A CONTRIBUTION TO MODELLING OF IBNR CLAIMS(*)

by

Ragnar Norberg
University of Oslo

Abstract

The mechanisms governing occurrence and notification of claims are pictured by a basic stochastic model judged to be fairly realistic in a number of practical situations. IBNR-reserves are composed in a number of different cases obtained by variation of the levels of specificity of model and run-off data. The reserves are obtained by established principles of mathematical statistics and range from empirical Bayes methods, both exact and linear (credibility), to methods based on models that do not include latent random variables. The present work is mainly of a theoretical nature; an empirical follow up study is in preparation.

Key words: IBNR; Model variations; Various prediction bases; Direct and indirect business.

(*) A first draft of this paper was presented at the Second Oberwolfach Seminar in Risk Theory, September 1982.

1. Introduction

1.1. Background and purpose of the present study

A. The problem of establishing provisions for IBNR (Incurred But Not Reported) claims has been a "hot subject" in actuarial circuits for more than a decade now. The literature on this topic has shown a marked trend from rather straightforward methods based on crude models with little structure, often with no stochasticity in them, to models and methods of an ever-increasing degree of sophistication. This pattern of development is hardly peculiar to actuarial research, but is certainly typical of it and reflects the conditions under which it is operating: the actuary is a decision-maker compelled to produce, currently and within narrow deadlines, decisions about premiums, reserves, retentions,... At first he will often have to decide to the best of his intuitive abilities. Then, if the same kind of problem presents itself repeatedly, he will look for some method, that is, a device that automatizes the production of current decisions. And if at some instance there is time left for afterthought, he may try to express his ideas and knowledge of the nature of the problem in a model and search for an optimal method, or at least one that performs well.

B. The present paper advocates the reverse ordering of these activities by demonstrating how the method for IBNR-reserving results from established principles of mathematical statistics when a model has been chosen and the purpose of the reserve has been defined precisely in the terms of this model. Moreover, the model framework presented here is proposed as a reasonable candidate description of the process governing occurrence and notification of claims in a number of classes of insurance business, in particular those subject to fluctuations in collective risk conditions acting on all individual risks simultaneously.

Once the model has been specified, a further circumstance that is decisive of the choice of method is the statistics that can be entered into the prediction. We shall distinguish between direct insurance, where one usually can observe both the number of claims and the individual claim amounts, and reinsurance, where one will typically have access only to certain total claims amounts.

1.2. Outline of the paper and a word of guidance to the reader

A. Section 2 describes the basic model underlying all the special cases treated in the succeding sections. In section 3 a number of different principles of IBNR-reserving are proposed. In sections 4 through 11 IBNR-predictors are constructed by various specifications of the model and the statistical data. Section 12 offers a survey of a selection of previous IBNR-studies related to the present one. In the final section 13 some lines of further development of the theory are indicated. For ease of reference, some selected results - mostly well known matters from risk theory - have been placed in an appendix.

B. As the scope of the present study is fairly broad, the presentation is organized in a manner that allows for a selective reading. The primary purpose is to present an assembly of methods for establishing IBNR-reserves. However, as the subject offers an exellent opportunity to discuss some general problems of modelling, a number of paragraphs and items have been included that are mainly of an educative nature. Such parts of the text are marked by an asterisk, and so are those parts concerned with pure technicalities or theoretical elaboration beyond what is required for an understanding of the principal message. Thus, earthbound readers seeking a quick way to results should simply avoid the stars.

2. Definitions and basic model assumptions

2.1. Notational conventions

A. Scalars are denoted by ordinary italics. Matrices and vectors are written in boldface. When speaking of a vector $\mathbf{x}$, we shall always have a column vector in mind. Row vectors are marked by a prime signifying transposition, e.g. $\mathbf{x}'$. By $\mathrm{diag}(\alpha_i)_{i=1,\ldots,m}$ is meant the m×m matrix with the indicated elements on the principal diagonal and zeros elsewhere.

B. Let $\mathbf{x} = (x_r, x_{r+1}, \ldots, x_s)'$ be a vector with entries numbered consecutively from r to s (a segment of the integers). The vector consisting of the t-r+1 first elements of $\mathbf{x}$ is written

$$\mathbf{x}_{\leq t} = (x_r, \ldots, x_t)',$$

and the sum of these elements is denoted by

$$x_{\leq t} = \sum_{j=r}^{t} x_j .$$

Analogously we also write $\mathbf{x}_{>t} = (x_{t+1}, \ldots, x_s)'$ and $x_{>t} = \sum_{j=t+1}^{s} x_t$.

C. Let $\mathbf{X}$ and $\mathbf{Y}$ be random vectors of dimension m and n, respectively. We denote by $\mathrm{Cov}(\mathbf{X}, \mathbf{Y}')$ the m×n matrix which has $\mathrm{Cov}(X_i, Y_j)$ as its (i,j)-entry. In particular we write $\mathrm{Var}\,\mathbf{X} = \mathrm{Cov}(\mathbf{X}, \mathbf{X}')$.

Let X and Y be random elements, X scalar-valued.

Whenever it exists, the conditional h'th central moment of X, given Y, is denoted by $M_h(X|Y)$, that is,

$$M_1(X|Y) = E(X|Y),$$
$$M_h(X|Y) = E[\{X-E(X|Y)\}^h|Y];\ h=2,3,... \tag{2.1}$$

2.2. The structure of the data

A. The Lexis type of diagram shown in figure 1 is a handy tool for visualization of data on occurrence and notification of insurance claims. Calendar time is measured along the horizontal axis, and development time (the time elapsing between occurrence and notification of a claim) is measured along the vertical axis. Thus a claim occurred at time s and reported at time t is represented by a diagonal line connecting the points (s,0) and (t,t-s). The "cohort" of claims occurred in year j can be traced along the band limited by the diagonals originating in the

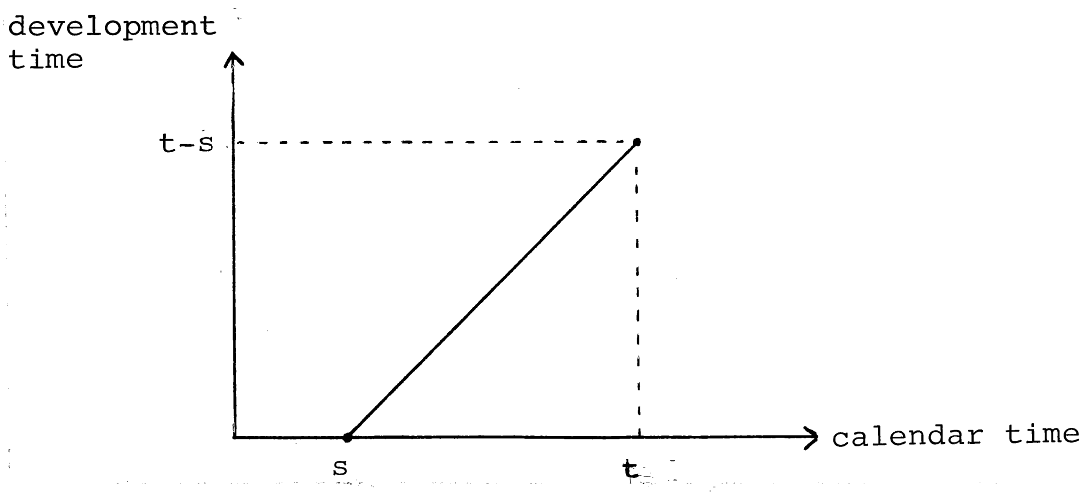

Figure 1. Lexis diagram with representation of a claim occurred at time s and reported at time t.

points (j-1,0) and (j,0), see figure 2. Quantities relating to the d-th year of development of that cohort are marked off in the parallelogram with corners (j-1+d,d-1), (j+d,d), (j+d,d+1), and (j-1+d,d). (The choice of the year as time unit is merely a matter of terminology. For long-tailed business, like marine, product liability, and accident, where claims may be reported several years after their occurrence, one year may be a suitable time unit. For short-tailed business a quarter of a year or a month may be more appropriate. Another piece of terminological convenience: when speaking of occurrence of a claim, we really mean occurrence of the event that gives rise to the claim.)

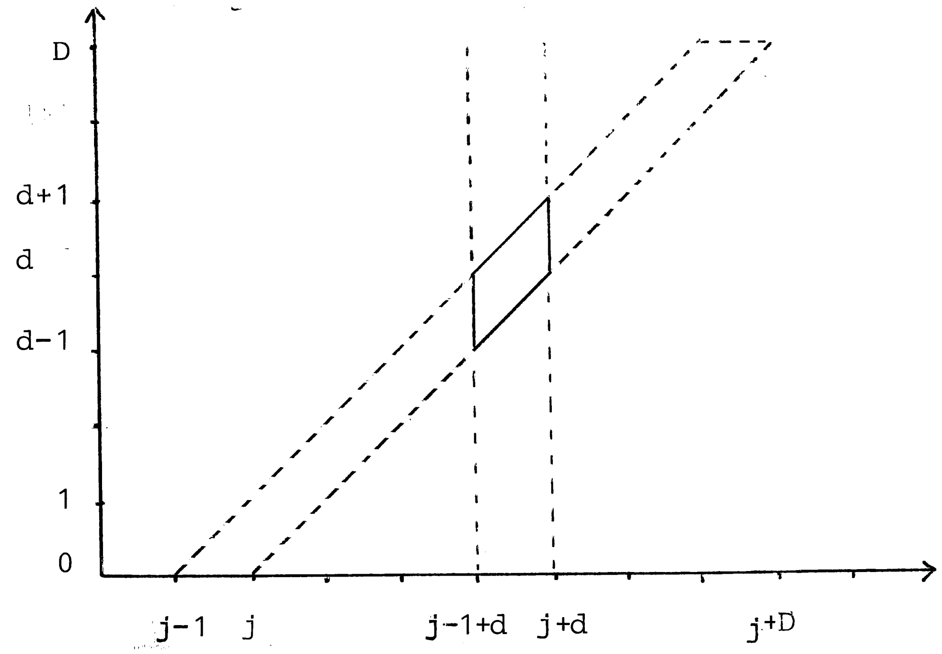

Figure 2. Lexis diagram with the parallelogram representing claims occurred in year j and reported in year j+d.

B. We consider one class of business and introduce the following quantities relating to the parallelogram in figure 2:

K_{jd}, the number of claims that occur in year j and are reported d years later, in year $j+d$,

Y_{jdk}, the size of the k-th of those claims that occur in year j and reported in year $j+d$; $k=1,2,\ldots$; (2.2)

and, defining $Y_{jd0} = 0$,

$S_{jd} = \sum_{k=0}^{K_{jd}} Y_{jdk}$, the total amount paid in respect of claims that occur in year j and are reported in year $j+d$. (2.3)

The domains of the indices are

$j=1,2,\ldots$ and $d=0,1,\ldots,D$,

respectively, D being the maximum time that can elapse between occurrence and notification of a claim.

2.3. Basic model assumptions

A. We make the following assumptions about the stochastic mechanism that generates the quantities defined above.

With each year j is associated a positive quantity p_j measuring the amount of risk exposed, e.g. the number of risks or risk years, the total mileage (in motor insurance), the premium of the direct business (in reinsurance), or some other appropriate measure of the volume of the business transacted in year j. The p_j's are observable and are viewed as nonrandom.

We also attach to each year j a pair of quantities $\Xi_j = (T_j, \Psi_j)$ representing the latent general risk conditions in that year, where T_j (capital Greek τ) acts upon the number of claims and Ψ_j acts upon the single claim amounts. (The decomposition of Ξ_j into two components is just a matter of notational convenience: it implies no assumptional restriction as long as nothing has been said about the relation between the two components. The quantities Ψ_j may be scalar- or vector-valued or even more general.) In keeping with the standard way of modelling fluctuating basic probabilities (see e.g. Beard et. al. (1969)), the Ξ_j's are conceived as unobservable random elements, and it is assumed that

I. $\Xi_1, \Xi_2, \ldots$ are i.i.d. $\sim U$

(independent and identically distributed in accordance with some distribution function U).

As our conditional model for fixed Ξ_j we adopt an extended version of the traditional generalized Poisson law. More specifically, we assume that T_j is a positive quantity, and that conditional on $(T_j, \Psi_j) = (\tau_j, \psi_j)$, the total number of claims occurring in year j is Poisson distributed with parameter $p_j \tau_j$. By way of example, one may interpret τ_j as the integral over the time interval $[j-1, j)$ of a basic claim intensity acting on each of the p_j risk units throughout that interval. About the notifications we make the simplifying assumption that single claims are reported independently of one another, each with a probability π_d of being reported d years after its occurrence. From these assumptions we gather:

II. Conditionally, given $(T_j,\Psi_j) = (\tau_j,\phi_j)$, the K_{jd}'s are mutually independent, and

$$K_{jd} \sim Po(\pi_d p_j \tau_j) \quad ; \; d=0,\ldots,D.$$

Here $Po(\kappa)$ denotes the Poisson distribution with parameter κ.

Next we make assumptions about the claim amounts.

III. Conditionally, given $(T_j,\Psi_j) = (\tau_j,\phi_j)$, the amounts Y_{jdk} are mutually independent and independent of the claim numbers K_{jd}, and

$$Y_{jdk} \; ; \; k=1,2,\ldots; \text{ are i.i.d.} \sim G_d(\cdot|\phi_j); \; d=0,\ldots,D.$$

By fixed $\underset{\sim}{\pi}$, U and $G_0,\ldots,G_D$ the following assumption completes the specification of the joint distribution of the introduced random variables.

IV. Quantities referring to different years of occurrence are stochastically indpendent.

B. In practice the distributions $\underset{\sim}{\pi}$, and $G_0,\ldots,G_D$ are not known at the outset. Consequently, all parameters required in predictions of IBNR-outstandings have to be estimated from data. For this purpose we have to specify the sets of distributions that are possible a priori:

V. $\underset{\sim}{\pi} \in \underline{\Pi}$, $U \in \underline{U}$, $G_d \in \underline{G}_d$; $d=0,\ldots,D$.

The basic probability model I-IV together with the specifications in V constitute our statistical model.

3. General formulations of the reserving problem

3.1. IBNR-triangle, prediction basis, and statistical basis

A. Referring to figure 3, suppose we are presently at time J and are to forecast the contents of the IBNR-triangle. In particular we want to predict the total amount of IBNR claims,

$$R = \sum_{j=J-D+1}^{J} S_{j,>} \tag{3.1}$$

where

$$S_{j,>} = S_{j,>J-j} = \sum_{d>J-j} S_{jd} \tag{3.2}$$

is the amount of IBNR claims occurred in year j; $J-D+1 \leq j \leq J$.

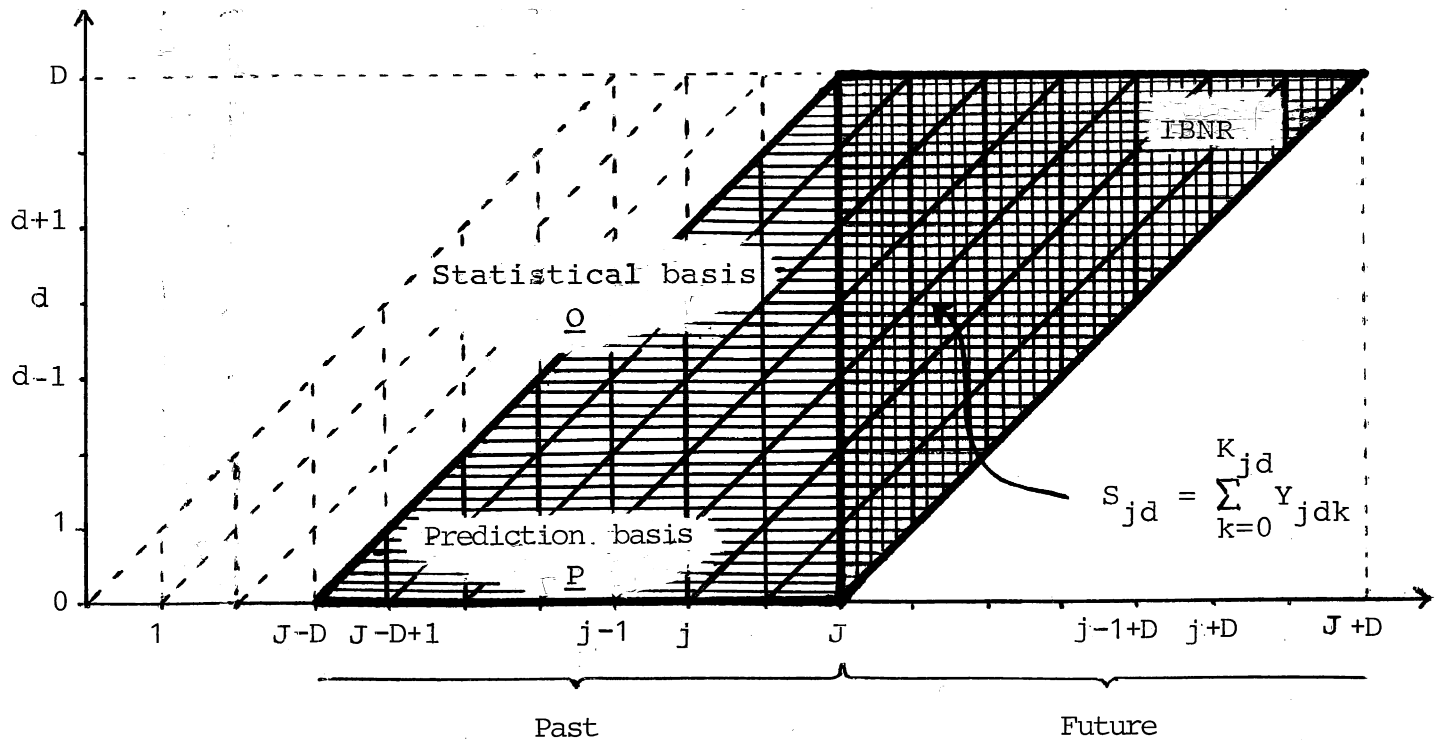

Figure 3. The IBNR-triangle (cross-hatching), the prediction basis (simple hatching), and the statistical basis (simple or no hatching)

Denote by $\underline{O}_j$ the data available by time J in respect of claims occurred in year j; $j=1,\ldots,J$. The statistical basis $\underline{O} = (\underline{O}_1,\ldots,\underline{O}_J)$ is made up of all observations available by time J. A special role in $\underline{O}$ is played by the (direct) prediction basis (or run-off triangle) $\underline{P} = (\underline{O}_{J-D+1},\ldots,\underline{O}_J)$, which consists of the statistical information from the not yet fully developed years.

The definition (3.2) illustrates a short-hand that will be used extensively in the following: when applying the notational device introduced in item 2.1.B to quantities like $S_{j,>J-j}$, $\underset{\sim}{K}_{j,\leq J-j}$, etc., we shall as a rule drop the obvious $J-j$ and simply write $S_{j,>}$, $\underset{\sim}{K}_{j,\leq}$, etc.

B. Taking items I-IV in the model as a basic framework, there are two circumstances that are decisive of the designation of the IBNR-reserve. In the first place it is the specificity of the statistical basis, that is, the kind of data contained in $\underline{O}$; in direct insurance one will typically have access to the basic quantities K_{jd} and Y_{jdk}, whereas in reinsurance one will often observe only the total amounts S_{jd} or possibly some even more summary statistics. In the second place it is the specificity of the model, that is, the extent to which the sets in V are specified by parametrization, assumptions of independence, etc.

3.2. Outline of sections 4-11

We are going to investigate a number of special cases, each of which will be treated in accordance with the following disposition.

1. Description of the case. The statistical basis $\underline{O}$ and the model elements $\underline{U}$ and $\underline{G}_d$; d=0,...,D; are specified. (It is assumed throughout that $\underset{\sim}{\Pi} = \{\underset{\sim}{\pi};\ \pi_d>0$ for all d and $\sum_{d=0}^{D}\pi_d = 1\}$.)

2. Prediction by known parameters. If the estimable parameters were known, we would select a predictor in the class of all functions depending on these parameters and on the direct prediction basis $\underline{P}$. (By the independence assumption IV, $\underline{P}$ would then contain all relevant statistical information.) For a given $\underline{P}$ it is the set of available (i.e. estimable) parameters that constrains the choice of predictor.

Consider first the case where the joint distribution of $\underline{P}$ and R is fully known, so that a full posterior analysis can be accomplished. A commonly used measure of the performance of a predictor $\check{R}$ is the expected squared error,

$$E(\check{R}-R)^2. \tag{3.3}$$

(We do not care to indicate explicitly that the expectation depends on $\underset{\sim}{\pi}$, U, $G_0,\ldots,G_D$). We introduce the conditional central moments (recall the principle of notation in (2.1))

$$M_{hj} = M_h(S_{j,>}|\underline{O}_j) \quad ;\ h=1,2,3. \tag{3.4}$$

The optimal predictor in terms of (3.3) is

$$\tilde{R} = E(R|\underline{O}) = \sum_{j=J-D+1}^{J} M_{1j}\ , \tag{3.5}$$

the second equality being a consequence of the independence assumption IV.

It may be argued that criterion (3.3) does not express perfectly the object of claims reserving since it implies that understating liabilities by a certain amount is equally undesirable as overstating them by the same amount. In fact, overreserving seems to be preferred by most claims departments and is certainly preferred by regulatory authorities, whose main concern is the adequacy of reserves to meet liabilities. An IBNR-reserve reflecting a cautious attitude is obtained by adding to the conditional expected value in (3.5) a safety loading depending on the conditional variance of R, given $\underline{O}$. By virtue of assumption IV, the general form of this reserve is

$$\tilde{R} = \sum_{j=J-D+1}^{J} M_{1j} + f\left(\sum_{j=J-D+1}^{J} M_{2j}\right), \tag{3.6}$$

where the M_{hj} are the conditional central moments defined in (3.4) and f is the square root or some other non-negative and non-decreasing function.

Another prudent principle, which has an obvious justification, consists in providing a reserve $\tilde{R}$ equal to the $(1-\varepsilon)$-fractile of the predictive distribution, that is,

$$P(R \leq \tilde{R} | \underline{P}) = 1-\varepsilon. \tag{3.7}$$

If calculation of the fractile in (3.7) is laborious, one could use some approximation method that employs only the first three moments of the distribution. One such method, which is very handy, is the so-called NP-approximation described in Beard et. al. (1984). It states that the $(1-\varepsilon)$-fractile of a distribution can be approximated by

$$\mu_1 + c_1\mu_2^{\frac{1}{2}} + c_2\mu_3/\mu_2 \,, \tag{3.8}$$

where μ_h is the h'th central moment of the distribution; $h = 1,2,3$; c_1 is the upper ε-fractile of the standard normal distribution, and $c_2 = (c_1^2-1)/6$. Again by virtue of assumption IV, the reserve delivered by this principle is

$$\tilde{R} = \sum_j M_{1j} + c_1(\sum_j M_{2j})^{\frac{1}{2}} + c_2\sum_j M_{3j}/\sum_j M_{2j} \,, \tag{3.9}$$

where the M_{hj}'s are given by (3.4) and all summations range over $j=J-D+1,\ldots,J$.

Next consider the case where the joint distribution of $\underline{P}$ and R is not fully known (or, rather, is not estimable from $\underline{O}$). Then the reserves defined by (3.5)-(3.7) and (3.9) typically depend on unknown parameters and are, therefore, not feasible. If, however, we know certain unconditional moments up to second order, we can instead of (3.5) use a credibility predictor $\tilde{R}$, which, roughly speaking, minimizes (3.3) as $\check{R}$ ranges in the class of all inhomogeneous linear functions of certain statistics depending on $\underline{P}$. By (A.18) in appendix A.3, the general form of $\tilde{R}$ is

$$\tilde{R} = \sum_{j=J-D+1}^{J} \tilde{S}_{j,>} \,, \tag{3.10}$$

where $\tilde{S}_{j,>}$ is some credibility predictor of $S_{j,>}$ based on $\underline{O}_j$.

By adding to (3.10) a security loading depending on the unconditional variance of R, we obtain a reserve of the form

$$\tilde{R} = \sum_{j=J-D+1}^{J} \tilde{S}_{j,>} + f(\sum_{j=J-D+1}^{J} \operatorname{Var} S_{j,>}). \tag{3.11}$$

Sometimes it is possible by credibility methods to approximate M_{2j} in (3.4) by a function $\tilde{M}_{2j}$ that depends on certain higher order unconditional moments. Then, if these moments are known, we can construct the following credibility analogue to (3.6):

$$\tilde{R} = \sum_{j=J-D+1}^{J} \tilde{S}_{j,>} + f\Big(\sum_{j=J-D+1}^{J} \tilde{M}_{2j}\Big). \tag{3.12}$$

(If the argument of $f(\cdot)$ becomes negative, we replace it by 0.)

If, furthermore, a credibility approximation $\tilde{M}_{3j}$ of M_{3j} can be arranged, then a "credibility approximated NP-approximation" is obtained by instead of (3.9) using

$$\tilde{R} = \sum_j \tilde{S}_{j,>} + c_1\Big(\sum_j \tilde{M}_{2j}\Big)^{\frac{1}{2}} + c_2 \sum_j \tilde{M}_{3j} \Big/ \sum_j \tilde{M}_{2j} \ . \tag{3.13}$$

3. Parameter estimation. An estimation procedure is briefly indicated. Parameter estimation problems will not be focussed at in this paper.

Upon replacing the parameters appearing in any one of the reserves in (3.5)-(3.7), (3.9)-(3.13) by their estimators, we finally obtain a genuine reserve $\tilde{R}^*$, which normally will be asymptotically equivalent to $\tilde{R}$ in the sense that $\tilde{R}^*/\tilde{R}$ tends to 1 in probability as J increases. Often we shall not care to mention this final step explicitly in special cases since that would amount to little more than merely repeating the phrases above. Exceptions are made only in those cases where an explicit and appealing formula of $\tilde{R}^*$ is obtained.

4. Comments. Notable features of the situation are briefly pointed out.

4. Prediction based on numbers of claims and single claims amounts when varying latent risk conditions are not modelled as random variables; a preparatory study

4.1. Description of the case

A. Let the available observations be

$$\underline{O} = \{K_{jd}, Y_{jdk};\ k=0,\ldots,K_{jd};\ d=0,\ldots,D(j);\ j=1,\ldots,J\}, \qquad (4.1)$$

where we have introduced

$$D(j) = \min(D, J-j).$$

Thus we have access to the complete history of the individual claims as recorded by the direct insurance business.

As all quantities in (4.1) are assumed known by time J, we have to accomodate definition (2.2) to claims that are reported, but not settled at that time. For these we must in practice let Y_{jdk} be the sum of the payments made up to time J and the provision made at time J to meet payments that will fall due in the future.

B. In this first case to be studied we apparently step aside from our basic model framework by leaving out assumption I in paragraph 2.3. Instead the latent risk conditions are represented by nonrandom parameters

$$\xi_j = (\tau_j, \phi_j) \qquad ;\ j=1,2,\ldots$$

Assumptions II-V are retained as before, with the modification that we drop the conditioning clause in II and III and replace the specification $U \in \underline{U}$ in V by $\tau_j > 0$ and $\phi_j \in \underline{\Psi}$; j=1,2,... (Speaking of the ϕ_j's as "parameters" does not necessarily imply that the families $\underline{G}_d$ are parametric in the sense that the set $\underline{\Psi}$ is of finite dimension.)

C. We shall examine this reduced model in some detail for several reasons. In the first place, some may prefer the point of view taken here, that the ξ_j's are non-random, and to those the results in this section present an interest of their own. (Very plausibly they will, however, change their opinion after having read the comments in paragraph 4.4.) In the second place, comparison of the results obtained here to those obtained in the full model gives rise to a number of instructive comments. In the third place, the calculations made here are needed in some of the following sections.

4.2. Prediction by known parameters

A. The present model specifies no stochastic dependence between past and future. Consequently, prediction by known parameters reduces to calculations in the marginal distribution of R. Hence we set out to determine this distribution.

$\underline{B}^*$. We pause here to supply a motivation of assumption II in paragraph 2.3. As is standard in risk theory, it is assumed that the total number of claims occurred in year j, $K_{j,\leq D}$, is distributed in accordance with $Po(p_j\tau_j)$. Combining this with the assumptions about the claims reporting described just before assumption II, we obtain for any $k_{jd} = 0,1,\dots$ and $k_j = \sum_{d=0}^{D} k_{jd}$ that

$$P(\bigcap_{d=0}^{D} K_{jd} = k_{jd}) = P(\bigcap_{d=0}^{D} K_{jd} = k_{jd} | K_{j,\leq D} = k_j) P(K_{j,\leq D} = k_j)$$

$$= \left(\frac{k_j!}{\prod_{d=0}^{D} k_{jd}!} \prod_{d=0}^{D} \pi_d^{k_{jd}}\right) \frac{(p_j \tau_j)^{k_j}}{k_j!} e^{-p_j \tau_j}$$

$$= \prod_{d=0}^{D} \left\{\frac{(\pi_d p_j \tau_j)^{k_{jd}}}{k_{jd}!} e^{-\pi_d p_j \tau_j}\right\},$$

which is just assumption II in the conditional model.

C. the marginal distribution of S_{jd} defined by (2.3) is generalized Poisson with frequency parameter $\pi_d p_j \tau_j$ and claim size distribution $G_d(\cdot|\psi_j)$. In short-hand we write

$$S_{jd} \sim g.Po(\pi_d p_j \tau_j, G_d(\cdot|\psi_j)).$$

For $S_{j,>}$ defined by (3.2) we have, by the result (A.15) in appendix A.2, that

$$S_{j,>} \sim g.Po(\pi_{>J-j} p_j \tau_j \;, G_{>J-j}(\cdot|\psi_j)),$$

with

$$G_{>J-j}(\cdot|\psi_j) = \pi_{>J-j}^{-1} \sum_{d>J-j} \pi_d G_d(\cdot|\psi_j).$$

The expression for the cumulative distribution function is

$$P(S_{j,>} \leq x) = \sum_{k=0}^{\infty} \frac{(\pi_{>J-j} p_j \tau_j)^k}{k!} e^{-\pi_{>J-j} p_j \tau_j} G_{>J-j}^{k*}(x|\psi_j), \qquad (4.2)$$

where "k*" designates k-th convolution. By (A.7), (A.8), (A.12), and (A.13) in appendix A.2, the first three central moments of $S_{j,>}$ are

$$\mu_{hj} = \pi_{>J-j} p_j \tau_j \int y^h dG_{>J-j}(y|\phi_j)$$
$$= p_j \tau_j \sum_{d>J-j} \pi_d \int y^h dG_d(y|\phi_j) \quad ; \; h=1,2,3. \tag{4.3}$$

Likewise we obtain for the total amount of IBNR claims in (3.1) that

$R \sim g.Po(\kappa, H)$,

with

$$\kappa = \sum_{j=J-D+1}^{J} \pi_{>J-j} p_j \tau_j \tag{4.4}$$

and

$$H(\cdot) = \kappa^{-1} \sum_{j=J-D+1}^{J} \pi_{>J-j} p_j \tau_j G_{>J-j}(\cdot|\phi_j). \tag{4.5}$$

The first three moments of R are obtained by summation of the moments in (4.3);

$$\textstyle\sum_{j=J-D+1}^{J} \mu_{hj} \quad ; \; h=1,2,3.$$

The cumulative distribution function of R is

$$P(R \leq r) = e^{-\kappa} \sum_{k=0}^{\infty} \frac{\kappa^k}{k!} H^{k*}(r). \tag{4.6}$$

D. We have now determined all elements required in the different IBNR-reserves defined in section 3.

In the present model the conditioning with respect to $\underline{P}$ drops out, and the reserve (3.5) reduces to

$$\tilde{R} = \sum_{j=J-D+1}^{J} \mu_{1j} , \tag{4.7}$$

the μ_{1j} being defined by (4.3).

The reserve (3.6) becomes

$$\tilde{R} = \sum_{j=J-D+1}^{J} \mu_{1j} + f(\sum_{j=J-D+1}^{J} \mu_{2j}).$$

Application of principle (3.7) requires numerical calculation of the tail of the distribution function (4.6). A uniform ε-approximation of this function is obtained by including the n first terms in the sum on the right hand side of (4.6), where n is the smallest integer satisfying $e^{-\kappa}\sum_{k=0}^{n}\kappa^k/k! \geq 1-\varepsilon$. If κ is large so that a large number of terms is required, then the recursive procedure proposed by Panjer (1981) may reduce the computational work substantially. Alternatively one could use the NP-approximation (3.9) with the M_{hj}'s replaced by the unconditional moments in (4.3).

4.3. Parameter estimation

A. Estimation of the parameters $\underset{\sim}{\pi}$ and (τ_j,ϕ_j); $j = 1,2,\ldots,J$; is based on the joint distribution of the observations in (4.1), which is given by

$$P(K_{jd} = k_{jd}\ ,\ y_{jdk}<Y_{jdk}\leq y_{jdk}+dy_{jdk}\ ;\ k=1,\ldots,k_{jd}\ ;$$
$$d=0,\ldots,D(j);\ j=1,\ldots,J)$$

$$= \prod_{j=1}^{J}\prod_{d=0}^{D(j)}\{\frac{(\pi_d p_j \tau_j)^{k_{jd}}}{k_{jd}!}\, e^{-\pi_d p_j \tau_j}\prod_{k=1}^{k_{jd}} dG_d(y_{jdk}|\phi_j)\} \tag{4.8}$$

$$= \left(\prod_{j=1}^{J} p_j^{k_{j,\leq D(j)}} \prod_{d=0}^{D(j)} \frac{1}{k_{jd}!}\right)\left(\prod_{d=0}^{D} \pi_d^{k_{\leq J-d,d}}\right)\left(\prod_{j=1}^{J} \tau_j^{k_{j,\leq D(j)}}\right) \tag{4.9}$$

$$e^{-\sum_{j=1}^{J}\pi_{\leq D(j)}p_j\tau_j} \prod_{j=1}^{J} \prod_{d=0}^{D(j)} \prod_{k=1}^{k_{jd}} dG(y_{jdk}|\phi_j) \ ;$$

$y_{jdk}\geq 0$; $k=1,\ldots,k_{jd}$; $k_{jd}=0,1,\ldots$;

$d=0,\ldots,D(j)$; $j=1,2,\ldots$

It is obvious how to interpret the above statements and expressions when $k_{jd} = 0$. (Here and elsewhere the dependence of P, E, Var, etc. on the parameters is suppressed in the notation.)

Inspection of the likelihood given by (4.8) tells us that inference about the π_d's and τ_j's should be based on the marginal distribution of the K_{jd}'s, whereas for each j inference about ϕ_j should be based on the conditional distribution of the Y_{jdk}'s, given $\underset{\sim}{K}_{j,\leq D(j)}$ (which is "ancillary" for ϕ_j).

<u>B</u>. The parameters $\underset{\sim}{\pi}$ and $\tau_1,\ldots,\tau_J$ can be estimated by maximizing the logarithm of the likelihood of the $\underset{\sim}{K}_{j,\leq D(j)}$'s subject to the constraint $\sum_{d=1}^{D}\pi_d = 1$. From (4.9) it is seen that the essential part of the Lagrange function for this problem is

$$L = \sum_{d=0}^{D} K_{\leq J-d,d}\log \pi_d + \sum_{j=1}^{J} K_{j,\leq D(j)}\log \tau_j - \sum_{j,d;\, j+d\leq J} \pi_d p_j \tau_j - c\sum_{d=0}^{D}\pi_d \ , \tag{4.10}$$

where c is the Lagrange multiplier. (It is assumed that $D\leq J$ since otherwise the parameters are not identifiable.) The maximum likelihood estimators $\underset{\sim}{\pi}^*$, $\tau_1^*,\ldots,\tau_J^*$ are the solution of the following equations, where (4.11) and (4.12) result from equating to zero the derivatives of L with respect to the τ_j's and the

π_d's, respectively, and (4.13) is the side condition:

$$K_{j,\leq D(j)} = p_j \tau_j^* \pi_{\leq D(j)}^* \qquad ; \ j=1,\ldots,J; \qquad (4.11)$$

$$K_{\leq J-d,d} = \Big(\sum_{j=1}^{J-d} p_j \tau_j^* - c\Big)\pi_d^* \qquad ; \ d=0,\ldots,D; \qquad (4.12)$$

$$\sum_{d=0}^{D} \pi_d^* = 1. \qquad (4.13)$$

These equations possess no explicit solution and have to be solved by numerical methods.

C*. If D is small compared to $J-D$, then the following simple procedure will be nearly as efficient as the full maximum likelihood procedure described above. First find the maximum likelihood estimator of $\underset{\sim}{\pi}$ and $\tau_1,\ldots,\tau_{J-D}$ based on the numbers of claims $\underset{\sim}{K}_{j;\leq D}$ for the fully developed years $j = 1,\ldots,J-D$. Instead of (4.11)-(4.13) we then get the equations

$$K_{j;\leq D} = p_j \tau_j^* \qquad ; \ j=1,\ldots,J-D;$$

$$K_{\leq J-D,d} = \Big(\sum_{j=1}^{J-D} p_j \tau_j^* - c\Big)\pi_d^* \qquad ; \ d=0,\ldots,D;$$

$$\sum_{d=0}^{D} \pi_d^* = 1,$$

which in case $K_{\leq J-D,\leq D} > 0$ possess the explicit and intuitively appealing solution

$$\tau_j^* = K_{j,\leq D}/p_j \qquad ; \ j = 1,\ldots,J-D; \qquad (4.14)$$

$$\pi_d^* = K_{\leq J-D,d}/K_{\leq J-D,\leq D} \qquad ; \ d = 0,\ldots,D; \qquad (4.15)$$

and $c = 0$. Next estimate $\tau_{J-D+1},\ldots,\tau_J$ by maximizing the likelihood of $\underset{\sim}{K}_{j,\leq J-j}$ for each of the not fully developed years $j=J-D+1,\ldots,J$ under the assumption that the π_d's are known, and finally insert the π_d^*'s from (4.15). The resulting estimators are

$$\tau_j^* = K_{j,\leq J-j}/\pi_{\leq J-j}^* p_j \qquad ; \; j=J-D+1,\ldots,J. \qquad (4.16)$$

The estimator $\underset{\sim}{\pi}^*$ defined by (4.15) is consistent as $\sum_{j=1}^{J-D} p_j \to \infty$. Consistency of the individual τ_j^*'s would require in addition that $p_j \to \infty$ for each j.

D. We now turn to the problem of estimating the claim size parameters ϕ_j. Each particular specification of the families $\underline{G}_d$; $d=0,\ldots,D$; (or, equivalently, of $\underline{\Psi}$) would require an analysis of its own. Usually estimators ϕ_j^* can be obtained by standard methods, hence our further remarks shall be held in general terms.

E*. The families $\underline{G}_d$ may be either parametric ($\underline{\Psi}$ finite-dimensional) or non-parametric ($\underline{\Psi}$ of infinite dimension). In any case the $G_d(\cdot|\phi_j)$'s of past book years, $j+d\leq J$, can always be estimated from $Y_{jd1},\ldots,Y_{jdK_{jd}}$ by standard methods for samples of i.i.d. observations when $K_{jd}>0$. This is, however, of little interest in the present context since our concern is to predict the future. The model has to be structured in such a manner that the future $G_d(\cdot|\phi_j)$'s; $d>J-j$; can be estimated from the observed Y_{jdk}'s; $d\leq J-j$; for each $j=J-D+1,\ldots,J$. This means, roughly speaking, that ϕ_j has to be identifiable from $G_0(\cdot|\phi_j),\ldots,G_{J-j}(\cdot|\phi_j)$ for each j, which is usually the case in parametric situations. An alternative way of making future claim size distributions estimable from past observations is treated under the next item.

F. Consider the special case where the risk conditions governing the claim sizes are invariable over time, that is, all ϕ_j have the same value ϕ. Then each $G_d(\cdot|\phi)$ can be estimated from all the Y_{jdk}'s from the years $j=1,\ldots,J-d$. A distribution-free estimator of $G_d(\cdot|\phi)$ is the empirical distribution function G_d^* based on all the Y_{jdk}'s; $k=1,\ldots,K_{jd}$;$K_{jd}>0$; $j=1,\ldots,J-d$. We have, with a selfexplaining notation,

$$\int y dG_d^*(y) = S_{\leq J-d,d}/K_{\leq J-d,d} \,. \tag{4.17}$$

The assumption that all the ϕ_j's are equal may seem unsuitable in the absence of a similar assumption about the τ_j's. Nevertheless it is often adopted in theoretical studies of the case with no delays $(D = 0)$, and we shall work with it in some of the sections below.

G. Genuine predictions are obtained upon replacing the parameters appearing in the formulas of paragraph 4.2 by their estimators. In general no closed formula in terms of past observations can be arranged when the unrestricted maximum likelihood estimators given by (4.11)-(4.13) are used.

If we instead employ the simple estimators (4.14) and (4.15) together with some estimators $\phi_{J-D+1}^*,\ldots,\phi_J^*$, we obtain the following expression for the estimated expected value predictor in (4.7):

$$\tilde{R}^* = \sum_{j=J-D+1}^{J} K_{j,\leq J-j} \frac{1}{K_{\leq J-D,\leq J-j}} \sum_{d>J-j} K_{\leq J-D,d} \int y dG_d(y|\phi_j^*); \tag{4.18}$$

h=1,2,3.

$\underline{H}^*$. Let us look a little closer into the special case discussed under item F above, where the single claim amounts are not influenced by varying risk conditions. Inserting (4.17) into (4.18), yields the easily interpretable formula

$$\tilde{R}^* = \sum_{j=J-D+1}^{J} K_{j,\leq J-j} \sum_{d>J-j} \frac{K_{\leq J-D,d}}{K_{\leq J-D,\leq J-j}} \frac{S_{\leq J-d,d}}{K_{\leq J-d,d}} . \qquad (4.19)$$

4.4. Comments*

$\underline{A}^*$. First we add one further remark on the model specified in 4.1.B. Informally, one might say that modelling $\xi_1, \xi_2, \ldots$ as nonrandom parameters is consistent with assumption I in the basic model with U "diffuse" or "non-informative".

Another point of view is that we operate in the full model, but confine ourselves to methods that rest entirely on the information contained in the conditional distribution for given Ξ_j and thus do not utilize the fact that the Ξ_j are i.i.d. random elements. The resulting methods remain perfectly meaningful also in the full model, but they are not optimal. Roughly speaking, their performance is poorer the more informative U is.

$\underline{B}^*$. As remarked already in paragraph 4.2, past observations are of no use in prediction of the future in the present model when the parameters are considered as known. They come into play only in paragraph 4.3, where they are used to estimate the parameters; it is the structure imposed on the parameters that now bridges past and future and enables us to predict the latter.

$\underline{C}^*$. A remarkable feature of the present model is that the volumes p_j essentially drop out of the analysis; they may be absorbed in the τ_j's as these range in all of $\underline{R}_+$. This fact is reflected also by the absence of the p_j's in the predictions (4.18) and (4.19). We conclude that if different years are not made comparable through the introduction of assumption I or some other way, then information about the amounts of risk exposure will be of no value. It may be felt that the irrelevance of measuring the size of the business is a shortcoming of the present model.

$\underline{D}^*$. In item 4.3.E it was mentioned that predictions are possible only if ϕ_j can be estimated from past observations at each stage of development of year j. A similar remark applies also to the parameters governing the numbers of claims. We have assumed that the probability distribution $\underset{\sim}{\pi}$ of the delay period is the same for all occurrence years. If we had not made this assumption and instead introduced a $\underset{\sim}{\pi}_j$ for each year j, we should be unable to predict the number of IBNR-claims. This is seen upon replacing π_d in (4.8) by π_{jd} : then only the frequency parameters $\pi_{jd}\tau_j$; $j+d\leq J$; are identifiable from the distribution of the past observations, and nothing could be inferred as to future K_{jd}'s; $j+d>J$.

When a new parameter ξ_j is introduced for each year j, each ξ_j has to be estimated from the claims data of year j alone. The accuracy of the estimators may be poor if the risk exposure is not great, especially at early stages of development. From the log likelihood (4.10) we easily obtain the asymptotic variances of the estimators defined by (4.11)-(4.13):

$$\text{as.Var } \tau_j^\star = \tau_j/\pi_{\leqslant D(j)}p_j \,, \tag{4.20}$$

$$\text{as.Var } \pi_d^\star = \pi_d/\sum_{j=1}^{J-d} p_j\tau_j \;.$$

As could be expected, $\underset{\sim}{\pi}^\star$ is consistent by increasing J, roughly speaking, whereas $\tau_j^\star$ is consistent only by increasing exposure in year j. This is a prise we have to pay for not being willing to specify any kind of connection between the risk conditions in different years.

$\underline{E}^\star$. The necessity of establishing some such connection appears even more clear when we face the problem of tariffication. In fact, the present model renders no possibility of fixing the premium level for a future year by statistical methods.

$\underline{F}^\star$. The circumstances mentioned in items B-E are inevitable consequences of our model assumptions. To the extent that they are incompatible with our intuition and conceptions about the nature of the IBNR-phenomenon, they point to deficiencies of the present reduced model. These will be overcome when we turn to the full model by including the i.i.d.-assumption I, which establishes a relation between the risk conditions in different years. But first we shall see in section 5 how some of the problems can be remedied within the present fixed-parameter-approach by introducing more assumptions, viz. that basic risk conditions remain unchanged from one year to another.

5. Prediction based on numbers of claims and single claim amounts by permanent risk conditions

5.1. Description of the case

A. The data $\underline{O}$ is the same as in the previous section.

B. The model in item 4.1.B is retained, but we now introduce the additional assumption that the risk conditions are invariable over time, that is, all ξ_j's are equal. Let $\xi = (\tau, \phi)$ denote their common value. This assumption may be suitable for instance in direct accident insurance when the number of risks or risk years are taken as volumes p_j .

C. By inspection of (4.8), it is seen that the relevant parameters now are ϕ and

$$\rho_d = \pi_d \tau \qquad ; \ d = 0, \ldots, D. \tag{5.1}$$

5.2. Prediction by known parameters

Predictions are made as in paragraph 4.2. Formulas (4.3)-(4.5) now become

$$\begin{aligned} \mu_{hj} &= p_j \rho_{>J-j} \int y^h dG_{>J-j}(y|\phi) \\ &= p_j \sum_{d>J-j} \rho_d \int y^h dG_d(y|\phi) \qquad ; \ h=1,2,3; \end{aligned} \tag{5.2}$$

$$\kappa = \sum_{j=J-D+1}^{J} p_j \rho_{>J-j} \ ,$$

$$H(\cdot) = \kappa^{-1} \sum_{j=J-D+1}^{J} p_j \sum_{d>J-j} \rho_d G_d(\cdot|\phi).$$

5.3. Parameter estimation

A. Upon inserting $\tau_j = \tau$ and $\phi_j = \phi$ and introducing the ρ_d's from (5.1), the essential part of (4.8) reduces to

$$\left(\prod_{d=0}^{D} \rho_d^{k_{\leqslant J-d,d}} e^{-p_{\leqslant J-d}\rho_d}\right)\left\{\prod_{d=0}^{D}\prod_{j=1}^{J-d}\prod_{k=1}^{k_{jd}} dG_d(y_{jdk}|\phi)\right\}.$$

The maximum likelihood estimator of ρ_d is readily found to be

$$\rho_d^{*} = K_{\leqslant J-d,d}/p_{\leqslant J-d} \qquad ; \ d=0,\ldots,D. \qquad (5.3)$$

B. Estimation of ϕ goes as in item 4.3.F.

C. We mention here only one example of a genuine prediction. Inserting the estimators (4.17) and (5.3) into (5.2)for $h=1$, we obtain the estimated expected value predictor

$$\tilde{R}^{*} = \sum_{j=J-D+1}^{J} p_j \sum_{d>J-j} S_{\leqslant J-d,d}/p_{\leqslant J-d} \,.$$

This result should be compared with (4.19).

5.4. Comments*

A*. It is noteworthy that the volumes p_j play an essential role in the present case, confer the comment in item 4.4.C.

B*. Another important feature of the present specification of the model is that the set of parameters, $\rho_0,\ldots,\rho_D,\phi$, does not increase as J increases. The maximum likelihood estimator of τ

is $\tau^* = \sum_{d=0}^{D} \rho_d^*$, with ρ_d^* defined by (5.3). Its variance is

$$\mathrm{Var}\ \tau^* = \tau \sum_{d=0}^{D} \pi_d / p_{\leqslant J-d} , \qquad (5.4)$$

which should be compared with (4.20). The expression in (5.4) tends to 0 as $p_{\leqslant J-D}$ increases, and in the present model it is always smaller than the expression in (4.20) for $j \leqslant J-D$. This observation points to the necessity of specifying parsimonious models with as few parameters as possible; if the risk conditions can be assumed to be virtually constant over time, then the introduction of a new ξ_j for each year j represents an extravagancy that has to be paid for by a loss of efficiency of the estimators.

6. Prediction based on numbers of claims and single claims amounts when single claim amounts are not affected by fluctuations in basic risk conditions and $\underline{U}$ is parametric

6.1. Description of the case

A. The statistical basis is the complete data $\underline{O}$ given by (4.1).

B. We now return to the full model in paragraph 2.3, with basic risk conditions in different years represented by random variables as specified in assumption I. We assume, however, that only the number of claims are subject to such fluctuations. This means that all Ψ_j's have the same value ψ, which then becomes a parameter of the distributions. Accordingly, U is now taken to be the common distribution of the random variables T_j .

In the present section we deal with the situation where $\underline{U}$ is a parametric class of distributions, that is,

$$\underline{U} = \{U_{\underset{\sim}{\alpha}} \ ; \ \underset{\sim}{\alpha} \in \underline{A}\}$$

for some open set $\underline{A} \subset \underline{R}^m$.

C. All the parameters $\underset{\sim}{\pi}$, $\underset{\sim}{\alpha}$, and ψ can be estimated from the data, hence any one of the principles of IBNR claim reservation presented in paragraph 3.2 can be employed.

6.2. Prediction by known parameters

A. We first derive the predictive distribution of R. Due to assumption IV we have only to determine, for each j, the conditional distribution of $S_{j,>}$ for given $\underline{O}_j$. From (4.9) it is seen that $K_{j,\leq}$ is sufficient for T_j in the Bayesian sense. Hence the only thing that is required is

$$P(S_{j,>} \leq x | K_{j,\leq} = m)$$
$$= \int_{\tau_j=0}^{\infty} P(S_{j,>} \leq x | T_j = \tau_j, K_{j,\leq} = m) dU_j(\tau_j|m), \qquad (6.1)$$

where $U_j(\cdot|m)$ is the conditional distribution of T_j, given $K_{j,\leq} = m$. As the conditional distribution of $K_{j,\leq}$, given $T_j = \tau_j$, is $Po(\pi_{\leq J-j} p_j \tau_j)$, we find

$$dU_j(\tau_j|m) = \frac{\tau_j^m e^{-\pi_{\leq J-j} p_j \tau_j} dU_{\underset{\sim}{\alpha}}(\tau_j)}{\int_{\tau=0}^{\infty} \tau^m e^{-\pi_{\leq J-j} p_j \tau} dU_{\underset{\sim}{\alpha}}(\tau)} . \qquad (6.2)$$

By the conditional independence of the K_{jd}'s for given T_j (assumption II), we can replace the first factor appearing under the integration sign in (6.1) by the expression on the right of (4.2), with $\phi_j = \phi$ for all j. We then obtain

$$P(S_{j,>} \leq x | K_{j,\leq} = m) = \sum_{k=0}^{\infty} q_j(k|m) G_{>J-j}^{k*}(x|\phi), \qquad (6.3)$$

where

$$q_j(k|m) = P(K_{j,>} = k | K_{j,\leq} = m)$$

$$= \frac{(\pi_{>J-j} p_j)^k}{k!} \frac{\int_{\tau=0}^{\infty} \tau^{k+m} e^{-p_j \tau} dU_{\underset{\sim}{\alpha}}(\tau)}{\int_{\tau=0}^{\infty} \tau^m e^{-\pi_{\leq J-j} p_j \tau} dU_{\underset{\sim}{\alpha}}(\tau)} \quad ; \; k=0,1,\ldots \qquad (6.4)$$

Finally we have to form the convolution of the distributions (6.3) for $j=J-D+1,\ldots,J$ to obtain the predictive distribution of R. When this has been accomplished, we can calculate fractiles of this distribution and a reserve by principle (3.7).

Each particular specification of $\underline{U}$ requires an analysis of its own, and the computational work may be extensive. We close this paragraph with an example of a family $\underline{U}$ that leads to tractable closed formulas for the counting probabilities in (6.4).

Example (the gamma case). Let $\underline{U}$ be the family of gamma distributions given by $\underset{\sim}{\alpha} = (\gamma,\delta) \in \underline{R}_+^2$ and

$$dU_{\underset{\sim}{\alpha}}(\tau)/d\tau = \begin{cases} \dfrac{\delta^{\gamma}}{\Gamma(\gamma)}\,\tau^{\gamma-1}e^{-\delta\tau}\,; & \tau>0; \\ 0\,; & \tau\leq 0. \end{cases} \tag{6.5}$$

By inspection of (6.2), we see that now also $U_j(\cdot|m)$ is a gamma distribution, namely with parameters $(\gamma+m, \delta+\pi_{\leq J-j}p_j)$, hence (6.4) becomes

$$q_j(k|m) = \binom{\gamma+m+k-1}{k}\left(\frac{\delta+\pi_{\leq J-j}p_j}{\delta+p_j}\right)^{\gamma+m}\left(\frac{\pi_{>J-j}p_j}{\delta+p_j}\right)^k\;;\; k=0,1,\ldots; \tag{6.6}$$

a negative binomial distribution. In this case (6.3) can be calculated by the recursive procedure of Panjer (1981).

B. We are going to construct the reserves (3.5) and (3.6), and as the distribution involved in the above analysis may be cumbersome to calculate, it is of interest to construct also the the approximate fractile reserve (3.9), which involves only the predictive moments defined by (3.4).

Our starting point is (4.3), which in the present model becomes (confer (2.1))

$$M_h(S_{j,>}|T_j) = a_{hj}T_j \quad ; \; h=1,2,3; \tag{6.7}$$

where

$$a_{hj} = \pi_{>J-j}p_j \int y^h dG_{>J-j}(y|\phi). \tag{6.8}$$

We need also the noncentral posterior moments (confer (6.2))

$$B_{hj} = E(T_j^h|K_{j,\leqslant})$$
$$= \frac{\int \tau^{K_{j,\leqslant}+h} e^{-\pi_{\leqslant J-j}p_j\tau} dU_{\underset{\sim}{\alpha}}(\tau)}{\int \tau^{K_{j,\leqslant}} e^{-\pi_{\leqslant J-j}p_j\tau} dU_{\underset{\sim}{\alpha}}(\tau)} \quad ; \; h=1,2,3; \tag{6.9}$$

and the relations

$$E(S_{j,>}^h|K_{j,\leqslant}) = E\{E(S_{j,>}^h|T_j)|K_{j,\leqslant}\} \quad ; \; h=1,2,3; \tag{6.10}$$

the latter being a consequence of the conditional independence of $S_{j,>}$ and $K_{j,\leqslant}$ for fixed T_j.

C. Consider first the simple expected value predictor (3.5). The term M_{1j} in (3.4) now reduces to $E(S_{j,>}|K_{j,\leqslant})$, and by successive application of (6.10) and (6.7),

$$M_{1j} = a_{1j}B_{1j}, \tag{6.11}$$

where a_{1j} and B_{1j} are defined by (6.8) and (6.9).

Example (continued). The h-th noncentral moment of the gamma distribution (6.5) is readily seen to be $\{\delta^\gamma/\Gamma(\gamma)\}\{\Gamma(\gamma+h)/\delta^{\gamma+h}\} = (\gamma+h-1)^{(h)}/\delta^h$. Thus, since $U_j(\cdot|m)$ is the gamma distribution with parameters $(\gamma+m, \delta+\pi_{\leqslant J-j}p_j)$,

$$B_{hj} = (\gamma+K_{j,\leqslant}+h-1)^{(h)}/(\delta+\pi_{\leqslant J-j}p_j)^h \quad ; \; h=1,2,3. \tag{6.12}$$

Specifically, (6.11) assumes the simple form

$$M_{1j} = a_{1j}(\gamma+K_{j,\leqslant})/(\delta+\pi_{\leqslant J-j}p_j),$$

which is to be entered into (3.5).

$\underline{D}^*$. To provide the reserve (3.6), we need also the second order predictive moments

$$\begin{aligned} M_{2j} &= M_2(S_{j,>}|K_{j,\leqslant}) \\ &= E(S_{j,>}^2|K_{j,\leqslant}) - M_{1j}^2 . \end{aligned} \tag{6.13}$$

Using in succession (6.10), (A.1) in appendix A.1, (6.7), and (6.9), we find that

$$\begin{aligned} E(S_{j,>}^2|K_{j,\leqslant}) &= E\{M_2(S_{j,>}|T_j) + M_1^2(S_{j,>}|T_j)|K_{j,\leqslant}\} \\ &= a_{2j}B_{1j} + a_{1j}^2 B_{2j} . \end{aligned} \tag{6.14}$$

Entering (6.11) and (6.14) into (6.13) yields

$$M_{2j} = a_{1j}^2(B_{2j}-B_{1j}^2) + a_{2j}B_{1j} , \tag{6.15}$$

where the elements on the right hand side are defined by (6.8) and (6.9).

The reserve (3.6) is now obtained by substitution of (6.11) and (6.15).

<u>Example (continued)</u>. By use of (6.12), we find that (6.15) in the gamma case becomes

$$M_{2j} = (\gamma+K_{j,\leqslant})\{a_{1j}^2 + a_{2j}(\delta+\pi_{\leqslant J-j}p_j)\}/(\delta+\pi_{\leqslant J-j}p_j)^2 .$$

$\underline{E}^*$. Finally, to construct the reserve (3.9), we need the third order predictive moments

$$M_{3j} = M_3(S_{j,>}|K_{j,\leqslant})$$
$$= E(S^3_{j,>}|K_{j,\leqslant}) - 3E(S^2_{j,>}|K_{j,\leqslant})M_{1j} + 2M^3_{1j} \; . \qquad (6.16)$$

The latter equality results from (A.4) in appendix A.1. By successive use of (6.10), (A.2) in appendix A.1, (6.7), and (6.9), we get

$$E(S^3_{j,>}|K_{j,\leqslant}) = E\{M_3(S_{j,>}|T_j) + 3M_2(S_{j,>}|T_j)M_1(S_{j,>}|T_j)$$
$$+ M^3_1(S_{j,>}|T_j)|K_{j,\leqslant}\}$$
$$= a_{3j}B_{1j} + 3a_{1j}a_{2j}B_{2j} + a^3_{1j}B_{3j} \; . \qquad (6.17)$$

Substituting (6.11), (6.15), and (6.17) into (6.16), we obtain after some trivial rearrangements that

$$M_{3j} = a_{3j}B_{1j} + 3a_{1j}a_{2j}(B_{2j} - B^2_{1j})$$
$$+ a^3_{1j}(B_{3j} - 3B_{1j}B_{2j} + 2B^3_{1j}), \qquad (6.18)$$

where the a_{hj}'s and B_{hj}'s are defined by (6.8) and (6.9).

Assembling the M_{hj}'s from (6.11), (6.15), and (6.18), we can now determine the reserve (3.9).

Formula (6.18) does not simplify in the gamma case, so we do not pursue the example here.

6.3. Parameter estimation

A. The joint distribution of the observations is obtained by integrating the conditional probability (4.9) over the joint distribution of the T_j's. Estimators of $\underset{\sim}{\pi}$ and $\underset{\sim}{\alpha}$ are obtained by maximizing the likelihood of the number of claims, the essential part of which is seen to be

$$\left(\prod_{d=0}^{D}\pi_d^{K_{\leq J-d,d}}\right)\prod_{j=1}^{J}\int_{\tau_j=0}^{\infty}\tau_j^{K_{j,\leq D(j)}}e^{-\pi_{\leq D(j)}p_j\tau_j}dU_{\underset{\sim}{\alpha}}(\tau_j). \tag{6.19}$$

Another estimation method is proposed in item 7.3.C below.

<u>Example (continued)</u>. In the gamme case (6.5) the expression (6.19) reduces to

$$\left(\prod_{d=0}^{D}\pi_d^{K_{\leq J-d,d}}\right)\delta^{J\gamma}\prod_{j=1}^{J}\frac{(\gamma+K_{j,\leq D(j)}-1)^{(K_{j,\leq D(j)})}}{(\delta+\pi_{\leq D(j)}p_j)^{\gamma+K_{j,\leq D(j)}}}.$$

Maximization under the constraint $\sum_{d=0}^{D}\pi_d = 1$ has to be performed by numerical methods.

<u>B</u>. Estimation of ψ goes as in item 4.3.F.

<u>6.4. Comments*</u>

<u>A*</u>. With paragraph 4.2 in mind we note that in the present model past and future are stochastically related through their joint dependence on the latent T_j's. As opposed to the analysis based on the model of section 4, $\underset{\sim}{P}$ now plays a role in the IBNR-prediction also when the parameters are considered as known; $K_{j,\leq}$ gives a pointer to the value of T_j and, thereby, also to the numbers of future notifications, $\underset{\sim}{K}_{j,>}$.

On the other hand, when it comes to genuine predictions, $\underset{\sim}{P}$ plays a central role also in the model of section 4. In fact, in that model the estimate (4.16) of τ_j rests entirely on the claims experience of year j alone (apart from the fact that $\underset{\sim}{\pi}$ is estimated by statistics from all the fully developed years). In the present model T_j is estimated partly from all claims experience of year j and (through $\underset{\sim}{\alpha}^*$) partly from that of other years. This circumstance has, of course, its root in the assumption that all T_j's stem from the same distribution. By this assumption the risk conditions certainly vary from one year to another, but they are not completely uncomparable as they were in the model of section 4; one can learn something about the present year by looking at what happened in former years.

$\underline{B}^*$. The volumes p_j play a significant role in the present model, recall item 4.4.C.

$\underline{C}^*$. The number of parameters is dramatically reduced as compared to the situation in section 4, confer item 4.4.D. Instead of introducing a new frequency parameter for each year, we now have only the parameter $\underset{\sim}{\alpha}$ of the distribution that generates the T_j's.

7. Prediction based on numbers of claims and single claim amounts when single claim amounts are not affected by fluctuations in basic risk conditions and $\underline{U}$ is nonparametric

7.1. Description of the case

A. The situation is the same as in section 6, see items 6.A and 6.B, except that $\underline{U}$ is now nonparametric.

B. In this case estimation of U and the functionals appearing in the reserves constructed in section 6 is not feasible in general. We can, however, still estimate all parameters required in credibility predictors based on the sufficient (in Bayes sense) statistics $K_{j,\leq}$; j=J-D+1,...,J. The parameters in question turn out to be ϕ, $\underset{\sim}{\pi}$, and the unconditional moments

$$\nu_h = ET_j^h \qquad ; \; h=1,2,\ldots,m; \tag{7.1}$$

where m depends on the choice of reserving formula. All ν_h's displayed in the following are assumed to exist.

C. Apart from ϕ the number of parameters is D+1+m.

7.2. Prediction by known parameters

A. In each item of this paragraph we assume that the known parameters are ϕ, $\underset{\sim}{\pi}$, $\nu_1,\ldots,\nu_m$, where m is the number of ν_h's needed in the analysis. As U is not fully specified, a full posterior analysis cannot be accomplished, and we have to resort to credibility methods.

B. The credibility predictor of T_j based on $K_{j,\leqslant}$ is

$$\tilde{T}_j = \zeta_j \hat{T}_j + (1-\zeta_j)\beta, \tag{7.2}$$

where

$$\hat{T}_j = K_{j,\leqslant}/\pi_{\leqslant J-j}p_j \ , \tag{7.3}$$

$$\zeta_j = \lambda(\lambda+\beta/\pi_{\leqslant J-j}p_j)^{-1} \ , \tag{7.4}$$

$$\beta = ET_j = \nu_1 \ , \quad \lambda = \text{Var } T_j = \nu_2 - \nu_1^2 \ , \tag{7.5}$$

and ν_1 and ν_2 are defined by (7.1). (The reparametrization (7.5) is made to facilitate reference to well known credibility formulas.) Formula (7.2) is demonstrated in item E below. It follows from (6.7) that the credibility predictor of $S_{j,>}$ is

$$\tilde{S}_{j,>} = a_{1j}\tilde{T}_j \ , \tag{7.6}$$

where a_{1j} is defined by (6.8). Thus the credibility IBNR-forecast (3.10) becomes

$$\tilde{R} = \sum_{j=J-D+1}^{J} a_{1j}\tilde{T}_j \ .$$

By use of (6.7),

$$\begin{aligned} \text{Var}S_{j,>} &= \text{Var}(a_{1j}T_j) + E(a_{2j}T_j) \\ &= a_{1j}^2\lambda + a_{2j}\beta, \end{aligned} \tag{7.7}$$

with λ and β defined by (7.5). From (7.6) and (7.7) we obtain the reserve by principle (3.11),

$$\tilde{R} = \sum a_{1j}\tilde{T}_j + f(\lambda\sum a_{1j}^2 + \beta\sum a_{2j}),$$

all sums extending over $j=J-D+1,\ldots,J$.

C*. Following the ideas of paragraph 3.2, we shall construct a more sophisticated reserve of the form (3.12). For this purpose we need, in addition to (7.6), also some kind of approximations of the predictive second order moments in (6.15). We propose to replace B_{1j} and B_{2j} on the right hand side of (6.15) by credibility approximations. Now the credibility approximation of B_{hj} happens to coincide with that of T_j^h as is seen from the identity

$$E\{T_j^h - C(K_{j,\leq})\}^2 = E\mathrm{Var}(T_j^h|K_{j,\leq}) + E\{E(T_j^h|K_{j,\leq}) - C(K_{j,\leq})\}^2 .$$

Having already the credibility approximation (7.2) of T_j, we only need an approximation of T_j^2. In item E below the credibility formula based on $K_{j,\leq}^{(2)}$ is shown to be

$$\tilde{T_j^2} = \eta_j \hat{T_j^2} + (1-\eta_j)\nu_2 , \tag{7.8}$$

where

$$\hat{T_j^2} = K_{j,\leq}^{(2)}/(\pi_{\leq J-j}p_j)^2 , \tag{7.9}$$

$$\eta_j = (\nu_4-\nu_2^2)\{\nu_4-\nu_2^2 + 4\nu_3/\pi_{\leq J-j}p_j + 2\nu_2/(\pi_{\leq J-j}p_j)^2\}^{-1} .$$

By the proposed recipe, we approximate (6.15) by

$$\tilde{M}_{2j} = a_{1j}^2(\tilde{T_j^2}-\tilde{T}_j^2) + a_{2j}\tilde{T}_j , \tag{7.10}$$

with $\tilde{T}_j$ and $\tilde{T_j^2}$ defined by (7.2) and (7.8), respectively. Principle (3.12) can now be applied with $\tilde{S}_{j,>}$ and $\tilde{M}_{2j}$ given by (7.6) and (7.10).

D*. We shall pursue further the ideas of the previous item and arrange also a variation of the reserve formula (3.13). From (6.18) it is seen that, in addition to the already established credibility approximations of B_{1j} and B_{2j}, we need also to approximate B_{3j} or, equivalently, T_j^3. In item E below we demonstrate that the credibility predictor of T_j^3 based on $K_{j,\leq}^{(3)}$ is

$$\tilde{T_j^3} = \rho_j \hat{T_j^3} + (1-\rho_j)\nu_3 , \tag{7.11}$$

where

$$\hat{T_j^3} = K_{j,\leq}^{(3)}/(\pi_{\leq J-j}p_j)^3 , \tag{7.12}$$

$$\rho_j = (\nu_6-\nu_3^2)\{\nu_6-\nu_3^2 + 9\nu_5/\pi_{\leq J-j}p_j + 18\nu_4/(\pi_{\leq J-j}p_j)^2$$
$$+ 6\nu_3/(\pi_{\leq J-j}p_j)^3\}^{-1} .$$

Approximate third order predictive moments $\tilde{M}_{3j}$ are now obtained upon replacing B_{1j}, B_{2j}, and B_{3j} in (6.18) by $\tilde{T}_j$, $\tilde{T_j^2}$, and $\tilde{T_j^3}$ from (7.2), (7.8), and (7.11). Finally, insert the $\tilde{M}_{hj}$'s in (3.13) to obtain an "approximate NP-approximation" of the upper ε-fractile of the predictive distribution of R.

E*. We shall sketch the calculations leading to the credibility formulas (7.2), (7.8), and (7.11). We need the relations

$$EK_{jd} = \pi_d p_j \nu_1 \tag{7.13}$$

and

$$EK_{j,\leq}^{(h)} = (\pi_{\leq J-j}p_j)^h \nu_h \qquad ; \ h=1,2,\ldots; \tag{7.14}$$

which result from (A.6) in appendix A.2 and the fact that, conditional on $T_j = \tau$, we have $K_{jd} \sim Po(\pi_d p_j \tau)$ and $K_{j,\leqslant} \sim Po(\pi_{\leqslant J-j} p_j \tau)$. Putting $T_j^{\hat{1}} = \hat{T}_j$ and recalling the definitions (7.3), (7.9), and (7.12), we have by (7.14) that

$$ET_j^h = ET_j^{\hat{h}} = \nu_h , \qquad (7.15)$$

and, by (A.6) in appendix A.1,

$$\text{Var } E(T_j^{\hat{h}} | T_j) = \text{Var } T_j^h$$

$$= \nu_{2h} - \nu_h^2 .$$

By use of (7.15) and the easy identities

$$k^2 = k^{(2)} + k,$$

$$(k^{(2)})^2 = k^{(4)} + 4k^{(3)} + 2k^{(2)} ,$$

$$(k^{(3)})^2 = k^{(6)} + 9k^{(5)} + 18k^{(4)} + 6k^{(3)} ,$$

we find that

$$\text{Var } T_j^{\hat{h}} = E(K_{j,\leqslant}^{(h)})^2/(\pi_{\leqslant J-j} p_j)^{2h} - \nu_h^2$$

$$= \begin{cases} \nu_2 + \nu_1/\pi_{\leqslant J-j} p_j - \nu_1^2 ; & h=1; \\ \nu_4 + 4\nu_3/\pi_{\leqslant J-j} p_j + 2\nu_2/(\pi_{\leqslant J-j} p_j)^2 - \nu_2^2 ; & h=2; \\ \nu_6 + 9\nu_5/\pi_{\leqslant J-j} p_j + 18\nu_4/(\pi_{\leqslant J-j} p_j)^2 & \\ \quad + 6\nu_3/(\pi_{\leqslant J-j} p_j)^3 - \nu_3^2 ; & h=3. \end{cases}$$

On identifying M and X in (A.17) in appendix A.3 with T_j^h and $T_j^{\hat{h}}$, respectively, we obtain (7.2), (7.8), and (7.11).

The credibility approximations derived in this section are not optimal in general. They can be improved upon by including more than one factorial of $K_{j,\leqslant}$ in the formulas. Such problems are treated by Neuhaus (1985).

7.3. Parameter estimation

A. Estimations of $\underset{\sim}{\pi}$ and the ν_h's can be performed by some moment method based on the sufficient statistics $K_{\leq J-d,d}$; $d=0,\ldots,D$; and $K_{j,\leq D(j)}$; $j=1,\ldots,J$. A convenient starting point are the following relations, which result from (7.13) and (7.14):

$$EK_{\leq J-d,d} = \pi_d p_{\leq J-d}\nu_1 \qquad ; \; d=0,\ldots,D; \qquad (7.16)$$

$$EK^{(h)}_{j,\leq D(j)} = (\pi_{\leq D(j)}p_j)^h \nu_h \; ; \; h=1,2,\ldots; \quad j=1,\ldots,J. \qquad (7.17)$$

If D is small compared to $J-D$, a particularly simple procedure can be arranged. First base estimators of the ν_h's on (7.17) for the fully developed years $j = 1,\ldots,J-D$, for which $\pi_{\leq D(j)} = 1$. A class of unbiased estimators is given by

$$\nu^*_h = \sum_{j=1}^{J-D} w_{hj}K^{(h)}_{j,\leq D} / \sum_{j=1}^{J-D} w_{hj}p^h_j \qquad ; \; h=1,2,\ldots; \qquad (7.18)$$

where the w_{hj} are some positive weights. When ν^*_1 is found, (7.16) motivates that π_d be estimated by

$$\pi^*_d = K_{\leq J-d,d}/p_{\leq J-d}\nu^*_1 \qquad ; \; d=0,\ldots,D. \qquad (7.19)$$

$\underline{B}^*$. An alternative procedure could be to start from the maximum likelihood estimates for the conditional model in section 4, either those in (4.11)-(4.13) based on all available observations or the simpler ones in (4.14)-(4.15). Consider those given by (4.11)-(4.13) and rebaptize each τ^*_j as T^*_j in accordance with the present model assumptions. Then $\underset{\sim}{\pi}^*$ is obtained directly by solving (4.11)-(4.13), and estimators of the two first moments

ν_1 and ν_2 can be based on the asymptotic properties (by increasing p_j's) of the T_j^*'s;

$$\text{as.}E(T_j^*|T_j) = T_j \ , \tag{7.20}$$

and (confer (4.20))

$$\text{as.Var}(T_j^*|T_j) = T_j/\pi_{\leq D(j)}p_j \ . \tag{7.21}$$

From (7.20) and (7.21) we find

$$ET_j^* \approx ET_j = \nu_1 \ , \tag{7.22}$$

$$\begin{aligned} E(T_j^*)^2 &= \text{Var } T_j^* + E^2T_j^* \\ &= \text{VarE}(T_j^*|T_j) + \text{EVar}(T_j^*|T_j) + E^2T_j^* \\ &\approx \text{Var } T_j + ET_j/\pi_{\leq D(j)}p_j + \nu_1^2 \\ &= \nu_2 + \nu_1/\pi_{\leq D(j)}p_j \ . \end{aligned} \tag{7.23}$$

A class of asymptotic moment method estimators based on (7.22) and (7.23) is given by

$$\nu_1^* = \sum_{j=1}^{J} w_{1j}T_j^* \ ,$$

$$\nu_2^* = \sum_{j=1}^{J} w_{2j}(T_j^*)^2 - \nu_1^* \sum_{j=1}^{J} w_{2j}/\pi_{\leq D(j)}^* p_j \ ,$$

where for each $h = 1,2$ the w_{hj}'s are positive weights summing to 1.

$\underline{C}^*$. As an alternative to the laborious maximum likelihood procedure presented in paragraph 6.3, one could use moment methods based on the simple estimators (7.18) and (7.19).

Example (the gamma case continued). Estimators of (γ,δ) are obtained by substitution of ν_1^* and ν_2^* from (7.18) and (7.19) into the relations

$$\nu_1 = \gamma/\delta, \quad \nu_2 = (1+\gamma)\gamma/\delta^2,$$

which yields

$$\gamma^* = (\nu_2^*/\nu_1^{*2} - 1)^{-1}, \quad \delta^* = \gamma^*/\nu_1^* .$$

7.4. Comments*

A*. The credibility formulas derived in paragraph 7.2 shed more light on the comment made in 4.4.B. For instance, the empirical counterpart of (7.2) obtained by inserting estimators for the parameters occurring in (7.3)-(7.5) shows clearly how the experience from occurrence year j is balanced against the experience from other years.

B*. Referring to the discussion in item 4.4.C, we notice that the significance of the p_j's in the present model is clearly exhibited by formulas (7.2) and (7.4); by increasing p_j the weight attached to the experience in year j increases, as one should expect.

8. Prediction based on total claim amounts in the unrestricted framework model

8.1. Description of the case

A. The available data is now assumed to be $\underline{O} = \{S_{jd};\ d=0,\ldots,D(j);\ j=1,\ldots,J\}$, which is typical of a reinsurance business written on an underwriting year basis.

B. No restrictions are imposed on the families of distributions $\underline{U}$ and $\underline{G}_d$; $d=0,\ldots,D$; except that the moments indicated below are assumed to exist.

C. In the present case the joint distribution of $\tilde{S}_j = (S_{j0},\ldots,S_{jD})'$ is not estimable in general. It would be if the p_j's were equal, which is not likely to occur in practice. We can, however, estimate the moments of the distribution of $\tilde{S}_j$ for each j. This circumstance is due to the distributional structure inherent in the basic model assumptions I-IV.

Introduce

$$B_{hjd} = \pi_d T_j \int y^h dG_d(y|\Psi_j) \quad ; \quad h=1,\ldots,4;\ d=0,\ldots,D\ ;\ j=1,2,\ldots \tag{8.1}$$

The moments up to fourth order of the S_{jd}'s turn out to depend on the following basic parameters.

		Number of parameters
1st order parameters:		
$\beta_d = EB_{1jd}$	; $\forall d$.	$D+1$
2nd order parameters:		
$\beta_{de} = E(B_{1jd}B_{1je})$	; $d \leq e$;	$D+1 + \binom{D+1}{2}$
$\beta_{d^2} = EB_{2jd}$	; $\forall d$.	$D+1$

3rd order parameters*:

$$\beta_{def} = E(B_{1jd}B_{1je}B_{1jf}) \quad ; \; d \leq e \leq f \qquad D+1 + (D+1)D + \binom{D+1}{3}$$

$$\beta_{de^2} = \begin{cases} E(B_{1jd}B_{2je}) & ; \; d \neq e \; ; \\ 3E(B_{1jd}B_{2jd}) & ; \; d = e \; ; \end{cases} \qquad (D+1)^2$$

$$\beta_{d^3} = EB_{3jd} \quad ; \; \forall d. \qquad D+1$$

4th order parameters*:

$$\beta_{defg} = E(B_{1jd}B_{1je}B_{1jf}B_{1jg}) \quad ; \; d \leq e \leq f \leq g \; ; \qquad D+1 + (D+1)D + \binom{D+1}{2} + (D+1)\binom{D}{2} + \binom{D+1}{4}$$

$$\beta_{def^2} = \begin{cases} 6E(B^2_{1jd}B_{2jd}) & ; \; d=e=f \; ; \\ 3E(B_{1jd}B_{1je}B_{2je}) & ; \; d \neq e = f \; ; \\ E(B_{1jd}B_{1je}B_{2jf}) & ; \; d<e, \; d \neq f, \; e \neq f \; ; \end{cases} \qquad D+1 + (D+1)D + (D+1)\binom{D}{2}$$

$$\beta_{(de)^2} = E(B^2_{1jd}B_{2je} + B^2_{1je}B_{2jd}); \; d<e \; ; \qquad \binom{D+1}{2}$$

$$\beta_{d^2e^2} = E(B_{2jd}B_{2je}) \quad ; \; d<e \; ; \qquad \binom{D+1}{2}$$

$$\beta_{de^3} = \begin{cases} E(3B^2_{2jd} + 4B_{1jd}B_{3jd}) & ; \; d=e \; ; \\ E(B_{1jd}B_{3je}) & ; \; d \neq e \; ; \end{cases} \qquad (D+1)^2$$

$$\beta_{d^4} = E \; B_{4jd} \quad ; \; \forall d. \qquad D+1$$

Let $n_h(D)$ be the total number of parameters of order h or less. We find that $n_2(D) = (D+1)(3+D/2)$, $n_3(D) = n_2(D)+(D+1)(3+2D+D^{(2)}/6)$, and $n_4(D) = n_3(D)+(D+1)(4+9D/2+D^{(2)}+D^{(3)}/24)$, and calculate $n_2(0) = 3$, $n_2(1) = 7$, $n_2(2) = 12$, $n_2(3) = 18$, $n_2(4) = 25$,

$n_2(5) = 33$, $n_2(6) = 42$, ... ; $n_3(0) = 6$, $n_3(1) = 17$, $n_3(2) = 34$, $n_3(3) = 58$, ... ; $n_4(0) = 10$, $n_4(1) = 34$, $n_4(2) = 79$, ...

The order restrictions appearing in the definitions of the β's are, of course, not essential. They have been introduced only for the purpose of keeping an account with the number of distinct parameters. By symmetry, $\beta_{ed} = \beta_{de}$ etc.

D. In the next paragraph we demonstrate the following formulas for the moments.

1st order moments:

$$ES_{jd} = p_j\beta_d \qquad ; \forall d. \tag{8.2}$$

2nd order moments:

$$ES^2_{jd} = p^2_j\beta_{dd} + p_j\beta_{d^2} \qquad ; \forall d ; \tag{8.3}$$

$$E(S_{jd}S_{je}) = p^2_j\beta_{de} \qquad ; d<e. \tag{8.4}$$

3rd order moments*:

$$ES^3_{jd} = p^3_j\beta_{ddd} + p^2_j\beta_{dd^2} + p_j\beta_{d^3} \qquad ; \forall d ; \tag{8.5}$$

$$E(S_{jd}S^2_{je}) = p^3_j\beta_{dee} + p^2_j\beta_{de^2} \qquad ; d\neq e ; \tag{8.6}$$

$$E(S_{jd}S_{je}S_{jf}) = p^3_j\beta_{def} \qquad ; d<e<f ; \tag{8.7}$$

4th order moments*:

$$ES^4_{jd} = p^4_j\beta_{dddd} + p^3_j\beta_{ddd^2} + p^2_j\beta_{dd^3} + p_j\beta_{d^4} \qquad ; \forall d ; \tag{8.8}$$

$$E(S_{jd}S^3_{je}) = p^4_j\beta_{deee} + p^3_j\beta_{dee^2} + p^2_j\beta_{de^3} \qquad ; d\neq e ; \tag{8.9}$$

$$E(S_{jd}^2 S_{je}^2) = p_j^4 \beta_{ddee} + p_j^3 \beta_{(de)^2} + p_j^2 \beta_{d^2 e^2} \quad ; \ d<e \ ; \qquad (8.10)$$

$$E(S_{jd} S_{je} S_{jf}^2) = p_j^4 \beta_{deff} + p_j^3 \beta_{def^2} \quad ; \ d<e, \ f \notin \{d,e\}; \qquad (8.11)$$

$$E(S_{jd} S_{je} S_{jf} S_{jg}) = p_j^4 \beta_{defg} \quad ; \ d<e<f<g. \qquad (8.12)$$

(As in the case of the β's above, the order restrictions can trivially be removed.)

$\underline{E}^*$. We shall prove the formulas (8.2)-(8.12). Upon replacing τ and EY^h in appendix A.2 by $\pi_d p_j T_j$ and $\int y^h dG_d(y|\Psi_j)$ and introducing the B_{hjd}'s from (8.1), we obtain from (A.8)-(A.11) that

$$E(S_{jd}|\Xi_j) = p_j B_{1jd} \ ,$$
$$E(S_{jd}^2|\Xi_j) = p_j^2 B_{1jd}^2 + p_j B_{2jd} \ ,$$

$$E(S_{jd}^3|\Xi_j) = p_j^3 B_{1jd}^3 + 3p_j^2 B_{1jd} B_{2jd} + p_j B_{3jd} \ ,$$

$$E(S_{jd}^4|\Xi_j) = p_j^4 B_{1jd}^4 + 6p_j^3 B_{1jd}^2 B_{2jd} + p_j^2 (3B_{2jd}^2 + 4B_{1jd} B_{3jd}) + p_j B_{4jd}.$$

Using these expressions, we find the following formulas, which are just the ones given in (8.2)-(8.12).

$$ES_{jd} = E(p_j B_{1jd}) \quad ; \ \forall d \ ;$$

$$ES_{jd}^2 = E(p_j^2 B_{1jd}^2 + p_j B_{2jd}) \quad ; \ \forall d \ ;$$

$$E(S_{jd} S_{je}) = E(p_j B_{1jd} p_j B_{1je}) \quad ; \ d<e \ ;$$

$$ES_{jd}^3 = E(p_j^3 B_{1jd}^3 + 3p_j^2 B_{1jd} B_{2jd} + p_j B_{3jd}) \quad ; \ \forall d \ ;$$

$$E(S_{jd} S_{je}^2) = E\{p_j B_{1jd} (p_j^2 B_{1je}^2 + p_j B_{2je})\} \quad ; \ d \neq e \ ;$$

$$E(S_{jd}S_{je}S_{jf}) = E(p_jB_{1jd}p_jB_{1je}p_jB_{1jf}) \qquad ; \; d<e<f \; ;$$

$$ES^4_{jd} = E\{p^4_jB^4_{1jd} + 6p^3_jB^2_{1jd}B_{2jd} + p^2_j(3B^2_{2jd} + 4B_{1jd}B_{3jd}) + p_jB_{4jd}\}; \; \forall d;$$

$$E(S_{jd}S^3_{je}) = E\{p_jB_{1jd}(p^3_jB^3_{1je} + 3p^2_jB_{1je}B_{2je} + p_jB_{3je})\} \qquad ; \; d \neq e \; ;$$

$$E(S^2_{jd}S^2_{je}) = E\{(p^2_jB^2_{1jd} + p_jB_{2jd})(p^2_jB^2_{1je} + p_jB_{2je})\} \qquad ; \; d<e \; ;$$

$$E(S_{jd}S_{je}S^2_{jf}) = E\{p_jB_{1jd}p_jB_{1je}(p^2_jB^2_{1jf} + p_jB_{2jf})\} \qquad ; \; d<e, \; f \notin \{d,e\} \; ;$$

$$E(S_{jd}S_{je}S_{jf}S_{jg}) = E(p_jB_{1jd}p_jB_{1je}p_jB_{1jf}p_jB_{1jg}) \qquad ; \; d<e<f<g.$$

8.2. Prediction by known parameters

A. As the joint distribution of the S_{jd}'s is not fully specified, a full posterior analysis is not feasible. When the 1st and 2nd order parameters are known, we can, however, employ the principles (3.10) and (3.11).

To construct the credibility predictor of $S_{j,>}$ based on $\underset{\sim}{S}'_{j,<} = (S_{j0},\ldots,S_{j,J-j})$, we pick from (8.2)-(8.4) the moments

$$ES_{j,>} = p_j\beta_{>J-j} \,, \tag{8.13}$$

$$\mathrm{Cov}(S_{j,>},\underset{\sim}{S}'_{j,<}) = \{\sum_{e>J-j}(p^2_j\beta_{de} - p^2_j\beta_d\beta_e)\}'_{d\leq J-j} = p^2_j \, \underset{\sim}{\gamma}'_{J-j} \,, \tag{8.14}$$

$$\mathrm{Var}\, \underset{\sim}{S}_{j,<} = (p^2_j\beta_{de} + \delta_{de}p_j\beta_{d2} - p^2_j\beta_d\beta_e)_{d,e\leq J-j} = p^2_j \, \underset{\sim}{\Sigma}_{j;J-j}, \tag{8.15}$$

$$E\underset{\sim}{S}_{j,<} = p_j \, \underset{\sim}{\beta}_{<J-j} \,, \tag{8.16}$$

where we have introduced

$$\gamma'_{J-j} = \{\sum_{e>J-j} (\beta_{de} - \beta_d \beta_e)\}'_{d \leq J-j} , \tag{8.17}$$

$$\underset{\sim}{\Sigma}_{j;J-j} = (\beta_{de} - \beta_d \beta_e + \delta_{de} p_j^{-1} \beta_{d2})_{d,e \leq J-j} . \tag{8.18}$$

(Note that γ'_d depends only on the stage of development and need not be calculated anew for each occurrence year.) By application of formula (A.16) in appendix A.3, we obtain from (8.13)-(8.16) the credibility predictor

$$\tilde{S}_{j,>} = p_j \beta_{>J-j} + \gamma'_{J-j} \underset{\sim}{\Sigma}_{j;J-j}^{-1} (\underset{\sim}{S}_j - p_j \underset{\sim}{\beta})_{\leq J-j} , \tag{8.19}$$

where γ'_{J-j} and $\underset{\sim}{\Sigma}_{j;J-j}$ are defined by (8.17) and (8.18). Finally insert (8.19) into (3.10) to obtain the credibility IBNR-predictor.

From (8.2)-(8.4) we get

$$\begin{aligned}
\text{Var } S_{j,>} &= ES^2_{j,>} - E^2 S_{j,>} \\
&= \sum_{d,e>J-j} E(S_{jd} S_{je}) - (\sum_{d>J-j} ES_{jd})^2 \\
&= \sum_{d,e>J-j} (p_j^2 \beta_{de} + \delta_{de} p_j \beta_{d2}) - \sum_{d,e>J-j} p_j \beta_d p_j \beta_e \\
&= p_j^2 \sum_{d,e>J-j} (\beta_{de} - \beta_d \beta_e) + p_j \sum_{d>J-j} \beta_{d2} .
\end{aligned} \tag{8.20}$$

On inserting (8.19) and (8.20) into (3.11), we obtain a reserve with a security loading.

<u>B</u>*. We can arrange a recursive algorithm for calculation of $\underset{\sim}{\Sigma}_{j;J-j}^{-1}$. Let $\sigma_{j;de}$ denote the (d,e)-element in $\underset{\sim}{\Sigma}_{j;D}$. For each $d = 1,\dots,D$ partition $\underset{\sim}{\Sigma}_{j;d}$ into

$$\underset{\sim}{\Sigma}_{j;d} = \begin{pmatrix} \underset{\sim}{\Sigma}_{j;d-1} & \underset{\sim}{\sigma}_{j;d} \\ \underset{\sim}{\sigma}'_{j;d} & \sigma_{j;dd} \end{pmatrix} , \tag{8.21}$$

where

$$\underset{\sim}{\sigma}'_{j;d} = (\sigma_{j;d0}, \dots, \sigma_{j;d,d-1}) .$$

The inverse of $\underset{\sim}{\Sigma}_{j;d}$ is partitioned correspondingly;

$$\underset{\sim}{\Sigma}^{-1}_{j;d} = \begin{pmatrix} \underset{\sim}{A}_{j;d} & \underset{\sim}{b}_{j;d} \\ \underset{\sim}{b}'_{j;d} & c_{j;d} \end{pmatrix} . \tag{8.22}$$

When p_j is known, the matrices $\Sigma^{-1}_{j,d}$; $d=0,1,\ldots,D$; may be calculated recursively as follows. Once $\underset{\sim}{\Sigma}^{-1}_{j;d-1}$ has been found, first calculate the auxiliary quantities

$$\underset{\sim}{v}_{j;d} = \underset{\sim}{\Sigma}^{-1}_{j;d-1} \underset{\sim}{\sigma}_{j;d} ,$$
$$w_{j;d} = \underset{\sim}{\sigma}'_{j;d} \underset{\sim}{v}_{j;d} , \tag{8.23}$$

and then

$$c_{j;d} = (\sigma_{j;dd} - w_{j;d})^{-1} , \tag{8.24}$$

$$b_{j;d} = - c_{j;d} \underset{\sim}{v}_{j;d} , \tag{8.25}$$

$$\underset{\sim}{A}_{j;d} = \underset{\sim}{\Sigma}^{-1}_{j;d-1} - \underset{\sim}{b}_{j;d} \underset{\sim}{v}'_{j;d} , \tag{8.26}$$

which determine $\underset{\sim}{\Sigma}^{-1}_{j;d}$ by (8.22). The recursion is initiated by

$$\underset{\sim}{\Sigma}^{-1}_{j;0} = c^{-1}_{j;00} .$$

The proof rests on the results in appendix A.4. Identify $\underset{\sim}{A}$ in (A.19) with $\underset{\sim}{\Sigma}_{j;d}$ in (8.21). Then (A.22) and (A.23) specialize to (8.24) and (8.25), respectively, and (A.20) becomes

$$\underset{\sim}{A}_{j;d} = (\underset{\sim}{\Sigma}_{j;d-1} - \underset{\sim}{\sigma}_{j;d} \sigma^{-1}_{j;dd} \underset{\sim}{\sigma}'_{j;d})^{-1}. \tag{8.27}$$

Upon identifying $\underset{\sim}{A}$ and $\underset{\sim}{b}$ in (A.25) with $\underset{\sim}{\Sigma}_{j;d-1}$ and $\underset{\sim}{\sigma}_{j;d} \sigma^{-\frac{1}{2}}_{j;dd}$ in (8.27), we obtain (8.26).

When access is being had to powerful computer equipment with standard programs for matrix inversion, it may not be worth while implementing the algorithm (8.23)-(8.26). It also ought to be said that in empirical Bayes situations, where parameter estimates are currently updated along with the emergence of fresh data, recursion formulas valid for fixed parameter values are of little practical value.

$\underline{C}^*$. To apply principle (3.12), we need, in addition to the credibility predictor (8.19), also a credibility approximation of the second central predictive moment

$$M_2(S_{j,>}|\underset{\sim}{S}_{j,\leq}) = E(S^2_{j,>}|\underset{\sim}{S}_{j,\leq}) - E^2(S_{j,>}|\underset{\sim}{S}_{j,\leq}). \tag{8.28}$$

The best linear approximation to the first term on the right of (8.28) is just the credibility predictor of $S^2_{j,>}$. To construct the credibility predictor based on $\underset{\sim}{S}_{j,\leq}$, we compile from (8.3)-(8.4) that

$$\begin{aligned} ES^2_{j,>} &= \sum_{d,e>J-j} E(S_{jd}S_{je}) \\ &= \sum_{d,e>J-j} (p_j^2\beta_{de} + \delta_{ef}p_j\beta_{d^2}) \\ &= p_j\,\delta_j\,, \end{aligned} \tag{8.29}$$

where

$$\delta_j = p_j \sum_{d,e>J-j} \beta_{de} + \sum_{d>J-j} \beta_{d^2}\,, \tag{8.30}$$

and from (8.2)-(8.7) that

$$\begin{aligned} \mathrm{Cov}(S^2_{j,>},\underset{\sim}{S}'_{j,\leq}) &= E(\sum_{e,f>J-j} S_{je}\,S_{jf}\,\underset{\sim}{S}'_{j,\leq}) - ES^2_{j,>}E\underset{\sim}{S}'_{j,\leq} \\ &= \{\sum_{e,f>J-j} (p_j^3\beta_{def} + \delta_{ef}p_j^2\beta_{de^2})\}'_{d=0,\ldots,J-j} \\ &\quad - (p_j^2 \sum_{e,f>J-j} \beta_{ef} + p_j \sum_{e>J-j} \beta_{e^2})p_j\underset{\sim}{\beta}'_{\leq J-j} \\ &= p_j^2\,\underset{\sim}{\varepsilon}'_j\,, \end{aligned} \tag{8.31}$$

where

$$\underset{\sim}{\varepsilon}_j' = \{p_j \sum_{e,f>J-j} (\beta_{def} - \beta_d\beta_{ef}) + \sum_{e>J-j} (\beta_{de^2} - \beta_d\beta_{e^2})\}'_{d\leq J-j} \quad (8.32)$$

From (8.29), (8.31), (8.15), (8.16), and (A.16) in appendix A.3 we obtain the credibility approximation

$$\tilde{E}(S^2_{j,>}|\underset{\sim}{S}_{j,\leq}) = p_j\delta_j + \underset{\sim}{\varepsilon}_j' \underset{\sim}{\Sigma}^{-1}_{j;J-j}(\underset{\sim}{S}_j - p_j\underset{\sim}{\beta})_{\leq J-j} \, , \quad (8.33)$$

where δ_j and $\underset{\sim}{\varepsilon}_j'$ are given by (8.30) and (8.32).

The credibility approximation of the conditional mean appearing in the second term in (8.28) is just $\tilde{S}_{j,>}$, hence (8.28) is approximated by

$$\tilde{M}_{2j} = \tilde{E}(S^2_{j,>}|\underset{\sim}{S}_{j,\leq}) - \tilde{S}^2_{j,>} \, , \quad (8.34)$$

where the terms on the right are defined by (8.19) and (8.33).

The required IBNR-reserve is now obtained upon inserting (8.19) and (8.34) into (3.12).

$\underline{D}^*$. To apply principle (3.13), we have to approximate predictive moments up to third order. The first two moments are approximated by (8.19) and (8.34). In addition we need some credibility approximation of the predictive third central moment, which by (A.4) in appendix A.1 is

$$\begin{aligned} M_{3j} = M_3(S_{j,>}|\underset{\sim}{S}_{j,\leq}) &= E(S^3_{j,>}|\underset{\sim}{S}_{j,\leq}) \\ &- 3E(S^2_{j,>}|\underset{\sim}{S}_{j,\leq})E(S_{j,>}|\underset{\sim}{S}_{j,\leq}) \quad (8.35) \\ &+ 2E^3(S_{j,>}|\underset{\sim}{S}_{j,\leq}). \end{aligned}$$

The best linear approximation of the first term on the right of (8.35) is the credibility predictor of $S^3_{j,>}$. The credibility predictor based on $\underset{\sim}{S}_{j,\leq}$ involves the moments $ES^3_{j,>}$, $Cov(S^3_{j,>}, \underset{\sim}{S}'_{j,\leq})$, and those in (8.15) and (8.16). From (8.5)-(8.7) we find (all sums indicated range over indices $d,e,f > J-j$)

$$\begin{aligned}
ES^3_{j,>} &= \sum_{d,e,f>J-j} E(S_{jd}S_{je}S_{jf}) \\
&= \sum_d ES^3_{jd} + 3\sum_{d\neq e} E(S_{jd}S^2_{je}) + \sum_{d\neq e,d\neq f,e\neq f} E(S_{jd}S_{je}S_{jf}) \\
&= \sum_d (p^3_j\beta_{ddd} + p^2_j\beta_{dd^2} + p_j\beta_{d^3}) \\
&\quad + 3\sum_{d\neq e}(p^3_j\beta_{dee} + p^2_j\beta_{de^2}) + \sum_{d\neq e,d\neq f,e\neq f} p^3_j\beta_{def} \\
&= p_j\,\rho_j\,,
\end{aligned} \tag{8.36}$$

where

$$\begin{aligned}
\rho_j &= p^2_j \sum_{d,e,f>J-j}\beta_{def} + p_j\Big(\sum_{d>J-j}\beta_{dd^2} \\
&\quad + 3\sum_{d,e>J-j,d\neq e}\beta_{de^2}\Big) + \sum_{d>J-j}\beta_{d^3}\,.
\end{aligned} \tag{8.37}$$

From (8.5)-(8.12) we find for any $d \leq J-j$ and $e,f,g > J-j$ that

$$\begin{aligned}
Cov(S_{jd}, S_{je}S_{jf}S_{jg}) &= E(S_{jd}S_{je}S_{jf}S_{jg}) - ES_{jd}E(S_{je}S_{jf}S_{jg}) \\
&= \begin{cases}
p^4_j\beta_{deee}+p^3_j\beta_{dee^2}+p^2_j\beta_{de^3} - p_j\beta_d(p^3_j\beta_{eee}+p^2_j\beta_{ee^2}+p_j\beta_{e^3}); & e=f=g\,; \\
p^4_j\beta_{deff} + p^3_j\beta_{def^2} - p_j\beta_d(p^3_j\beta_{eff}+p^2_j\beta_{ef^2}) & ; \quad e\neq f=g\,; \\
\text{similar expressions when } f\neq e=g \text{ or } e=f\neq g\,, & \\
p^4_j\beta_{defg} - p_j\beta_d p^3_j\beta_{efg} & ; \quad e\neq f,\ e\neq g,\ f\neq g;
\end{cases} \\
&= p^2_j[p^2_j(\beta_{defg}-\beta_d\beta_{efg}) + p_j\{\delta_{efg}(\beta_{dee^2}-\beta_d\beta_{ee^2}) \\
&\quad + (1-\delta_{ef})\delta_{fg}(\beta_{def^2}-\beta_d\beta_{ef^2}) + (1-\delta_{ef})\delta_{eg}(\beta_{dfg^2}-\beta_d\beta_{fg^2}) \\
&\quad + (1-\delta_{fg})\delta_{ef}(\beta_{dge^2}-\beta_d\beta_{ge^2})\} + \delta_{efg}(\beta_{de^3}-\beta_d\beta_{e^3})].
\end{aligned}$$

It follows that

$$\mathrm{Cov}(S^3_{j,>}, \underset{\sim}{S}'_{j,\leq}) = \{\sum_{e,f,g>J-j} \mathrm{Cov}(S_{jd}, S_{je}S_{jf}S_{jg})\}_{d\leq J-j} = p_j^2\, \underset{\sim}{g}'_j \,, \tag{8.38}$$

where

$$\underset{\sim}{g}'_j = [p_j^2 \sum_{e,f,g>J-j} (\beta_{defg} - \beta_d\beta_{efg}) + p_j\{\sum_{e>J-j} (\beta_{dee^2} - \beta_d\beta_{ee^2}) + 3\sum_{e,f>J-j;e\neq f} (\beta_{def^2} - \beta_d\beta_{ef^2})\} + \sum_{e>J-j} (\beta_{de^3} - \beta_d\beta_{e^3})]'_{d\leq J-j}. \tag{8.39}$$

By (8.15), (8.16), (8.36), (8.38), and (A.16) in appendix A.3, the required credibility approximation is

$$\tilde{E}(S^3_{j,>}|\underset{\sim}{S}_{j,\leq}) = p_j\rho_j + \underset{\sim}{g}'_j\, \underset{\sim}{\Sigma}^{-1}_{j;J-j}(\underset{\sim}{S}_j - p_j\underset{\sim}{\beta})_{\leq J-j} \,, \tag{8.40}$$

with ρ_j and $\underset{\sim}{g}'_j$ defined by (8.37) and (8.39).

Upon replacing the conditional expected values occurring on the right of (8.35) by their credibility approximations, we obtain the approximate third order predictive moment

$$\tilde{M}_{3j} = \tilde{E}(S^3_{j,>}|\underset{\sim}{S}_{j,\leq}) - 3\tilde{E}(S^2_{j,>}|\underset{\sim}{S}_{j,\leq})\tilde{S}_{j,>} + 2\tilde{S}^3_{j,>} \,, \tag{8.41}$$

the single terms in which are defined by (8.19),(8.33), and (8.40).

An approximate NP-approximation of the $(1-\varepsilon)$-fractile IBNR-reserve is now obtained by entering (8.19),(8.34) and (8.41) into (3.13).

8.3. Parameter estimation

A. We shall construct a class of simple weighted least squares estimators of the 1st and 2nd order parameters. Let w_{jd}; $d=0,\dots,D$; $j=1,\dots,J-d$; and w_{jde}; $0\leq d\leq e\leq D$; $j=1,\dots,J-e$; be some positive constants. A set of unbiased estimators of the parameters up to 2nd order is given by

$$\beta^{*}_{d} = \sum_{j=0}^{J-d} w_{jd}\, S_{jd} \,/ \sum_{j=0}^{J-d} w_{jd}\, p_j \qquad ; \quad d=0,\ldots,D \; ; \tag{8.42}$$

$$\beta^{*}_{de} = \sum_{j=0}^{J-e} w_{jde}\, S_{jd}\, S_{je} \,/ \sum_{j=0}^{J-e} w_{jde}\, p_j^2 \qquad ; \quad 0 \leq d < e \leq D \; ; \tag{8.43}$$

$$\beta^{*}_{dd} = (a_{2d}A_{2d} - a_{3d}A_{1d})/(a_{4d}a_{2d} - a_{3d}^2) \quad ; \quad d=0,\ldots,D \; ; \tag{8.44}$$

$$\beta^{*}_{d^2} = (a_{4d}A_{1d} - a_{3d}A_{2d})/(a_{4d}a_{2d} - a_{3d}^2) \quad ; \quad d=0,\ldots,D \; ; \tag{8.45}$$

where the a_{hd}'s and A_{hd}'s are defined by

$$a_{hd} = \sum_{j=1}^{J-d} w_{jdd}\, p_j^h \qquad ; \quad h=2,3,4 \quad ;$$

$$A_{hd} = \sum_{j=1}^{J-d} w_{jdd}\, p_j^h\, S_{jd}^2 \qquad ; \quad h=1,2 \; ; \; d=0,\ldots,D \; .$$

The unbiasedness of the estimators in (8.42) and (8.43) is a direct consequence of (8.2) and (8.4). The estimators in (8.44) and (8.45) are constructed by the technique of least squares based on the linear regression (8.3); for each d we minimize the weighted sum of squared deviations

$$\sum_{j=1}^{J-d} w_{jdd}(S_{jd}^2 - p_j^2\beta_{dd} - p_j\beta_{d^2})^2 . \tag{8.46}$$

It is well known (and easy to check) that the least squares estimators are those in (8.44) and (8.45) and that they are unbiased.

The question of how to specify the weights w_{jd} and w_{jde} is discussed in item B below. Let it suffice here to state, what is intuitively obvious, that in the case where all the p_j are equal one should lay equal emphasis on statistics from different years, that is, use the uniform weights $w_{jd} = w_{jde} = 1$. Often in practice the p_j's do not vary much, so that uniform weights will produce reasonably efficient estimates.

B*. We shall discuss the choice of weights in the estimations (8.42)-(845). The estimator β_d^* defined by (8.42) is recognized as the weighted least squares estimator obtained by minimizing

$$\sum_{j=1}^{J-d} (w_{jd}/p_j)(S_{jd} - p_j\beta_d)^2 .$$

It is well known that the optimal choice of weights w_{jd}/p_j, in the sense of minimizing $E(\beta_d^* - \beta_d)^2$, is $w_{jd}/p_j = (\mathrm{Var}\ S_{jd})^{-1}$, which by (8.2) and (8.3) is equivalent to

$$w_{jd} = \{p_j(\beta_{dd} - \beta_d^2) + \beta_{d^2}\}^{-1} \quad ; \quad j=1,\ldots,J-j. \tag{8.47}$$

Likewise, β_{de}^* defined by (8.43) is the weighted least squares estimator that minimizes

$$\sum_{j=0}^{J-e} (w_{jde}/p_j^2)(S_{jd}S_{je} - p_j^2\beta_{de})^2 .$$

The optimal weights are $w_{jde}/p_j^2 = \{\mathrm{Var}(S_{jd}S_{je})\}^{-1}$, by (8.10) and (8.4),

$$w_{jde} = \{p_j^2(\beta_{ddee} - \beta_{de}^2) + p_j\beta_{(de)^2} + \beta_{d^2e^2}\}^{-1} \; ; \; j=1,\ldots,J-e. \tag{8.48}$$

The optimal choice of weights in (8.46) is $w_{jdd} = (\mathrm{Var}\ S_{jd}^2)^{-1}$ or, by (8.3) and (8.8),

$$\begin{aligned} w_{jdd} = \{&p_j^4(\beta_{dddd} - \beta_{dd}^2) + p_j^3(\beta_{ddd^2} - 2\beta_{dd}\beta_{d^2}) \\ &+ p_j^2(\beta_{dd^3} - \beta_{d^2}^2) + p_j\beta_{d^4}\}^{-1} \quad ; \quad j=1,\ldots,J-d. \end{aligned} \tag{8.49}$$

As the optimal weights depend on the parameters, no uniformly optimal choice can be made. Yet the formulas (8.47)-(8.49) are useful in our search for a good weighting; any set of weights that are not of the general form given by these formulas, with appropriate values of the β's, cannot be optimal at any parameter point. A reasonable procedure could be to specify a set of parameter values β_d^0, β_{de}^0 , etc. for which we want the estimators to perform well, and to use the weights (8.47)-(8.49) corresponding to these values. (A variation of this idea is to pick values β_d^0, β_{de}^0 ,... that are judged as "likely to be close to the true values of β_d, β_{de},... ". This would, however, imply a willingness to specify a prior distribution on the parameter space, and to act in accordance with this attitude, we should estimate the parameters by Bayesian methods.)

As an example, consider the problem of specifying the weights w_{jd} in (8.42). The choice $w_{jd} = 1$; j=1,...,J-d; which gives $\beta_d^\star = \Sigma\, S_{jd}/\Sigma\, p_j$, is nearly optimal if the terms $p_j(\beta_{dd} - \beta_d^2)$ are small compared to β_{d2}. On the other hand, the weights $w_{jd} = p_j^{-1}$; j=1,...,J-d; are nearly optimal if the terms $p_j(\beta_{dd} - \beta_d^2)$ are large as compared to β_{d2}. As a hold in deliberations of this kind, note that $\beta_{dd} - \beta_d^2 = \mathrm{Var}\, E(S_{jd}/p_j|\Xi_j)$ measures the magnitude of the fluctuations in basic risk conditions from one year to another, whereas $\beta_{d2} = p_j E\, \mathrm{Var}(S_{jd}/p_j|\Xi_j)$ measures the (average) pure random variation around the expected result by fixed risk conditions for a portfolio with unit risk exposure.

At any rate, the uniform weights $w_{jd} = w_{jde} = 1$ are close to the optimal ones when the exposures p_j do not vary too much, confer the closing remark in the previous paragraph.

<u>C</u>*. The relations (8.5)-(8.12) specify linear regressions of cross products of 3rd and 4th order of the S_{jd}'s on parameters of 3rd and 4th order. These parameters may, therefore, be estimated by linear methods as described by Norberg (1982). The resulting estimates are needed if we want to use the predictors constructed in

8.2.C and 8.2.D above. They may also be utilized to construct empirical generalized least squares estimators of the 1st and 2nd order parameters by the method proposed by Norberg (1982).

D. By insertion of the parameter estimators from paragraph 8.3 into the formulas for the reserves derived in paragraph 8.2, we will not obtain any compact and appealing expressions, and the formulas shall not be displayed here.

8.4. Discussion

A. As was pointed out in paragraph 8.1, the number of parameters quickly becomes large as D increases. Already for $D = 4$ the number of parameters required to construct the simplest reserving formula (8.19) amounts to 25, which is prohibitive if the statistical basis $\underline{O}$ is scanty, as it often is. Thus the present analysis based on the unrestricted framework model, is directly applicable mainly to situations where D is small. Therefore, it becomes a central issue to specify additional assumptions V that on the one hand are sufficiently rigid to reduce the number of parameters to a manageable level and on the other hand are flexible enough to provide a realistic description of the situation at hand.

In the following section we shall analyze the case where all the Ψ_j are assumed to be equal, and it will turn out that this restriction brings about a substantial reduction of the number of parameters.

Another possibility is to assume that the parameters β_d, β_{de}, etc. are certain parametric functions of the indices d,e,f,g. Specifically, by letting the β's be linear functions of some smaller set of basic parameters, the expected values in (8.2)-(8.12) will remain linear functions of these parameters, and the estimation can still be based on linear regression techniques. For instance we could assume that the β's are linear or quadratic functions of the indices or some scalar functions thereof, e.g. $\beta_d = \alpha + \beta d + \gamma d^2$ or $\beta_d = \alpha + \beta e^{-d}$. A next step would then be to develop test procedures for further reduction of the parameter space, e.g. to test whether γ can be deleted. The choice of parametrization and reduction hypotheses will depend on our a priori knowledge in each particular instance of application, and we shall not pursue these ideas further here.

All formulas for reserves by known parameters obtained in paragraph 8.2 are, of course, valid for any particular specification V, whereas the problem of parameter estimation will depend entirely on the assumptions made.

$\underline{B}^*$. In the present analysis based on the total claim amounts S_{jd} in the unrestricted framework model, the assumption that $\underset{\sim}{\pi}$ is independent of Ξ_j is of no significance. The parameter structure will remain unchanged if we drop this assumption.

9. Prediction based on total claim amounts when single claim amounts are not affected by fluctuations in basic risk conditions

9.1. Description of the case

A. The observable quantities $\underline{O}$ are those specified in 8.1.A.

B. In addition to the basic model assumptions I-IV it is now assumed that the single claim amounts are independent of variations in the basic risk conditions, that is, all Ψ_j are equal to some fixed parameter $\phi \in \underline{\Psi}$.

C. The B_{hjd}'s defined by (8.1) are now of the form

$$B_{hjd} = T_j \alpha_{hd} , \tag{9.1}$$

where

$$\alpha_{hd} = \pi_d \int y^h \, dG_d(y|\phi) \qquad ; \; h=1,\ldots,4 \; ; \; d=0,\ldots,D. \tag{9.2}$$

We put $\nu_h = ET_j^h$ as before, and introduce

$$\kappa_h = \nu_h / \nu_1^h \qquad ; \; h=2,3,\ldots \tag{9.3}$$

$$\eta_{hd} = \nu_1 \alpha_{hd} \qquad ; \; h=1,\ldots,4 \; ; \; d=0,\ldots,D. \tag{9.4}$$

Upon inserting (9.1) into the defining expressions in item 8.1.C, we find that the β's depend on the basic parameters in (9.3) and (9.4) as follows:

1st order parameters:

$$\beta_d = \eta_{1d} \qquad ; \; \forall d \; ; \tag{9.5}$$

2nd order parameters:

$$\beta_{de} = \kappa_2 \eta_{1d} \eta_{1e} \quad ; \ d \leqslant e \ ; \qquad (9.6)$$

$$\beta_{d^2} = \eta_{2d} \quad ; \ \forall d \ ; \qquad (9.7)$$

3rd order parameters*:

$$\beta_{def} = \kappa_3 \eta_{1d} \eta_{1e} \eta_{1f} \quad ; \ d \leqslant e \leqslant f \ ;$$

$$\beta_{de^2} = \begin{cases} \kappa_2 \eta_{1d} \eta_{2e} & ; \ d \neq e \ ; \\ 3\kappa_2 \eta_{1d} \eta_{2d} & ; \ d = e \ ; \end{cases}$$

$$\beta_{d^3} = \eta_{3d} \quad ; \ \forall d \ ;$$

4th order parameters*:

$$\beta_{defg} = \kappa_4 \eta_{1d} \eta_{1e} \eta_{1f} \eta_{1g} \quad ; \ d \leqslant e \leqslant f \leqslant g \ ;$$

$$\beta_{def^2} = \begin{cases} 6\kappa_3 \eta_{1d}^2 \eta_{2d} & ; \ d = e = f \ ; \\ 3\kappa_3 \eta_{1d} \eta_{1e} \eta_{2e} & ; \ d \neq e = f \ ; \\ \kappa_3 \eta_{1d} \eta_{1e} \eta_{2f} & ; \ d < e, \ d \neq f, \ e \neq f \ ; \end{cases}$$

$$\beta_{(de)^2} = \kappa_3 (\eta_{1d}^2 \eta_{2e} + \eta_{1e}^2 \eta_{2d}) \quad ; \ d < e \ ;$$

$$\beta_{d^2 e^2} = \kappa_2 \eta_{2d} \eta_{2e} \quad ; \ d < e \ ;$$

$$\beta_{de^3} = \begin{cases} \kappa_2 (3\eta_{2d}^2 + 4\eta_{1d} \eta_{3d}) & ; \ d = e \ ; \\ \kappa_2 \eta_{1d} \eta_{3e} & ; \ d \neq e \ ; \end{cases}$$

$$\beta_{d^4} = \eta_{4d} \quad ; \ \forall d \ .$$

The β's are now tied together as they are all functions of the basic parameters κ_h and η_{hd}. Define the order of a basic parameter to be h if it is uniquely determined by values of β's of order h or less and h is the smallest number with this property. Then, by inspection of the above table of β's, always starting from the top, we obtain the following classification of the basic parameters.

		Number of parameters
1st order parameters:		
η_{1d}	; d=0,...,D ;	D+1
h-th order parameters ; h=2,3,4 :		
κ_h , η_{hd}	; d=0,...,D .	D+2

In total there are $n_h(D) = h(D+2)-1$ basic parameters of order h or less. If, for instance, $D = 5$, the number of parameters that have to be estimated in order to establish the simplest reserves based on 1st and 2nd order parameters, is $n_2(5) = 13$. In the unrestricted model of section 8 we found $n_2(5) = 33$. This illustrates what can be gained by working into the model any a priori insight one might be in possession of.

9.2. Prediction by known parameters

A. All results established in subsection 8.2 carry over to the present case. However, due to the structure now possessed by the β's, the expressions simplify to closed formulas that are easy to interpret and compute.

By substitution of (9.5)-(9.7), $\underset{\sim}{\beta}$ and the parameter functions in (8.17) and (8.18) now assume the forms

$$\underset{\sim}{\beta} = \underset{\sim}{\eta}_1 = (\eta_{10},\ldots,\eta_{1D})', \tag{9.8}$$

$$\underset{\sim}{\gamma}'_{J-j} = \eta_{1,>J-j}(\kappa_2-1)\underset{\sim}{\eta}'_{1,\leq J-j}\ , \tag{9.9}$$

$$\underset{\sim}{\Sigma}_{j;J-j} = \{(\kappa_2-1)\underset{\sim}{\eta}_{1,\leq J-j}\underset{\sim}{\eta}'_{1,\leq J-j} + p_j^{-1}\mathrm{diag}(\eta_{2d})_{d\leq J-j}\}. \tag{9.10}$$

The matrix $\underset{\sim}{\Sigma}_{j;J-j}$ is easily inverted by aid of (A.25) in appendix A.4, and some simple calculations lead to

$$(\kappa_2-1)\underset{\sim}{\eta}'_{1\leq J-j}\ \underset{\sim}{\Sigma}^{-1}_{j;J-j}(\underset{\sim}{S}_j - p_j\underset{\sim}{\eta}_1)_{\leq J-j} = p_j(\tilde{Q}_j - 1), \tag{9.11}$$

where

$$\tilde{Q}_j = \frac{(\kappa_2-1)\Sigma_{d\leq J-j}\ S_{jd}\eta_{1d}/\eta_{2d} + 1}{(\kappa_2-1)p_j\Sigma_{d\leq J-j}\ \eta_{1d}^2/\eta_{2d} + 1}\ . \tag{9.12}$$

(It is easy to check that $\tilde{Q}_j$ is the credibility approximation of $Q_j = T_j/\nu_1$ based on $\underset{\sim}{S}_{j,\leq}$.) On inserting (9.8)-(9.10) into (8.19) and then substituting (9.11), we obtain the credibility formula

$$\tilde{S}_{j,>} = p_j\eta_{1,>J-j}\tilde{Q}_j\ , \tag{9.13}$$

where $\tilde{Q}_j$ is given by (9.12).

Formula (8.20), which is needed for reserving by principle (3.11), now reduces to

$$\mathrm{Var}\ S_{j,>} = p_j^2(\kappa_2-1)\eta_{1,>J-j}^2 + p_j\eta_{2,>J-j}\ .$$

<u>B</u>*. Next we turn to the results in item 8.2.C on linear approximation of predictive moments of 1st and 2nd order. Substituting the expressions in 9.1.C for the β's, we find that the quantities in (8.30) and (8.32) now become

$$\delta_j = p_j\kappa_2\eta^2_{1,>J-j} + \eta_{2,>J-j} ,$$

$$\underset{\sim}{\varepsilon}'_j = \{p_j \sum_{e,f>J-j}(\kappa_3\eta_{1d}\eta_{1e}\eta_{1f} - \eta_{1d}\kappa_2\eta_{1e}\eta_{1f})$$

$$+ \sum_{e>J-j}(\kappa_2\eta_{1d}\eta_{2e} - \eta_{1d}\eta_{2e})\}_{d\leqslant J-j}$$

$$= \{p_j(\kappa_3-\kappa_2)\eta^2_{1,>J-j} + (\kappa_2-1)\eta_{2,>J-j}\}\underset{\sim}{\eta}'_{1,\leqslant J-j}. \qquad (9.14)$$

By inspection of (8.33), (9.14), and (9.11), it is seen that the quantity in (9.12) once more plays a key role. We easily gather that

$$\tilde{E}(S^2_{j,>}|\underset{\sim}{S}_{j,\leqslant}) = p_j(\kappa_2-1)^{-1}[p_j(\kappa_2^2-\kappa_3)\eta_1^2$$
$$+ \{p_j(\kappa_3-\kappa_2)\eta^2_{1,>J-j} + (\kappa_2-1)\eta_{2,>J-j}\}Q_j]. \qquad (9.15)$$

Upon entering (9.13) and (9.15) into (8.34) we find an expression for $\tilde{M}_{2j}$, which together with $\tilde{S}_{j,>}$ from (9.13) deliver a reserve by principle (3.12).

<u>C</u>*. Proceeding as in item B above, we find that (8.37) and (8.39) in the present case reduce to

$$\rho_j = p_j^2\kappa_3\eta^3_{1,>J-j} + 3p_j\kappa_2\eta_{1,>J-j}\eta_{2,>J-j} + \eta_{3,>J-j} ,$$

$$\underset{\sim}{\sigma}'_j = \{p_j^2(\kappa_4-\kappa_3)\eta^3_{1,>J-j} + 3p_j(\kappa_3-\kappa_2)\eta_{1,>J-j}\eta_{2,>J-j}$$
$$+ (\kappa_2-1)\eta_{3,>J-j}\}\underset{\sim}{\eta}'_{1,\leqslant J-j} .$$

Substitution of these expressions in (8.40) yields

$$\tilde{E}(S^3_{j,>}|\underset{\sim}{S}_{j,\leqslant}) = p_j(\kappa_2-1)^{-1}[p_j\{p_j(\kappa_3\kappa_2-\kappa_4)\eta^3_{1,>J-j}$$
$$+ 3(\kappa_2^2-\kappa_3)\eta_{1,>J-j}\eta_{2,>J-j}\} + \{p_j^2(\kappa_4-\kappa_3)\eta^3_{1,>J-j} \qquad (9.16)$$
$$+ 3p_j(\kappa_3-\kappa_2)\eta_{1,>J-j}\eta_{2,>J-j} + (\kappa_2-1)\eta_{3,>J-j}\}Q_j].$$

The elements in (8.41) are now given by (9.13), (9.15), and (9.16).

9.3. Parameter estimation

A. A class of consistent estimators of the 1st and 2nd order parameters appearing on the right of (9.5)-(9.7) is given by

$$\eta^{\star}_{1d} = \beta^{\star}_{d} , \tag{9.17}$$

$$\kappa^{\star}_{2} = \sum_{d<e} w_{de}\beta^{\star}_{de} \Big/ \sum_{d<e} w_{de}\beta^{\star}_{d}\beta^{\star}_{e} , \tag{9.18}$$

$$\eta^{\star}_{2d} = \beta^{\star}_{d2} , \tag{9.19}$$

where the $\beta^{\star}$'s are picked from (8.42)-(8.45) and the w_{de} are weights that sum to 1. The estimators (9.17) and (9.19) are trivially motivated by (9.5) and (9.7). The estimator (9.18) is obtained by inserting estimators for all parameters in (9.6), save κ_2, and forming a weighted sum.

B. A more refined procedure than the one proposed in item A would be to apply weighted least squares techniques to the non-linear regressions

$$ES_{jd} = p_j\eta_{1d} ,$$

$$E(S_{jd}S_{je}) = p_j^2\kappa_2\eta_{1d}\eta_{1e} + \delta_{de}p_j\eta_{2d} \qquad ; \quad 0\leqslant d\leqslant e\leqslant D.$$

Also higher order moments can be estimated by methods similar to those presented here. We shall not dwell upon the question of optimality properties of estimators.

9.4. Comments

The assumption that claim amounts are not affected by variations in basic risk conditions, may be judged as not fully realistic in a given situation. It is nevertheless of interest as an approximation hypothesis; it provides a means of a substantial reduction of the parameter space.

10. Prediction based on total claim amounts by permanent risk conditions

10.1. Description of the case

A. The statistical basis is still the one defined in 8.1.A.

B. We now assume that the basic risk conditions are not subject to fluctuations from one year to another, that is, we drop the basic assumption I and assume that all (T_j, Ψ_j)'s are equal to some constant (τ, ψ). Thus the model is the same as in section 5.

C. Putting $T_j = \tau$ into the formulas in item 9.1.C, we find that all the κ_h's in (9.3) now become equal to 1, whereas the η_{hd}'s in (9.4) become

$$\eta_{hd} = \tau\alpha_{hd} \quad ; \quad h=1,2,3 \quad ; \quad d=0,\ldots,D \; ; \qquad (10.1)$$

with α_{hd} defined by (9.2); it turns out that the fourth order parameters η_{4d} are no longer needed. The η_{hd}'s in (10.1) are now the basic parameters of the model; in total there are $3(D+1)$ of them.

The formulas for the β's displayed in 9.1.C are still valid, only that all the κ_h's are equal to 1.

10.2. Prediction by known parameters

As the exact distribution of the S_{jd}'s is unknown, we have to resort to reserving methods that utilize only some first moments The central moments up to third order of S_{jd} are easily seen to be

$$\mu_{hj} = p_j\eta_{h,>J-j} \quad ; \quad h=1,2,3. \qquad (10.2)$$

The formula (10.2) may be picked e.g. from (5.2) by translation to the present parametrization.

Reserves may now be constructed by any of the principles (3.5), (3.6), and (3.9) upon replacing the M_{hj}'s by the μ_{hj}'s in (10.2).

10.3. Parameter estimation

Estimation of the parameters in (10.1) is straightforward by moment methods along the same lines as in paragraph 8.3. In the present case everything becomes simpler, of course, and we skip the details.

10.4. Comments

The present model is well structured, with a small number of parameters, and represents, therefore, one interesting answer to the problem discussed in item 8.4.A. On the other hand, it is clear that the present model is not suitable in situations where fluctuations in basic risk conditions may contribute substantially to the total risk, as is likely to be the case e.g. in product liability insurance and marine insurance.

11. Prediction based on total claim amounts as per accounting year in the unrestricted framework model

11.1. Description of the case

A. In this final case to be studied we shall discuss briefly a situation met with in a number of lines of reinsurance, where the only statistics are the total claims paid in each accounting year. Thus $\underset{\sim}{O} = \{S_j;\ j=1,\dots,J\}$, where

$$S_j = \sum_{h=j-D}^{j} S_{h,j-h} \quad . \tag{11.1}$$

Roughly speaking, the past is observed along the columns in figure 3, and only through total sums. The upper left triangle in the figure should now be included in the statistical basis.

B. The analysis will be based on the unrestricted framework model specified in item 8.1.B.

C. Since the statistical basis (11.1) is far more summary than that in section 8, the necessity of a parsimonious specification of the model is now even more pressing. In practical applications one will have to reduce the parameter space, either by introducing the assumptions of section 9 or some assumptions of the kind mentioned in item 8.4.A, or a combination of the two. At any rate, the general formulas below will remain valid in all special cases.

11.2. Prediction by known parameters

The moments needed in a credibility predictor of a future S_m; $J<m\leq J+D$; based on S_j; $j=m-D,\dots,J$; are

$$ES_k = \sum_{h=k-D}^{k} p_h\beta_{k-h} \qquad ; \quad k=m-D,\dots,J,m; \tag{11.2}$$

$$\mathrm{Cov}(S_\ell, S_k) = \sum_{h=\ell-D}^{k} \{p_h^2(\beta_{k-h,\ell-h} - \beta_{k-h}\beta_{\ell-h}) + \delta_{\ell k} p_h \beta_{(k-h)2}\}; \qquad (11.3)$$

$$k \leq \ell;\ k,\ell = m-D,.,J,m;$$

where the parameters on the right are defined in item 8.1.C.

The credibility predictor is now obtained from the general formula (A.16) in appendix A.3, with $M = S_m$ and $\underset{\sim}{X} = (S_{m-D},\ldots,S_J)'$. The expressions become messy and shall not be displayed here.

11.3. Parameter estimation

In principle the 1st and 2nd order parameters can be estimated by moment methods based on (11.2) and (11.3). It is, however, of limited interest to carry through this analysis in the full framework model, confer the remark in item 11.1.C above.

11.4. Comments

Our description of the data in item 11.1.A was intentionally superficial at one point: the definition of the S_j's implies that the statistical basis rhombe in figure 3 be extended to a rectangle by including the upper left triangle. Actually this is the typical form of the data in so-called "short cut" reinsurance business kept on an accounting year basis. But then the problem arises, does the reinsurer really know the volumes p_j for the years $j = -D+1,\ldots\ldots,-1$, which are needed in the analysis of the observations $S_1,\ldots,S_{D-1}$? The answer is likely to be "no". In fact, the p_j's are usually not directly observed at all; one will only have access to the more summary "earned premium" in each accounting year.

The problems pointed out here and in item 11.1.C suggest that the present case may put a limit to the practical applicability of the micro-theory approach advocated in the present work. It may be that some cruder "non-explaining fit-model" is more apt in situations with scanty data and little knowledge about the underlying processes.

12. A view to related literature

12.1. Models with nonrandom basic risk conditions

A. Provisions for IBNR claims have been established by accountants long before mathematical models were created for the purpose. The multifarious attempts of today's actuaries to forecast IBNR-liabilities by aid of stochastic models seem to have their origin in papers by Verbeek (1972) and Straub (1972).

Verbeek (1972) treats only the numbers of claims and assumes that $K_{jd} \sim Po(\lambda_{j+d}\mu_d)$. Here the λ_{j+d}'s and μ_d's are fixed parameters, which are estimated by the maximum likelihood method.

Verbeek's multiplicative model is extended and applied to the total claim amounts by Taylor (1977), who assumes (the present author's interpretation) that the conditional expected value of S_{jd}, given the total number $K_{j,\leq D}$ of claims occurred in year j, is of the form $\kappa_j\lambda_{j+d}\mu_d$.

Verbeek's contemporaries Kramreiter and Straub (1973) (see also Straub (1972)) start from the total claim amount $S_{j,\leq d}$ of year j as known by the end of development year d; $d=0,\dots,D$; $j=1,2,\dots$ Under various assumptions about the 1st and 2nd order moments of these quantities they predict R (essentially) by the unbiased homogeneous linear function $\Sigma_{j=1}^{J}\Sigma_{d=0}^{D(j)} a_{jd}S_{j,\leq d}$ that minimizes the expected squared error. Their framework model specifies that the moments are of the form $ES_{j,\leq d} = p_j\alpha_d$ and $Cov(S_{j,\leq d}, S_{j,\leq e}) = p_j\alpha_{de}$, which accords with the model in sections 5 and 10 above. The number of parameters is $(D+1)(D/2+2)$. Further structure is added by assuming that $S_{j,\leq d+1} = \Lambda_{j,d+1}S_{j,\leq d} + p_j\Delta_{j,d+1}$, where $(\Lambda_{jd},\Delta_{jd})$; $j=1,2,\dots$; are i.i.d., and all Λ_{jd}'s, Δ_{jd}'s, and S_{j0}'s are mutually independent. The number of parameters is then reduced to $4D+2$.

B. Hoem (1973) analyses a model that comes out of the one in section 4 above if the total numbers of claims occurred in each year are regarded as fixed parameters, which essentially means that he operates in the conditional model, given $K_{j,\leq D}$; j=1,2,... In that it specifies assumptions about the joint distribution of the time lapse between occurrence and notification and the claim size for each single claim, Hoem's work is a pioneering one in the tradition of micro-modelling IBNR claims.

A more ambitious attempt in the same direction is made by Bühlmann, Schnieper, and Straub (1980). They treat the problem of claims reserving in its entirety, modelling the frequency of claims and - for each single claim -the time lapse from occurrence until notification and the succeeding stream of payments up to final settlement. As far as pure IBNR-aspects are concerned, their model is the one for permanent risk conditions studied in section 5 above, extended with an assumption of exponential monetary inflation. Their supplementary set of assumptions as to how the amounts Y_{jdk} of claims reported in development year d decompose into amounts Y_{jdke} (say) paid in development years $e = d, d+1, \ldots$ is only one among a number of possibilities.

Yet another notable contribution in the vein of micro-theory is Reid's (1978) paper.

12.2. Models with random basic risk conditions

A. In all works reviewed above the basic risk conditions are represented by fixed parameters, either invariable over time as in sections 5 and 10 of the present paper, or depending on occurrence year as in section 4. However, already Verbeek (1972) remarks that there may be reasons to prefer a variable K_{jd} having a fluctuating basic probability structure. He abstains from such a model in view of the generally small numbers of observations available. It is pertinent to recall here that in going from the model of section 5 with a fixed frequency parameter τ to the model of paragraph 6 with gamma-distributed T_j's , the number of parameters is increased only by 1.

B. The idea of representing fluctuating basic risk conditions by random variables, which is at the base of the present work, is not new to actuaries. It was brought into the context of IBNR claims reserving by de Vylder (1982). He assumes that the vectors $\underset{\sim}{S}_j = (S_{j0},\ldots,S_{jD})'$ are of the form $\Psi_j \underset{\sim}{S}_j^0$, where the Ψ_j's are i.i.d. and independent of the $\underset{\sim}{S}_j^0$'s, which are also mutually independent with common expected value $\underset{\sim}{v}$ and covariance matrices of the form

$$\text{Var}\ \underset{\sim}{S}_j^0 = p_j \rho \underset{\sim}{I} \qquad ; \ j=1,2,\ldots \qquad (12.1)$$

We arrive at this model if we assume that the between years fluctuations in basic risk conditions affect only the claim sizes through latent "claims cost indexes" Ψ_j and that the "deflated" total claim amounts $\underset{\sim}{S}_j^0$ are independent of the Ψ_j's and have moments as specified above.

The assumption (12.1) implies that all S^0_{jd} ; $d=0,\ldots,D$; are equally variable. Although mathematically convenient, this assumption is hardly appropriate as an a priori description of the IBNR-process. A reasonable way of relaxing assumption (12.1) could be to replace $\rho \underset{\sim}{I}$ by $\text{diag}(\rho_0,\ldots,\rho_D)$. In order to limit the number of parameters, the ρ_d's could be taken as some simple parametric functions of d, e.g. $\rho_d = \alpha + \beta d$.

Also in the recent work by de Jong and Zehnwirth (1983) basic risk conditions are represented by random quantities, viz. as a stochastic process in the framework of Kalman-filtering. In other respects, however, their angle of attack is quite different from the one of the present paper; instead of composing a micro-theory from some conceptions of the evolution of the claim process, they fit a model that, hopefully, is sufficiently flexible to reflect the main features of the process.

12.3. Further references

Extensive surveys of works on claims reserving are given by van Eeghen (1981) and Taylor (1983).

13. Some final comments on the theory and suggested issues for further study

A. One can easily think of circumstances that may influence the IBNR-development in some lines of insurance and that have not been taken into account in the basic model I-IV. This problem is universal. A model is not an attempt to describe all features of a phenomenon in their right proportions; modelling necessarily means magnifying some features and leaving others out, and a good model is one that magnifies the essentials and neglects the less important details.

It is the intent of this section to indicate some possible ways of extending the model I-IV to make it more realistic. Throughout we must, however, keep in mind what has been emphasized repeatedly in the previous discussions, that improved realism can only be gained at the sacrifice of model parsimony, that is, by increasing the number of parameters.

B. One obvious way of introducing more flexibility in the basic model is to let the probability distribution $\underset{\sim}{\pi}$ depend on Ξ_j , thus allowing for a dependence between the number and type of claims and their pattern of development. (As was pointed out already in item 8.4.B, this relaxation of our assumptions would not change the structure of the moments of the S_{jd}'s in the unrestricted framework model. In other cases it may, however, complicate matters a great deal.) A first attempt in this direction could be to replace $\underset{\sim}{\pi}T_j$ by a random vector $\underset{\sim}{\Lambda}_j$ and, possibly, add some assumptions about the moments, e.g. that the components Λ_{jd} are independent and have expected values that are simple parametric functions of d.

C. The reader may have observed that there is a lack of symmetry in our presentation of the different cases; the unrestricted framework model has been analysed only in conjunction with the statistical basis consisting of the S_{jd}'s and not with the complete records on numbers of claims and single claim amounts. It is an issue for further studies to find a specification of the joint distribution of T_j and Ψ_j that yields a tractable analysis in the latter case. We are here facing the old problem of credibility for severity treated by Hewitt (1970), Jewell (1973), and Bühlmann (1974), only more complex due to the inclusion of IBNR-effects.

As a first step one could consider the case where T_j and Ψ_j are independent. (In passing we note that this assumption would not bring about any simplification of the parameter structure of the $\underset{\sim}{S}_j$'s, given in paragraph 8.1.)

A pragmatic way of circumventing the severity problem in practice would be simply to employ the reserving formulas in section 8 based on the total claim amounts S_{jd}, deliberately sacrificing the details of information contained in the K_{jd}'s and Y_{jdk}'s.

D. We can, of course, not bring our discussions to a decent conclusion without having commented on inflation, a pet subject of people concerned with IBNR-problems.

It is the present author's firm opinion that, if it can be avoided, inflationary effects should not be worked into the model. When inflation can be reasonably well determined from exogeneous sources, like index numbers of prices, then one should apply the analyses presented above to the price adjusted quantities. In the present context of claims reserving one would then, of course, have to make a skilled guess concerning the future development of prices.

However, in some lines of insurance the level of claim costs may develop more or less independently of general price indexes. For instance, liability insurance claims may be subject to a special inflationary effect caused by a trend towards more victim-oriented judicial decisions. In such cases it may be necessary to model the mechanism of inflation and to estimate it endogeneously from the claim statistics itself.

Assume now that only the individual claim amounts are affected by inflation (in the liability insurance example one could imagine that also the number of claims is shoved up by a changed court ruling). A simple way of modelling inflation is to introduce a price index ω_j ; $j=1,2,\ldots$; and assume that the deflated amounts $Y'_{jdk} = Y_{jdk}/\omega_j$; $j=1,2,\ldots$; $k=1,2,\ldots$; are i.i.d. $\sim G_d$; $d=0,\ldots,D$. Further simplification is attained by letting ω_j be described by some simple parametric function of j, e.g. $\omega_j = \omega^j$ or $\omega_j = \omega' + \omega''j$. Then one can still estimate the distributions G_d and the ω_j's by traditional methods for location/scale models.

Consider now reinsurance, where the S_{jd}'s are the only observable claim statistics. Then the p_j's will typically be premium incomes, and it is reasonable to assume that they follow the same pattern of inflation as the S_{jd}'s. More specifically, we assume that the deflated amounts $S'_{jd} = S_{jd}/\omega_{j+d}$ and $p'_j = p_j/\omega_j$ satisfy the framework model assumptions and thus have the moment structure given in paragraph 8.1, with $\beta_d, \beta_{de}, \ldots$ replaced by $\beta'_d, \beta'_{de}, \ldots$, say. It is easily seen that in the case of exponential growth of inflation, $\omega_j = \omega^j$, also the nominal quantities S_{jd} and p_j will fit into the moment structure in paragraph 8.1; the price indexes will be absorbed into the parameters β. In this case, therefore, we do not have to be much concerned with the inflation problem.

References

Beard, R., Pentikäinen, T., and Pesonen, E. (1984). Risk Theory, 3rd edition. Chapman and Hall, London.

Bühlmann, H. (1974). A comparison of three credibility formulae using multidimensional techniques. ASTIN Bull. 7, 203-207.

Bühlmann, H., Schnieper, R., and Straub, E. (1980). Claims reserves in casualty insurance based on a probabilistic model. Mitt. Ver. Schweiz. Vers. Math. 80, 21-45.

De Jong, P. and Zehnwirth, B. (1983). Claims reserving, state-space models and the Kalman filter. J. Inst. Act. 110, 157-181.

De Vylder, F. (1982): Estimation of IBNR claims by credibility theory. Insurance: Math. Econ. 1, 35-40.

Hewitt, C.C., Jr. (1970). Credibility for severity. Proc. Cas. Actuarial Soc. 57, 148-171. Discussion in Proc. Cas. Actuarial Soc. 58, 25-31.

Hoem, J.M. (1973). Fondsdannelse i skadeforsikring. Chapter VII in Skadeforsikringsmatematikk, lecture notes, Inst. of Math., Univ. of Oslo. (In Norwegian.)

Jewell, W.S. (1973). Multi-dimensional credibility. ORC 73-7, Operations Research Center, University of California, Berkeley.

Kramreiter, H. and Straub, E. (1973). On the calculation of IBNR reserves II. Mitt. Ver. Schweiz. Vers. Math. 73, 177-190.

Neuhaus, W. (1985). Choice of statistics in linear Bayes estimation. Scand. Actuarial J. 1985, 1-26.

Norberg, R. (1980). Empirical Bayes credibility. Scand. Actuarial J. 1980, 177-194.

Norberg, R. (1982). On optimal parameter estimation in credibility. Insurance: Math. Econ. 1, 73-89.

Panjer, H.H. (1981). Recursive calculation of a family of compound distributions. ASTIN Bull. 12, 22-26.

Reid, D.H. (1978). Claim reserves in general insurance (with discussion). J. Inst. Act. 105, 211-296.

Straub, E. (1971). On the calculation of IBNR reserves. Prize-winning papers in the Boleslaw-Monic-Fund Competition 1971. Nederlandse Reassurantie Groep N.V., Amsterdam 1972.

Taylor, G.C. (1977). Separation of inflation and other effects from the distribution of non-life insurance claim delays. ASTIN Bull. 9, 219-230.

Taylor, G.C. (1983). A survey of claims analysis in non-life insurance. E.S. Knight & Co., 71 York Street, Sydney N.S.W. 2000, Australia.

Van Eeghen, J. (1981). Loss reserving methods. Surveys of actuarial studies nr. 1. Nationale-Nederlanden N.V.

Verbeek, H.G. (1972). An approach to the analysis of claims experience in motor liability excess of loss reinsurance. ASTIN Bull. 6, 195-202.

Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics. Wiley, New York.

Appendix

A.1. Relations between moments

Let X be a real random variable. Provided they exist, denote by λ_h and μ_h the noncentral and central moment of order h, that is,

$$\lambda_h = EX^h \qquad ; \; h=1,2,\dots ;$$

$$\mu_1 = EX \; , \quad \mu_h = E(X-\mu)^h \qquad ; \; h=2,3,\dots$$

By definition, $\lambda_1 = \mu_1$. Furthermore, the moments up to third order are related by the following identities, which are easily verified:

$$\lambda_2 = \mu_2 + \mu_1^2 \; , \tag{A.1}$$

$$\lambda_3 = \mu_3 + 3\mu_2\mu_1 + \mu_1^3 \; , \tag{A.2}$$

$$\mu_2 = \lambda_2 - \lambda_1^2 \; , \tag{A.3}$$

$$\mu_3 = \lambda_3 - 3\lambda_2\lambda_1 + 2\lambda_1^3 \; , \tag{A.4}$$

$$\mu_4 = \lambda_4 - 4\lambda_3\lambda_1 + 6\lambda_2\lambda_1^2 - 3\lambda_1^4 . \tag{A.5}$$

A.2. Some properties of Poisson distributions

Assume that $K \sim Po(\tau)$, that is,

$$P(K=k) = \frac{\tau^k}{k!} e^{-\tau} \qquad ; \; k=0,1,\dots$$

The h-th factorial of K is the product $K^{(h)} = K(K-1)\dots(K-h+1)$; $h=0,1,\dots$ The h-th factorial moment of K is

$$EK^{(h)} = \sum_{k=0}^{\infty} k^{(h)} \frac{\tau^k}{k!} e^{-\tau}$$

$$= \tau^h \sum_{k=h}^{\infty} \frac{\tau^{k-h}}{(k-h)!} e^{-\tau}$$

$$= \tau^h. \qquad (A.6)$$

Assume that $Y_1, Y_2, \ldots$ are i.i.d. ~ G and that they are independent of K. The random variable

$$S = \sum_{k=1}^{K} Y_k ,$$

which is defined as 0 when K = 0, has a generalized Poisson distribution, and we write $S \sim g.Po(\tau, G)$.

Assume that G possesses finite moments up to order 4, and put

$$\alpha_h = \tau \int y^h dG(y) \qquad ; h=1,2,3,4 . \qquad (A.7)$$

Then the first four moments of S exist and are given by

$$ES = \alpha_1 , \qquad (A.8)$$

$$ES^2 = \alpha_1^2 + \alpha_2 , \qquad (A.9)$$

$$ES^3 = \alpha_1^3 + 3\alpha_1\alpha_2 + \alpha_3 , \qquad (A.10)$$

$$ES^4 = \alpha_1^4 + 6\alpha_1^2\alpha_2 + 3\alpha_2^2 + 4\alpha_1\alpha_3 + \alpha_4 , \qquad (A.11)$$

or, by use of (A.3)-(A.5) and (A.8)-(A.11),

$$\text{Var } S = \alpha_2 , \qquad (A.12)$$

$$E(S-ES)^3 = \alpha_3 , \qquad (A.13)$$

$$E(S-ES)^4 = \alpha_4 + 3\alpha_2^2 . \qquad (A.14)$$

To prove (A.11), write

$$ES^4 = E(\sum_{k=1}^{K} Y_k)^4$$

$$= EE(\sum_{k,\ell,m,n} Y_k Y_\ell Y_m Y_n | K),$$

where the latter sum ranges over all k,ℓ,m,n between 1 and K. In this sum there are $K^{(4)}$ terms of the form $Y_k Y_\ell Y_m Y_n$ with $k,\ell,m,$ and n all different, $(4!/2!1!1!)K^{(3)}$ terms of the form $Y_k Y_\ell Y_m^2$ with $k,\ell,$ and m all different, $\binom{4}{2}\binom{K}{2} = 3K^{(2)}$ terms of the form $Y_k^2 Y_\ell^2$ with $k \neq \ell$, $4K^{(2)}$ terms of the form $Y_k Y_\ell^3$ with $k \neq \ell$, and K terms of the form Y_k^4. Thus, since the Y_k's are independent of K,

$$ES^4 = E\{K^{(4)}E^4Y + 6K^{(3)}E^2Y\ EY^2 + 3K^{(2)}(EY^2)^2 + 4K^{(2)}EYEY^3 + K\ EY^4\},$$

and by use of (A.6) we arrive at (A.11). The expressions in (A.8)-(A.10) are obtained by similar arguments, only simpler.

Convolutions of generalized Poisson distributions are generalized Poisson; if $S_1,\ldots,S_n$ are independent random variables, and $S_i \sim g.Po(\tau_i, G_i)$; $i=1,\ldots,n$; then

$$\left.\begin{aligned} &\sum_{i=1}^{n} S_i \sim g.Po(\tau, G), \\ &\text{with} \\ &\tau = \sum_{i=1}^{n} \tau_i\ ,\quad G = \tau^{-1}\sum_{i=1}^{n} \tau_i G_i\ . \end{aligned}\right\} \qquad \text{(A.15)}$$

A.3. Linear predictors and credibility formulas

Let M be a real random variable and $\underset{\sim}{X}$ a random column vector of dimension n, both assumed to be square integrable. Consider the class of inhomogeneous linear functions of $\underset{\sim}{X}$,

$$\underline{M} = \{\check{M} = g_0 + \underset{\sim}{g}'\underset{\sim}{X}\ ;\ g_0 \in \underline{R},\ \underset{\sim}{g} \in \underline{R}^n,\ g_0 \text{ and } \underset{\sim}{g} \text{ constants}\}.$$

The element in $\underline{M}$ that minimizes $E(\check{M}-M)^2$ is

$$\tilde{M} = EM + Cov(M,\underset{\sim}{X}')(Var\underset{\sim}{X})^{-1}(\underset{\sim}{X}-E\underset{\sim}{X}). \tag{A.16}$$

For a proof of (A.16), see e.g. Norberg (1980).

If X is realvalued and $M = E(X|\Xi)$ for some random element Ξ, then (A.16) assumes the form of a credibility weighted mean,

$$\tilde{M} = \zeta X + (1-\zeta)EX, \tag{A.17}$$

where the credibility weight ζ is given by

$$\zeta = Var\ E(X|\Xi)/Var\ X.$$

For each $i=1,\dots,I$ let M_i be a square integrable real random variable and $\tilde{M}_i$ its $\underline{M}$-approximation defined by replacing M by M_i in (A.17). By the linearity of the operators E and Cov, it follows that the best $\underline{M}$-approximation of $M = \sum_i M_i$ is

$$\tilde{M} = \sum_i \tilde{M}_i . \tag{A.18}$$

A.4. Two results on matrices

$\underline{A}$. Let $\underset{\sim}{A}$ be a nonsingular $n\times n$ matrix. Decompose $\underset{\sim}{A}$ and its inverse into

$$\underset{\sim}{A} = \begin{pmatrix} \underset{\sim}{A}_{11} & \underset{\sim}{A}_{12} \\ \underset{\sim}{A}_{21} & \underset{\sim}{A}_{22} \end{pmatrix}, \qquad \underset{\sim}{A}^{-1} = \begin{pmatrix} \underset{\sim}{A}^{11} & \underset{\sim}{A}^{12} \\ \underset{\sim}{A}^{21} & \underset{\sim}{A}^{22} \end{pmatrix}, \tag{A.19}$$

where $\underset{\sim}{A}_{11}$ and $\underset{\sim}{A}^{11}$ are of order $p\times p$ $(p<n)$. On inserting the

right hand side expressions in (A.19) into the defining relation $\underset{\sim}{A}\underset{\sim}{A}^{-1} = \underset{\sim}{I}$ and multiplying blockwise, we easily obtain

$$\underset{\sim}{A}^{11} = (\underset{\sim}{A}_{11} - \underset{\sim}{A}_{12}\underset{\sim}{A}_{22}^{-1}\underset{\sim}{A}_{21})^{-1} , \quad (A.20)$$

$$\underset{\sim}{A}^{21} = - \underset{\sim}{A}_{22}^{-1}\underset{\sim}{A}_{21}\underset{\sim}{A}^{11} , \quad (A.21)$$

$$\underset{\sim}{A}^{22} = (\underset{\sim}{A}_{22} - \underset{\sim}{A}_{21}\underset{\sim}{A}_{11}^{-1}\underset{\sim}{A}_{12})^{-1} , \quad (A.22)$$

$$\underset{\sim}{A}^{12} = - \underset{\sim}{A}_{11}^{-1}\underset{\sim}{A}_{12}\underset{\sim}{A}^{22} . \quad (A.23)$$

B. Let $\underset{\sim}{A}$ be a nonsingular n×n matrix, D a p×p matrix, and $\underset{\sim}{B}$ an n×p matrix. If $\underset{\sim}{A}^{-1}$ has already been calculated and p is much smaller than n, then the matrix $(\underset{\sim}{A} + \underset{\sim}{B}\underset{\sim}{D}\underset{\sim}{B}')^{-1}$, whenever it exists, can conveniently be calculated by use of the classic identity

$$(\underset{\sim}{A} + \underset{\sim}{B}\underset{\sim}{D}\underset{\sim}{B}')^{-1} = \underset{\sim}{A}^{-1} - \underset{\sim}{A}^{-1}\underset{\sim}{B}\underset{\sim}{D}(\underset{\sim}{I} + \underset{\sim}{B}'\underset{\sim}{A}^{-1}\underset{\sim}{B}\underset{\sim}{D})^{-1}\underset{\sim}{B}'\underset{\sim}{A}^{-1} . \quad (A.24)$$

In particular, when $p = 1$, $\underset{\sim}{D} = -1$, and $\underset{\sim}{A}$ is symmetric, (A.24) reduces to

$$\begin{aligned}(\underset{\sim}{A} - \underset{\sim}{b}\underset{\sim}{b}')^{-1} &= \underset{\sim}{A}^{-1} + (1 - \underset{\sim}{b}'\underset{\sim}{A}^{-1}\underset{\sim}{b})^{-1}\underset{\sim}{A}^{-1}\underset{\sim}{b}\underset{\sim}{b}'\underset{\sim}{A}^{-1} \\ &= \underset{\sim}{A}^{-1} + (1 - \underset{\sim}{b}'\underset{\sim}{d})^{-1}\underset{\sim}{d}\underset{\sim}{d}' ,\end{aligned} \quad (A.25)$$

with

$$\underset{\sim}{d} = \underset{\sim}{A}^{-1}\underset{\sim}{b} .$$

For the sake of completeness, and in the absence of a suitable reference, we prove (A.24). Put

$$\underset{\sim}{C} = (\underset{\sim}{A} + \underset{\sim}{B}\underset{\sim}{D}\underset{\sim}{B}')^{-1} .$$

By definition, we have

$$(A + BDB')C = I ,$$

which is equivalent to

$$C + A^{-1}BDB'C = A^{-1}. \qquad (A.26)$$

Premultiply in (A.26) by B' to get

$$B'C + B'A^{-1}BDB'C = B'A^{-1}$$

or, equivalently,

$$B'C = (I + B'A^{-1}BD)^{-1}B'A^{-1}. \qquad (A.27)$$

Substituting (A.27) back into (A.26), we arrive at (A.24). To complete the proof, it remains to establish that the matrix $I + B'A^{-1}BD$ is invertible if and only if $A + BDB'$ is, which is equivalent to asserting that their determinants vanish simultaneously. This follows by use of the identity $|I + MN| = |I + NM|$ (see e.g. Zellner, 1971, p.231), which gives $|A + BDB'|$ $= |A||I + A^{-1}BDB'| = |A||I + B'A^{-1}BD|$.