MODELLING MORTGAGE INSURANCE CLAIMS EXPERIENCE: A CASE STUDY

BY GREG TAYLOR

December 1991
Revised June 1993
A Coopers & Lybrand research report

ABSTRACT

Mortgage insurance indemnifies a mortage lender against loss on default by the borrower. The sequence of events leading to a claim under this type of insurance is relatively complex, depending not only on the credit worthiness of the borrower but also on a number of external economic factors.

Prominent among these external factors are the loan to valuation ratio of the insured loan, the disposable income of the borrower, and movements in property values. A broad theoretical model of the functional dependencies of claim frequency and average claim size on these variables is established in Sections 6 and 7. Section 8 fits these models, extended by other "internal" variables such as the geographic location of the mortgaged property, to a real data set.

Section 9 compares the fitted model with the data, and finds an acceptable fit despite extreme fluctuations in the claims experience recorded in the data set.

KEYWORDS

Mortgage insurance; housing price index; loan to valuation ratio; regression.

1. INTRODUCTION

Mortgage insurance indemnifies a mortgage lender against loss on default by the borrower. The typical sequence of events leading to the invocation of the indemnity is as follows.

The amount of the mortgage is repayable by a sequence of instalments, perhaps monthly, over a period of some years, up to perhaps 25 or in a few cases more. If a borrower fails to meet one or more of these instalments, arrears collection procedures will be instigated. If it appears that the borrower is experiencing financial difficulties which threaten his capacity to pay the scheduled instalments, the lender's initial response will usually be to attempt rehabilitation of the borrower, possibly by some form of rescheduling of the debt repayment.

In many cases this will render the borrower's difficulties temporary. In other less fortunate cases it will become clear that the borrower is quite unable to repay the debt. The lender will then force sale of the mortgaged property, and retain that part of the sale proceeds required to discharge the remaining debt. In the majority of sales, the proceeds will be sufficient for this purpose, but if they are not the mortgage insurance indemnity is invoked to reimburse the lender for the shortfall.

It is an elementary observation that inflation of property values reduces the call on mortgage insurance; the proceeds of property sales cover a greater proportion of the corresponding debts. It is also clear from the above description that a loan needs to go through several stages (healthy → in arrear → property under management → sale of property) before a mortgage insurance claim arises, and each of these stages involves some delay. As will be discussed in Section 3, each of them also depends on its own specific economic factors.

For these reasons, the underlying process generating mortgage insurance claims is complex and dependent on several variables which are exogenous to the insurance portfolio. Consequently, mortgage insurance run-off arrays, whether in terms of numbers or amounts of claims, exhibit very different characteristics from those of other lines of business. A striking example of this is given in Section 2.

These different characteristics necessitate rather different modelling techniques. The purpose of the present paper is to illustrate these techniques by means of a case study. Since this study is specific to a particular portfolio, it cannot be claimed that the modelling techniques illustrated are generally applicable. It is hoped, however, that they are fairly generally indicative of the **type** of modelling which needs to be attempted.

2. NUMERICAL EXAMPLE: PRELIMINARY DISCUSSION

The following data are given as an indication of the difficulties likely to arise if a mortgage insurance portfolio is subjected to conventional run-off analysis. More detail of the data on which this paper is based appears in Appendices E and G.

Year of loan advance	Number of claims, per 10,000 loan advances, emerging in development year (a)										
	0	1	2	3	4	5	6	7	8	9	10
1980					30	18	6	0	0	0	6
1981				116	42	31	5	0	0	0	
1982			54	27	45	36	13	13	4		
1983		25	20	20	23	9	0	3			
1984	0	13	24	55	35	5	0				
1985	1	21	134	68	15	6					
1986	0	17	30	4	2						
1987	3	1	0	2							
1988	0	0	5								
1989	0	0									
1990	0										

(a) **Development year** is defined as year of emergence of claim minus year of loan advance.

https://doi.org/10.2143/AST.24.1.2005084 Published online by Cambridge University Press

Let the term relative claims frequency denote the number of claims per 10,000 loan advances. If C_{ij} denotes the relative claim frequency in development year j of year of advance i, and A_{ij} denotes the age-to-age factor:

(2.1)
$$A_{ij} = \sum_{k=0}^{j+1} C_{ik} \Big/ \sum_{k=0}^{j} C_{ik}'$$

then the following table of age-to-age factors is obtained.

Year of loan advance i	Age-to-Age factor in development year j =				
	1	2	3	4	5
1984	2.86	2.50	1.38	1.04	1.00
1985	7.12	1.44	1.07	1.03	
1986	2.71	1.08	1.05		
1987	1.00	1.50			

The great instability in these ago-to-age factors is evident in the sense of variability within a development year. The basic reason for the instability is clear from the first table. It is the apparent correlation between relative claim frequency and year of emergence of claim, i.e. with the number of the diagonal in the table. Such a data structure suggests application of the separation method (TAYLOR, 1977, 1986), with the model structure:

(2.2)
$$E[C_{ij}] = r_j \lambda_{i+j}.$$

The separation method yields the following parameter estimates.

j	$\hat{r}_j$	k	$\hat{\lambda}_k$
0	0.00		
1	0.06		
2	0.20		
3	0.22		
4	0.14	1984	366
5	0.11	1985	167
6	0.03	1986	195
7	0.03	1987	350
8	0.02	1988	196
9	0.00	1989	48
10	0.20	1990	29

This produces the following comparison between observed and fitted relative claim frequencies.

Year of loan advance	Observed and fitted (shown in bold type) relative claim frequency in development year																							
	0		1		2		3		4		5		6		7		8		9		10		Total	
1980									30	**52**	18	**18**	6	**6**	0	**9**	0	**3**	0	**0**	6	**6**	60	**94**
1981							116	**79**	42	**24**	31	**21**	5	**11**	0	**5**	0	**1**	0	**0**			195	**140**
1982					54	**72**	27	**36**	45	**28**	36	**38**	13	**6**	13	**1**	4	**0**					193	**181**
1983			25	**21**	20	**33**	20	**42**	23	**50**	9	**21**	0	**1**	3	**1**							101	**169**
1984	0	**1**	13	**9**	24	**38**	55	**76**	35	**28**	5	**5**	0	**1**									131	**159**
1985	1	**1**	21	**11**	134	**69**	68	**42**	15	**7**	6	**3**											245	**133**
1986	0	**1**	17	**20**	30	**38**	4	**10**	2	**4**													53	**73**
1987	3	**1**	1	**11**	0	**9**	2	**6**															6	**28**
1988	0	**1**	0	**3**	5	**6**																	5	**9**
1989	0	**0**	0	**2**																			0	**2**
1990	0	**0**																					0	**0**

The table indicates that the separation method achieves a reasonable fit. No formal goodness-of-fit statistics are examined, because this model is later discarded. The difficulty is that, despite the reasonableness of the fit, the sequence of **escalation index numbers** λ_k is peculiar by normal standards. Until some explanation of this peculiarity is found, it is impossible to produce any reliable projection of the sequence into future years.

One of the major objectives of subsequent sections of this paper will therefore be to obtain such an explanation. The discussion of this aspect of the modelling problem is taken up in Section 3.

3. THE PROCESS OF CLAIM OCCURRENCE

3.1. Major financial factors

As pointed out in Section 1, a loan must traverse several stages of financial deterioration before producing a mortgage insurance claim. These stages are subject to different financial influences. Of these separate influences, two are of particular prominence:

(a) the onset of financial difficulties for the borrower; and
(b) in the event of forced sale, the extent to which the sale proceeds repay the outstanding loan.

These two factors are discussed in the following two sub-sections.

3.2. Onset of borrower's financial difficulties

Despite its importance in a borrower's budget, the mortgage payment instalment will nevertheless be to some extent a residual item in that budget. It will rank after tax and consumer expenditure on necessities (food, clothing, etc.). In addition, most past loans have been of a type whereby the amount of instalment varies with variations in current day interest rates.

It appears, therefore, that a reasonable measure of the degree of financial pressure on mortgage borrowers would be provided by an estimate of the average residual income after allowance for tax, consumer expenditure and mortgage instalment. This residual income, called here the **home affordability index** (HAI), was constructed in the following form:

Home affordability index = average weekly gross household income
minus
tax
minus
consumer expenditure
minus
mortgage instalment,

expressed as a percentage of gross income.

A baseline distribution of gross household income over these categories of expenditure was derived from a 1988/89 household expenditure survey (HES) conducted by the Australian Bureau of Statistics. The items of expenditure for this base year were adjusted to other years in various ways, indicated by the following table.

Item of income or expenditure	Adjustment from year to year according to
Gross household income	Average weekly earnings
Tax	Average weekly earnings (a)
Consumer expenditure	Consumer price index
Mortgage instalments	Average weekly earnings (b) Mortgage interest rates (b)

(a) Preliminary investigation indicated little variation in the effective average tax rate over the period concerned.
(b) The average amount of a new loan was assumed to change in proportion with average weekly earnings. These loans were assumed repayable over periods of 20 years, and the average mortgage instalment calculated on the basis of the most common interest rate charged in the year concerned in respect of the loan portfolio under analysis.

The component time series used in the construction of the HAI (at year end) are set out as Appendix F.

The resulting HAI (at mid-year) is as set out in the following table.

The rather irregular progression of this index is seen in Appendix F to derive from quite reasonable component indexes. Each of these components may be projected over future years, producing a rationally based projection of HAI. This situation may be contrasted with that which arises on application of "black box" estimates of past claims escalation, as in Section 2, and in which no guidance as to future escalation is available.

Year	Home affordability index
1979	100.0
1980	104.8
1981	111.9
1982	101.7
1983	104.1
1984	128.9
1985	128.3
1986	101.7
1987	87.4
1988	90.6
1989	81.5
1990	81.2

3.3. Recovery of outstanding loan on forced sale

The HAI of Section 3.2 provides an indication of the likelihood that an individual borrower will experience financial difficulty in a particular year. However, such difficulty, while a necessary condition, **is not** sufficient for the emergence of a mortgage insurance claim. It is quite possible the borrower's difficulties are such as to force sale of the property, but that property values will be sufficient for the entirety of the outstanding loan amount to be recovered by the lender.

Whether or not this is the case will depend mainly on movements in property values between the date of advance of the loan and the date of the forced sale. In Sydney these movements may be estimated by reference to the **Housing Price Index** (HPI) computed and published by Residex Pty Limited. The following table was derived from that index with slight modification.

Year ended 30 June	Housing price index (Sydney) at mid-year (30/6/79 = 100)
1980	115.3
1981	145.1
1982	158.6
1983	158.4
1984	168.2
1985	177.2
1986	182.4
1987	191.5
1988	245.8
1989	363.5
1990	430.7

https://doi.org/10.2143/AST.24.1.2005084 Published online by Cambridge University Press

Evidently, the greater the increase in value of properties generally, the less the chance that forced sale of a particular property will lead to a loss to the mortgage lender.

3.4. Lags in claims process

While movements in the HAI (Section 3.2) and HPI (Section 3.3) have been identified as major variables in the frequency of mortgage insurance claims, it is to be expected that there will be a lag between cause and effect in each case.

Information from the company operating the mortgage insurance portfolio discussed in this paper was that, broadly:

(a) the average period between mortgage instalments falling in arrears and the property being taken under management (if indeed this latter occurred) was about 6 months; and
(b) the average period between taking a property under management and effecting its sale was also about 6 months.

On the basis of this information, it might be reasonable to expect lags of:

(a) 12 months between movements in the HAI and the consequent movement in claim frequency; and
(b) 6 months between a movement in the HPI and its consequent movement in claim frequency.

Thus, it has been assumed in subsequent modelling that a claim frequency experienced during year t is dependent upon:

(a) the value of the home affordability index at the **middle** of year $t-1$; and
(b) the value of the HPI at the **end** of year $t-1$.

Examination of alternatives suggested that this choice of lags provided about the best fit of model to data. Further detail on the incorporation of the HAI and HPI in the model is given in Section 6.2.

4. DATA

4.1. Variables affecting claims experience

Section 3 identified the HAI and HPI as likely to be major explanatory variables of **claim frequency**. Other variables in this category include:

(a) the proportion of the original property value advanced by way of mortgage, i.e. the **loan to valuation ratio** (LVR);
(b) the geographic area of the mortgaged property (described in more detail in Section 4.2);
(c) the agreed term of the mortgage loan;

https://doi.org/10.2143/AST.24.1.2005084 Published online by Cambridge University Press

(d) the type of property mortgaged (e.g. new house, old unit, land only, etc.);
(e) the financial type of the loan (e.g. reducible loan with variable interest, interest only instalments with fixed interest rate, etc.).

In addition, it is likely that claims experience will vary with **development year**, even in the absence of movements in the HAI and HPI. This would reflect a process of natural selection operating on each year's mortgage advances, whereby the poorest risks succumb to financial pressures relatively early, and the remainder survive the mortgage term.

It is clear that the major variable affecting **claim size** will be the **size of the original loan**. In addition, the explanatory variables (a) to (e) of claim frequency potentially affect claim size also.

4.2. Form of data

As the tables of Section 2 indicate, claims experience relates to the period 1984 to 1990. In fact, the 1984 experience covers only 7 months of that year.

Data supplied in respect of these claims consisted of a claim by claim tabulation, recording in each case the relevant variables identified in Section 4.1:

(a) year of advance;
(b) amount of loan;
(c) value of property;
(d) geographic area of property;
(e) term of loan;
(f) type of property;
(g) financial type of loan;
(h) year of emergence of claim.

The tabulated geographic area was the postal code of the property. These codes were grouped into 14 broad urban and rural regions within the states of New South Wales and Australian Capital Territory, as follows:

> Metropolitan regions 1 to 5; Canberra (6); Newcastle (7); Wollongong (8); Central Coast (9); North Coast (10); South Coast (11); Blue Mountains (12); Southern Highlands (13); Other (14).

The exposure base for the study consisted of all loans advanced over the years 1980 to 1990 inclusive. These were recorded, loan by loan, according to the variables (a) to (g) listed above as potentially affecting claim frequency.

As the data described above constitute a unit record file, it is not practical to present the full detail here. It is not even practical to tabulate cells of data since there are 1499 exposure cells. However, Appendix G gives a tabulation of exposures and claims according to year of advance and development year. It is to be stressed that, while such a tabulation is interesting, it omits a great deal of the raw data.

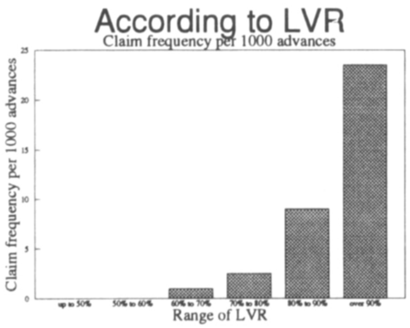
5. EXPLORATORY DATA ANALYSIS

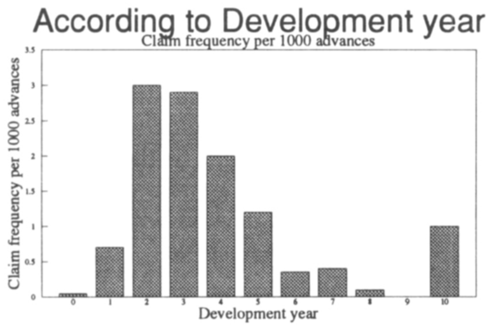
5.1. Claim frequency

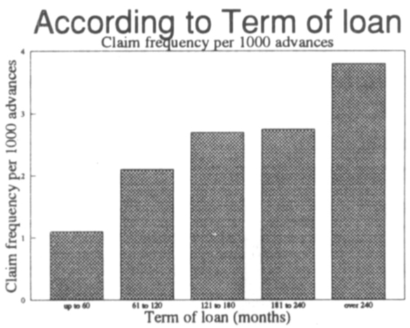
Section 4.1 identified a number of characteristics of individual loans (such as LVR, term of loan, etc.) which might have a bearing on the likelihood of those loans leading to claims. These characteristics will be referred to here as **risk variables**.

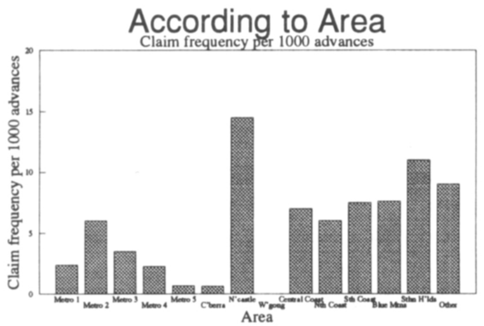
Initially, data concerning claim numbers were analysed according to the risk variables listed in Section 4.1. This provided initial guidance concerning the types of loans which were subject to high or low risk of default.

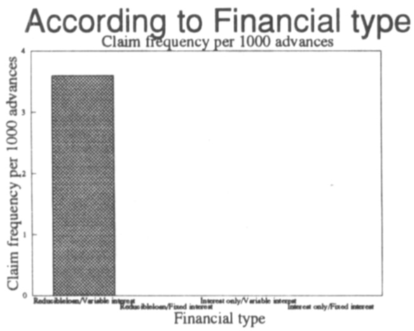
The results of this analysis are summarized in the following sequence of bar charts.

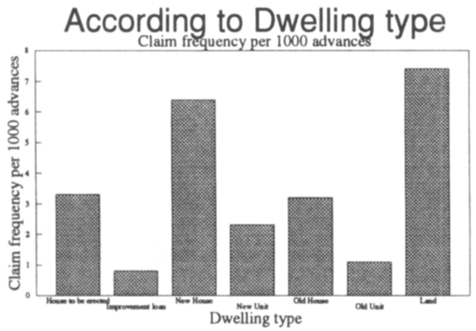












https://doi.org/10.2143/AST.24.1.2005084 Published online by Cambridge University Press

These charts raise the following possibilities:

(a) claim frequency peaks in the second, third and fourth years after the year of advance;
(b) claim frequency increases dramatically with increasing loan to valuation ratio (LVR);
(c) claim frequency increases significantly with increasing term of loan;
(d) certain geographic areas experience conspicuously higher or lower claim frequencies than average;
(e) defaults appear to be confined totally to reducible loans carrying a variable interest rate;
(f) claim frequency appears highest in relation to land, higher in relation to new properties than old, and lowest in relation to improvement loans.

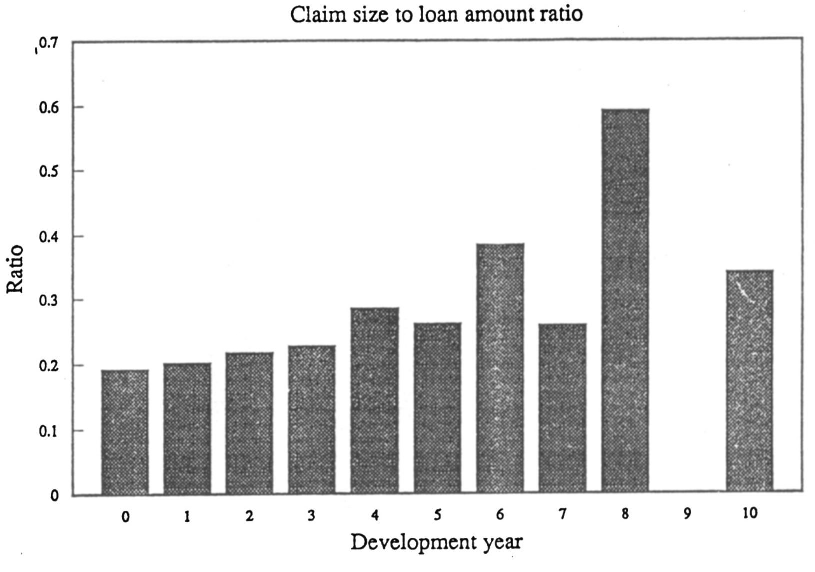
As stated, these are raised as **possibilities** only, rather than conclusions. Without further analysis, it would be impossible to determine whether all of these variables affect the default risk directly, or some of them are merely correlated with the genuinely operative risk variables.

For example, it might be the case that term of loan has no bearing on default risk, but appears to be relevant because LVR does have such a bearing and long terms are associated with high LVRs.

The question of possible correlation between risk variables is remarked upon further in Section 8.1.

5.2. Claim size

Initially, data concerning claim sizes were analysed according to the risk varibles listed in Section 4.1. This provided initial guidance concerning the



https://doi.org/10.2143/AST.24.1.2005084 Published online by Cambridge University Press

types of loans which led to larger or smaller losses when default occurred. The detailed results of this analysis are set out in Appendix D. The results indicate little variation in claim size with any of the risk variables except development year. The variation of claim size with development year is graphed in the preceding chart.

The chart suggests that the greater the time elapsed between advance of loan and default, the greater the **claim size to loan amount ratio**, i.e. the greater the loss on default expressed as a proportion of the original advance. This result is confirmed by formal regression analysis, as described in Section 8.2.

Since growth in property value generally increases with development year, this chart is consistent with the predicted form (7.2) of model.

6. FORM OF CLAIM FREQUENCY MODEL

6.1. General

In the following the basic units of tabulation of claims data will be referred to as **cells**. A cell will consist of an item of data associated with a particular combination of year of advance, development year, and any sub-set of the risk variables identified in Section 4.1.

It is reasonable that the total effect of risk variables on claim frequency should be **multiplicative**, i.e.

(6.1) expected relative claim frequency = function (development year, HAI, HPI)
×
function (risk variables, e.g. LVR, geographic area, etc.).

The form of the first of the two functions on the right will be discussed in Section 6.2. As far as the second function is concerned, a reasonable first approximation would consist of the product of a factor in respect of each of the risk variables present. Equation (6.1) then becomes:

(6.2) expected relative claim frequency = function (development year, HAI, HPI)
×
factor dependent on LVR
×
factor dependent on geographic area
×
etc.

Interactions between the factors making up this product could be added if necessary.

Expected relative claim frequency (per loan advanced) is adjusted by a factor of 7/12 in all cells whose experience relates to 1984. This allows for the fact that the data include only 7 months' claims (Section 4.2).

Some of the risk variables identified in Section 4.1, e.g. financial type of loan, are categorical by nature. Others, e.g. LVR, are continuous by nature. It was convenient for exploratory analysis of the data to convert all variables (i.e. risk variables, not HAI and HPI) to categorical form. Details appear in Section 5.1. The categorical form of data was retained in the final modelling, described in Section 8.1.

6.2. Dependence on development year and economic variables

Preliminary analysis (Section 5.1) indicated that relative claim frequency, expressed as a function of development year, was generally consistent with the shape of a **Hoerl curve**. Appendix B provides a theoretical underpinning of this observation. Consequently, the model adopted for relative claim frequency in the absence of any other effects took the form:

$$\text{const.} \times (j + \tfrac{1}{2})^{\alpha} \exp(-cj), \tag{6.3}$$

where j represents development year.

The modification of (6.3) by HAI and HPI raises some questions. Consider HAI first.

As noted in Section 3.2, the HAI may be regarded as a measure of the average borrower's residual income after payment of mortgage instalment. An individual borrower will experience difficulties in payment of mortgage instalment if this residual income turns negative. The frequency with which this occurs in the event of movements of HAI will depend on the distribution of individual residual incomes, rather than just the average of this distribution represented by HAI. There is virtually no information available in respect of this distribution.

There is, however, some evidence that individual gross incomes are subject to a Paretian distribution (MANDELBROT, 1960).

If a similar assumption is made about residual incomes after payment of mortgage instalment (i.e. HAI), then Appendix A demonstrates that, to first approximation, logged claim frequency will contain a term linear in $R(i+j)/R(i)$, where i denotes year of advance, j development year, and $R(t)$ the HAI experienced in year t. Allowance for the one year lag in the effect of HAI, as discussed in Section 3.4, modifies this term to $R(i+j-1)/R(i)$ (1 in the case $j = 0$).

Because of the approximations leading to this result in Appendix A, an alternative linear term involving

$$\log[R(i+j-1)/R(i)] \qquad \text{for} \qquad j \geq 1;$$

or

$$0, \qquad \text{for} \qquad j = 0, \tag{6.4}$$

https://doi.org/10.2143/AST.24.1.2005084 Published online by Cambridge University Press

was tried. This latter form produced a slightly better fitting regression than the unlogged ratio, and has been adopted henceforth. In fact, both alternatives produced quite similar results.

Appendix B, particularly (B.10), demonstrates that, under seemingly reasonable assumptions about the accumulation of the amount of mortgage debt on default, and about property values on resale, claim frequency should also contain the following factor involving LVR and HPI:

$$L^{v}[H(i+j)/H(i)]^{-v}, \qquad v \text{ const.} > 0,$$

where L denotes LVR and $H(t)$ the HPI experienced in year t. In order to accommodate the lag in the effect of HPI discussed in Section 3.4, this last expression should be modified to the following:

$$L^{v}[H(i+j-½)/H(i)]^{-v}, \qquad j \geq 1;$$

or

(6.5) $$L^{v}, \qquad j = 0,$$

where $H(t-½)$ is interpreted a the HPI experienced at the **end** of year $t-1$.

Note that (6.5) indicates that claim frequency should include the **same power** of both LVR and HPI. However, this result was derived in Appendix B on the assumption that LVR affected the proportion of principal outstanding at default, but not the risk of default itself. In practice, it is likely that LVR is correlated with the ability of the borrower to meet financial commitments, in which case it intrinsically affects the risk of default. For this reason, (6.5) should be generalized to the following:

$$L^{\lambda}[H(i+j-½)/H(i)]^{-v}, \qquad j \geq 1;$$

or

(6.6) $$L^{\lambda}, \qquad j = 0.$$

Combination of (6.2) to (6.4) and (6.6) yields the following model:

(6.7) expected relative claim frequency in development year j of year advance i

$$= \text{const.} \times (j+½)^{\alpha} \exp(-cj)$$
$$\times L^{\lambda}[R(i+j-1)/R(i)]^{-\rho}\ [H(i+j-½)/H(i)]^{-v}$$
$$\times \text{ factor dependent on geographic area}$$
$$\times \text{ etc. for } j \geq 1,$$

and with the two square bracketed terms removed in the case $j = 0$.

Let $\mu(i,j)$ denote the expected relative claim frequency (6.7), and $E(i)$ the number of loans advanced in year i. Let $N(i,j)$ denote the number of claims emerging in development year j of year of advance i. Then the claim frequency model adopted was:

(6.8) $$N(i,j) \sim \text{Poisson}\,[E(i)\,\mu(i,j)].$$

It should be noted that the precise form of dependency of relative claim frequency on LVR and HPI in (6.7) relies upon distributional assumptions made in Appendix B. If these assumptions were varied, the form of (6.7) would change. Consequently, an alternative to (6.7) is considered in Section 8.1, in which the terms involving LVR and HAI are replaced by:

$$\exp(\lambda L)\ \exp[-\nu\, H(i+j-½)/H(i)].$$

This alternative model turns out to be inferior to (6.7).

7. FORM OF AVERAGE CLAIM SIZE MODEL

Appendix C shows that, on the same seemingly reasonable assumptions as in Appendix B (referred to in relation to the development of (6.5)), the average claim size in respect of loans advanced in year i should progress over development years according to the following parametric form:

$$E[Q(i,j)] = \text{const.} \times H(i+j)/H(i), \tag{7.1}$$

where

$Q(i,j)$ = the claim ratio (i.e. ratio of claim size to original loan size) experienced in development year j of year of advance i;

$H(t)$ = HPI experienced during year t.

One may note the interesting effect whereby average claim size **increases** with development year even though outstanding principal is decreasing. Clearly this result derives from the assumptions made in Appendices B and C. Different assumptions would lead to a different parametric form in (7.1). However, an examination of the development of Appendix C indicates that the property of increasing $E[Q(i,j)]$ with $H(i+j)$ derives only from an assumption that the variable γ has a **decreasing failure rate**, where $\gamma = \alpha/\beta$ and

α = a random variable representing the factor by which outstanding principal has been enlarged after default by arrears of principal and interest and any other costs,

β = a random variable representing the factor by which the property value has been reduced by the forced nature of the sale and the associated expenses.

While there is no particular evidence concerning the failure rate of γ, it is interesting to note that the seemingly reasonable assumption of a Pareto distribution leads to the result (7.1) which is found in Section 8.2 to accord with experience, at least to the extent that the claim ratio trends upward with increasing property factor. However, because the Pareto assumption may be a little too specific, it is reasonable to widen the model (7.1) to the following:

$$Q(i,j) = a + b\, H(i+j)/H(i) + \text{error term}, \tag{7.2}$$

where approximately

(7.3) $$\text{error term} \sim N(0, \sigma^2).$$

The appropriateness of this error term is discussed further in Section 8.2.

8. FITTING THE MODEL

8.1. Claim frequency

By (6.7) and (6.8),

(8.1) $$\begin{aligned} \log E[N(i,j)] = {} & \log E(i) + \text{const.} + \alpha \log (j + \tfrac{1}{2}) - cj \\ & + \lambda \log L - p \log [R(i+j-1)/R(i)] \\ & - \nu \log [H(i+j-\tfrac{1}{2})/H(i)] \\ & + \text{term dependent on geographic area} \\ & + \text{etc.}, \; j \geq 1, \end{aligned}$$

with the two square bracketed terms on the right omitted for the case $j = 0$. This linear form, subject to the error structure (6.8), was fitted to the data using GLIM (Generalised Linear Interactive Modelling) (Royal Statistical Society, 1987). Various combinations of the potential explanatory variables listed in Section 4.1 were tried, and the main results are reported in the next table but one.

Geographic area		
Original coding (a)	First aggregation	Second aggregation
1, 4	Area 1	AREA 1
3	Area 3	
5	Area 4	
6	Area 5	
2	Area 2	AREA 2
7, 10-12	Area 6	AREA 3
9, 14	Area 7	
13	Area 9	
8	Area 8	AREA 4

(a) As set out in Section 4.2.

The results of the trial regressions are displayed in the following table.

Variable	Coefficient in variable at left (a) in Regression No.						
	1	2	3	4	5	6	7
Regression constant	−9.505	−12.18	−10.50	−9.848	−12.90	−5.776	−5.943
Development year	−1.093	−1.143	−1.218	−1.097	−1.096	−1.119	−0.8536
Log (development year + ½)	4.908	5.066	4.558	4.906	4.903	5.076	4.505
LVR (d)	1.100	1.144	0.994	1.100	1.099		
Log (LVR)						8.93	8.413
Log (home affordability factor) (b)							−2.158
Property growth factor (c)	−3.039	−3.070	−2.036	−3.017	−3.015		
Log (property growth factor)						−4.636	−5.658
Indicator variables (f):							
AREA 2				0.52	0.52	0.53	0.5131
AREA 3				0.87	0.87	0.87	0.8772
AREA 4				−5.24	−5.24	−5.25	−7.254*
Area 2	0.60						
Area 3	0.16*						
Area 4	−0.35*						
Area 5	−0.26*						
Area 6	1.05						
Area 7	1.15						
Area 8	−5.33*						
Area 9	0.81						
60 ≤ Term < 120 months		3.74*					
120 ≤ Term < 180 months		2.95*					
180 ≤ Term < 240 months		2.00*					
240 ≤ Term		2.74*			3.06*		
Dwelling:							
Improvements & increases			1.33*				
All other than improvements, increases & land only			0.64*				
Dwelling type missing			7.05*				
Deviance (e)	854	549	632	611	610	593	527

(a) Dependent variable in regression log (claim frequency), as in (8.1).
An asterisk attached to a coefficient in the table indicates that this coefficient differs from zero by less than 2 standard errors.

(b) The home affordability factor is the ratio of values of HAI appearing in (8.1).

(c) The property growth factor is the ratio of values of HPI appearing in (8.1).

(d) The variable referred to here is in fact

$$10 \times \text{LVR} - 3.5.$$

The variable log (LVR) uses the genuine LVR, though grouped in ranges of 10 percentage points width. Each such range is represented by its mid-value.

(e) Deviance is a measure of goodness of fit, related to the log likelihood ratio of the model. A lower deviance implies a better fit.

(f) The variables Area *k* and AREA *m* have already been described as 0-1 indicator variables. The variables listed subsequently in the table are also of the 0-1 indicator type, taking the value 1 if the loan is subject to the risk variable displayed, 0 otherwise.

By (6.8) and (8.1), the model is multivariate Poisson with multiplicative structure of the mean. GLIM fits this by maximum likelihood. Note that the logarithmic form of (8.1) is no more than a convenience of expression. It could equally have been written in its unlogged (multiplicative) form. In particular, (8.1) does **not** imply that the observations $N(i, j)$ (many of which are zero) are to logged.

For the interpretation of this table, special reference should be made to geographic area of the mortgaged property. On the strength of the chart of Section 5.1, a number of areas, seemingly similar in claim frequency and/or physically contiguous, were aggregated. The areas at this initial level of aggregation were denoted by "Area k". These were 0-1 variables, taking the value 1 if the property lay in the relevant area, 0 otherwise.

Regression 1 in the table indicated that further aggregation was possible. The new variables resulting from this aggregation were denoted by "AREA m", and were 0-1 variables, each of which consisted of the sum of the relevant variables Area k. The key to the two aggregations is as shown in the previous table but one.

It may be noted that the trial regressions included alternative versions of (8.1) in which the terms dependent on LVR and HPI were replaced by their respective unlogged forms, as discussed at the end of Section 6.2. These alternatives were, however, inferior to (8.1) in terms of fit.

Regression 7 provided the best fit of model to data, and was adopted as the final model. This final model, expressed in non-symbolic form, was as follows:

(8.2)

CLAIM FREQUENCY = (per 1000 advances) IN DEVELOPMENT YEAR t	$2.624\,(t+\tfrac{1}{2})^{4.505}\exp(-0.8536\,t)$
	$\times$
	$(\text{LVR})^{8.413}$
	$\div$
	$[(\text{HOME AFFORDABILITY FACTOR})^{2.158}$
	$\times$
	$(\text{PROPERTY GROWTH FACTOR})^{5.658}]$
	$\times$
	$\begin{cases} 1 \text{ if AREA 1} \\ 1.670 \text{ if AREA 2} \\ 2.404 \text{ if AREA 3} \\ 0.0007 \text{ if AREA 4} \end{cases}$

where

HOME AFFORDABILITY FACTOR and PROPERTY GROWTH FACTOR are the ratios involving H and R respectively in (8.1).

The formula in the box indicates that claim frequency:

(a) moves sharply upward with increasing LVR;

(b) moves sharply downward as property values or disposable incomes after mortgage instalments increase;
(c) varies significantly by geographic area, exhibiting a particularly low value in the Wollongong district.

Because of correlations of the type discussed at the end of Section 5.1, not all of the risk variables exhibited a significant effect on claim frequency.

8.2. Average claim size

The form of the model was suggested in Section 7 as the following:

$$Q(i,j) = a + b\,H(i+j)/H(i) + \text{error term}, \tag{7.2}$$

where approximately

$$\text{error term} \sim N(0, \sigma^2). \tag{7.3}$$

This model appears unnatural to the extent that the normal error term would permit claim sizes to be negative. This would be avoided by the inclusion of an error term which was by nature positive. An example would be a lognormal error term, as would be incorporated in an alternative model of the form:

$$\log Q(i,j) = \log a + b \log [H(i+j)/H(i)] + \text{error term}, \tag{8.3}$$

where

$$\text{error term} \sim N(0, \sigma^2). \tag{8.4}$$

Equivalently,

$$Q(i,j) = a[H(i+j)/H(i)]^b \times \text{error term}, \tag{8.5}$$

where

$$\text{error term} = \text{lognormal}\,(0, \sigma^2). \tag{8.6}$$

Note that both forms (7.2) and (8.5) accommodate the theoretical form (7.1).

Ordinary regression produced the following two alternative models.

Parameter	Unlogged model (a)	Logged model (b)
a	0.1622	0.1555
b	0.0494	0.3083
σ^2	0.0257	0.8676

(a) This is the model described by (7.2) and (7.3). Of the 425 observed claim ratios, 2 large values have been excluded as outliers.
(b) This is the model described by (8.3) and (8.4).

In fact, neither of the two models considered in the preceding table produced an ideal fit to the data. Their respective residuals are tabulated in the following table.

Values of standardized residuals	Relative frequency of standardized residual in	
	Unlogged model	Logged model
	%	%
less than −3	0	1
−3 to −2	0	3
−2 to −1	12	8
−1 to 0	47	32
0 to 1	24	44
1 to 2	10	12
2 to 3	5	0
more than 3	1	0
Total	100	100

These two tabulations of standardized residuals are very much reflections of each other about the origin. While the unlogged model is somewhat skewed to the right, the logged model is about equally skewed to the left. This suggests that the correct model lies somewhere between normal and log normal. Such a model might be of the form (7.2), but with the error term strictly positive and skewed to the right but less so than log normal.

Note that the fitted values of claim ratios, according to the two alternative models, are:

(8.7)
$$EQ(i,j) = a + bH(i+j)/H(i) \text{ for unlogged model;}$$
(8.8)
$$= a[H(i+j)/H(i)]^b \exp(\tfrac{1}{2}\sigma^2) \text{ for logged model.}$$

In the event, (8.8) produced a rather heavy upward bias, about 18% in total, in fitted values of claim amount relative to observed amounts. The form of this comparison was exactly as reported in Section 9.2, but with the unlogged model used there replaced by the logged.

This result appears to indicate that the exponential scaling factor in (8.8) is not robust against the non-normality in the error term of (8.4), as was demonstrated in the above table of standarized residuals.

On the other hand, Section 9.2 indicates that the unlogged model provides an adequate fit, and accordingly it was adopted.

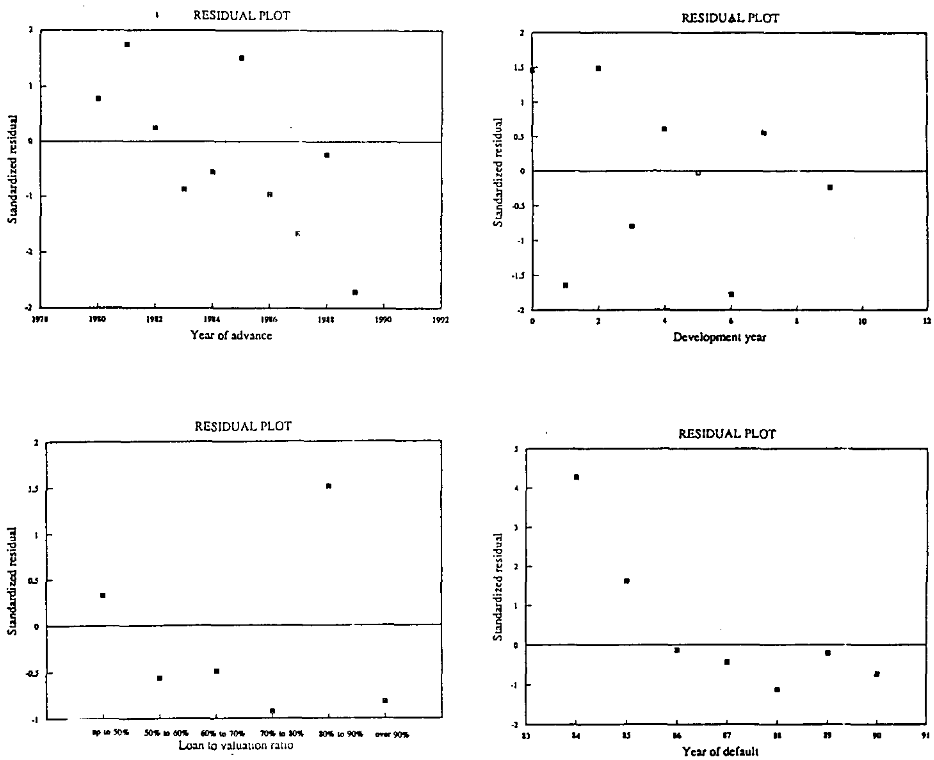
9. MODEL VERIFICATION

9.1. Claim frequency

The model adopted in Section 8.1 has been used to compute standardized residuals according to several variables. The resulting residual plots appear below. Note that each residual relates to the aggregation of all experience at the value of the independent variable displayed. For example, the first residual in the first plot may be obtained from the second table of the present sub-section as:

$$(8-6)/\sqrt{6} = 0.8.$$

A plot of the residuals of all cells (taken over all explanatory variables) would not be helpful since the great majority of cells contain very small expectations.



These plots appear generally satisfactory in terms of magnitude, with the exception of year of default 1984. This one anomaly, in the relatively distant past, involves relatively few claims (see first table below) and is insufficient to invalidate the model.

The plot against year of advance contains a downward trend. If included in the model, year of advance appears as a highly significant explanatory variable; other things equal, claim frequency declines by 29 % as between each year of advance and the next. Naturally, the effects of the other explanatory variables, particularly those which are time dependent, change.

While this model provides a superior fit to the data, the abstract nature of the year of advance effect is problematic. It might be interpreted as a factor representing improvement in underwriting. However, in this case, the total improvement over the decade of underwriting would be almost 97%, which might strain credulity.

It seems more likely that year of advance is acting as a proxy for some other unidentified explanatory variable(s). When this variable is omitted from the model, its effect is largely captured by the other explanatory variables.

Moreover, an examination of the fitted numbers of claims (using the model which omits year of advance effect) against the data suggests that the apparent trend in the residuals may not be particularly meaningful (see second table below).

The following table displays the actual and model numbers of claims underlying the above plot of standardized residuals by experience year.

Period	Number of claims emerging	
	Actual	Model
1984 (7 months)	28	13
1985	32	24
1986	53	54
1987	168	174
1988	103	115
1989	21	22
1990	20	24
Total	425	425

The table illustrates the close agreement between actual and model numbers of claims for all experience years except 1984, despite the extreme fluctuations in numbers of claims.

More detailed information is given by the following table which tabulates experience and model simultaneously by year of advance and development year, and from which the above table may be derived.

Year of loan advance	Observed and fitted (shown in bold type) number of claims in development year											
	0	1	2	3	4	5	6	7	8	9	10	Total
1980					3 **1.8**	3 **1.5**	1 **1.2**	0 **1.2**	0 **0.3**	0 **0.0**	1 **0.0**	8 **6**
1981				13 **4.5**	8 **4.8**	6 **4.4**	1 **4.9**	0 **1.4**	0 **0.1**	0 **0.0**		28 **20**
1982			7 **4.9**	6 **7.6**	10 **8.7**	8 **11.4**	3 **3.5**	3 **0.3**	1 **0.1**			38 **37**
1983		5 **1.6**	7 **5.3**	7 **8.8**	8 **14.7**	3 **5.2**	0 **0.5**	1 **0.2**				31 **36**
1984	0 **0.1**	7 **4.3**	13 **15.5**	30 **37.7**	19 **16.8**	3 **1.8**	0 **0.8**					72 **77**
1985	1 **0.3**	16 **16.2**	104 **86.6**	53 **56.7**	12 **7.6**	5 **3.8**						191 **171**
1986	0 **0.2**	14 **17.1**	24 **24.6**	3 **4.8**	2 **3.1**							43 **50**
1987	3 **0.3**	1 **6.2**	0 **2.7**	2 **2.5**								6 **12**
1988	0 **0.4**	0 **2.7**	8 **5.6**									8 **9**
1989	0 **0.2**	0 **7.1**										0 **7**
1990	0 **0.3**											0 **0**

The following table presents these results in the same format as in Section 2, enabling comparison of the present set of results with those from the separation method.

Year of loan advance	Observed and fitted (shown in bold type) relative claim frequency in development year											
	0	1	2	3	4	5	6	7	8	9	10	Total
1980					30 **18**	18 **9**	6 **7**	0 **7**	0 **2**	0 **0**	6 **0**	60 **43**
1981				116 **41**	42 **25**	31 **23**	5 **26**	0 **7**	0 **1**	0 **0**		195 **122**
1982			54 **38**	27 **34**	45 **39**	36 **51**	13 **16**	13 **1**	4 **0**			193 **179**
1983		25 **8**	20 **16**	20 **26**	23 **43**	9 **15**	0 **1**	3 **1**				101 **109**
1984	0 **0**	13 **8**	24 **28**	55 **69**	35 **31**	5 **3**	0 **1**					131 **140**
1985	1 **0**	21 **21**	134 **111**	68 **73**	15 **10**	6 **5**						245 **220**
1986	0 **0**	17 **21**	30 **30**	4 **6**	2 **4**							53 **62**
1987	3 **0**	1 **6**	0 **3**	2 **3**								6 **12**
1988	0 **0**	0 **2**	5 **3**									5 **5**
1989	0 **0**	0 **6**										0 **6**
1990	0 **0**											0 **0**

9.2. Average claim ratio

For each claim in the experience, a fitted value of its claim ratio was calculated according to (8.7) using the values of *a* and *b* tabulated in Section 8.2. Each of these claim ratios was multiplied by the associated amount of its loan, to produce a fitted claim size.

Observed and fitted claim sizes were then summarized in 2-way tabulations by year of advance and development year. These tabulations are displayed in Appendix E, and reduced to their corresponding 1-way tabulations below.

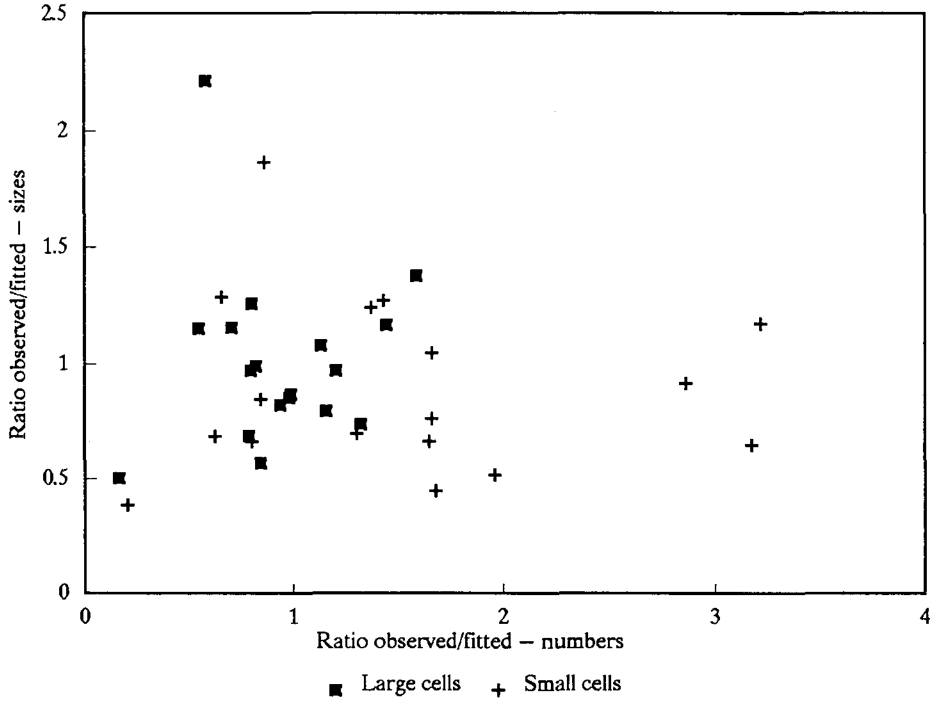
Year of advance	Amount of claims			Development year	Amount of claims		
	Observed	Fitted	Ratio: Observed / fitted		Observed	Fitted	Ratio: Observed / fitted
	$ 000	$ 000	%		$ 000	$ 000	%
1980	51	70	73	0	32	46	70
1981	294	312	94	1	425	471	90
1982	398	374	106	2	1750	1844	95
1983	354	323	110	3	1051	1133	93
1984	632	642	98	4	674	642	105
1985	1931	2063	94	5	321	301	107
1986	425	472	90	6	47	38	124
1987	46	69	67	7	31	35	88
1988	259	222	117	8	56	28	199
1989	0	0		9	0	0	
1990	0	0		10	1	7	14
Total	4388	4545	97		4388	4545	97

https://doi.org/10.2143/AST.24.1.2005084 Published online by Cambridge University Press

It should be particularly noted that the fitted amounts of claims, according to the above description are **conditional** upon the observed numbers of claims. This is a proper approach to examination of the fit of the average claim size model. Agreement between model and data appears satisfactory.

It is useful to carry out some check that the common dependence of the claim frequency and claim size models on the HPI does not lead to unwanted correlation between the two. That this does not in fact occur is indicated by the following scatter plot of the observed fitted ratios of average claim size against a similar ratio for number of claims.

Each point represents a particular combination of year of advance and development year. To give a simple indication of the significance of the plotted points, they are divided into "large cells" and "small cells". The former are those cells containing a fitted number of claims in excess of 5; otherwise the cell is "small".



9.3. Loan sizes associated with claims

While Section 9.2 models the claim size which will arise from a particular loan size **if a claim occurs**, it provides no indication of which loan sizes are likely to lead to claims.

There is no particular reason to believe that the sizes of loans associated with claims will be representative of the entire portfolio of loans advanced. Indeed,

https://doi.org/10.2143/AST.24.1.2005084 Published online by Cambridge University Press

the table below indicates that, on average, it is the larger loans that lead to claims.

Care is needed here, however, as the model of claim frequency in Section 9.1 conditions on LVR and other risk factors, for which average loan sizes may differ from the portfolio average, and so without further analysis it is not clear to what extent the inclusion of these factors in the model will effectively select average loan sizes above the portfolio average. This question is also examined in the following table.

Year of advance	As a percentage of portfolio average loan size		
	average loan size associated with past claims (a)		average loan size weighted by model numbers of future claims (b)
	%		%
1980	135	(8)	96
1981	144	(28)	102
1982	119	(38)	101
1983	116	(31)	102
1984	85	(72)	102
1985	95	(191)	102
1986	144	(43)	103
1987	97	(6)	100
1988	241	(8)	98
Average	109 (c)	(425)	102 (d)

(a) The numbers of claims on which the ratios are based are shown in parenthesis. For each year of advance, the average size of loans associated with **recorded** claims has been calculated and related to the portfolio average (for that year of advance).
(b) For each combination of year of advance and risk variables, the average loan advanced and model claim frequency (according to the model of Section 8.1) are calculated. The average loan advanced, weighted by model claim frequency, is then calculated for each year of advance.
(c) Average of the entries in the column, weighted by numbers of claims shown in parenthesis.
(d) Unweighted average of the entries in the column.

The table suggests that the average loan size associated with claims of a particular cell for a particular year of advance is about 7% higher than the overall average loan size for the cell.

Thus, a forecast of future claim amount for a particular cell of development year j of year of advance i would be computed as:

$$1.07 \times \text{average loan size in year of advance } i \\ \times \hat{N}(i,j)\,\hat{Q}(i,j),$$

where $\hat{N}(i,j)$, $\hat{Q}(i,j)$ are estimates of $N(i,j)$ and $Q(i,j)$ from Sections 9.1 and 9.2.

An alternative approach to the above would be to include loan size as an explanatory variable in the claim frequency model of Section 8.1. This might be awkward in practice, however, because it would increase very considerably the number of data cells entering into the regressions of Section 8.1.

10. CONCLUSION

Section 8 fits models to the claim frequency and claim ratio in the mortgage insurance portfolio examined. Section 9 verifies that these models provide a reasonable fit to the data.

The models therefore can be, and indeed have been, used to estimate the liability for claims still to emerge in respect of past years of loan advance. In order to carry out this estimation, one needs to project future values of the HAI and HPI. This in turn requires projection of incomes, tax rates, mortgage interest rates and growth in property values. Projections such as these are, problems of substance in their own right, but are beyond the scope of the present paper.

11. ACKNOWLEDGEMENT

I should like to acknowledge the computing assistance provided by Mr A. J. Greenfield in the preparation of this paper.

APPENDIX A

DEPENDENCE OF CLAIM FREQUENCY ON HOME AFFORDABILITY INDEX

Let X denote the random variable representing the proportion of an individual's income required for tax, consumption and mortgage instalment. Assume this variable to be Pareto distributed, i.e. with p.d.f.:

(A.1) $$f(x) = kx^{-\alpha-1}, \; k \text{ const.}$$

The borrower will experience financial difficulties if $X \geq 1$, which occurs with probability:

(A.2) $$P[X \geq 1] = kx^{-\alpha}/\alpha|_{x=1}.$$

Now, suppose that X shifts by a factor of c to $X' = cX$. Then the probability (A.2) shifts to

(A.3) $$P[X' \geq 1] = P[X \geq 1/c] = kx^{-\alpha}/\alpha|_{x=1/c}.$$

Comparison of (A.2) and (A.3) shows that the probability (A.2) has shifted by a factor of c^α. Now note that the scale shift of X to cX must shift the mean of X by a factor of c:

(A.4) $$E[X'] = cE[X].$$

Let

$$Y = 1 - X,$$

https://doi.org/10.2143/AST.24.1.2005084 Published online by Cambridge University Press

and note that

(A.5) $$E[Y] \;\alpha\; \text{HAI}.$$

Then the factor by which HAI changes when X changes to X' is:

(A.6) $$R = \{1-E[X']\}/\{1-E[X]\}$$
$$= (1-c\mu)/(1-\mu),$$

where

$$\mu = E[X].$$

Inversion of (A.6) yields:

(A.7) $$c = [1-R(1-\mu)]/\mu.$$

Thus, the shift in HAI by a factor of R causes the frequency with which borrowers experience difficulties to shift by a factor of:

(A.8) $$c^{\alpha} = \{[1-R(1-\mu)]/\mu\}^{\alpha}.$$

Now, it is convenient to analyse log (claim frequency), which will depend on log (frequency of borrower's difficulties), and (A.8) shows that this latter will depend on an additive term of:

$$\log c^{\alpha} = \alpha \log \{[1-R(1-\mu)]/\mu\}$$
$$\sim -\alpha R(1-\mu) + \text{const.},$$

for small values of $(1-\mu)R$.

Thus, to first approximation, the model of expected log (claim frequency) should include a linear term in R, the ratio by which HAI has changed since advance of the loan(s) in question.

APPENDIX B

DEPENDENCE OF CLAIM FREQUENCY ON HOUSING PRICE INDEX, LVR AND DEVELOPMENT YEAR

Consider a loan taken at time $t = 0$. Let $V(t)$ be the value of the associated property at time t, and $P(t)$ the amount of principal then outstanding. Then

(B.1) $$V(t) = V(0)[H(t)/H(0)],$$

(B.2) $$P(t) = P(0)f(t),$$

where

$H(t)$ = HPI at time t;

$f(t)$ = proportion of principal still to be repaid at time t.

By (B.1) and (B.2),

(B.3) $$P(t)/V(t) = Lf(t)\,H(0)/H(t),$$

where

(B.4) $$L = P(0)/V(0) = \text{loan to valuation ratio}.$$

Suppose that the borrower has encountered financial difficulties at some time $s < t$. At time t sale of the property is forced. At that point, the debt in respect of the loan will be $P(t)\,\alpha(t)$, where

$\alpha(t)$ = a random variable representing the factor by which outstanding principal has been enlarged by arrears of principal and interest and any other costs.

Similarly, the net proceeds of the sale of the property will be $V(t)\,\beta(t)$, where

$\beta(t)$ = a random variable representing the factor by which the property value has been reduced by the forced nature of the sale and the associated expenses.

Then the ratio of outstanding debt to sale proceeds is:

(B.5) $$X(t) = \gamma(t)\,P(t)/V(t),$$

where

(B.6) $$\gamma(t) = \alpha(t)/\beta(t).$$

By (B.3) and (B.5),

(B.7) $$X(t) = L[H(t)/H(0)]^{-1} f(t)\,\gamma(t).$$

A claim will occur if $X(t) > 1$, i.e. if

(B.8) $$\gamma(t) > [H(t)/H(0)]\,[Lf(t)]^{-1}.$$

Now suppose that $\gamma(t)$ is Pareto distributed with d.f.

(B.9) $$F(\gamma) = 1-(\gamma/a)^{-\nu}, \qquad \gamma > a,$$

assumed independent of t. Then, by (B.8), the probability of occurrence of a claim is:

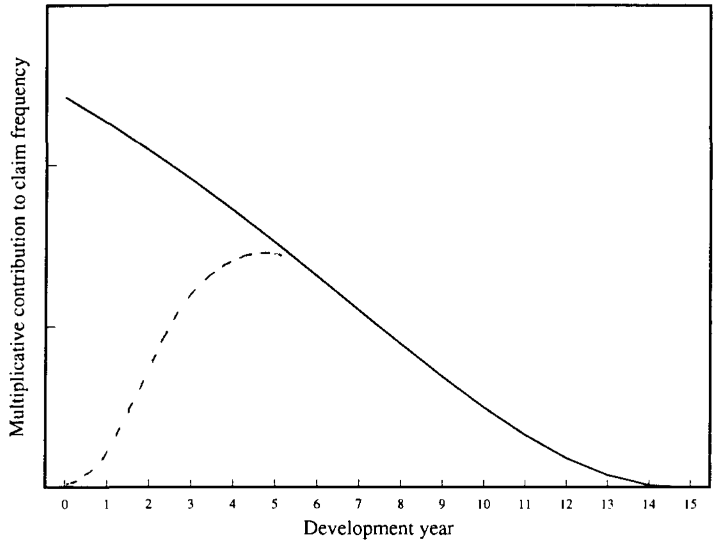
(B.10) $$P[X(t) > 1] = \{a f(t)\,L[H(t)/H(0)]^{-1}\}^{\nu}.$$

Thus, expected claim frequency varies as a power of $L[H(t)/H(0)]^{-1}$. Note also that claim frequency for policies of a particular term n varies over development years t by a factor of

(B.11) $$[f(t)]^{\nu} \quad \alpha \quad [a_{\overline{n-t|}}]^{\nu},$$

which has the shape illustrated by the solid line in the following diagram.

However, note the above assumption that the distribution of the factor $\gamma(t)$ is independent of t. While perhaps largely true, it will break down as $t \to 0$ as the screening procedures of the lender force claim frequency toward zero. Hence, the curve (B.11) of frequency over development year will be modified for small t in the manner indicated by the broken line in the diagram.



When allowance is made for the variety of original terms *n*, the dependence of claim frequency on development year is seen to be represented by a weighted average of curves of the type illustrated in the diagram.

APPENDIX C

DEPENDENCE OF AVERAGE CLAIM SIZE ON HOUSING PRICE INDEX

As noted just prior to (B.8), the financial difficulties of a borrower will lead to a claim if $X(t)$, as defined there, exceeds 1. In fact, by the same argument as led to that result, the amount of the claim will be

$$\text{(C.1)} \qquad \begin{aligned} A(t) &= \alpha(t)\, P(t) - \beta(t)\, V(t) \\ &= \beta(t)\, V(t)\, [X(t)-1]. \end{aligned}$$

Note that $\beta(t)$ and $\gamma(t)$ (and hence $X(t)$) will **not** be independent, even if $\alpha(t)$ and $\beta(t)$ are. For general random variables Y and Z, let μ_Y and μ_Z denote their means, v_Y and v_Z their coefficients of variation, and ρ_{YZ} their correlation. It is straightforward to demonstrate that:

$$\text{(C.2)} \qquad E[YZ] = \mu_Y \mu_Z (1 + \rho_{YZ} v_Y v_Z).$$

By (C.1) and (C.2),

$$\text{(C.3)} \qquad E[A(t)] = V(t)\, E[X(t)-1]_+\, \mu_\beta (1 + \rho_{\beta X} v_\beta v_X),$$

where $E[Y]_+$ denotes $E[Y | Y > 0]$.

Now, by (B.5)

$$\text{(C.4)} \qquad E[X(t)-1]_+ = E[\gamma(t) - V(t)/P(t)]_+\, P(t)/V(t).$$

https://doi.org/10.2143/AST.24.1.2005084 Published online by Cambridge University Press

By the Pareto assumption (B.9),

(C.5) $$E[\gamma(t) - V(t)/P(t)]_+ = [V(t)/P(t)]\, \nu/(\nu - 1),$$

whence (C.3) and (C.4) yield:

(C.6) $$E[A(t)] = V(t)\, \mu_\beta (1 + \rho_{\beta X} v_\beta v_X)\, \nu/(\nu - 1)$$
$$\propto V(0)\, H(t)/H(0) \qquad \text{[by (B.1)]}$$

if μ_β, v_β, v_X and $\rho_{\beta X}$ are the assumed independent of t.

Thus, the expected average claim size is directly proportional to property values, all other things equal. This has the interesting effect of causing average claim size in respect of a group of identical policies usually to **increase** with development year even though outstanding principal is decreasing.

APPENDIX D

EXPLORATORY ANALYSIS OF CLAIM SIZE

D1. Variation of claim ratio with loan to valuation ratio

Loan to valuation ratio	Number of claims	Claim to loan ratio		95% confidence limits (a)	
		Sample mean	Sample standard deviation	Lower	Upper
up to 50%	1	55.8%			
50 to 60%	1	56.9%			
60 to 70%	8	23.3%	13.7%	11.8%	34.8%
70 to 80%	36	23.9%	19.2%	17.4%	30.4%
80 to 90%	189	22.9%	18.4%	20.3%	25.6%
over 90%	191	23.5%	15.6%	21.3%	25.7%

(a) These are the symetric t-distribution confidence limits. Where the sample size is less than 2, the confidence limits do not exist.

D2. Variation of claim ratio with term

Term	Number of claims	Claim to loan ratio		95% confidence limits (a)	
		Sample mean	Sample standard deviation	Lower	Upper
months					
60 to 119	3	36.4%	14.1%	1.3%	71.4%
120 to 179	16	34.8%	29.8%	18.9%	50.7%
180 to 239	55	28.4%	20.2%	22.9%	33.9%
240 & more	352	22.0%	15.6%	20.4%	23.7%

(a) See Appendix D1.

D3. Variation of claim ratio with area

Area	Number of claims	Claim to loan ratio		95% confidence limits (a)	
		Sample mean	Sample standard deviation	Lower	Upper
M1, M4	29	16.5%	11.7%	12.0%	20.9%
M2	63	21.2%	15.0%	17.5%	25.0%
M3	77	16.5%	12.6%	13.7%	19.4%
M5	5	25.8%	14.8%	7.5%	44.1%
Canberra	4	23.1%	13.0%	2.4%	43.8%
Coastal	100	24.6%	18.2%	21.0%	28.2%
Newcastle	32	31.7%	17.2%	25.6%	37.9%
Wollongong	0				
Other	116	27.5%	19.4%	23.9%	31.1%

(a) See Appendix D1.

D4. Commentary

All pairs of confidence limits in Appendices D1 to D3 straddle the overall mean of 23.4% except in four cases. All four of these cases relate to area of residence, and are found in Appendix D3.

APPENDIX E

COMPARISON OF OBSERVED AND FITTED CLAIM AMOUNTS

The following are the amounts of claim **observed** in respect of each combination of year of advance and development year.

Year of advance	Amount of claims observed in development year										
	0	1	2	3	4	5	6	7	8	9	10
	$	$	$	$	$	$	$	$	$	$	$
1980					28522	13349	7873	0	0	0	1009
1981				115151	69711	105156	3724	0	0	0	
1982			71488	29799	102851	81026	35484	20827	56169		
1983		60085	71469	61801	85959	64416	0	10110			
1984	0	45337	68811	325411	180820	11766	0				
1985	9591	161743	1060021	474840	179612	44976					
1986	0	150351	219581	28174	26638						
1987	22882	7054	0	15810							
1988	0	0	258976								
1989	0	0									
1990	0										

https://doi.org/10.2143/AST.24.1.2005084 Published online by Cambridge University Press

The following are the amounts of claims **fitted** to each combination of year of advance and development year by the procedure described in Section 9.2.

Year of advance	Amount of claims fitted in development year										
	0	1	2	3	4	5	6	7	8	9	10
	$	$	$	$	$	$	$	$	$	$	$
1980					27287	25853	9332	0	0	0	7380
1981				125940	91833	84727	9687	0	0	0	
1982			56280	43406	129344	70032	19012	27658	28253		
1983		51324	96763	63585	74571	29094	0	7572			
1984	0	68421	121228	258339	167683	26301	0				
1985	14819	185929	1089849	576994	130423	64647					
1986	0	151670	258058	41149	20740						
1987	30697	13995	0	23866							
1988	0	0	221693								
1989	0	0									
1990	0										

Each cell in this table is of the form:

actual number of claims

×

fitted average claim size.

Hence comparison of the table with the previous one examines only variation of experience from model amounts of claim.

An alternative version of the preceding table consists of cells of the form:

fitted number of claims

×

fitted average claim size.

This table is as follows.

Year of advance	Amount of claims fitted in development year										
	0	1	2	3	4	5	6	7	8	9	10
	$	$	$	$	$	$	$	$	$	$	$
1980					16472	13202	11077	**0**	**0**	**0**	52
1981				44040	55444	61935	47805	**0**	**0**	**0**	
1982			39396	55278	111986	99883	22086	2637	2910		
1983		15962	73512	80326	136558	50459	**0**	1408			
1984	**0**	41551	144634	324560	148532	15693	**0**				
1985	4668	188718	907194	617384	82395	49662					
1986	**0**	185146	264079	66099	31805						
1987	3131	86881	**0**	29785							
1988	**0**	**0**	153966								
1989	**0**	**0**									
1990	**0**										

For cells in which where are no claims observed, the procedure of Section 9.2 does not produce a fitted average claim size. These cells, **indicated in bold,** have been assigned a fitted amount of claims equal to zero.

APPENDIX F

HOME AFFORDABILITY INDEX

Year (as at 31 December)	Economic indicators			Household expenditure					
								Residual income	
	Average weekly earnings	Consumer price index	Mortgage interest rates (a)	Gross household income (b)	Tax (b)	Consumer expenditure (b)	Mortgage instalment (b)	Amount	As percentage of gross income
	$		p.a.	$ per week	$ per week	$ per week	$ per week	$ per week	
1978	224.35	82.4	11.50%	562.74	118.28	326.21	64.40	53.85	9.569%
1979	246.00	91.1	11.50%	617.05	129.70	360.65	70.61	56.08	9.089%
1980	278.25	100.0	12.00%	697.94	146.70	395.89	82.26	73.10	10.473%
1981	315.90	110.2	14.50%	792.38	166.55	436.27	107.18	82.39	10.397%
1982	346.70	123.4	15.50%	869.64	182.79	488.52	123.78	74.54	8.572%
1983	375.90	130.9	14.00%	942.88	198.19	518.22	124.22	102.26	10.846%
1984	405.40	136.0	13.50%	1016.88	213.74	538.41	130.41	134.33	13.210%
1985	428.20	147.5	15.00%	1074.07	225.76	583.93	149.07	115.30	10.735%
1986	450.85	161.4	15.50%	1130.88	237.70	638.96	160.96	93.25	8.246%
1987	477.70	173.7	14.50%	1198.23	251.86	687.66	162.07	96.64	8.066%
1988	521.65	187.7	14.25%	1308.47	275.03	743.08	174.68	115.68	8.841%
1989	560.75	203.0	17.25%	1406.55	295.64	803.65	217.77	89.48	6.362%
1990	600.68	213.0	15.50%	1506.69	316.69	843.24	214.46	132.30	8.781%

(a) The most common interest rates applying to loans in the mortgage insurance portfolio under analysis.

(b) These four columns were derived in a consistent manner from the HES, as described in Section 3.2.

APPENDIX G

DATA

The data described in Section 4.2 are summarized in the following table. This should be considered in conjunction with the qualification set out in the final paragraph of Section 4.2.

Year of advance	Number of loans advanced	Number of claims (a) recorded in development year										
		0	1	2	3	4	5	6	7	8	9	10
1980	1700					3	3	1	0	0	0	1
1981	1917				13	8	6	1	0	0	0	
1982	2231			7	6	10	8	3	3	1		
1983	3426		5	7	7	8	3	0	1			
1984	5496	0	7	13	30	19	3	0				
1985	7787	1	16	104	53	12	5					
1986	8077	0	14	24	3	2						
1987	9910	3	1	0	2							
1988	17646	0	0	8								
1989	11878	0	0									
1990	13614	0										

(a) Development year is defined as year of emergence of claim minus year of loan advance. Claims emerging in 1984 represent the experience of only 7 months.