THE LINEAR MARKOV PROPERTY IN CREDIBILITY THEORY

By Thomas Witting

ETH Zürich

ABSTRACT

We study the linear Markov property, i.e. the possibility of basing the credibility estimator on data of the most recent time period without loss of accuracy. Necessary and sufficient conditions are derived generally. The meaning of the linear Markov property is also discussed in different experience rating and loss reserving models.

KEYWORDS

Linear Markov property; linear sufficiency; credibility.

1. INTRODUCTION

A fundamental question in credibility theory is that of upon which statistic of the available data the credibility estimator should be based. A very general treatment of this problem and a survey of other approaches can be found in NEUHAUS (1985). We consider the special case of data ordered with respect to time. Is it then possible to reduce the data to those of the last time period without diminishing the accuracy of the credibility estimator? If this is the case, then we have defined the linear Markov property. This principle is introduced generally and discussed in some important models of risk theory. We give some sufficient and necessary conditions which are useful in situations when the linear Markov property is not obvious. In most cases the linear Markov property results in a considerable reduction of the number of normal equations which it is necessary to solve to derive the credibility estimator explicitly.

This paper is in a way a summary of the first part of the author's PhD thesis which is taken sometimes as a reference. A copy of this thesis can be obtained from the author.

2. CREDIBILITY ESTIMATION AND LINEAR MARKOV PROPERTY

2.1. *General Assumptions and Notation*

In the present paper it is generally assumed that random variables are square integrable, i.e. all (mixed) second moments exist and are finite. The transpose of a matrix A is A^T. (Random) vectors are in boldface and have to be interpreted as column vectors, i.e. $x = (x_1, \ldots, x_n)^T$ is a vector with n components.

I_n is the $n \times n$-unit matrix and δ_{ij} the Kronecker symbol.

For random vectors $X = (X_1, \ldots, X_n)^T$ and $Y = (Y_1, \ldots, Y_m)^T$ we use the following symbols and terminology:

P_X for the probability distribution of X.

$P_{X|Y=y}$ for the conditional probability distribution of X given $[Y = y]$ and $P_{X|Y}$ for the corresponding stochastic kernel.

$E[X] = (E[X_1], \ldots, E[X_n])^T$ for the expected value of X.

$E[X|Y] = (E[X_1|Y], \ldots, E[X_n|Y])^T$ for the conditional expected value of X given Y.

$C[X, Y] = E[(X - E[X])(Y - E[Y])^T]$ for the joint covariance matrix of X and Y.

$C[X] = C[X, X]$ for the covariance matrix of X.

It is generally assumed that all symmetric covariance matrices $C[X]$ appearing in the text are positive definite, i.e. the inverse $C[X]^{-1}$ exists. The regularity of $C[X]$ is equivalent to linear independence of the "vectors" $1, X_1, \ldots, X_n$ in the linear space $L_2(R)$ of all square integrable real random variables. For a proof see e.g. WITTING (1986). In particular all random variables appearing in the text are not degenerate. All equations between random variables should be understood in the sense of L_2-equivalence.

2.2. *Credibility Estimation*

We want to estimate the real random variable Y with help of the n-dimensional random vector X which represents the available data. It is well known that $g^*(X) = E[Y|X]$ is the optimal estimator in the sense of minimizing the expected squared loss $E[(g(X) - Y)^2]$ in the class of all measurable functions $g(x)$. Because $E[Y|X]$ can be calculated explicitly by a closed formula only in a few special cases the estimation problem is simplified: we look for the optimal estimator of Y only in the class of (inhomogeneous) linear estimators

$$g(X) = a_0 + \sum_{i=1}^{n} a_i X_i.$$

This optimal estimator exists, is uniquely determined and interpreted as the orthogonal projection of Y onto the $n + 1$-dimensional subspace of $L_2(R)$ which is generated by $1, X_1, \ldots, X_n$. Therefore we denote it $\hat{E}[Y|X]$. $\hat{E}[Y|X]$ is called the *credibility estimator* of Y given X.

The orthogonal principle can be formulated in a probabilistic manner as follows:

(1)
$$E[\hat{E}[Y|X]] = E[Y]$$
$$C[Y - \hat{E}[Y|X], X] = \mathbf{0}.$$

If the credibility estimator is written in the form

$$\hat{E}[Y|X] = a_0 + \sum_{i=1}^{n} a_i X_i$$

(1) is equivalent to

(2)
$$a_0 = E[Y] - \sum_{i=1}^{n} a_i E[X_i]$$

$$\sum_{i=1}^{n} a_i \operatorname{Cov}(X_i, X_k) = \operatorname{Cov}(Y, X_k) \qquad k = 1, \ldots, n.$$

This linear system of normal equations for determining the coefficients $a_0, a_1, \ldots, a_n$ has a unique solution because of our general assumption that $C[X]$ is positive definite. There is no guarantee for being able to calculate $C[X]^{-1}$ explicitly although this may be useful in theoretical situations. However, a recursive algorithm for the inversion of $C[X]$ exists always (see e.g. NORBERG (1985)).

If a random vector $Y = (Y_1, \ldots, Y_m)^T$ has to be estimated, we define $\hat{E}[Y|X] = (\hat{E}[Y_1|X], \ldots, \hat{E}[Y_m|X])^T$ and confirm the property

$$E[(\hat{E}[Y|X] - Y)^T(\hat{E}[Y|X] - Y)] = \min_{g} E[(g(X) - Y)^T(g(X) - Y)].$$

The minimum is taken over the class of all functions $g(x) = a + Ax$ with m-dimensional vector a and $m \times n$-matrix A. The generalization of (1) to this case is obvious. Finally we get the well-known formula

(3)
$$\hat{E}[Y|X] = C[Y, X]C[X]^{-1}(X - E[X]) + E[Y].$$

2.3. *Linear Sufficiency*

We consider again the problem of estimating Y by means of X. For many statistical problems one can restrict the investigation to decision functions which depend only through a "sufficient" statistic $T(x)$ on the original observation x. Here we call a statistic $T(x)$ *sufficient* if

(4)
$$P_{Y|X} = P_{Y|T(X)}$$

This corresponds with the Bayesian definition of sufficiency if Y is interpreted as a "prior variable".

In the credibility situation one should manage only with linear statistics and the knowledge of second-order moments. This fact suggests a slight change of the meaning of sufficiency in our case.

DEFINITION: The linear statistic $T(x)$ (which is formally a linear mapping $T: R^n \to R^r$ with $r < n$) is called *linear sufficient* if

(5)
$$\hat{E}[Y|X] = \hat{E}[Y|T(X)].$$

REMARKS: (i) By comparing the system (X, Y) with the corresponding system $(\tilde{X}, \tilde{Y})$ which is normally distributed with the same second-order moment structure it can be proved that for linear statistics $T(x)$ the implication (4) $\Rightarrow$ (5) is valid.

(ii) We can restrict the investigation to homogeneous linear mappings $T(x)$, because a possible inhomogeneous part has to be adapted anyway afterwards when $\hat{E}[Y \mid T(X)]$ is calculated.

(iii) The concept of linear sufficiency has been already introduced into statistical literature, but only in the context of estimation for linear models; e.g. DRYGAS (1983, 1985).

LEMMA 1. *Let $r < n$ and A a full rank $r \times n$-matrix. The statistic $T(x) = Ax$ is linear sufficient if, and only if*

$$C[Y, X]\{I_n - A^T(AC[X]A^T)^{-1}AC[X]\} = 0.$$

PROOF. Without loss of generality we may assume $E[X] = \mathbf{0}, E[Y] = \mathbf{0}$. Then it follows that $T(x)$ is linear sufficient.

$$\underset{(3)}{\Leftrightarrow} \hat{E}[Y \mid X] = C[Y, T(X)]C[T(X)]^{-1}T(X)$$

$$\underset{(1)}{\Leftrightarrow} C[Y - C[Y, T(X)]C[T(X)]^{-1}T(X), X] = 0$$

(6) $\qquad \Leftrightarrow C[Y, X] = C[Y, T(X)]C[T(X)]^{-1}C[T(X), X]$

$$\Leftrightarrow C[Y, X] = C[Y, X]A^T(AC[X]A^T)^{-1}AC[X]. \qquad\qquad \text{QED}$$

EXAMPLE. ($r = m = 1$).

Let A be the $1 \times n$-matrix and E the $n \times n$-matrix whose elements are all equal to 1. We assume that the random variables $X_1, \ldots, X_n$ are exchangeable relative to Y, i.e. $P_{(X_1, \ldots, X_n, Y)} = P_{(X_{\pi(1)}, \ldots, X_{\pi(n)}, Y)}$ for all permutations π of $1, \ldots, n$.

In SUNDT (1979), Theorem 1, it is shown that this condition implies the linear sufficiency of the statistic $T(x) = Ax = \Sigma x_i$. This implication can also be derived from Lemma 1, for it follows from the exchangeability condition with appropriate constants c, d and e that:

$$C[X] = d(E + cI_n), C[Y, X] = eA.$$

This implies together with the simple relationships $EE = nE, AE = nA$ and $A^TA = E$ that:

$$AC[X]A^T = dn(n + c), A^TAC[X] = d(n + c)E.$$

From this follows

$$C[Y, X]\{I_n - A^T(AC[X]A^T)^{-1}AC[X]\} = eA(I_n - 1/nE) = e(A - A) = \mathbf{0}.$$

2.4. *Linear Markov Property*

In the present paper we consider mainly a special case of linear sufficiency, namely the linear Markov property.

Now, there are given n information **vectors** of dimension l

$$X_1 = (X_{11}, \ldots, X_{l1})^T, \ldots, X_n = (X_{1n}, \ldots, X_{ln})^T$$

from which the random vector $Y = (Y_1, \ldots, Y_m)^T$ shall be estimated. We patch the complete information together to the $n \times l$-dimensional vector $X = (X_1^T, \ldots, X_n^T)^T$.

DEFINITION. The sequence $X_1, \ldots, X_n, Y$ is called *linear Markovian (l.M.)* if $\hat{E}[Y \mid X] = \hat{E}[Y \mid X_n]$.

The linear Markov property is equivalent to the linear sufficiency of the statistic $T(x) = x_n$ and makes it possible to reduce the complete information to the information of the last period.

In the language of NEUHAUS (1985) it means that the secondary statistic $(X_1^T, \ldots, X_{n-1}^T)$ may be excluded from the basic statistic X without loss. The linear Markov property can be characterized by a relation between the second-order moments:

LEMMA 2. *The sequence* $X_1, \ldots, X_n, Y$ *is l.M. if, and only if*

(7) $C[Y, X_i] = C[Y, X_n] C[X_n]^{-1} C[X_n, X_i]$ for $i = 1, \ldots, n - 1$.

The proof follows as special case from Lemma 1 with $T(x) = x_n$, because (6) is then equivalent to (7).

Now we define the linear Markov property also for **processes**:

DEFINITION. Let X_i be a l-dimensional random vector for all $i \in N$. The stochastic (vector-)processes $(X_i)_{i \in N}$ is called *linear Markovian (l.M.)* if the sequences $X_1, \ldots, X_n, X_{n+k}$ are l.M. for all $n, k \in N$.

REMARKS. (i) We consider a 1-dimensional process $(X_i)_{i \in N}$.
Then $(X_i)_{i \in N}$ is l.M. if, and only if the following relation is valid with $c_{i,k} = \mathrm{Cov}(X_i, X_k)$:

(8) $c_{n+k,i} c_{n,n} = c_{n+k,n} c_{n,i}$ for $i, k, n \in N$ with $i < n$.

FELLER (1966) shows that the ordinary Markov property is characterized by (8) for a Gaussian process $(X_i)_{i \in N}$. In this special case the ordinary and the linear Markov property do coincide.

PAPOULIS (1965) shows the corresponding result for the optimal homogeneous linear estimation. In that case we would have to define $\hat{E}[Y \mid X_1, \ldots, X_n]$ as an orthogonal projection from Y onto the linear subspace generated by $X_1, \ldots X_n$. Then (8) is valid with $c_{i,k} = E[X_i X_k]$.
(ii) For a standard normal variable Z and arbitrary i.i.d. variables Z_1, Z_2 it follows that the sequence $X_1 = Z^2, X_2 = Z, Y = Z^2$ is Markovian in the ordinary sense but not l.M. The sequence $X_1 = Z_1, X_2 = Z_1 + Z_2, Y = Z_1 \cdot Z_2$ is l.M. but not Markovian in the ordinary sense.

These two examples show that the ordinary Markov property does not imply the linear Markov property and vice versa.

The following lemma gives some helpful necessary and sufficient conditions to detect the linear Markov property directly by inspection of the covariance structure of the process.

LEMMA 3. *The following conditions (7)', (9) and (10) are equivalent to the linear Markov property of the process* $(X_i)_{i \in N}$:

$$(7)' \qquad\qquad C[X_{n+k}, X_i] = C[X_{n+k}, X_n] C[X_n]^{-1} C[X_n, X_i]$$

for $i, k, n \in N$ *with* $i < n$.
There exists a sequence $(\Lambda_i)_{i \in N}$ *of regular* $l \times l$-*matrices with*

$$(9) \qquad\qquad C[X_j, X_i] = \left(\prod_{k=i+1}^{j} \Lambda_k \right) C[X_i] \qquad for\ i \leqslant j.$$

There exist sequences $(A_i)_{i \in N}$ *and* $(B_i)_{i \in N}$ *of regular* $l \times l$-*matrices with*

$$(10) \qquad\qquad C[X_j, X_i] = B_j A_i \qquad for\ i \leqslant j.$$

where

$$\prod_{k=r}^{m} C_k := \begin{cases} C_m \ldots C_r & for\ r \leqslant m \\ I_l & for\ r > m. \end{cases}$$

PROOF. Because of Lemma 2 the linear Markov property of the process $(X_i)_{i \in N}$ is equivalent to condition (7)'. Therefore it suffices to show:

$(9) \Rightarrow (10) \Rightarrow (7)' \Rightarrow (9);$

(9) implies (10) with

$$B_j = \prod_{k=1}^{j} \Lambda_k \text{ and } A_i = \left(\prod_{k=1}^{i} \Lambda_k \right)^{-1} C[X_i];$$

(10) implies (7)' by means of the relation

$$C[X_{n+k}, X_n] C[X_n]^{-1} C[X_n, X_i] = B_{n+k} A_n A_n^{-1} B_n^{-1} B_n A_i$$
$$= B_{n+k} A_i$$
$$= C[X_{n+k}, X_i];$$

(7)' implies (9) with $\Lambda_k = C[X_k, X_{k-1}] C[X_{k-1}]^{-1}$, for it follows with $i < j$:

$$C[X_j, X_i] \underset{(7)'}{=} C[X_j, X_{j-1}] C[X_{j-1}]^{-1} C[X_{j-1}, X_i]$$

$$= \Lambda_j C[X_{j-1}, X_i]$$

$$\underset{\text{(induction)}}{=} \Lambda_j \ldots \Lambda_{i+1} C[X_i] \qquad\qquad\qquad\qquad \text{QED}$$

NOTATION. The sequence of $l \times l$-matrices $(A_i)_{i \in N}$ in (10) is called a *l.M.-factor*. A_i is fixed uniquely to the extent of multiplication from the left of a $l \times l$-matrix independent of i.

Analogously with the segment (10) $\Rightarrow$ (7)$'$ in the proof of Lemma 3 the following result for finite sequences can be shown:

LEMMA 4. *Let (10) be valid for $1 \leqslant i \leqslant j \leqslant n$ whereby $A_1, \ldots, A_n$ is the beginning of the l.M.-factor and also*

(11) $\qquad C[Y, X_i] = BA_i$ *for $1 \leqslant i \leqslant n$ with a $l \times l$-matrix B.*

Then the sequence $X_1, \ldots, X_n, Y$ is l.M.

2.5. *Componentwise Linear Markov Property*

We use the same notation as in 2.4.

DEFINITION. The sequence $X_1, \ldots, X_n, Y$ is called *componentwise linear Markovian (c.l.M.)* if the sequences $X_{k1}, \ldots, X_{kn}, Y$ are l.M. for all $k = 1, \ldots, l$.

If the components, i.e. the rows of the $l \times n$-matrix $(X_1, \ldots, X_n)$, are independent, it holds that:

$$X_1, \ldots, X_n, Y \text{ is l.M.} \Leftrightarrow X_1, \ldots, X_n, Y \text{ is c.l.M.}$$

This equivalence is evident by Lemma 2, because in the case of independence the matrices $C[X_n]^{-1}$ and $C[X_n, X_i]$ in (7) are diagonal.

Generally no direction of this equivalence is valid, for it holds with two independent real random variables Z_1 and Z_2:

The sequence $X_1 = (0, Z_2)^T, X_2 = (Z_1 + Z_2, 0)^T, Y = Z_1$ is c.l.M. but not l.M.

The sequence $X_1 = (Z_1, Z_2)^T, X_2 = (Z_2, Z_1)^T, Y = Z_1 + Z_2$ is l.M. but not c.l.M.

In the situation of insurance the independence of components is not always fulfilled. As an example one should imagine the components to be claim numbers and totals of claims. Reflecting on the better handling of the c.l.M.-property we are looking for an additional condition that the c.l.M.-property implies the l.M.-property even in the case of dependent components.

To solve this problem we consider two vector valued components. So let $X_i = (Z_i^T, N_i^T)^T$ with l_1-dimensional random vector Z_i and l_2-dimensional random vector N_i and $l_1 + l_2 = l$ ($1 \leqslant i \leqslant n$).

LEMMA 5. *Let the following four conditions be valid:*

The sequence $Z_1, \ldots, Z_n, Y$ is l.M.
The sequence $N_1, \ldots, N_n, Y$ is l.M.
The sequence $Z_1, \ldots, Z_n, N_n$ is l.M.
The sequence $N_1, \ldots, N_n, Z_n$ is l.M.

Then the sequence $X_1, \ldots, X_n, Y$ is l.M.

To prove this lemma condition (7) has to be checked with help of inversion of the matrix $C[X_n]$. This is somewhat tedious and can be found in WITTING (1986), p.33−36.

3. THE LINEAR MARKOV PROPERTY IN SOME EXPERIENCE RATING MODELS

In the following models our starting point is always a real stochastic process $(Y_i)_{i \in N}$ with covariance structure given by $c_{i,k} = \mathrm{Cov}(Y_i, Y_k)(i, k \in N)$. Thereby Y_i may be interpreted as claim number or total of claims during the period i. At the end of period n the net premium P_{n+1} for the next period will be fixed by $P_{n+1} = \hat{E}[Y_{n+1} \mid Y_1, \ldots, Y_n]$. Let $(X_i)_{i \in N}$ be a "linear cumulated transform" of the process $(Y_i)_{i \in N}$, i.e.

$$X_i = \sum_{k=1}^{i} a_k Y_k$$

with appropriate coefficients $a_1, \ldots, a_i$ $(i \in N)$.

Let us further assume that the process $(X_i)_{i \in N}$ has the covariance structure given by

(12) $\mathrm{Cov}(X_i, X_j) = g_i + f_i f_j$ for $i \leqslant j$ with $g_i > 0$ and $f_{i+1} - f_i \neq 0$.

With help of the multiplicative decomposition criterion (10) of Lemma 3 the following equivalence can easily be verified:

(13) $(X_i)_{i \in N}$ is I.M. $\Leftrightarrow$ The quotient f_i / g_i is independent of i.

Indication of the proof of the " $\Rightarrow$ " part: From (10) it follows that the fractions

$$\frac{g_i + f_i f_j}{g_1 + f_1 f_j} \quad \text{and} \quad \frac{1 + (f_i / g_i) f_j}{1 + (f_1 / g_1) f_j}$$

do not depend on j. This can only be true if f_i / g_i is independent of i because of our assumption in (12) that f_j depends on j.

3.1. *The Model of Jewell*

JEWELL (1975) considers the covariance structure given by $c_{i,k} = \delta_{ik} \gamma_k + \alpha_i \alpha_k$ with appropriate numbers $\gamma_k > 0$ and $\alpha_i \neq 0$ $(i, k \in N)$. This covariance structure is shown by Jewell to yield an explicit solution of the normal equations (2). Under which conditions has the transformed process $(X_i)_{i \in N}$ the linear Markov property? In Jewell's model we have

$$\mathrm{Cov}(X_i, X_j) = \sum_{m=1}^{i} a_m^2 \gamma_m + \left(\sum_{m=1}^{i} a_m \alpha_m \right) \left(\sum_{k=1}^{j} a_k \alpha_k \right).$$

So (12) is fulfilled with

$$g_i = \sum_{m=1}^{i} a_m^2 \gamma_m \text{ and } f_i = \sum_{m=1}^{i} a_m \alpha_m.$$

We conclude:

$$(X_i)_{i \in N} \text{ is I.M. } \underset{(13)}{\Leftrightarrow} \frac{\displaystyle\sum_{m=1}^{i} a_m \alpha_m}{\displaystyle\sum_{m=1}^{i} a_m^2 \gamma_m} \text{ is independent of } i.$$

(14) $\Leftrightarrow \dfrac{\alpha_m}{a_m \gamma_m}$ is independent of m.

Therefore in Jewell's model the statistic $\Sigma(\alpha_k/\gamma_k)y_k$ is linear sufficient and the premium becomes

$$P_{n+1} = \hat{E}\left[Y_{n+1} \mid \sum_{k=1}^{n} \frac{\alpha_k}{\gamma_k} Y_k\right].$$

3.2. The Classical Credibility Model

It is obvious that the classical Bühlmann–Straub model (BÜHLMANN and STRAUB, 1970) in the ordinary formulation is a special case of Jewell's model with known numbers α_m and γ_k. We choose a cumulative view onto the model. Considering one risk unit we denote by

X_i, the cumulated total of claims up to the end of period i;

Θ, a random risk parameter describing the unknown characteristics of the risk unit;

p_i, a known cumulated measure of volume up to period i.

It is assumed that for given $[\Theta = \theta]$ the process $(X_i)_{i \in N}$ has independent increments, and $E[X_i \mid \Theta = \theta] = \mu(\theta)p_i$ with a measurable function $\mu(.)$ independent of i. From these assumptions it follows that:

$$\text{Cov } (X_i, X_j) = E[\text{Var}[X_i \mid \Theta]] + p_i p_j \text{Var}[\mu(\Theta)] \text{ for } i \leqslant j.$$

With (13) we conclude:

(15) $\qquad (X_i)_{i \in N}$ is l.M. $\Leftrightarrow \dfrac{p_i}{E[\text{Var}[X_i \mid \Theta]]}$ is independent of i.

In this case the l.M.-factor is (p_i).

It should be mentioned that (15) is fulfilled in the classical credibility model of Bühlmann and Straub.

3.3. The Model of Shur

SHUR (1972) considers the following model. The variables $Y_1, Y_2, \ldots$ have all the same expected value μ and the same variance σ^2, and the covariance structure is given by

(16) $\qquad c_{i,k} = \rho^{|i-k|}\sigma^2 \text{ with } 0 \leqslant \rho \leqslant 1.$

Hence the correlation between the total losses of two different periods decreases geometrically with the number of periods separating them. By inversion of the matrix

$$(c_{i,k})_{i,k=1}^{n} = \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\ \rho & 1 & \rho & \ldots & \rho^{n-2} \\ \vdots & & & & \vdots \\ \rho^{n-1} & \rho^{n-2} & & \ldots & 1 \end{pmatrix} \sigma^2$$

and application of (3) one gets the credibility formula

$$(17) \qquad\qquad P_{n+1} = \rho Y_n + (1 - \rho)\mu.$$

Of course, this formula implies the linear Markov property of the process $(Y_i)_{i\in N}$. But this fact can already be detected directly from (16), because

$$c_{i,j} = \mathrm{Cov}(Y_i, Y_j) = \rho^{-i}\rho^j\sigma^2 \text{ for } i \leqslant j$$

fulfils the multiplicative decomposition criterion (10).

Following this way the inversion of the matrix $(c_{i,k})$ becomes unnecessary for the calculation of formula (17).

3.4. *A Model with Claim Numbers and Individual Claim Amounts*

Let us consider one risk unit with the following notations:

N_i is the cumulative number of claims up to the end of period i.
$Z^{(k)}$ is the amount of the kth individual claim. (It is assumed that these claims are numbered according to their order of occurrence.)

$X_i = \sum_{k=1}^{N_i} Z^{(k)}$ is the cumulative total of claims up to the end of period i.

Θ denotes a random risk parameter describing the unknown characteristics of the risk unit.

We make the following assumptions:
(A1) Given $[\Theta = \theta]$ the random variables $Z^{(1)}, Z^{(2)}, \ldots$ are i.i.d.
(A2) Given $[\Theta = \theta]$ the stochastic processes $(N_i)_{i\in N}$ and $(Z^{(k)})_{k\in N}$ are independent.

PROBLEM. Which are sufficient conditions such that the process $(X_i)_{i\in N}$ resp. the 2-dimensional process $((X_i, N_i)^T)_{i\in N}$ is l.M.? This would simplify the premium

$$P_{n+1} = \hat{E}[X_{n+1} - X_n \mid X_1, \ldots, X_n]\,\text{resp.}$$

$$P_{n+1} = \hat{E}[X_{n+1} - X_n \mid X_1, \ldots, X_n, N_1, \ldots, N_n]$$

as usual, namely

$$P_{n+1} = \hat{E}[X_{n+1} - X_n \mid X_n] \text{ resp. } P_{n+1} = \hat{E}[X_{n+1} - X_n \mid X_n, N_n].$$

CONDITION I. Given $[\Theta = \theta]$ $(N_i)_{i\in N}$ is an inhomogeneous Poisson process, $\lambda_i g(\theta)$ being the Poisson parameter of N_i. Thereby $g(.)$ is a measurable function independent of i.

NOTE. It is not required in assumptions (A1)-(A2) and condition I that claim numbers and claim amounts are independent. They have only to be conditionally independent.

We get from assumptions (A1) and (A2) and Condition I: $(X_i)_{i\in N}$ is a process

with conditional independent increments (given Θ), and

$$E[X_i \mid \Theta] = \lambda_i g(\Theta) E[Z^{(1)} \mid \Theta], \; \mathrm{Var}[X_i \mid \Theta] = \lambda_i g(\Theta) E[Z^{(1)}]^2 \mid \Theta].$$

Therefore condition (15) is valid with $p_i = \lambda_i$ and the process $(X_i)_{i \in N}$ (and also the process $(N_i)_{i \in N}$) is l.M. with l.M.-factor (λ_i). It remains to prove the l.M.-property of the 2-dimensional process $((X_i, N_i)^T)_{i \in N}$. Because of Lemma 5 it suffices to check two 1-dimensional conditions (for fixed n):

(18) The sequence $X_1, \ldots, X_n, N_n$ is l.M.

(19) The sequence $N_1, \ldots, N_n, X_n$ is l.M.

It is true for $1 \leqslant i \leqslant n$ that:

$\mathrm{Cov}(X_i, N_n) = \mathrm{Cov}(N_i, X_n)$

$$= E\Big[\mathrm{Cov}\Big(\sum_{k=1}^{N_i} Z^{(k)}, N_n \mid \Theta\Big)\Big] + \mathrm{Cov}\Big(E\Big[\sum_{k=1}^{N_i} Z^{(k)} \mid \Theta\Big], E[N_n \mid \Theta]\Big)$$

$$= E[E[Z^{(1)} \mid \Theta] \mathrm{Var}[N_i \mid \Theta]] + \lambda_i \lambda_n \mathrm{Cov}(E[Z^{(1)} \mid \Theta]g(\Theta), g(\Theta))$$

(A1),(A2)

$$= \lambda_i \{ E[E[Z^{(1)} \mid \Theta]g(\Theta)] + \lambda_n \mathrm{Cov}(E[Z^{(1)} \mid \Theta]g(\Theta), g(\Theta)) \}$$

cond.I

$$= \lambda_i \times \text{term which is independent of } i.$$

Because each of the processes $(X_i)_{i \in N}$ and $(N_i)_{i \in N}$ has the l.M.-factor (λ_i) we get (18) and (19) by application of Lemma 3 (criterion (10)) and Lemma 4. The process $((X_i, N_i)^T)_{i \in N}$ is actually l.M.

Now we replace the Poisson assumption (condition I) by the hypothesis that the counting process $(N_i)_{i \in N}$ is l.M. and claim numbers and claim amounts are independent (level-2 assumption):

CONDITION II.

(A3) $(N_i)_{i \in N}$ is l.M. with l.M.-factor $(E[N_i])$.
(A4) $(N_i)_{i \in N}$ and Θ are independent.

REMARKS. (i) We have lost the convenient property that the increments of the process $(N_i)_{i \in N}$ resp. $(X_i)_{i \in N}$ are independent given Θ. Therefore it is not possible to apply the classical credibility model and condition (15) any longer.
(ii) Condition I implies (A3).
We need the further notation:

$$v_Z = E[\mathrm{Var}[Z^{(1)} \mid \Theta]] \text{ and } w_Z = \mathrm{Var}[E[Z^{(1)} \mid \Theta]].$$

Then we get from the assumptions (A1)-(A4):

$\mathrm{Cov}(X_i, X_j) = E[\mathrm{Cov}(X_i, X_j \mid \Theta, (N_k)_{k \in N})]$

$$+ \mathrm{Cov}(E[X_i \mid \Theta, (N_k)_{k \in N}], E[X_j \mid \Theta, (N_k)_{k \in N}])$$

$$= E[\, N_i \, \text{Var}[\, Z^{(1)} \mid \Theta] \,] + \text{Cov}(N_i E[\, Z^{(1)} \mid \Theta],\, N_j E[\, Z^{(1)} \mid \Theta])$$
(A1),(A2)

$$= E[\, N_i] \, v_Z + \text{Cov}(N_i,\, N_j)\{ w_Z + m_Z{}^2 \} + E[\, N_i] E[\, N_j] \, w_Z$$
(A4)

$$= E[\, N_i] \times \text{factor which depends only on } j \text{ (and not on } i) \ (i \leqslant j).$$
(A3),(10)

Applying criterion (10) again we obtain that the process $(X_i)_{i \in N}$ is l.M. with l.M.-factor $(E[\, N_i])$. Furthermore it is true for $i \leqslant n$:

$$\text{Cov}(X_i,\, N_n) = \text{Cov}(N_i,\, X_n) = \text{Cov}(N_i,\, E[\, X_n \mid \Theta,\, (N_k)_{k \in N}])$$

$$= \text{Cov}(N_i,\, N_n E[\, Z^{(1)} \mid \Theta]) = \text{Cov}(N_i,\, N_n) m_Z$$
(A1),(A2)   (A4)

$$= E[\, N_i] \times \text{factor independent of } i.$$

Analogously with condition I the l.M.-property of the process $((X_i, N_i)^T)_{i \in N}$ follows.

4. THE LINEAR MARKOV PROPERTY IN SOME LOSS-RESERVING MODELS

The problem of estimating the ultimate loss reserve will not be presented with full rigour. Our only aim is to indicate the role of the l.M.-property in the most important loss-reserving models with credibility character.

The usual loss-reserving terminology is assumed to be known. Let Y_{ij} be the total of claims of accident year j which is reported during the development year i. Thereby we assume that each individual claim of accident year j is settled at its full amount immediately, i.e. there are no IBNER-claims resp. the IBNER-part is already contained in Y_{ij} as estimation.

The statistician considers each of the processes $(Y_{ij})_{i \in N}$ up to a certain time $n(j)$. For constituting the reserve he has to estimate the random variable

$$R_j = Y_{n(j)+1, j} + \ldots + Y_{\infty j}.$$

Because of the usual assumption of independent accident years it remains to evaluate

$$\hat{R}_j = \hat{E}[R_j \mid Y_{1j}, \ldots, Y_{n(j), j}] \text{ for each } j.$$

Modelling the development process $(Y_{ij})_{i \in N}$, different well-known experience rating models can be used.

4.1. *The Model of de Vylder*

DE VYLDER (1982) bases the development process on a special case of the (noncumulative) classical credibility model of Bühlmann–Straub. Therefore the covariance structure is contained in the model of Jewell. As described in Section

3.1 one may gain by linear transformation of the development process a process $(X_{ij})_{i \in N}$ which is l.M. That is, the reserve estimation becomes $\hat{R}_j = \hat{E}[R_j \mid X_{n(j),j}]$.

4.2. *The Model of Norberg*

NORBERG (1985) constructs a micro-model with claim numbers and individual claim amounts similar to the experience rating model in Section 3.4 with condition I. However, the distribution of the individual claim amounts may also depend on the reporting year. The resulting covariance structure of the development process becomes too complicated for calculating the theoretical credibility estimator up to an explicit formula. Therefore Norberg proposes numerical evaluation of the credibility estimator. In Norberg's model the cumulated claim number process is l.M. because of the Poisson assumption. This fact caused the present author to consider credibility estimators for the IBNR-claims in a distribution-free loss-reserving model where the Poisson assumption is replaced by the linear Markov property. This assumption is shown to be natural if the delay distribution does not depend on the hidden risk characteristics of the accident year (WITTING, 1986).

4.3. *The Model of Kramreiter and Straub*

Let us consider for fixed j the process $(X_{ij})_{i \in N}$ of the cumulative burning costs. KRAMREITER and STRAUB (1973) discuss the optimal unbiased homogeneous linear estimator of R_j with given statistical basis $X_{1j}, \ldots, X_{n(j),j}$ in a distribution-free model. "Optimal" means that the expected squared loss is minimized. This estimator exists and is uniquely determined. Kramreiter and Straub write the covariance structure in the form $\mathrm{Cov}(X_{ij}, X_{mj}) = c_{im}/p_j$, whereby p_j is a known volume measure of accident year j.

The most general covariance structure given by Kramreiter and Straub for which explicit calculation of the optimal homogeneous linear estimator remains possible is

$$c_{im} = c_i \prod_{k=i+1}^{m} \lambda_k \qquad \text{for } i \leqslant m,$$

where $(\lambda_i)_{i \in N}$ is a real sequence.

Because of criterion (9) in Lemma 3 this is exactly the linear Markov property of the process $(X_{ij})_{i \in N}$, which appears now as the actual assumption of the Kramreiter–Straub model.

General Remark on the Linear Markov Property

In the present paper we have only treated the case of a stochastic process ordered with respect to time. One may imagine the linear Markov property also with respect to other orders. An example for that is the recent paper of BÜHLMANN and JEWELL (1986), who have used the linear Markov property for recursive calculation of the credibility estimator in a general hierarchical model.

REFERENCES

BÜHLMANN, H. and JEWELL, W. S. (1986) Hierarchical credibility. Preprint for the Summer School 1986 of the Association of Swiss Actuaries.

BÜHLMANN, H. and STRAUB, E. (1970) Glaubwürdigkeit für Schadensätze. *Bulletin of the Association of Swiss Actuaries* **70/1**, 111–133.

DE VYLDER, F. (1982) Estimation of IBNR-claims by credibility theory. *Insurance: Mathematics and Economics* **82/1**, 21–45.

DRYGAS, H. (1983) Sufficiency and completeness in the general Gauss–Markov model. *Sankhya* A **45**, 88–98.

DRYGAS, H. (1985) Linear sufficiency and some applications in multilinear estimation. *Journal of Multivariate Analysis* **16**, 71–84.

FELLER, W. (1966) *An Introduction to Probability and its Applications*, Vol. 2. Wiley, New York.

JEWELL, W. S. (1975) Two classes of covariance matrices giving simple linear forecasts. *IIASA Research Memorandum* **75/17**.

KRAMREITER, H. and STRAUB, E. (1973) On the calculation of IBNR reserves II. *Bulletin of the Association of Swiss Actuaries* **73**, 177–190.

NEUHAUS, W. (1985) Choice of statistics in linear Bayes estimation. *Scandinavian Actuarial Journal* **85/1**, 1–26.

NORBERG, R. (1985) A contribution to modelling of IBNR claims. *Statistical Research Report of the University of Oslo* **85/4**.

PAPOULIS, A. (1965) *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, Tokyo.

SHUR, W. (1972) Discussion of Margolin's "Credibility of group insurance claim experience". *Transactions of the Society of Actuaries* **23**, 240–251.

SUNDT, B. (1979) Credibility and relative exchangeability. *Scandinavian Actuarial Journal* **79**, 45–51.

WITTING, T. (1986) Über Kredibilitätsschätzungen von reinen IBNR-Schäden. Dissertation ETH Nr. 8042.

T. WITTING

ETH Zürich, Mathematikdepartement HG G 36.1, Rämistr. 101, CH-8092 Zürich.