims in Missing Casebinders	Number of Missing Casebinders	Claim Rate
1979	5	39	12.8%
1981	25	64	39.0%
1983	17	74	23.0%

As a result, the overall claim rate of each of the three samples is too low. Because the missing casebinders are not concentrated in any HUD area office, we do not have a good feel for how these cases would alter the relationship between investor and owner-occupant claim rates in any endorsement year.

(2) In constructing the Beta distributions, we implicitly assumed that the claim experience of all the mortgages within a given endorsement year/occupant type grouping was mutually, stochastically independent. We know that in practice this assumption is not true because claim rates are dependent on local, regional, and national economic conditions. Unfortunately, it is not apparent how to model the dependency structure within our stochastic (Beta-binomial) models.

Similar problems arise in the construction of the replicates used to make the jackknife estimates. It is not clear to us to what extent the violation of the assumption of independent replicates affects the results.

(3) Since our "proxy" definition is based on the loan-to-

value ratio and mortgage amount of individual mortgages, we are only able to analyze mortgages which had data on both these characteristics in our database. The percentages of mortgages which could not be classified in this fashion are:

Endorsement Year	Percentage of Cases Lacking Mortgage Amount and/or Loan-to-Value Ratio
1979	15.4%
1981	21.3%
1983	18.9%

We have assumed that these cases are "missing at random" i.e., that the proportion of investor loans, claims, etc., is approximately the same among those "missing" as it is among the population as a whole. The estimates of the number of actual investor loans with loan-to-value ratios above 80% are based only on the mortgages for which both characteristics are present. Hence, these estimates should probably be increased somewhat to account for this other type of "missingness."

(4) We have in the past encountered a small number of data entry problems with the database used in this study--OFA's A-43 Single-Family Insurance System. We are unaware of any large-scale systematic study of the data quality of the A-43 system. Our feeling is that there are a relatively low number of errors and that these have little or no impact on the results of this study. Nevertheless, this is a potential concern to us.

<u>References</u>:

Bak, T.J., Herzog, T.N., and Middaugh, D.A., "15 Million Mortgages: The FHA Experience," <u>Mortgage Banking</u>, November, 1984.

Efron, B., <u>The Jackknife, the Bootstrap, and Other Resampling Plans</u>, SIAM, Philadelphia, 1982.

Deming, W.E., <u>Out of the Crisis</u>, M.I.T., Cambridge, 1986.

Herzog, T.N., <u>An Introduction to Stochastic Simulation, Part 3 Study Note</u>, Society of Actuaries, Chicago, 1984.

__________, <u>An Introduction to Bayesian Credibility</u>, Part 4 Study Note, Casualty Actuarial Society, New York, 1985.

__________ and Stasulli, D.C., <u>Claims by HUD Field Office: Investor Loans vs. Non-Investor Loans</u>, Actuarial Branch, HUD, 1987.

__________ and Fogel, J.R., <u>Mortgage Insurance Risk--Investor and Low-Downpayment Loans</u>, Actuarial Branch, HUD, 1987.

Hogg, R.V. and Craig, A.T., <u>Introduction to Mathematical Statistics</u>, Fourth Edition, Macmillan, New York, 1978.

Lindley, D.V., <u>Introduction to Probability and Statistics From a Bayesian Viewpoint</u>, Part 2, Cambridge University Press, New York, 1970.

Mosteller, F. and Tukey, J.W., <u>Data Analysis and Regression</u>, Addison-Wesley, Reading, Massachusetts, 1977.

Novick, M.R. and Grizzle, J.E., "A Bayesian Approach to the Analysis of Data from Clinical Trials," <u>Journal of the American Statistical Association</u>, Vol. 60, pages 81-96, 1965.

Table 1

Sampled Section 203(b) 30-Year Term Level Payment Mortgages
Satisfying Proxy Definition of Investor Loan

		Investor Loans			Occupant Loans			
Endorsement Year	Number of Casebinders In Sample	Number of Claims	Number of Endorsements	Claim Rate	Number of Claims	Number of Endorsements	Claim Rate	Number of Missing Casebinders
1979	1,986	11	352	3.1%	76	1,595	4.8%	39
1981	2,094	281	1,105	25.4%	140	925	15.1%	64
1983	2,049	82	1,061	7.7%	42	914	4.6%	74

Table 2

Claim Rates on Section 203(b) 30-Year Term
Level Payment Loans

Endorsement Year	Sampled Investor Loans*	"Proxy" Definition**	All Section 203(b) 30-Year Term Level Payment Loans
1979	3.1%	4.2%	6.4%
1981	25.4%	23.1%	16.2%
1983	7.7%	7.2%	6.1%

* From column 5 of Table 1 of this work.

** From Tables 10, 12, and 14 of Herzog and Stasulli[1987].

TABLE 3

SAMPLED SECTION 203(B) 30-YEAR TERM LEVEL PAYMENT MORTGAGES
SATISFYING PROXY DEFINITION OF INVESTOR LOAN FOR 1981

REGION	NUMBER OF CASEBINDERS IN REGION	INVESTOR LOANS			OCCUPANT LOANS			NUMBER OF MISSING CASEBINDERS
		NUMBER OF ENDORSEMENTS	NUMBER OF CLAIMS	CLAIM RATE	NUMBER OF ENDORSEMENTS	NUMBER OF CLAIMS	CLAIM RATE	
1	16	6	0	.00%	9	0	.00 %	1
2	141	13	0	.00	121	7	5.79	7
3	144	64	4	6.25	71	11	15.49	9
4	362	138	25	18.12	215	27	12.56	9
5	232	73	23	31.51	153	32	20.92	6
6	293	182	51	28.02	102	18	17.65	9
7	50	16	3	18.75	34	4	11.76	0
8	119	62	15	24.19	50	12	24.00	7
9	548	403	111	27.54	135	23	17.04	10
10	189	148	49	33.11	35	6	17.14	6
TOTALS	2094	1105	281	25.43	925	140	15.14	64

TABLE 4

SAMPLED SECTION 203(B) 30-YEAR TERM LEVEL PAYMENT MORTGAGES
SATISFYING PROXY DEFINITION OF INVESTOR LOAN FOR 1983

REGION	NUMBER OF CASEBINDERS IN REGION	INVESTOR LOANS NUMBER OF ENDORSEMENTS	INVESTOR LOANS NUMBER OF CLAIMS	INVESTOR LOANS CLAIM RATE	OCCUPANT LOANS NUMBER OF ENDORSEMENTS	OCCUPANT LOANS NUMBER OF CLAIMS	OCCUPANT LOANS CLAIM RATE	NUMBER OF MISSING CASEBINDER
1	23	5	0	.00%	18	0	.00%	0
2	93	34	0	.00	56	0	.00	3
3	193	95	4	4.21	94	0	.00	4
4	388	228	14	6.14	136	7	5.15	24
5	288	95	12	12.63	186	11	5.91	7
6	238	140	15	10.71	89	7	7.87	9
7	75	27	1	3.70	47	0	.00	1
8	167	85	8	9.41	79	7	8.86	3
9	437	259	15	5.79	164	8	4.88	14
10	147	93	13	13.98	45	2	4.44	9
TOTALS	2049	1061	82	7.73	914	42	4.60	74

Table 5

Posterior Distribution of the Proportion
of Claims on Investor Loans of Table 1

Endorsment Year	Mode	Mean	Standard Deviation
1979	2.7%	2.9%	0.92%
1981	25.3%	25.4%	1.31%
1983	7.6%	7.6%	0.82%

Table 6

Posterior Distribution of the Proportion
of Claims on Owner-Occupant Loans of Table 1

Endorsement Year	Mode	Mean	Standard Deviation
1979	4.7%	4.7%	0.53%
1981	15.0%	15.0%	1.18%
1983	4.4%	4.5%	0.69%

Table 7

Estimated Standard Deviations for 1981 Mortgages

Type of Mortgagor	Jackknife Estimate Number of Replicates 7	8	9	10	11	12	13	Beta-Binomial Model
Investor	1.14%	0.51%	1.63%	1.34%	1.07%	1.23%	1.22%	1.31%
Owner-Occupant	1.31%	1.31%	0.78%	1.23%	0.96%	1.28%	0.85%	1.18%