WORKSHOP

A MIXED MODEL FOR LOSS RATIO ANALYSIS

By M. Y. El-Bassiouni

*Department of Statistics and Insurance, Faculty of Commerce,
Kuwait University, Kuwait*

Abstract

The model introduced may be treated as a mixed two-way analysis of variance with fixed company effects and random time effects. Further, the risk volumes are integrated into the model in such a way that the unexplained variance is inversely proportional to the risk volume of each company. The proposed model is used to analyze loss ratio data from the general insurance market in Kuwait. The maximum likelihood estimates of the structural parameters are obtained. These estimates are then used to compute the loss ratios and solvency margins for the four domestic insurance companies.

1. introduction

Loss ratios play a very important role in risk theory. For one thing, they are used in credibility analysis to predict future losses, which is pertinent to ratemaking, and for another, they are used to compute solvency margins.

The traditional approach is to assume that the loss ratios follow a beta distribution, whose two parameters are then estimated by the method of moments, e.g., see De Wit and Kastelijn (1980). This approach was criticized by Ramlau-Hansen (1982) who assumed the loss ratios to follow a lognormal distribution. The shape of the lognormal curve is appealing in this context and it has been applied before to model loss ratio data, see Hunter (1980). The model introduced by Ramlau-Hansen may be treated as a two-way analysis of variance with random company and time effects. It may also be viewed as a parametric credibility model with seasonal random factor, see Sundt (1979).

The assumption of random-effects for the companies under study is appropriate when they are considered to be a random sample of the companies in the population. However, if they constitute the whole population or if they are considered to represent themselves, they should be assumed to have fixed-effects. To deal with such a situation, we propose here a mixed two-way

analysis of variance model with fixed company effects and random time effects.

RAMLAU-HANSEN also assumed that the unexplained variance is constant and independent of the risk volume of each company. However, since a small portfolio would usually mean large fluctuations in loss ratios, and therefore a large company would need a lower solvency margin than a small company, we adopt here the more realistic assumption that the unexplained variance is inversely proportional to the risk volume, e.g., earned premiums, of each company.

The mixed model is introduced in Section 2, while the maximum likelihood estimators of the parameters are developed in Section 3. Then, in Section 4, the proposed model is used to estimate the loss ratios and solvency margins for the four national insurance companies in Kuwait using data from the general insurance market.

Regarding notation, we use I_m to denote the $m \times m$ identity matrix, H' and $\text{tr}\,(H)$ to denote the transpose and trace, respectively, of a matrix H, and $N_m(\mu, \Sigma)$ to denote the m-dimensional multivariate normal distribution with mean vector μ and covariance matrix Σ.

2. THE MODEL

Let X_{ij} and p_{ij} denote the loss ratio and the earned premiums, respectively, for company i in year j. Set $Y_{ij} = \ln X_{ij}$ and assume that

$$Y_{ij} = \alpha_i + \beta_j + e_{ij}, \qquad i = 1, \ldots, a, \quad j = 1, \ldots, b,$$

where the α_i are unknown fixed constants, while the β_j and e_{ij} are mutually independent normal random variables having zero means and variances θ_2 and θ_1/p_{ij}, respectively. Thus, the parameter space is given by

$$\Theta = \{\alpha_1, \ldots, \alpha_a, \theta_1, \theta_2 : \alpha_i \in \mathscr{R}, \ i = 1, \ldots, a, \theta_1 > 0, \theta_2 \geq 0\}.$$

Now, for company i in year j, the loss ratio is given by

$$(1) \qquad\qquad L_{ij} = \exp\{\alpha_i + .5\,(\theta_2 + \theta_1/p_{ij})\}.$$

Further, the upper limit of the loss ratio, at the probability level $(1 - \varepsilon)$, is given by

$$(2) \qquad\qquad U_{ij} = \exp\{\alpha_i + \Phi^{-1}(1 - \varepsilon)\sqrt{\theta_2 + \theta_1/p_{ij}}\},$$

where Φ denotes the cumulative standard normal distribution function. Hence, the solvency margin is obtained from

$$(3) \qquad\qquad S_{ij} = \max\{0, (U_{ij} + E_i - 100)\},$$

where E_i denote the expense ratio for company i.

3. THE MAXIMUM LIKELIHOOD ESTIMATORS

Let $n = ab$ and write the model in matrix form as

$$Y = A\alpha + B\beta + e,$$

where $Y, A, \ldots, e$ are defined in the obvious manner. In particular,

$$A = \begin{pmatrix} 1_b & 0 & \ldots & 0 \\ 0 & 1_b & \ldots & 0 \\ . & . & \ldots & . \\ 0 & 0 & \ldots & 1_b \end{pmatrix},$$

and

$$B = (I_b, \ldots, I_b)'.$$

Then, define the diagonal matrix

$$P = \operatorname{diag}(p_{11}, \ldots, p_{ab}).$$

Therefore, under the model assumptions, we have that

$$Y \sim N_n(A\alpha, \theta_1 P^{-1} + \theta_2 BB').$$

To simplify the discussion in the sequel we consider the transformation $P^{1/2} Y$. Thus,

$$P^{1/2} Y \sim N_n(P^{1/2} A\alpha, \Sigma),$$

where

$$\Sigma = \theta_1 I_n + \theta_2 P^{1/2} BB' P^{1/2}.$$

We now give an explicit expression for Σ^{-1}. Let

$$p_{+j} = \sum_{i=1}^{a} p_{ij}, \qquad j = 1, \ldots, b,$$

and define

$$\rho_j = \theta_2 / (\theta_1 + \theta_2 p_{+j}), \qquad j = 1, \ldots, b.$$

It is easily verified that

$$\Sigma^{-1} = (1/\theta_1) [I_n - P^{1/2} B\Lambda B' P^{1/2}],$$

where $\Lambda = \operatorname{diag}(\rho_1, \ldots, \rho_b)$.

Following HARVILLE (1977), the likelihood equation for α is given by

(4) $$A' P^{1/2} \Sigma^{-1} P^{1/2} A\alpha = A' P^{1/2} \Sigma^{-1} P^{1/2} Y.$$

Now, let

$$p_{i+} = \sum_{j=1}^{b} p_{ij}, \qquad i = 1, \ldots, a,$$

and define H to be the $a \times a$ matrix whose elements are given by

$$H_{rs} = p_{r+} - \sum_{j=1}^{b} p_j p_{rj}^2, \qquad r = s = 1, \ldots, a,$$

$$= - \sum_{j=1}^{b} p_i p_{rj} p_{sj}, \qquad r \neq s.$$

Further, define h to be the $a \times 1$ vector whose elements are given by

$$h_r = \sum_{j=1}^{b} p_{rj} \left[Y_{rj} - p_j \sum_{i=1}^{a} p_{ij} Y_{ij} \right], \qquad r = 1, \ldots, a.$$

It is easily shown that equation (4) reduces to

(5) $$\alpha = H^{-1} h.$$

Since the maximum likelihood estimators of θ_1 and θ_2 take no account of the loss in degrees of freedom resulting from estimating α, the vector of fixed company effects, we consider the restricted maximum likelihood approach of PATTERSON and THOMPSON (1971) to estimate the variance components, see HARVILLE (1977) for more details. To this end, let $Z = Y - B\beta^*$, where

$$\beta^* = \theta_2 B' P^{1/2} \Sigma^{-1} P^{1/2} (Y - A\alpha),$$
$$= (\theta_2/\theta_1) [I_b - B' PBA] B' P(Y - A\alpha).$$

Consequently, $Z_{ij} = Y_{ij} - \beta_j^*$, where

$$\beta_j^* = p_j \sum_{i=1}^{a} p_{ij} (Y_{ij} - \alpha_i), \qquad j = 1, \ldots, b.$$

Furthermore, define

$$S = I_n - P^{1/2} A (A' PA)^{-1} A' P^{1/2}.$$

Thus, the restricted likelihood equation for θ_1 is given by

$$\theta_1 = Y' P^{1/2} SP^{1/2} Z/(n-a),$$
$$= [Y' PZ - Y' PA (A' PA)^{-1} A' PZ]/(n-a),$$
$$= \left[\sum_{i=1}^{a} \sum_{j=1}^{b} p_{ij} Y_{ij} Z_{ij} - \sum_{i=1}^{a} \left(\sum_{j=1}^{b} p_{ij} Y_{ij} \right) \times \right.$$

(6) $$\left. \times \left(\sum_{j=1}^{b} p_{ij} Z_{ij} \right) \middle/ p_{i+} \right] \middle/ (n-a).$$

On the other hand, let

$$T = [I_b + (\theta_2/\theta_1) B' P^{1/2} SP^{1/2} B]^{-1},$$
$$= (\theta_1/\theta_2) A [I_b + B' PAH^{-1} A' PBA].$$

Then,

$$\text{tr } (T) = (\theta_1/\theta_2) \left[\sum_{j=1}^{b} p_j + \text{tr } (H^{-1} G) \right],$$

where G is the $a \times a$ matrix whose elements are given by

$$G_{rs} = \sum_{j=1}^{b} p_j^2 p_{rj} p_{sj}, \qquad r, s = 1, \ldots, a.$$

It can be shown that the restricted likelihood equation for θ_2 is given by

(7) $$\theta_2 = \sum_{j=1}^{b} \beta_j^{*2}/[b - \text{tr } (T)].$$

The likelihood equations (5)-(7) must be solved simultaneously for $\hat{\alpha}, \hat{\theta}_1$, and $\hat{\theta}_2$. However, except in some special cases, e.g., $p_{ij} = p$, or $p_{ij} = p_i$, say, explicit solutions of the above likelihood equations do not exist. Nevertheless, the form of the equations suggests the following iterative procedure, due to HENDERSON (1973). Set $\theta = (\theta_1, \theta_2)'$ and let $\theta^{(k)}, k = 1, 2, \ldots$, denote the value produced by the procedure on its kth iteration. Furthermore, for any quantity f which is a function of θ, we use $f^{(k)}$ to denote the value of f at $\theta = \theta^{(k)}$. Hence, we start the iteration by substituting an initial guess $\theta^{(0)}$ into

(8) $$\theta_1^{(k+1)} = \left[\sum_{i=1}^{a} \sum_{j=1}^{b} p_{ij} Y_{ij} Z_{ij}^{(k)} - \sum_{i=1}^{a} \left(\sum_{j=1}^{b} p_{ij} Y_{ij} \right) \times \right.$$
$$\left. \times \left(\sum_{j=1}^{b} p_{ij} Z_{ij}^{(k)} \right) \middle/ p_{i+} \right] \middle/ (n-a)$$

and

(9) $$\theta_2^{(k+1)} = \sum_{j=1}^{b} [\beta_j^{*\,(k)}]^2/[b - \text{tr } (T^{(k)})],$$

and continue the iteration until $\theta^{(k+1)}$ is sufficiently close to $\theta^{(k)}$ in some norm.

HARVILLE (1977) showed that, if the above iterative procedure is started with strictly positive values for the variance components, then at no point can the values for the variance components ever become negative. It should also be noticed that this procedure must not be started with a zero value for any variance component, since the value for that component would then continue to be zero throughout the iterative procedure.

If we happen to have any prior information about θ, then we could use it to formulate an initial guess for θ. Otherwise, we could use the usual ANOVA estimators, obtained from (6) and (7) assuming that $P = I_n$, as initial guesses.

To this end, let $Y_{i+} = \sum\limits_{j=1}^{b} Y_{ij}$, $Y_{+j} = \sum\limits_{i=1}^{a} Y_{ij}$, and $Y_{++} = \sum\limits_{i=1}^{a} \sum\limits_{j=1}^{b} Y_{ij}$.

Now, define

$$R_2 = (1/a) \sum_{j=1}^{b} Y_{+j}^2 - Y_{++}^2/n,$$

and

$$R_1 = \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij}^2 - (1/b) \sum_{i=1}^{a} Y_{i+}^2 - R_2.$$

Hence, the initial estimates are given by

(10) $\theta_1^{(0)} = R_1/[(b-1)(a-1)]$,

and

(11) $\theta_2^{(0)} = R_2/(b-1) - \theta_1^{(0)}]/a$.

However, equation (11) may produce negative estimates which do not belong in the parameter space Θ. The usual practice of fixing negative estimates of variance components at zero would not be useful here since, as indicated above, we should not start the iterative procedure with a zero value for θ_2. Instead, an initial guess in the interior of Θ should be used to start the iterative procedure.

Thus, the procedure of computing the maximum likelihood estimates of the parameters starts by obtaining initial estimates of the variance components, e.g. from (10) and (11). These estimates are then substituted into (5) to estimates α. This estimate of α along with the initial estimates of θ_1 and θ_2 are then substituted into (8) and (9) to obtain $\theta^{(1)}$. This iterative process is to be continued until we achieve convergence after m iterations, say, at which time we get $\hat{\theta} = \theta^{(m)}$ and $\hat{\alpha} = \alpha^{(m)} = \alpha(\theta^{(m)})$.

The estimated loss ratios and solvency margins, for company i in year j, are obtained by substituting $\hat{\alpha}_i$, $\hat{\theta}_1$, $\hat{\theta}_2$ and p_{ij} into (1)-(3).

4. AN APPLICATION

Four domestic companies are operating in the Kuwaiti general insurance market, namely, Kuwait, Ahlia, Gulf and Warba. Since Warba started operating in 1978, we limited our data to the period 1978-1986. The incurred loss ratios during this period are given in Table 1, along with the associated data on earned premiums in millions of Kuwaiti dinars.

It must be more realistic to assume that the four domestic companies have fixed effects. Thus, the mixed model seems to be more suitable for the analysis of this set of data.

TABLE I

INCURRED LOSS RATIOS (LR) AND EARNED PREMIUMS (EP)

Company	Kuwait		Ahlia		Gulf		Warba	
Year	LR	EP	LR	EP	LR	EP	LR	EP
1978	63	12.5	65	8.6	68	5.0	48	2.7
1979	67	12.1	68	9.6	66	5.7	84	5.6
1980	63	13.6	73	10.3	63	7.2	55	6.5
1981	58	15.8	78	11.9	67	8.5	56	7.2
1982	63	18.3	73	12.4	72	9.1	62	7.6
1983	67	20.2	54	12.8	89	9.2	53	7.9
1984	71	21.4	73	13.8	68	10.0	59	8.9
1985	68	18.6	74	12.6	72	10.0	58	8.3
1986	69	16.3	67	11.2	76	10.1	77	8.7

The initial estimates, computed from (10) and (11), were $\hat{\theta}_1^{(0)} = 0.01505$ and $\hat{\theta}_2^{(0)} = -0.00059$. Since $\hat{\theta}_2^{(0)}$ is negative, several initial guesses, all of which are in the interior of Θ, were used to start the iterative procedure and, in all cases, the procedure converged to the following estimates: $\hat{\theta}_1 = 0.12146$ and $\hat{\theta}_2 = 0$. These estimates were computed from (8) and (9) so that they are correct to 5 decimals. Then, they were substituted into (5) to get

$$\hat{\alpha} = (4.184, \ 4.237, \ 4.271, \ 4.114).$$

The loss ratios and solvency margins were computed from (1)-(3) using the above estimates along with the earned premiums of 1986. Further, it was assumed that $\varepsilon = 0.001$ and $E_i = 30\%$. The results appear in Table 2.

TABLE 2

ESTIMATED LOSS RATIOS AND SOLVENCY MARGINS

Company	Loss Ratio	Solvency Margin
Kuwait	65.9	15.1
Ahlia	69.6	24.7
Gulf	72.0	29.6
Warba	61.6	17.3

REFERENCES

HARVILLE, D. A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**, 320–338.
HENDERSON, C. R. (1973) Maximum likelihood estimation of variance components. Unpublished Manuscript.
HUNTER, R. (1980) Flucutation reserves — simulation of German method for calculating fluctuation reserves. *General Insurance Conference, Cambridge, available by the Institute of Actuaries' Office.*
PATTERSON, H. D. and THOMPSON, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554.
RAMLAU-HANSEN, H. (1982) An application of credibility theory to solvency margins — some comments on a paper by G. W. DE WIT and W. M. KASTELIJN. *ASTIN Bulletin* **13**, 37–45.
SUNDT, B. (1979) An Insurance model with collective seasonal random factors. *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker*, 57–64.
WIT, B. W. DE and KASTELIJN, W. M. (1980) The solvency margin in non-life insurance companies. *ASTIN Bulletin* **11**, 136–144.

Dr. M. Y. EL-BASSIOUNI
Department of Insurance and Statistics, Faculty of Commerce,
P.O. Box 5486 Safat, 13055 Kuwait.