UMVUE of the IBNR Reserve in a Lognormal Linear Regression Model [1]

by Louis G. Doray, Université de Montréal [2]

Abstract

In this paper, we first find an expression for the mean and the variance of the IBNR claims in a lognormal linear regression model, of which the chain ladder model is considered as a special case. We then derive the unique uniformly minimum variance unbiased estimator (UMVUE) and the maximum likelihood estimator (MLE) of those quantities and calculate the variance of the UMVUE of the mean of the IBNR claims; we also find an estimator not involving an infinite series, which provides an excellent approximation to the UMVUE of the mean of the IBNR claims. Finally, the claims experience of an insurance company is used to compare the various estimators of the IBNR reserve developed in the paper. Several tests and graphs are used to verify model assumptions.

Keywords: IBNR reserve; lognormal regression; UMVUE; maximum likelihood.

---

[1] The author gratefully acknowledges the financial support of the CAFIR research fund, Université de Montréal, and of the Natural Sciences and Engineering Research Council of Canada. Professor Stuart Klugman and the referees should be thanked for many suggestions which improved the paper.

[2] Dr Louis Doray, Département de mathématiques et statistique, Université de Montréal, P.O. Box 6128, Succursale centre-ville, Montréal, Québec, H3C 3J7 Canada

1 Introduction

In their comprehensive survey of loss reserving models, Van Eeghen (1981) and Taylor (1986) present many models useful in practice. However, the statistical properties of the estimators derived from some of those models have not been fully investigated, in particular their bias and mean square error. It is our intention to study those quantities for various estimators of the IBNR (incurred but not reported) reserve in a lognormal linear regression model.

Accurate estimation of the level of IBNR reserve for an insurance company is required because of its impact on multiple aspects of the operation of the company: investment policy, dividend declaration (Taylor (1986, p.211)) and tax payments. In addition, a systematic overestimation of the loss reserve would eventually lead to premiums too high and a bleak prospect for future new business, while an underestimation represents a serious threat to the solvency of the insurer. Moreover, accounting regulations require that financial statements provide a true and fair representation of the financial situation of the insurer. It is clear therefore that we should attempt to find an estimator of the IBNR reserve which is unbiased.

Taylor (1986, p.4) wrote "the problem of first moments will never be solved, except perhaps in the sense of producing estimates of outstanding claims which resemble minimum-variance estimators." The estimator of the first moment can be judged to be a good estimator or not only after finding its variance. In this paper, we develop an estimator of the IBNR reserve which possesses those desirable properties of unbiasedness and minimum variance. Other estimators can then be compared to this optimal one.

The paper is organized as follows. Section 2 presents the lognormal linear regression model; the stochastic chain ladder model can be obtained as a special case of this model with a proper choice of the design matrix. Section 3 gives the MLE's of the parameters and their properties. We then find expressions for the mean and variance of IBNR claims (section 4) and derive the UMVUE's of those quantities in terms of the hypergeometric function. It is shown how to find an approximation to this function, which does not require an infinite series (section 5). We then calculate the variance of the UMVUE of the mean of the IBNR claims (section 6). This quantity will provide a lower bound for the variance of all unbiased estimators of the IBNR reserve for this model. In section 7, we consider the MLE of the mean and the variance of IBNR claims. We compare the various estimators of the IBNR reserve presented in the paper, using the actual claims experience of an insurance company and we check model assumptions using several plots and tests (section 8). Finally, we present some concluding remarks.

2 A general model

The initial model presented in this section is similar to that of Kremer (1982). Let S_{ij}, i and $j = 1, \ldots, m$ be non-negative random variables which represent the cumulative amount of claims paid by development year j, for claims which occurred in accident year i. The S_{ij}'s could also denote the total claims incurred by development year j, for accident year i, which equal the total claims paid up to that date plus an estimate of the outstanding liability.

A subset of the upper triangle $(S_{ij}, i = 1, \ldots, m, j = 1, \ldots, m-i+1)$ is observed, the trapezium of data. We define Y_{ij} as the incremental claim amount

$$Y_{i,j+1} = S_{i,j+1} - S_{ij} \quad j \geq 1$$

$$Y_{i,1} = S_{i,1}.$$

We will assume that all Y_{ij}'s are positive. Starting with the multiplicative model used by De Vylder (1978), to which we add a multiplicative lognormal random error, we get

$$Y_{ij} = R_i \cdot C_j \cdot E_{ij} \tag{2.1}$$

where R_i is a row effect for accident year i and C_j is a column effect for development year j, whose product will correspond to the amount of claims for accident year i incurred (or paid) by development year j, and E_{ij} are independent, identically distributed (i.i.d.) lognormal random errors with parameters 0 and σ^2, denoted $LN(0, \sigma^2)$, and density

$$f(t; \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma t} e^{-\frac{1}{2}(\frac{\ln t}{\sigma})^2}, \qquad t > 0.$$

This implies that E_{ij} has mode $e^{-\sigma^2}$, median 1, and mean $e^{\sigma^2/2}$. There is equal probability that Y_{ij} be overstated or understated.

To know more about the properties of the lognormal distribution, the reader is referred to the books by Aitchison and Brown (1957), Johnson and Kotz (1970, chapter 14) and Crow and Shimizu (1988); these last two authors discuss at length various estimators for certain functions of the parameters.

Verrall and Li (1993) analyzed a model allowing negative incremental values, which can occur because of subrogation or salvage. A threshold parameter is introduced in the lognormal distribution. Cohen (1988) discusses the complications created by the introduction of this threshold in the estimation of the parameters from the data. Because the likelihood is infinite along a ridge, he considers moment estimators, modified moment estimators and local MLE's.

Taking the logarithm of both sides of equation (2.1), we get the lognormal linear model

$$Z_{ij} = \ln Y_{ij} = \alpha_i + \beta_j + \epsilon_{ij},$$

where ϵ_{ij} are i.i.d. $N(0, \sigma^2)$ random variables, $\alpha_i = \ln R_i$ and $\beta_j = \ln C_j$. We will term this model, the stochastic chain ladder model. This is the same set-up as Kremer (1982). A parameter, β_1 for example, must be set equal to 0, so the regression matrix is not singular. In matrix notation, this linear model can be represented as

$$Z = \ln Y = X\beta + \epsilon, \qquad \epsilon \sim MN(0, \sigma^2 I)$$

where

$$\begin{aligned} Y' &= (Y_{1,1}, \ldots, Y_{1,m}, Y_{2,1}, \ldots, Y_{2,m-1}, \ldots, Y_{m,1}) \\ Z' &= (Z_{1,1}, \ldots, Z_{1,m}, Z_{2,1}, \ldots, Z_{2,m-1}, \ldots, Z_{m,1}) \\ \epsilon' &= (\epsilon_{1,1}, \ldots, \epsilon_{1,m}, \epsilon_{2,1}, \ldots, \epsilon_{2,m-1}, \ldots, \epsilon_{m,1}) \end{aligned}$$

$$\beta' = (\alpha_1, \ldots, \alpha_m, \beta_2, \ldots, \beta_m),$$

and X is the design matrix.

The above design assumes that the full upper triangle is available to the analyst. Should some elements of it not be available, the corresponding elements should be removed from the vectors Z and ϵ, and the corresponding rows deleted from the matrix X.

In the next sections, we will develop the theory in terms of a general lognormal linear model, i.e. a model linear in the parameters, of the form

$$Z_o = \ln Y_o = X_o \beta + \epsilon_o, \tag{2.2}$$

where Z_o, Y_o and ϵ_o represent the oth element of vectors Z, Y and ϵ and X_o represents the oth row of the matrix X (o for observed). It follows that Y_o has a lognormal distribution $\mathrm{LN}(X_o\beta, \sigma^2)$. The stochastic chain ladder model is just a special case of this general model, with $X_o\beta = \alpha_i + \beta_j$.

3 Estimation of the parameters

The maximum likelihood estimators of β and σ^2 are

$$\hat{\beta} = (X'X)^{-1}X'Z$$

and

$$\hat{\sigma}^2 = \frac{1}{n}(Z - X\hat{\beta})'(Z - X\hat{\beta}),$$

where n is the dimension of the vector Z. Since $\hat{\sigma}^2$ is a biased estimator of σ^2, we also define the following unbiased estimator of σ^2,

$$\tilde{\sigma}^2 = \frac{1}{n-p}(Z - X\hat{\beta})'(Z - X\hat{\beta}) = \frac{SS_z}{n-p},$$

where p is the dimension of vector β and SS_z is the residual sum of squares.

The above results can be found in any standard book on linear models; see for example Graybill (1961). It is also well known that:

1- $\hat{\beta}$ has a multivariate normal distribution $MN(\beta, \sigma^2(X'X)^{-1})$,

2- $(n-p)\tilde{\sigma}^2/\sigma^2$ has a χ^2 distribution with $(n-p)$ degrees of freedom,

3- $\hat{\beta}$ and $\tilde{\sigma}^2$ are independent,

4- $(\hat{\beta}, SS_z)$ are jointly complete and sufficient statistics for the parameters (β, σ^2).

These results will be used in section 5 to find the UMVUE's of the mean and variance of IBNR claims.

4 Mean and variance of IBNR claims

The problem of estimating IBNR claims consists in forecasting the values of Y_{kl} over all (k, l) in the lower unobserved triangle, i.e. over $k = 2, \ldots, m$ and $l = m+2-k, \ldots, m$. From now on, the index u (u for unobserved) will refer to cell (k, l)

in this lower triangle. Let B be the prediction matrix (defined analogously to the regression matrix X), with uth row denoted b_u.

According to our model, the forecast Z_u is given by

$$\ln Y_u = Z_u = b_u\hat{\beta} + \epsilon_u,$$

In the stochastic chain ladder model, $b_u\hat{\beta} = \hat{\alpha}_k + \hat{\beta}_l$. The expected value of Z_u is $b_u\beta$, which is estimated by $b_u\hat{\beta}$. From now on, the subscript u will be dropped from b_u to simplify the notation.

To get an expression for the variance of Z_u, we note that $\hat{\beta}$ and ϵ_u are independent, since $\hat{\beta}$ is a function of past observations, while ϵ_u is a random error in a future observation.

Therefore, the variance of Z_u is

$$\begin{aligned} Var(Z_u) &= Var(b\hat{\beta}) + Var(\epsilon_u) \\ &= \sigma^2[1 + b(X'X)^{-1}b']. \end{aligned}$$

An unbiased estimate of this variance is $\tilde{\sigma}^2[1 + b(X'X)^{-1}b']$.

Taylor and Ashe (1983) used the terminology estimation error for $Var(b\hat{\beta})$ and statistical or random error for $Var(\epsilon_u)$. The estimation error arises in the estimation of the vector parameter $\hat{\beta}$ from the data, and the statistical error comes from the stochastic nature of model (2.2). There is a third error always present in the modelling process, the specification error (see Bartholomew (1975)), which arises as a result of using an inadequate model. The model may be wrongly specified in the sense that its assumptions may not hold in practice.

It follows that the distribution of Y_u is lognormal $(b\beta, \sigma^2[1+b(X'X)^{-1}b'])$. The mode, median and mean of Y_u are therefore:

$$\begin{aligned} \text{mode } (Y_u) &= e^{b\beta-\sigma^2[1+b(X'X)^{-1}b']} \\ \text{median } (Y_u) &= e^{b\beta} \\ E(Y_u) &= e^{b\beta+\frac{1}{2}\sigma^2[1+b(X'X)^{-1}b']}. \end{aligned}$$

The variance of Y_u is: $Var(Y_u) = e^{2b\beta+\sigma^2[1+b(X'X)^{-1}b']}(e^{\sigma^2[1+b(X'X)^{-1}b']}-1)$

It is now easy to find $E(\text{IBNR claims})$,

$$E(\text{IBNR claims}) = \sum_u E(Y_u) = \sum_u e^{b\beta+\frac{1}{2}\sigma^2[1+b(X'X)^{-1}b']}.$$

The variance of the IBNR claims can be expressed as

$$\begin{aligned} Var(\text{IBNR claims}) &= Var(\sum_u Y_u) \\ &= \sum_u Var(Y_u) + \sum_{u\neq u'}\sum_{u'} Cov(Y_u, Y_{u'}). \end{aligned}$$

We now just need to find $Cov(Y_u, Y_{u'})$. The vector $\mathbf{Y} = (Y_u)$ has a multivariate lognormal distribution with parameters $\mu = B\beta$ and $\sum$, denoted by $MLN(\mu, \sum)$.

Let u and u' represent two different cells of the lower unobserved triangle such that $Z_u = b\hat{\beta}+\epsilon_u$ and $Z_{u'} = c\hat{\beta}+\epsilon_{u'}$, and let $E(Z_u) = \mu_1 = b\beta$ and $E(Z_{u'}) = \mu_2 = c\beta$, $Var(Z_u) = \sigma_1^2 = \sigma^2[1+b(X'X)^{-1}b']$ and $Var(Z_{u'}) = \sigma_2^2 = \sigma^2[1+c(X'X)^{-1}c']$ and $\rho(Z_u, Z_{u'}) = \rho$.

The covariance between Y_u and $Y_{u'}$ is (see Crow and Shimizu (1988), p. 12)

$$Cov(Y_u, Y_{u'}) = e^{\mu_1+\mu_2+\frac{1}{2}(\sigma_1^2+\sigma_2^2)}(e^{\rho\sigma_1\sigma_2}-1).$$

This can be proved by using the moment generating function of a bivariate normal distribution with parameters $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, which is (see Hogg and Craig (1978), p.170)

$$M(t_1, t_2) = exp[t_1\mu_1 + t_2\mu_2 + \frac{1}{2}(\sigma_1^2 t_1^2 + \sigma_2^2 t_2^2 + 2\rho\sigma_1\sigma_2 t_1 t_2)]$$

and evaluating it at $t_1 = 1, t_2 = 1$. The correlation coefficient $\rho(Z_u, Z_{u'})$ is calculated from

$$\rho = \frac{\frac{1}{2}[(b+c)(X'X)^{-1}(b+c)' - b(X'X)^{-1}b' - c(X'X)^{-1}c']}{\sqrt{[1 + b(X'X)^{-1}b'][1 + c(X'X)^{-1}c']}}, \quad b \neq c. \tag{4.1}$$

This expression for ρ is derived from $Var(Z_u + Z_{u'}) = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$.

After simplification, we find that the variance of IBNR claims is

$$Var(\text{IBNR claims}) = \sum_u \left(e^{2b\beta + 2\sigma^2(1 + b(X'X)^{-1}b')} - e^{2b\beta + \sigma^2(1 + b(X'X)^{-1}b')} \right)$$

$$+ \sum_{u \neq u'} \sum_{u'} e^{(b+c)\beta + \sigma^2} [e^{\frac{1}{2}\sigma^2(b+c)(X'X)^{-1}(b+c)'} - e^{\frac{1}{2}\sigma^2 b(X'X)^{-1}b' + \frac{1}{2}\sigma^2 c(X'X)^{-1}c'}]. \tag{4.2}$$

Equation (4.2) shows that the covariance for each pair of elements in the lower triangle needs to be evaluated to find the variance of the IBNR claims.

The estimate of IBNR developed by Kremer (1982) is the sum of the medians of each of the Y_u's. He does not obtain an expression for the variance of predicted IBNR claims.

5 UMVUE's of the mean and variance of IBNR claims

The uniformly minimum variance unbiased estimators (UMVUE's) of the mean and variance of IBNR claims derived in the preceding section, will be constructed, using the method of Finney (1941), as applied by Shimizu (1988) to lognormal linear models.

Theorem 5.1: The unique UMVUE of the mean of IBNR claims is

$$\hat{\theta}_U^E = {}_0F_1(\frac{n-p}{2}; \frac{SS_z}{4}) \sum_u e^{b\hat{\beta}}, \tag{5.1}$$

where ${}_0F_1(\alpha; z)$ is the hypergeometric function

$${}_0F_1(\alpha; z) = \sum_{j=0}^{\infty} \frac{z^j}{j!(\alpha)_j}, \quad \text{with} \quad (\alpha)_j = \begin{cases} \alpha(\alpha+1)\cdots(\alpha+j-1) & j \geq 1 \\ 1 & j = 0, \end{cases}$$

which converges for all values of z.

Proof: Since $\hat{\beta}$ and SS_z are jointly complete and sufficient statistics, it follows from the vector version of the Lehmann-Scheffé theorem (Mood et al. (1974), p. 356) that if

$$E(Y_u) = E(e^{Z_u}) = e^{b\beta + \frac{1}{2}\sigma^2[1+b(X'X)^{-1}b']}$$

admits an unbiased estimator, which is a function of $(\hat{\beta}, SS_z)$, then this estimator is the UMVUE for $E(Y_u)$ and it is unique.

Since the statistics $\hat{\beta}$ and SS_z are independent and $E(e^{b\hat{\beta}}) = e^{b\beta + \frac{1}{2}\sigma^2 b(X'X)^{-1}b'}$, we just need to find a function of SS_z, $h(SS_z)$ satisfying

$$
\begin{aligned}
E(Y_u) &= E(e^{b\hat{\beta}} \times h(SS_z)) \\
e^{b\beta+\frac{1}{2}\sigma^2[1+b(X'X)^{-1}b']} &= (e^{b\beta+\frac{1}{2}\sigma^2 b(X'X)^{-1}b'}) \times E(h(SS_z)).
\end{aligned}
$$

This implies that

$$
\begin{aligned}
E[h(SS_z)] &= e^{\sigma^2/2} \\
&= \sum_{j=0}^{\infty} \frac{1}{j!}(\frac{\sigma^2}{2})^j.
\end{aligned}
$$

The variable SS_z/σ^2 follows a χ^2 distribution with $(n-p)$ degrees of freedom; for every positive integer j, we can calculate its jth moment,

$$
E(\frac{SS_z}{\sigma^2})^j = \int_0^\infty t^j \cdot \frac{(\frac{1}{2})^{\frac{n-p}{2}}}{\Gamma(\frac{n-p}{2})} \cdot t^{\frac{n-p}{2}-1} e^{-\frac{1}{2}t} dt = \frac{2^j \Gamma(\frac{n-p}{2}+j)}{\Gamma(\frac{n-p}{2})}.
$$

Therefore, the function $h(SS_z)$ is equal to

$$
\begin{aligned}
h(SS_z) &= \sum_{j=0}^{\infty} \frac{1}{j!} \frac{\Gamma(\frac{n-p}{2})}{2^j \Gamma(\frac{n-p}{2}+j)} (\frac{SS_z}{2})^j \\
&= \sum_{j=0}^{\infty} \frac{1}{j!} \frac{(\frac{SS_z}{4})^j}{(\frac{n-p}{2})_j} = {}_0F_1(\frac{n-p}{2}; \frac{SS_z}{4}),
\end{aligned}
$$

and the UMVUE for $E(Y_u)$ to

$$
\hat{\theta}_U^{EY} = e^{b\hat{\beta}}\ {}_0F_1(\frac{n-p}{2}; \frac{SS_z}{4}).
$$

It just remains to sum the preceding expression over all cells (k,l) to obtain the UMVUE of the mean of IBNR claims. The UMVUE is unique since the statistics $\hat{\beta}$ and SS_z are complete. $\square$

Verrall (1991) also looked at the problem of finding an unbiased estimator for the expected value of IBNR claims in a lognormal linear regression model, but finds the estimator

$$\sum_u e^{b\hat{\beta}} g_r(\frac{\tilde{\sigma}}{2}[1 - b(X'X)^{-1}b']),$$

where $g_r(t) = \sum_{k=0}^{\infty} \frac{r^k(r+2k)}{r(r+2)\ldots(r+2k)} \frac{t^k}{k!}$ and $r = n - p$. He obtained a different estimator because he added the estimation error to the predicted amount instead of the log predicted amount, as we did. But, as can be seen from the equations for $Var(Y_u)$ and $Var(Z_u)$, estimation error and statistical error can not be separated for Y_u, only for Z_u. For the same reason, a different estimator is obtained for Var(IBNR claims). In section 8, we will compare the numerical estimates obtained by the two methods on the same data set.

Although the calculation of the hypergeometric function ${}_0F_1(\alpha; z)$ in (5.1) involves an infinite series, the speed of convergence of the summation is extremely rapid. An example in section 8 will show that adding the first few terms of the series will give sufficient accuracy.

We will now derive an upper bound for $\hat{\theta}_U^E$, not involving the hypergeometric function. Since

$$\begin{aligned}
{}_0F_1(\frac{n-p}{2}; \frac{SS_z}{4}) &= \sum_{j=0}^{\infty} (\frac{1}{j!}) \frac{(\frac{SS_z}{4})^j}{(\frac{n-p}{2})_j} \\
&= \sum_{j=0}^{\infty} \frac{1}{j!} \frac{(\frac{SS_z}{2})^j}{(n-p)(n-p+2)\cdots(n-p+2(j-1))} \\
&< \sum_{j=0}^{\infty} \frac{1}{j!} \frac{(\frac{SS_z}{2})^j}{(n-p)^j} = e^{\frac{SS_z}{2(n-p)}},
\end{aligned}$$

it follows that

$$\hat{\theta}_U^E < e^{\frac{SS_z}{2(n-p)}} \sum_u e^{b\hat{\beta}} = \sum_u e^{(b\hat{\beta}+\tilde{\sigma}^2/2)}.$$

We define a new estimator for E(IBNR claims), $\hat{\theta}_1^E$, equal to this last expression. The estimator $\hat{\theta}_1^E$ will be a close upper bound for $\hat{\theta}_U^E$ if $(n-p)$ is large and SS_z is small. It should be noted that $\exp[b\hat{\beta} + \tilde{\sigma}^2/2]$ is the estimator of the mean of a lognormal distribution $LN(b\beta, \sigma^2)$ obtained by replacing the parameters β and σ^2 by their unbiased estimate.

Proceeding in the same way as for the mean, we can now construct the UMVUE for the variance of IBNR claims. Since the UMVUE for $e^{\gamma\sigma^2}$ is

$${}_0F_1(\frac{n-p}{2}, \frac{\gamma SS_z}{2}),$$

it follows from equation (4.2) that the unique UMVUE for the variance of IBNR claims is

$$\hat{\theta}_U^V = \sum_u e^{2b\hat{\beta}} \left\{ {}_0F_1(\frac{n-p}{2}; SS_z) - \ {}_0F_1(\frac{n-p}{2}; \frac{SS_z}{2}(1 - b(X'X)^{-1}b')) \right\}$$
$$+ \sum_{u \neq u'} \sum_{u'} e^{(b+c)\hat{\beta}} \left\{ {}_0F_1(\frac{n-p}{2}; \frac{SS_z}{2}) \right.$$
$$\left. - {}_0F_1(\frac{n-p}{2}; \frac{SS_z}{4}[2 + b(X'X)^{-1}b' + c(X'X)^{-1}c' - (b+c)(X'X)^{-1}(b+c)']) \right\}$$

An approximation for $\hat{\theta}_U^V$ can be obtained by removing the first $b(X'X)^{-1}b'$ and replacing the hypergeometric function by the exponential function, giving

$$\hat{\theta}_1^V = \left(e^{\frac{2SS_z}{n-p}} - e^{\frac{SS_z}{n-p}} \right) \sum_u e^{2b\hat{\beta}}$$
$$+ \sum_{u \neq u'} \sum_{u'} e^{(b+c)\hat{\beta}} \left(e^{\frac{SS_z}{n-p}} - e^{\frac{SS_z}{2(n-p)}[2+b(X'X)^{-1}b'+c(X'X)^{-1}c'-(b+c)(X'X)^{-1}(b+c)']} \right).$$

This approximation will be examined in section 8.

6 Variance of UMVUE of E(IBNR claims)

To derive the variance of the estimator $\hat{\theta}_U^E$, the UMVUE of the expected value of IBNR claims constructed in the preceding section, we will follow the method of Shimizu (1988).

First, let us define the generalized hypergeometric function

$$ {}_pF_q(\alpha_1, \ldots, \alpha_p; \beta_1, \ldots, \beta_q; z) = \sum_{j=0}^{\infty} \frac{(\alpha_1)_j \cdots (\alpha_p)_j}{(\beta_1)_j \cdots (\beta_q)_j} \frac{z^j}{j!}. $$

In general, this series will converge for all finite z if $p \leq q$, converge for $|z| < 1$ if $p = q + 1$ and diverge for all $z \neq 0$ if $p > q + 1$.

Using equation (2) in Erdélyi (1981), Vol. 1, p. 185, we note that

$$ [{}_0F_1(\alpha; z)]^2 = {}_1F_2(\frac{1}{2}(2\alpha - 1); \alpha, 2\alpha - 1; 4z). $$

Since

$$ \begin{aligned} E[{}_0F_1(\frac{n-p}{2}; \frac{SS_z}{4})]^2 &= E[{}_1F_2(\frac{n-p-1}{2}; \frac{n-p}{2}, n-p-1; SS_z)] \\ &= \sum_{j=0}^{\infty} \frac{1}{j!} \frac{(\frac{n-p-1}{2})_j}{(\frac{n-p}{2})_j (n-p-1)_j} 2^j (\frac{n-p}{2})_j (\sigma^2)^j \\ &= {}_1F_1(\frac{n-p-1}{2}; n-p-1; 2\sigma^2) \end{aligned} $$

and (see Shimizu (1988), p. 33)

$$ {}_1F_1(\alpha; 2\alpha; 2kz) = e^{kz} \; {}_0F_1\left(\alpha + \frac{1}{2}; -\frac{z^2}{4}\right), $$

it follows that

$$ Var(\hat{\theta}_U^{EY}) = e^{2b\beta + \sigma^2(1 + b(X'X)^{-1}b')} \left\{ e^{\sigma^2 b(X'X)^{-1}b'} \; {}_0F_1(\frac{n-p}{2}; -\frac{1}{4}\sigma^4) - 1 \right\}. $$

Shimizu (1988, p. 34) has proved that

$$E\{\ {}_0F_1(\frac{n-p}{2};\frac{a}{2}SS_z)\cdot\ {}_0F_1(\frac{n-p}{2};\frac{b}{2}SS_z)\}$$

$$= e^{(a+b)\sigma^2}\ {}_0F_1(\frac{n-p}{2};ab\sigma^4).$$

Therefore,

$$Cov(e^{b\hat{\beta}}\ {}_0F_1(\frac{n-p}{2};\frac{SS_z}{4}), e^{c\hat{\beta}}\ {}_0F_1(\frac{n-p}{2};\frac{SS_z}{4}))$$

$$= E(e^{(b+c)\hat{\beta}})\cdot e^{\sigma^2}\cdot\ {}_0F_1(\frac{n-p}{2};\frac{\sigma^4}{4}) - E(Y_u)\times E(Y_{u'}),$$

by independence of $\hat{\beta}$ and SS_z. It follows that the variance of $\hat{\theta}_U^E$ is

$$Var(\hat{\theta}_U^E) = \sum_u e^{2b\beta+\sigma^2(1+b(X'X)^{-1}b')}\left\{e^{\sigma^2 b(X'X)^{-1}b'}\ {}_0F_1(\frac{n-p}{2};-\frac{\sigma^4}{4}) - 1\right\}$$

$$+\sum_{u\neq u'}\sum_{u'} e^{(b+c)\beta}\left\{\begin{array}{l} e^{\sigma^2[1+\frac{1}{2}(b+c)(X'X)^{-1}(b+c)']}\ {}_0F_1(\frac{n-p}{2};\frac{\sigma^4}{4}) \\ -e^{\sigma^2[1+\frac{1}{2}b(X'X)^{-1}b'+\frac{1}{2}c(X'X)^{-1}c']} \end{array}\right\}. \qquad (6.1)$$

For sufficiently large n, asymptotic expansion to the order n^{-1} of ${}_0F_1(\frac{n-p}{2};\frac{\sigma^4}{4})$ gives

$${}_0F_1(\frac{n-p}{2};\frac{\sigma^4}{4}) = 1 + \frac{\sigma^4}{2(n-p)} + O(n^{-2}),$$

leading to an expression for the asymptotic variance of the estimator $\hat{\theta}_U^E$.

7 MLE's of the mean and variance of IBNR claims

The maximum likelihood estimators for $E(Y_u)$ and $Var(Y_u)$ are

$$\hat{\theta}_L^{EY} = e^{b\hat{\beta}+\frac{1}{2}\hat{\sigma}^2[1+b(X'X)^{-1}b']}$$

and

$$\hat{\theta}_L^{VY} = e^{2b\hat{\beta}+\hat{\sigma}^2[1+b(X'X)^{-1}b']} \left(e^{\hat{\sigma}^2[1+b(X'X)^{-1}b']} - 1 \right).$$

From Mood et al. (1974, p.285, theorem 2), the maximum likelihood estimator of E (IBNR claims), denoted by $\hat{\theta}_L^E$, will be

$$\hat{\theta}_L^E = \sum_u e^{b\hat{\beta}+\frac{1}{2}\hat{\sigma}^2[1+b(X'X)^{-1}b']}.$$

Verrall (1991) has considered an estimator similar to $\hat{\theta}_L^E$, but with $\hat{\sigma}^2$ replaced with $\tilde{\sigma}^2$,

$$\hat{\theta}_V^E = \sum_u e^{b\hat{\beta}+\frac{1}{2}\tilde{\sigma}^2[1+b(X'X)^{-1}b']}.$$

We can find a lower bound for $\hat{\theta}_V^E$. Since $(X'X)^{-1}$ is a positive definite matrix, $b(X'X)^{-1}b' > 0$ and we get

$$\hat{\theta}_V^E > \sum_u e^{b\hat{\beta}+\tilde{\sigma}^2/2}.$$

This last expression is just the estimator $\hat{\theta}_1^E$, introduced in section 5. We can therefore order three of the four estimators for E (IBNR claims) mentioned in this paper,

$$\hat{\theta}_U^E < \hat{\theta}_1^E < \hat{\theta}_V^E.$$

The above inequalities imply that

$$E(\hat{\theta}_U^E) < E(\hat{\theta}_1^E) < E(\hat{\theta}_V^E).$$

Hence, both the estimators $\hat{\theta}_1^E$ and $\hat{\theta}_V^E$ exhibit a positive bias, which can be evaluated by calculating the expected value of $\hat{\theta}_1^E$ and $\hat{\theta}_V^E$.

Proposition 7.1:

$$E(\ \hat{\theta}_1^E) = \left(1 - \frac{\sigma^2}{n-p}\right)^{-\frac{n-p}{2}} \sum_u e^{b\beta + \frac{1}{2}\sigma^2 b(X'X)^{-1}b'}$$

Proof:

$$E(\ \hat{\theta}_1^E) = \sum_u E(e^{b\hat{\beta}})E(e^{\tilde{\sigma}^2/2}),$$

by independence of $\hat{\beta}$ and SS_z.

Since SS_z/σ^2 follows a χ^2 distribution with $(n-p)$ degrees of freedom and $E(e^{\gamma SS_z})$ is its moment generating function evaluated at $\gamma\sigma^2$, we obtain

$$E(e^{\gamma SS_z}) = \left(\frac{1}{1-2\gamma\sigma^2}\right)^{\frac{n-p}{2}};$$

Setting $\gamma = 1/2(n-p)$ leads to the desired result.

Proposition 7.2:

$$E(\ \hat{\theta}_V^E) = \sum_u e^{b\beta + \frac{1}{2}\sigma^2 b(X'X)^{-1}b'} \left(1 - \frac{\sigma^2[1 + b(X'X)^{-1}b']}{n-p}\right)^{-\frac{n-p}{2}}$$

Proof: Proceed similarly as for proposition 7.1, with $\gamma = \frac{1+b(X'X)^{-1}b'}{2(n-p)}$.

Proposition 7.3:

$$E(\ \hat{\theta}_L^E) = \sum_u e^{b\beta + \frac{1}{2}\sigma^2 b(X'X)^{-1}b'} \left(1 - \frac{\sigma^2[1 + b(X'X)^{-1}b']}{n}\right)^{-\frac{n-p}{2}}$$

Proof: Same as the proof of proposition 7.2, with $\gamma = \frac{1+b(X'X)^{-1}b'}{2n}$.

The maximum likelihood estimator for Var (IBNR claims), $\hat{\theta}_L^V$, is similarly obtained by replacing β by $\hat{\beta}$ and σ^2 by $\hat{\sigma}^2$ in formula (4.2).

8 Numerical example

Table 1 presents the incurred claims (in thousands of dollars), of a general insurance company's liability line of business, over the accident years 1978-1987. Using the stochastic chain ladder model

$$\ln Y_{ij} = \alpha_i + \beta_j + \epsilon_{ij}$$

on the trapezium of incremental data derived from table 1, we will compare the various estimators for E(IBNR claims) and Var(IBNR claims) proposed in this paper.

Table 2 contains the maximum likelihood estimates of the parameters of the model, while figure 1 shows a studentized residual plot for the stochastic chain ladder model, with the residuals appearing in the same order as vector Z.

From the parameters estimates, we easily get the estimates for E(IBNR claims) appearing in table 3 ($\hat{\theta}_C^E$ is the chain ladder estimator of the reserve). The unbiased estimate for E(IBNR claims) calculated with Verrall's formula gives 23,579, while formula (5.1) gives a value of 24,403.

Convergence for the hypergeometric series ${}_0F_1(15; \frac{SS_z}{4}) = 1.0362525$ is obtained by adding only the first six terms of the series. Using the upper bound $e^{SS_z/60}$ instead of ${}_0F_1(15; \frac{SS_z}{4})$ introduces a relative error of less than 0.004%. The estimator $\hat{\theta}_1^E$ therefore provides an excellent approximation, in this example, to $\hat{\theta}_U^E$. The estimated standard deviation of $\hat{\theta}_U^E$ using formula (6.1) gives 3804.

Table 4 contains the various estimators for the standard deviation of IBNR claims. The estimator $\hat{\theta}_U^S$ is obtained by taking the square root of $\hat{\theta}_U^V$, but it should be noted that it will not be unbiased for the standard deviation of IBNR claims, since the minimum variance unbiasedness property of an estimator is not preserved, under a non-linear tansformation. The maximum likelihood estimator is, however, invariant under transformations.

Besides a residual plot, model assumptions can be checked with a normal quantile-quantile plot of the ordered studentized residuals versus the normal density quantiles. This Q-Q plot (see figure 2) shows an excellent fit. The Shapiro-Francia (1972) test statistic measuring the correlation between the points and the straight line, gives 0.985; the critical value at the 5% level for a sample of size of 45, is 0.950, also indicating a good fit.

Table 5 gives the studentized residuals; Cook and Weisberg (1982) show that they are identically distributed random variables with mean 0 and variance 1. They are correlated, with $Cov(r_i, r_j) = -v_{ij}/[(1 - v_{ii})(1 - v_{jj})]^{0.5}$, where V is the hat matrix $V = X(X'X)^{-1}X'$. With this particular design of the stochastic chain ladder model, calculations of the covariances among the residuals of a given accident or development year showed that they were all negatively correlated, making difficult their interpretation.

9 Conclusion

In this paper, we have developed new estimators for the IBNR reserve in a log-normal linear model, among them the minimum variance unbiased estimator, and we have proposed several tests and plots to verify model assumptions. Other lognormal linear models, besides the stochastic chain ladder model, for which the minimum variance unbiased estimator can be calculated include the model proposed by Zehnwirth (1990),

$$\ln Y_{ij} = \alpha + \beta \ln j + \gamma j + \iota(i + j - 2) + \epsilon_{ij},$$

which takes into account superimposed inflation.

Those models all assume the homoskedasticity of the errors. Further work should look at models with unequal variances. The parameters would be estimated by weighted regression, with the weights determined empirically. Models where the errors could be correlated from year to year should also be investigated. We have only looked at the information contained in the aggregate payments; models which can use other information (case estimates, number of claims) deserve further research.

10 References

Aitchison, J. and Brown, J.A.C. (1957). *The Lognormal Distribution.* Cambridge University Press, Cambridge.

Bartholomew, D.J. (1975). Errors of prediction for Markov chain models. *Journal of the Royal Statistical Society, Series B*, 37 444-456.

Cohen, C. (1988). Three-parameter estimation. In *Lognormal Distributions, Theory and Applications*. Marcel Dekker, Inc., New York.

Cook, R.D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* Chapman & Hall, London.

Crow, E.L. and Shimizu, K. (1988). *Lognormal Distributions, Theory and Applications.* Marcel Dekker, Inc., New York.

De Vylder, F. (1978). Estimation of IBNR claims by least squares. *Bulletin of the Association of Swiss Actuaries* 78, 247-254.

Erdélyi, A. (1981). *Higher transcendental functions.* Robert E. Krieger Publishing Company, Inc., Malabar.

Finney, D.J. (1941). On the distribution of a variable whose logarithm is normally distributed. *Journal of the Royal Statistical Society* 7, 155-161.

Graybill, F.A. (1961). *An introduction to linear statistical models*, Volume 1. McGraw Hill Book Company, Inc., New York.

Hogg, R.V. and Craig, A.T. (1978). *Introduction to Mathematical Statistics.* Macmillan Publishing Co., Inc., New York.

Johnson, N.L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 1 and 2, Houghton Mifflin Company, Boston.

Kremer, E. (1982). IBNR claims and the two-way model of ANOVA. *Scandinavian Actuarial Journal*, 47-55.

Mood, A.M., Graybill, F.A. and Boes, D.C. (1974). *Introduction to the theory of Statistics*. McGraw Hill Book Company, Inc., New York.

Shapiro, S.S. and Francia, R.S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67, 215-216.

Shimizu, K. (1988). Point Estimation. In *Lognormal Distributions, Theory and Applications*. Marcel Dekker, Inc., New York.

Taylor, G.C. (1986). *Claim reserving in non-life insurance*. North-Holland, Amsterdam.

Taylor, C.G. and Ashe, F.R. (1983). Second moments of estimates of outstanding claims. *Journal of Econometrics* 23, 37-61.

Van Eeghen, J. (1981). *Loss reserving methods*. Volume No. 1 in the series Surveys of actuarial studies, published by Nationale-Nederlanden N.V., Rotterdam.

Verrall, R.J. (1991). On the estimation of reserves from loglinear models. *Insurance: Mathematics and Economics* 10, 75-80.

Verrall, R.J. and Li, Z. (1993). Negative Incremental Claims: Chain Ladder and Linear Models. *Journal of the Institute of Actuaries* 120, 171-183.

Zehnwirth, B. (1990). *Interactive Claims Reserving Forecasting System*, Version 5.3, Manual Volume 1, Benhar Nominees Pty. Ltd., Australia.

Table 1: Claims Incurred

Accident year	Development year					
	1	2	3	4	5	6
1978	8489	9785	10709	11289	11535	11661
1979	12970	14766	16201	17060	17714	17979
1980	17522	20305	21774	22797	23220	23872
1981	21754	24338	25501	26284	27171	27526
1982	19208	21549	22769	23388	24229	24932
1983	19604	22073	23296	24543	25155	
1984	21922	24233	25374	26882		
1985	25038	28401	30545			
1986	32532	37006				
1987	39862					

Table 2: MLE's for the stochastic chain ladder model

Parameter	Estimate	Standard Error
α_1	9.0654	0.1387
α_2	9.6161	0.1387
α_3	9.8497	0.1387
α_4	9.8120	0.1387
α_5	9.8486	0.1387
α_6	9.8442	0.1454
α_7	9.9079	0.1553
α_8	10.1798	0.1709
α_9	10.4119	0.1990
α_{10}	10.5932	0.2670
β_2	-2.0277	0.1259
β_3	-2.5926	0.1316
β_4	-2.9081	0.1379
β_5	-3.3435	0.1454
β_6	-3.7737	0.1549
σ^2	0.0475	

Table 3: Estimates of the mean of IBNR claims

Estimator	Estimate
$\hat{\theta}_U^E$	24,403
$\hat{\theta}_1^E$	24,404
$\hat{\theta}_L^E$	24,677
$\hat{\theta}_V^E$	25,262
$\hat{\theta}_C^E$	23,919

Table 4: Estimate of the standard deviation of IBNR claims

Estimator	Estimate
$\hat{\theta}_U^S$	4667
$\hat{\theta}_1^S$	4786
$\hat{\theta}_L^S$	3984

Table 5: Studentized Residuals for stochastic chain ladder model

accident year	development year					
	1	2	3	4	5	6
1978	-0.083	0.567	1.567	0.913	-0.974	-2.089
1979	-0.639	-0.417	1.081	0.212	0.947	-1.205
1980	-0.344	0.479	0.155	-0.0495	-2.064	1.853
1981	0.769	0.319	-0.708	-1.069	1.436	-0.762
1982	0.0635	-0.274	-0.658	-2.276	1.032	2.204
1983	0.175	-0.0222	-0.640	0.871	-0.385	
1984	0.402	-0.640	-1.235	1.489		
1985	-0.252	-0.154	0.408			
1986	-0.123	-0.123				
1987	0					