Modelling mortgage insurance as a multi-state process

Greg Taylor
Taylor Fry Consulting Actuaries
University of Melbourne
University of New South Wales

Peter Mulquiney
Taylor Fry Consulting Actuaries

UNSW Actuarial Studies Research Symposium

11 November 2005
Mortgage insurance (aka Lenders mortgage insurance (LMI))

- Indemnifies lender against loss in the event of default by the borrower when collateral property is sold
- Loss would occur if sale price less associated costs is insufficient to meet outstanding loan principal
Mortgage insurance (cont’d)

- Policies have relatively unique properties
  - Single premium but multi-year coverage
  - Claims experience influenced by variables related to housing sector of economy
  - Claim occurs at a defined sequence of events
Earning of premium

- Accounting Standard requires that earning of premium be proportionate to incidence of risk
- Typical situation
  - Premium is earned in fixed percentages over the years of a policy’s life
    - E.g. 5% in Year 1, 15% Year 2, etc
  - True incidence of risk over lifetime of a cohort of policies will vary with economic conditions
    - Downturn in house prices likely to generate claims
Literature survey

- **Taylor (1994):** Modelled claims experience with a GLM with external economic variables as predictors
- **Ley & O’Dowd (2000):** Extended GLM to allow for changes in LMI market and products
- **Driussi & Isaacs (2000):** Overview of LMI industry, less concerned with modelling. Contains useful data
- **Kelly & Smith (2005):** Stochastic model of external economic predictors
Multi-state process leading to a claim

- **Healthy**
  - **In arrears**
    - **Property in possession**
      - **Borrower’s sale**
      - **Property sold**
        - **Loan discharged**
        - **Claim**
      - **Cured**
  - **Cured**

Transition matrix

<table>
<thead>
<tr>
<th></th>
<th>Healthy</th>
<th>In arrears</th>
<th>PIP</th>
<th>Sold</th>
<th>Loan discharged</th>
<th>Claim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>1-a</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In arrears</td>
<td>c</td>
<td>1-c-p-b</td>
<td>p</td>
<td>b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIP</td>
<td></td>
<td>1-s</td>
<td>s</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sold</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>d</td>
</tr>
<tr>
<td>Loan discharged</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1-d</td>
</tr>
<tr>
<td>Claim</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

- Each distinct probability requires a separate model
Required models

1. Probability Healthy $\rightarrow$ In arrears
2. Probability of cure of arrears
3. Probability In arrears $\rightarrow$ PIP
4. Probability PIP $\rightarrow$ Sold
5. Probability Sold $\rightarrow$ Claim
6. Distribution of claim sizes

• The case [In arrears $\rightarrow$ Sold] may be treated as:
  In arrears $\rightarrow$ PIP $\rightarrow$ Sold
  *Nil duration*
Structure of sub-models – transition probabilities

- GLMs for all 6 sub-models
- Five of them are probabilities

\[ Y^{(m)}_{ij} \sim \text{Bin} \left[ 1, 1 - \exp \{ u^{(m)}_{ij} \log \{ 1 - p^{(m)}_{ij} \} \} \right], \quad m=1,\ldots,5 \]

where

- \( Y^{(m)}_{ij} \) is binomial response of the i-th policy in the j-th calendar interval under the m-th model
  - e.g. the transition Healthy \( \rightarrow \) In arrears
- \( p^{(m)}_{ij} \) is the associated probability
- \( u^{(m)}_{ij} \) is the time on risk of the relevant transition

\[
\text{logit } p^{(m)}_{ij} = \sum_k \beta^{(m)}_k x_{ijk}
\]

with

- \( x_{ijk} \) predictors
- \( \beta^{(m)}_k \) their coefficients
Structure of sub-models – claim size

\[ Y^{(6)}_{ij} \sim EDF(\mu_{ij}, q) \]
\[ \log \mu_{ij} = \sum_k \beta^{(6)}_k x_{ijk} \]
\[ E[Y^{(6)}_{ij}] = \mu_{ij} \]
\[ \text{Var}[Y^{(6)}_{ij}] = (\varphi/w_{ij}) \mu_{ij}^q \]

- We found \( q=1.5 \) satisfactory
  - Right skewed
  - But shorter tailed than Gamma
Predictors

• Several categories
  • **Policy variables** (specific to individual policies)
    • **Static**, e.g. date of policy issue, Loan-to-valuation ratio (LVR)
    • **Dynamic**, e.g. number of quarters since transition into current status
  • **External economic variables** (common to all policies), e.g. interest rates, rates of housing price increases
  • **Manufactured risk indicators** (derivatives of the previous categories)
Manufactured risk indicators – an example

Potential claim size = Amount of arrears

less

Principal repaid

less

Original loan amount

×

\[ \text{Growth in borrower’s equity} \]

\[ \times \left[ \text{Housing price growth factor} \times \frac{(1-q)}{LVR} - 1 \right] \]

where

• Housing price growth relates to period from inception to the current date

• \( q \) = proportion of property value lost in deadweight costs on sale
Predictors (cont’d)

• There may be many potential predictors available in the data base, e.g.
  • State of Australia
  • Issue quarter
  • LVR
  • Stock price growth
  • etc

• We found a total of more than 30 statistically significant predictors over the 6 models
  • Many appear in more than one model
  • More than 20 in one of the models
Forecast claims experience

- We take the fitting of the GLM sub-models to claims experience as routine
  - No further comment on this
- Conventional form of forecasting future claims experience consists of plugging parameter estimates into models over future periods
- This procedure is undesirable on two counts
  - Not feasible computationally
  - Produces biased forecast
Computational feasibility

- **Example of evolution of a claim**

  - Commencement of loan \( h_1 \) quarters → In arrears \( a_1 \) quarters → Cured
  - Healthy \( h_2 \) quarters → In arrears \( a_2 \) quarters → Cured
  -...
  - Healthy \( h_n \) quarters → In arrears \( a_n \) quarters → PIP \( p_n \) quarters → Claim

- **Too many combinations for feasible computation**
Forecast bias

- Forecast liability (outstanding claims or premiums) takes form

\[ L^* = \sum_{i,j} L^{(j)}_i(\hat{\beta}, x^*_{ij}) \]

where

- \( L^{(j)}_i(. , .) = \) simulated liability cash flow of policy \( i \) in future calendar quarter \( j \)
- \( \hat{\beta} = \) parameter estimates
- \( x^*_{ij} = \) future values of predictors
Forecast bias (cont’d)

- Forecast liability (outstanding claims or premiums) takes form

\[ L^* = \sum_{i,j} L^{(j)}_i (\beta, x^*_{ij}) \]

\[ x^*_{ij}^T = [\xi^*_{i}^T, \zeta^*_{ij}^T, z^*_{ij}^T] \]

- Static policy variables (non-stochastic)
- Dynamic policy variables (non-stochastic)
- External economic variables (stochastic)
**Jensen’s inequality.** Let $f$ be a function that is convex downward. Let $X$ be a random variable. Then

$$E[f(X)] \geq f(E[X])$$

with equality if and only if either

- $f$ is linear; or
- the distribution of $X$ is concentrated at a single point
Jensen’s inequality.  Let \( f \) be a function that is convex downward.  Let \( X \) be a random variable.  Then

\[
E[f(X)] \geq f(E[X])
\]

with equality if and only if either

- \( f \) is linear; or
- the distribution of \( X \) is concentrated at a single point
Forecast bias (cont’d)

\[ L^* = \sum_{i,j} L^{(j)}(\beta, x^*_{ij}) \]
\[ = \sum_{i,j} L^{(j)}(\ldots, z^*_{ij}) \quad \text{[z stochastic]} \]

- \( L^* \) is convex downward in some of the components of \( z^*_{ij} \)
  - E.g. \( z^*_{ijk} = \) rate of increase of house price index
- By Jensen’s inequality
- \( E[L^*] \geq \sum_{i,j} L^{(j)}(\ldots, E[z^*_{ij}]) \)
  - Plugging expected values of external variables into forecast leads to downward bias
    - This result has recently been observed empirically by Kelly and Smith (2005)
  - For other lines of business, this is usually not significant because the external variables have much lower dispersion (e.g. superimposed inflation)
Forecast error

- Forecast model is fully stochastic
- Forecast error (MSEP) can be estimated, consisting of:
  - Specification error
  - (Model) parameter error
  - Process error
  - **Predictor error**
    - Due to future stochastic variation of predictors
Estimation of forecast error

• By means of an abbreviated form of bootstrap
  • We refer to it as a **fast bootstrap**
• Conventional bootstrapping (re-sampling) is not computationally feasible
Conventional bootstrap

- Data
  - Model
  - Parameter estimates
    - Fitted values
      - Residuals
        - Re-sample
          - Re-sampled residuals
            - Pseudo-data
              - Model
                - Pseudo-parameter estimates
                  - Pseudo-forecast
  - Forecast

Replicate
Fast bootstrap

Conventional bootstrap

Data → Model → Parameter estimates → Fitted values → Residuals → Re-sample → Re-sampled residuals → Pseudo-data → Model → Pseudo-parameter estimates → Forecast

Fast bootstrap

Data → Model → Parameter estimates → Forecast

Just sample assuming normally distributed with model estimates of mean and standard error
Computation

\[ L^* = \sum_{i,j} L^{(j)}(\beta, x^*_{ij}) \]

- Need to simulate for:
  - All values of \( i (=\text{policy}), j (=\text{future period}) \)
  - All stochastic components of \( x^*_{ij} \)
- This produces:
  - Central estimate
  - Process error
- Then need to re-simulate for each drawing of pseudo-parameters
- This produces:
  - Parameter error
  - Stochastic predictor error
References


