ACTUARIAL RESEARCH CLEARING HOUSE
1997 VOL. 1

THE CASE OF THE MISCODED FHA SINGLEFAMILY MORTGAGE RECORDS

THOMAS N. HERZOG, Ph.D., A.S.A.
FEDERAL HOUSING ADMIN.

The views expressed in this paper are those of the author and not necessarily those of any Federal government agency or the Government of the United States.

1. INTRODUCTION

In this work we describe an analysis of data anomalies on computer records on 246 federally-insured mortgages. Twenty-three of the corresponding mortgages were sampled and all were found to have identical amortization plans and to have been insured under the same insurance fund. The question posed, in probabilistic terms, is the following: How many of the remaining 223 mortgages are identical in these two aspects to the 23 sampled mortgages?

1.1 Background

The Federal Housing Administration (FHA) provides insurance on mortgages. The risk insured against is the inability of the borrower to make timely payments of principle and interest. The FHA consists of four insurance funds, (i) the Mutual Mortgage Insurance Fund (MMIF), (ii) the General Insurance Fund (GIF), (iii) the Special Risk Insurance Fund (SRIF), and (iv) the Cooperative Management Housing Insurance Fund (CMHIF). All of the insurance programs of these funds operate under the provisions of the National Housing Act of 1934 and subsequent statutes enacted by the Congress of the United States.

The Mutual Mortgage Insurance Fund consists of insurance on singlefamily home mortgages under the provisions of Section 203(b) of the National Housing Act. In addition, graduated payment singlefamily mortgages may be insured under the provisions of both Section 203(b) and Section 245(a). Finally, there is a small singlefamily coinsurance program which operates under the provisions of Section 203(b) and Section 244. Under the coinsurance program, the risk of insurance claims is shared by both the FHA and the coinsurer. The premium income on these mortgages is also shared.

Since September 1, 1983, all mortgages insured under the MMIF (except for those coinsured) have been subject to a "one-time premium" charge, due at the origination of the mortgage. Mortgages on condominium units may be insured under the GIF according to the provisions of Section 234, but not under the MMIF. Such mortgages are not subject to the one-time premium charge at origination, but instead pay an annual mortgage insurance premium over the life of the mortgage.

Each FHA-insured singlefamily mortgage is supposed to be represented by a single case record on FHA's singlefamily insurance database. This database includes a field called the ADP Section-of-the-Act Code which indicates the exact program under which each mortgage is insured. Another field in the database indicates whether the mortgage's borrower paid a one-time premium at the origination of the mortgage.

1.2 Statement of the Problem

In the course of some work on FHA's singlefamily coinsurance mortgage program, 1191 mortgage records were discovered to have ADP Section-of-the-Act Code 245, 545, 294, or 594, none of which have the required statutory authority. Of these 1191 mortgage records, 249 were recorded in FHA's singlefamily insurance data system as having a one-time premium amount.

A sample of casebinders on 26 of these 249 mortgages was examined (23 whose ADP Section-of-the-Act Code was 245 and all three whose ADP Section-of-the-Act Code was either 294 or 594). All 23 of the mortgages recorded under ADP Section-of-the-Act Code 245 should have been recorded under Section 203(b)/245(a) of the National Housing Act. This is the section that specifies graduated payment mortgages. Of the other three mortgages whose casebinders were examined, two were under Section 203(b)/245(a) and one was intended to be under Section 234(c)/244 (coinsurance on condominium units), for which there is no statutory authority. There are 223 other cases recorded on FHA's singlefamily insurance data system with ADP Section-of-the-Act Code 245 and a one-time premium amount. The question is as follows: Is it necessary to examine the other 223 casebinders, or could FHA just assume that all of the mortgages in question were insured under Section 203(b)/245(a)?[1]

The above question can be restated as a probabilistic question so that it can be answered in mathematical terms. A description of the methodological approach employed is given in the following section.

1.3 Overview of the Methodological Approach

We know of no frequentist/classical statistical approach that will lead us to an answer, so our only choice is to use Bayesian methods. We assume that

---

[1]Since FHA began insuring graduated payment singlefamily mortgages in 1977, it has insured more than 300,000 such mortgages under Section 203(b)/245(a). It is amazing that so few FHA singlefamily mortgage insurance records ended up under ADP Section-of-the-Act Code 245 because that would have been a natural error to make. Hopefully this scarcity of errors is an indication of the high degree of accuracy of this data element on other records in FHA's singlefamily insurance system.

the data are from a sequence of independent Bernoulli trials with probability of "success" equal to Θ. Thus we assume that we have statistical independence in the sense that the results of one case record do not influence those of any other case record. The likelihood function can be considered to be the binomial. We compute our results under the assumption that the prior density function of Θ is a beta density function $f(\theta|a,b)$. The beta is used for computational convenience because it is the conjugate prior density of the binomial density. We used four pairs of parameters for the beta density in order to test the sensitivity of results to the choice of parameters. We describe the results below.

2. BETA PRIOR DENSITIES AND CONDITIONAL PROBABILITIES

The prior density function, $f(\theta|a,b)$, of Θ is

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

The corresponding mean is $\frac{a}{a+b}$.

For $i = 1,2,\ldots,m+n$, with n and m positive integers, let X_i denote the result of the i^{th} trial where 1 denotes a "success" and 0 denotes a "failure." As shown in Section 8.2.1 of Herzog [1996], if (1) the likelihood of the data is binomial with r successes in n trials and (2) the prior density function of Θ is $f(\theta|a,b)$, then the posterior density function of Θ is $f(\theta|a+r,b+n-r)$. (The mean of the posterior density of Θ is $\frac{a+r}{a+b+n}$.)

In the ensuing derivation of the desired probability, we twice employ the result $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} = (\alpha+\beta-1)\binom{\alpha+\beta-2}{\alpha-1}$ for positive integers α and β. For $x = 0,1,\ldots,m$, we obtain

$$
\begin{aligned}
&P\left[\sum_{i=n+1}^{m+n} X_i = x \,\middle|\, \sum_{i=1}^{n} X_i = r\right] \\
&= \int_0^1 P\left[\sum_{i=n+1}^{m+n} X_i = x \,\middle|\, \Theta = \theta\right] \cdot f(\theta|a+r,b+n-r)\, d\theta \\
&= \int_0^1 \binom{m}{x} \cdot \theta^x (1-\theta)^{m-x} \cdot \frac{\Gamma(a+b+n)}{\Gamma(a+r)\cdot\Gamma(b+n-r)} \cdot \theta^{a+r-1}(1-\theta)^{b+n-r-1}\, d\theta \\
&= \binom{m}{x} \cdot \frac{\Gamma(a+b+n)}{\Gamma(a+r)\cdot\Gamma(b+n-r)} \cdot \int_0^1 \theta^{a+r+x-1}(1-\theta)^{b+n-r-m-x-1}\, d\theta \\
&= \binom{m}{x} \cdot \frac{\Gamma(a+b+n)}{\Gamma(a+r)\cdot\Gamma(b+n-r)} \cdot \frac{\Gamma(a+r+x)\cdot\Gamma(b+n-r+m-x)}{\Gamma(a+b+n+m)} \\
&= \binom{m}{x} \cdot \frac{(a+b+n-1)\cdot\binom{a+b+n-2}{a+r-1}}{1} \cdot \frac{1}{(a+b+n+m-1)\cdot\binom{a+b+n+m-2}{a+r+x-1}} \\
&= \binom{m}{x} \cdot \frac{(a+b+n-1)}{(a+b+n+m-1)} \cdot \frac{\binom{a+b+n-2}{a+r-1}}{\binom{a+b+n+m-2}{a+r+x-1}}\,. \qquad (1)
\end{aligned}
$$

The results for the special case at hand are summarized in Table 1 below. Here we have $n = 23$, $r = 23$, and $m = 223$, so the revised or posterior estimate of the mean of Θ is $\frac{a+r}{a+b+n} = \frac{a+23}{a+b+23}$, since the number of successes is $r = 23$ and the number of initial trials is $n = 23$. In order to assess the sensitivity of the results to different values of the parameters a and b, we employ four pairs of values of a and b.

Using Equation (1), we compute each of the following:

(1) The probability of 223 successes (or, equivalently, no failures) in 223 independent Bernoulli trials.

(2) The probability of at least 219 successes (or, equivalently, no more than 4 failures) in 223 independent Bernoulli trials.

(3) The probability of at least 214 successes (or, equivalently, no more than 9 failures) in 223 independent Bernoulli trials.

because the last integral is that of a constant times a beta density function. Thus we find that the initial probabilities $P\left[\sum_{i=1}^{23} X_i = x\right]$ do not depend on the value of x. In other words, if the prior density function of Θ is the uniform density function over [0,1], then the initial prob-abilities are equidistributed over the range of values $x = 0, 1, \ldots, 23$.

4. CONCLUSION

Based on our assumptions and the ensuing results of Table 1, the probability that all 223 case records represent Section 203(b)/245 mortgages ranges from .071 to .354. Restated, the probability ranges from .646 (which is $1-.354$) to .929 (which is $1-.071$) that at least one of the 223 case records represents a loan *not* insured under Section 203(b)/245(a), but most likely under Section 234(c)/244. The probability ranges from $1-.988 = .012$ to $1-.647 = .353$ that at least ten of the 223 case records represent loans *not* insured under Section 203(b)/245(a).

5. REFERENCE

Herzog, T.N., *Introduction to Credibility Theory*, Second Edition. Winsted: ACTEX Publications, 1996.