JIA 116 (1989) 513-528

TIME SERIES MODELS FOR INSURANCE CLAIMS

BY PROFESSOR A. C. HARVEY, B.A., M.Sc. AND C. FERNANDES
(of the Department of Statistics, London School of Economics and Political Science)

[Presented at the Seminar 'Applications of Mathematics in Insurance, Finance and Actuarial Work,' sponsored by the Institute of Mathematics and Its Applications, the Institute of Actuaries, and the Faculty of Actuaries, held at the Institute of Actuaries, 6–7 July 1989.]

ABSTRACT

The distribution of insurance claims in a given time period is usually regarded as a random sum. This paper sets up a time series model for the value of the claims and combines it with a model for the number of claims. Thus past observations can be used to make predictions of future values of the random sum, and the overall model ensures that they are discounted appropriately. It is shown that explanatory variables can be introduced into the model, and how it can be extended to handle several groups. The general approach is based on the recently developed structural time series methodology.

1. INTRODUCTION

A fundamental problem in insurance is predicting the distribution of the total value of claims in a given time period. Similar problems arise elsewhere. For example, we may be interested in the total expenditure on some category of consumer durables, such as cars, for a given group of the population. The essential feature of such problems is that the quantity of interest is a *random sum*.

In order to aid the exposition, it will be assumed that we are working in an insurance context. Let Y_{jt} denote the amount of the j-th claim at time t, where $j = 1, \ldots, N_t$ and $t = 1, \ldots, T$. The number of time periods for which observations are available is T, while the number of claims at time t, N_t, is, like Y_{jt}, a random variable. The total value of claims is the random sum

$$S_t = \sum_{j=1}^{N_t} Y_{jt}, \qquad t = 1, \ldots, T. \tag{1.1}$$

If the sizes of claims are mutually independent, and independent of the number of claims, the distribution function of S_t is given by

$$F(S_t) = \sum_{N_t=0}^{\infty} F^{N_t*}(Y_t)p(N_t) \tag{1.2}$$

where $F(Y_t)$ is the distribution function of the claims, $p(N_t)$ is the probability that the number of claims is N_t and the * denotes the convolution of N_t variables. Obtaining an analytic expression for $F(S_t)$ is not usually possible, except in a few very special cases. However, the moment generating function (MGF) of S_t can

always be obtained from the MGFs of Y_{jt} and N_t; see Gerber (1979),[5] page 12. Furthermore, it is not difficult to show that

$$E(S_t) = E(N_t) \cdot E(Y_{jt}), \tag{1.3}$$

as might be expected, while

$$\text{Var}(S_t) = \text{Var}(N_t) \cdot \{E(Y_{jt})\}^2 + E(N_t) \cdot \text{Var}(Y_{jt}). \tag{1.4}$$

There is a considerable literature in risk theory concerned with obtaining approximations to the distribution of S_t.

Estimating the parameters of the distribution of S_t requires that past data be used to estimate the parameters of the distributions of the size and number of claims. As a rule it is assumed that these parameters remain constant over time, although it has long been recognized that this may not be realistic; see, for example, the comments in Beard *et al.* (1984)[1] Ch. 6. This paper sets out methods for constructing and estimating time series models for the size and number of claims. Section 3 looks at a single group, while Section 4 extends the analysis to several groups. Section 2 reviews recent techniques in time series modelling which are appropriate for handling Gaussian and Poisson observations. It is these methods which form the basis for the statistical development in the later sections. In making these developments it is recognized that the number of time periods for which observations are available may be relatively small.

2. STRUCTURAL TIME SERIES MODELS

A structural time series model is one which is set up in terms of components of interest, such as trends, seasonals and cycles. Models of this kind are related to the ARIMA models popularized by Box and Jenkins (1976),[2] but their more natural interpretation has a number of advantages and they have now been used successfully in a wide range of situations; see, for example, Kitagawa and Gersch (1984),[10] Harvey and Durbin (1986)[8] and Harvey (1989).[6] Most of the work involving structural models has assumed normally distributed observations. However, while normality might be a reasonable assumption for a claims distribution, or at least its logarithm, it is not a reasonable assumption for the number of claims when such numbers are typically rather small. Recent work by Harvey and Fernandes (1989),[7] however, shows that the methodology of structural time series modelling can be extended to handle Poisson observations with a time-varying mean.

As Sections 3 and 4 will show, the structural time series models for Gaussian and Poisson observations can be used as the basis for modelling random sums.

2.1 *Stochastic Trends*

The simplest structural time series models consist of a stochastic trend component, μ_t, plus an irregular random disturbance term, ε_t.

The *random walk plus noise* or local level model is

$$y_t = \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0,\sigma_\varepsilon^2) \tag{2.1a}$$

$$\mu_t = \mu_{t-1} + \eta_t, \qquad \eta_t \sim \text{NID}(0,\sigma_\eta^2) \tag{2.1b}$$

where, as the notation indicates, the ε_ts are normally and independently distributed with mean zero and variance σ_ε^2, and η_t has a similar distribution with variance σ_η^2. Furthermore ε_t and η_t are mutually independent. Only the y_ts are observed; the trend, like the irregular term, is unobservable. However, the model, as it stands, is in state space form and as a result the Kalman filter can be used as the basis for computing optimal estimators of μ_t within the sample and for making predictions. Furthermore, it enables maximum likelihood (ML) estimators of the unknown hyperparameters, σ_η^2 and σ_ε^2, to be computed via the *prediction error decomposition*. All of this is well documented in the references already cited. Note that an important feature in motivating the random walk plus noise model is that the predictions are essentially formed as an exponentially weighted moving average (EWMA). The smoothing constant depends on the signal–noise ratio, $q = \sigma_\eta^2/\sigma_\varepsilon^2$.

In the local linear trend model, (2.1b) is replaced by

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim \text{NID}(0,\sigma_\eta^2) \\ \beta_t &= \beta_{t-1} + \zeta_t, & \zeta_t &\sim \text{NID}(0,\sigma_\zeta^2) \end{aligned} \tag{2.2}$$

where β_t is the stochastic slope. The forecasts from this model correspond to those obtained from the non-seasonal Holt–Winters recursions with suitably chosen smoothing constants.

The components μ_t in (2.1b) and (2.2) are known as *stochastic trends*. Deterministic trends emerge as a special case. Thus if, in (2.2), $\sigma_\eta^2 = \sigma_\zeta^2 = 0$, then the model reduces to the linear time trend

$$y_t = \mu_0 + \beta t + \varepsilon_t \tag{2.3}$$

where μ_0 and β are unknown intercept and slope parameters. Similarly if $\sigma_\eta^2 = 0$ in (2.1b), the observations are simply distributed about a constant mean.

Models of the above kind may be extended by adding other stochastic components such as seasonals and cycles.

2.2 *Time Series Models for Poisson Observations*

A model for Poisson observations which allows the mean, λ_t, to change stochastically over time can be constructed by analogy with (2.1). Let the observation at time t be drawn from a Poisson distribution,

$$p(N_t|\lambda_t) = \lambda_t^{N_t} e^{-\lambda_t}/N_t! \tag{2.4}$$

This corresponds to the measurement equation of (2.1a). However, rather than trying to formulate a transition equation analogous to (2.1b), we look to the properties of natural conjugate distributions of the type used in Bayesian

statistics. This approach was originated by Smith (1979)[15] but in the further development carried out by Harvey and Fernandes (1989)[7] the procedure is put within a classical framework by constructing a likelihood function and suggesting various diagnostic test statistics.

The conjugate prior for a Poisson distribution is the gamma distribution. Let $p(\lambda_{t-1}|N_{t-1})$ denote the PDF of λ_{t-1} conditional on the information at time $t-1$, namely the values of the first $t-1$ observations, denoted as N_{t-1}. Suppose that this distribution is gamma, that is $N_t \sim \Gamma(a,b)$, with PDF

$$p(\lambda;a,b) = \frac{e^{-b\mu}\lambda^{a-1}}{\Gamma(a)\,b^{-a}}, \qquad a,b>0 \tag{2.5}$$

with $\lambda = \lambda_{t-1}$, $a = a_{t-1}$ and $b = b_{t-1}$ where a_{t-1} and b_{t-1} are computed from the first $t-1$ observations. In model (2.1) with normally distributed observations, $\mu_{t-1} \sim N(m_{t-1}, p_{t-1})$ at time $t-1$ implies that $\mu_t \sim N(m_{t-1}, p_{t-1}+\sigma_\eta^2)$ at time $t-1$. In other words the mean of $\mu_t|N_{t-1}$ is the same as that of $\mu_{t-1}|N_{t-1}$ but the variance increases. The same effect can be induced in the gamma distribution by multiplying a and b by a factor less than one. We therefore suppose that $p(\lambda_t|N_{t-1})$ follows a gamma distribution with parameters $a_{t|t-1}$ and $b_{t|t-1}$ such that

$$a_{t|t-1} = \omega\, a_{t-1} \tag{2.6a}$$

$$b_{t|t-1} = \omega\, b_{t-1} \tag{2.6b}$$

where ω is a constant in the range $0<\omega\leqslant 1$. Then

$$E(\lambda_t|N_{t-1}) = a_{t|t-1}/b_{t|t-1} = a_{t-1}/b_{t-1} = E(\lambda_{t-1}|N_{t-1})$$

while

$$\mathrm{Var}(\lambda_t|N_{t-1}) = a_{t|t-1}/b_{t|t-1}^2 = \omega^{-1}\,\mathrm{Var}(\lambda_{t-1}|N_{t-1}).$$

The stochastic mechanism governing the transition of λ_{t-1} to λ_t is therefore defined implicitly rather than explicitly. However, it is possible to show that this mechanism is formally equivalent to a multiplicative transition equation of the form

$$\lambda_t = \omega^{-1}\lambda_{t-1}\eta_t,$$

where η_t has a beta distribution with parameters ωa_{t-1} and $(1-\omega)a_{t-1}$; see the discussion in Smith and Miller (1986).[16]

Once the observation N_t becomes available, the posterior distribution, $p(\lambda_t|N_t)$, is given by a gamma distribution with parameters

$$a_t = a_{t|t-1} + N_t \tag{2.7a}$$

$$b_t = b_{t|t-1} + 1. \tag{2.7b}$$

The initial prior gamma distribution, that is the distribution of λ_t at time $t=0$, tends to become diffuse, or non-informative, as $a, b \to 0$, although it is actually degenerate at $a=b=0$ with $\Pr(\lambda=0)=1$. However, none of this prevents the recursions (2.6) and (2.7) being initialized at $t=0$ with $a_0=b_0=0$. A proper distribution for λ_t is then obtained at time $t=\tau$ where τ is the index of the first non-zero observation. It follows that, conditional on N_τ, the joint density of the observations $N_{\tau+1}, \ldots, N_T$ is

$$p(N_{\tau+1}, \ldots, N_T; \omega) = \prod_{t=\tau+1}^{T} p(N_t|N_{t-1}). \tag{2.8}$$

The predictive PDF's are given by

$$p(N_t|N_{t-1}) = \int_0^\infty p(N_t|\lambda_t)\, p(\lambda_t|N_{t-1}) d\lambda_t \tag{2.9}$$

and for Poisson observations and a gamma prior this operation yields a negative binomial distribution

$$p(N_t|N_{t-1}) = \binom{a+N_t-1}{N_t} b^a (1+b)^{-(a+N_t)} \tag{2.10}$$

where $a = a_{t|t-1}$ and $b = b_{t|t-1}$ and

$$\binom{a+N-1}{N} = \frac{\Gamma(a+N)}{\Gamma(n+1)\Gamma(a)}$$

although since N is an integer, $\Gamma(N+1) = N!$ Hence the log-likelihood function for the unknown hyperparameter ω is

$$\log L(\omega) = \sum_{t=\tau+1}^{T} \{\log \Gamma(a_{t|t-1}+N_t) - \log N_t! - \log \Gamma(a_{t|t-1}) + a_{t|t-1} \log b_{t|t-1} - (a_{t|t-1}+N_t) \log (1+b_{t|t-1})\}. \tag{2.11}$$

It follows from the properties of the negative binomial that the mean and variance of the predictive distribution of N_{T+1} given N_T are respectively

$$\hat{N}_{T+1|T} = E(N_{T+1}|N_T) = a_{T+1|T}/b_{T+1|T} = a_T/b_T, \tag{2.12a}$$

and

$$\begin{aligned} \mathrm{Var}(N_{T+1}|N_T) &= a_{T+1|T}(1+b_{T+1|T})/b_{T+1|T}^2 \\ &= \omega^{-1} \mathrm{Var}(\lambda_T|N_T) + E(\lambda_T|N_T). \end{aligned} \tag{2.12b}$$

Repeated substitution from (2.3) and (2.4) shows that the forecast function is

$$\hat{N}_{T+1|T} = a_T/b_T = \sum_{j=0}^{T-1} \omega^j N_{T-j} \Big/ \sum_{j=0}^{T-1} \omega^j. \tag{2.13}$$

This is a weighted mean in which the weights decline exponentially. In large samples the denominator of (2.13) is approximately equal to $1/(1-\omega)$ when $\omega < 1$ and the forecasts can be obtained recursively by the EWMA scheme

$$\tilde{N}_{t+1|t} = (1-\lambda)\,\tilde{N}_{t|t-1} + \lambda\, N_t, \qquad t = 1, \ldots, T \tag{2.14}$$

where $\tilde{N}_{1|0}=0$ and $\lambda = 1-\omega$ is the smoothing constant. When $\omega = 1$, the right hand side of (2.13), is equal to the sample mean. Regarding this as an estimate of λ, the choice of zeroes as initial values for a and b in the filter is seen to be justified insofar as it yields the classical solution. In Bayesian terms it corresponds to a non-informative prior.

Details of multi-step prediction can be found in Harvey and Fernandes (1989).[7]

2.3 *Explanatory Variables*

Observable explanatory variables may be added to structural time series models. For a Gaussian model, such as (2.1), the first equation becomes

$$y_t = \mu_t + \boldsymbol{x}_t'\boldsymbol{\delta} + \varepsilon_t, \tag{2.15}$$

where $\boldsymbol{x}_t$ is a $k \times 1$ vector of explanatory variables and $\boldsymbol{\delta}$ is the corresponding $k \times 1$ vector of unknown parameters. If μ_t is removed (2.4) reduces to a classical linear regression model.

In a model with Poisson observations, but no dynamic structure, explanatory variables are introduced via a link function; see the discussion of the GLIM framework in McCullagh and Nelder (1983).[13] The exponential link function

$$\lambda_t = \exp(\boldsymbol{x}_t'\boldsymbol{\delta}) \tag{2.16}$$

ensures that λ_t remains positive. Note that it does not make sense to include a lagged dependent variable as an explanatory variable when the observations are small and discrete.

Explanatory variables can be introduced into time series models for count data as follows. As in (2.15), the level λ_t may be thought of as a component which has a separate effect from that of the explanatory variables in x_t, none of which is a constant. Suppose that

$$\lambda_{t-1} \sim \Gamma(a_{t-1}, b_{t-1})$$

and that, conditional on the information at time $t-1$,

$$\lambda_t \sim \Gamma(\omega a_{t-1}, \omega b_{t-1}).$$

This level component may be combined multiplicatively with an exponential link function for the explanatory variables so that the distribution of N_t, conditional on λ_t, is Poisson with mean

$$\lambda_t^+ = \lambda_t \exp(\boldsymbol{x}_t'\,\boldsymbol{\delta}). \tag{2.17}$$

It follows from the properties of the gamma distribution that, conditional on N_{t-1},

$$\lambda_t^+ \sim \Gamma(a_{t|t-1}^+, b_{t|t-1}^+)$$

where

$$a_{t|t-1}^+ = \omega\, a_{t-1} \qquad \text{and} \qquad b_{t|t-1}^+ \exp(-x_t' \delta) \tag{2.18}$$

respectively.

The log-likelihood of the observations is therefore as in (2.11) with $a_{t|t-1}$ and $b_{t|t-1}$ replaced by $a_{t|t-1}^+$ and $b_{t|t-1}^+$. This must be maximized with respect to ω and δ.

As regards updating, $\lambda_t^+ \sim \Gamma(a_t^+, b_t^+)$ where a_t^+ and b_t^+ are obtained from $a_{t|t-1}^+$ and $b_{t|t-1}^+$ via updating equations of the form (2.7). Therefore the posterior distribution of λ_t *is* $\Gamma(a_t, b_t)$, where a_t and b_t are given by

$$a_t = \omega\, a_{t-1} + N_t \tag{2.19a}$$

$$b_t = \omega\, b_{t-1} + \exp(x_t' \delta), \qquad t = \tau + 1, \ldots, T. \tag{2.19b}$$

Thus the only amendment as compared with the recursions in the previous subsection is the replacement of unity by $\exp(x_t' \delta)$ in the equation for b_t.

In the Gaussian case, the computational burden is eased considerably by the fact that δ may be estimated linearly by generalized least squares; see Kohn and Ansley (1985).[11] It is unfortunate that this is no longer possible for the Poisson model. However, it is sometimes possible to use estimates from the Gaussian model as starting values; the difficulty lies in how to handle zero observations when logarithms are being taken.

Another limitation to the Poisson model is that only the level can be allowed to be time-varying. Unlike in the Gaussian case, the time trend and seasonals must be deterministic, entering the model as explanatory variables. However, it is argued in Harvey and Fernandes (1989)[7] that Poisson observations are unlikely to contain enough information to enable changes in slope and seasonal effects to be picked up.

3. MODELLING THE SIZE AND NUMBERS OF CLAIMS

As in the introduction we consider a situation where there is a single group at risk and there are N_t claims at time t. It is assumed that the size of each individual claim is known. Estimation by maximum likelihood is proposed.

3.1 *Stochastic Trends in Claim Sizes*

Structural time series models for claim sizes may be developed by assuming that claims are lognormally distributed. According to Beard *et al.* (1984) Ch. 3,[1] such an assumption has been found to be reasonable in practice and it means that we can construct a Gaussian model for $\log(Y_{jt}) = y_{jt}$.

The observations satisfy the model

$$y_{jt} = \mu_t + \varepsilon_{jt}, \qquad j = 1, \ldots, N_t, \qquad t = 1, \ldots, T \tag{3.1}$$

where the ε_{jt} are mutually and serially uncorrelated disturbances with variance σ_ε^2 and μ_t is either formulated as a random walk, (2.1b), or, more generally, as a local linear trend model, (2.2). As noted in § 2.1, if $\sigma_\eta^2 = \sigma_\zeta^2 = 0$ in the local linear trend model, it collapses to a deterministic time trend. This is known as the Hachemeister model in the credibility theory literature; see Hachemeister (1975).[9] The stochastic generalization to the local linear trend model is a natural one. It has been suggested recently by Ledolter, Klugman and Lee (1989)[12] for credibility models, although the modelling framework within which they use it is somewhat different.

The model in (3.1) can be extended to include other stochastic components, such as seasonals, and explanatory variables. Handling these extensions poses no problems, and so attention is focused on (3.1). Indeed matters can be simplified further by restricting attention to the case where μ_t follows a random walk. The full model can be written as

$$\boldsymbol{y}_t = \boldsymbol{i}\,\mu_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathrm{NID}(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I}) \tag{3.2a}$$

$$\mu_t = \mu_{t-1} + \eta_t, \qquad \eta_t \sim \mathrm{NID}(0, \sigma_\eta^2) \tag{3.2b}$$

where $\mathbf{i}$ is an $N_t \times 1$ vector of ones. This is a special case of a dynamic factor model; see Fernández Macho *et al.* (1987)[4] and Harvey (1989)[6] Ch. 8. As such it can be handled by the Kalman filter and the exact likelihood function constructed. However, the special form of the state space model in (3.2a) means that (a) only a univariate Kalman filter need be run; and (b) numerical optimization only need be carried out with respect to a single parameter, the relative variance $q = \sigma_\eta^2/\sigma_\varepsilon^2$.

The univariate Kalman filter is based on the averages in each time period. Thus (3.2a) implies

$$\bar{y}_t = \mu_t + \bar{\varepsilon}_t, \qquad \bar{\varepsilon}_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2/N_t). \tag{3.3}$$

The fact that the Kalman filter applied to the full set of disaggregated observations does indeed reduce to a Kalman filter applied to (3.3) and (3.2b) can be shown as follows. Let $\sigma_\varepsilon^2 p_t$ denote the variance of μ_t conditional on the information at time t. The Kalman filter for (3.2) can be written simply as

$$m_t = m_{t-1} + p_{t|t-1}\,\boldsymbol{i}'\boldsymbol{F}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{i}m_{t-1}) \tag{3.4a}$$

$$p_t = p_{t|t-1} - p_{t|t-1}\,\boldsymbol{i}'\boldsymbol{F}_t^{-1}\boldsymbol{i}\,p_{t|t-1}, \tag{3.4b}$$

where m_t is the minimum mean square estimator of μ_t based on information at time t,

$$p_{t|t-1} = p_{t-1} + q \tag{3.5}$$

and

$$\boldsymbol{F}_t = \boldsymbol{I} + \boldsymbol{i}p_{t|t-1}\,\boldsymbol{i}' \tag{3.6}$$

By a standard matrix inversion lemma

$$\boldsymbol{F}_t^{-1} = \boldsymbol{I} - \{p_{t|t-1}/(1 + N_t p_{t|t-1})\}\boldsymbol{i}\boldsymbol{i}' \tag{3.7}$$

and so

$$m_t = m_{t-1} + \{N_t p_{t|t-1}/(1 + N_t p_{t|t-1})\}(\bar{y}_t - m_{t-1}). \tag{3.8}$$

This recursion and the corresponding recursion for p_t are precisely the expressions obtained by writing down the Kalman filter for the univariate aggregate model (3.3) and (3.2b).

The initial conditions for the Kalman filter must be computed from one of the observations at time $t = 1$. This yields

$$\mu_0 \sim N(y_{1*}, \sigma_\varepsilon^2) \tag{3.9}$$

where y_{1*} is the value of the selected observation. Since y_{1*} has been used in this way it should not be used in the subsequent calculations. Thus N_1 will be redefined as the actual number of observations at time $t = 1$ less one. It can be shown that the likelihood function and the estimates of the state are not affected by the choice of the initial observation. In more general models with d nonstationary elements in the state vector, an estimator of the state vector at $t = 0$ may be computed by taking an observation from each of the first d sets of observations and making the appropriate amendments to the values of $N_1, \ldots, N_d$.

If all the N_t's are regarded as being fixed, the log-likelihood function of the full set of observations on claims can be written as

$$\log L = -\frac{N}{2}\log 2\pi - \tfrac{1}{2}\sum_{t=1}^{T}\log |\boldsymbol{F}_t| - \frac{N}{2}\log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{T}\boldsymbol{v}_t' \boldsymbol{F}_t^{-1}\boldsymbol{v}_t \tag{3.10a}$$

where

$$\boldsymbol{v}_t = y_t - \boldsymbol{i}\, m_{t-1}, \qquad t = 1, \ldots, T \tag{3.10b}$$

and

$$N = \sum_{t=1}^{T} N_t. \tag{3.10c}$$

Thus conditional on q, the ML estimator of σ_ε^2 is

$$\tilde{\sigma}_\varepsilon^2 = \frac{1}{N}\sum_{t=1}^{T}\boldsymbol{v}_t' \boldsymbol{F}_t^{-1}\boldsymbol{v}_t. \tag{3.11}$$

However, using (3.7),

$$\tilde{\sigma}_\varepsilon^2 = \frac{1}{N}\sum_{t=1}^{T}\left[\sum_{j=1}^{N_t} v_{jt}^2 - \left\{\frac{N_t^2 p_{t|t-1}}{1+N_t p_{t|t-1}}\right\}\bar{v}_t^2\right] \tag{3.12}$$

where

$$v_{jt} = y_{jt} - m_{t-1}, \qquad j = 1, \ldots, N_t, \qquad t = 1, \ldots, T$$

and $\bar{v}_t$ is the mean of the v_{jt}'s. The likelihood function may therefore be concentrated with respect to σ_ε^2, leaving the following function to be minimized with respect to q:

$$S(q) = N \log \tilde{\sigma}_\varepsilon^2 + \sum_{t=1}^{T} \log (1 + N_t p_{t|t-1}). \tag{3.13}$$

The simplification in the determinental term arises as a corollary to the matrix inversion result in (3.7); see Rosenberg (1973)[14] page 416.

If σ_ε^2 and q were known, the Kalman filter would yield the mean and variance of the distribution of an observation at $T+1$ conditional on the information at time T. Thus,

$$y_{t,T+1} \sim N(m_{T+1|T},\ \sigma_\varepsilon^2\{p_{T+1|T}+1\}). \tag{3.14}$$

This result is approximately true when σ_ε^2 and q are replaced by their ML estimators.

3.2 *Modelling the Number of Claims*

A deterministic Poisson model for the number of claims can be set up within the GLIM framework by using the link function, (2.16). In addition to a constant term, it would seem sensible for the set of explanatory variables, x_t, to at least contain $\log P_t$ where P_t is the population at risk. If the corresponding δ parameter is set equal to unity, then the mean of the Poisson distribution is proportional to P_t. Other candidates for explanatory variables might be a time trend, seasonals and income; see the discussion in Beard *et al.* (1984[1] Ch. 6).

When N_t has a Poisson distribution, the distribution function of the value of claims, $F(S_t)$ in (1.2), is known as a compound Poisson distribution. Unfortunately, finding an analytic expression for $F(S_t)$ is not possible when the size of claims follows a lognormal distribution. Nevertheless, using the information available at time T, predictions for S_{T+1} together with any higher order moments, may be made by combining predictions from the Poisson count model with predictions from the lognormal claims model.

A stochastic time series model may be set up for the number of claims as described in Section 2. Instead of a constant term, the model contains a stochastic level and the mean of the distribution of N_{T+1} at time T is given by (2.13). The predictive distribution for the total value of claims is no longer a compound Poisson distribution, however, because the predictive distribution of N_t is not

Poisson if ω is less than one. As shown in Section 2, it is negative binomial for predictions one step ahead. Evaluating the moments of S_t can be carried out as usual, with the mean and variance given by (1.3) and (1.4). That is, the expected value of claims in the next time period is:

$$\underset{T}{E}(S_{T+1}) = \underset{T}{E}(N_{T+1}) \cdot \underset{T}{E}(Y_{j,T+1}) \tag{3.15a}$$

with a prediction MSE given by

$$\underset{T}{\text{Var}}(S_{T+1}) = \underset{T}{\text{Var}}(N_{T+1})\{\underset{T}{E}(Y_{j,T+1})\}^2 + \underset{T}{E}(N_{T+1})\underset{T}{\text{Var}}(Y_{j,T+1}) \tag{3.15b}$$

with

$$\underset{T}{E}(N_{T+1}) = a^+_{T+1|T}/b^+_{T+1|T} \tag{3.16a}$$

and

$$\underset{T}{\text{Var}}(N_{T+1}) = a^+_{T+1|T}(1 + b_T)/b^{+\,2}_{T+1|T}. \tag{3.16b}$$

For lognormal claims,

$$\underset{T}{E}(Y_{j,T+1}) = \exp(m_T + \tfrac{1}{2}\sigma_\varepsilon^2 p_{T+1|T}) \tag{3.17a}$$

and

$$\underset{T}{E}(Y^2_{j,T+1}) = \exp(2m_T + 2\sigma_\varepsilon^2 p_{T+1|T}). \tag{3.17b}$$

When $\omega = 1$, in the model for the number of claims, $F(S_{T+1})$ reverts to a compound Poisson distribution and because the estimated mean at time $T+1$ is equal to the variance, (3.15b) simplifies to

$$\underset{T}{\text{Var}}(S_{T+1}) = \underset{T}{E}(N_{T+1}) \cdot \underset{T}{E}(Y^2_{j,T+1}). \tag{3.18}$$

Maximum likelihood estimation of the parameters in the full random sum model can be obtained by first writing down the joint density function for the individual claims and the number of claims at time t. Thus if $\boldsymbol{y}_t$ is the $N_t \times 1$ vector containing y_{jt}, $j = 1, \ldots, N_t$ and $\boldsymbol{y}^*_t = (\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_1)$,

$$p(\boldsymbol{y}_t, N_t | N_{t-1}, \boldsymbol{y}^*_{t-1}) = p(\boldsymbol{y}_t | N_t, \boldsymbol{y}^*_{t-1}) \cdot p(N_t | N_{t-1}, \boldsymbol{y}^*_{t-1}). \tag{3.19}$$

As a rule the distribution of N_t will not depend on past values of the claims, in which case $\boldsymbol{y}^*_{t-1}$ can be dropped from the last term in (3.19). In any case, it follows

from (3.19) that the log-likelihood function can be written as the sum of the log-likelihood for the claims, (3.10a), and the log-likelihood for the counts, (2.11). If the models contain no parameters in common, ML estimation of each can be carried out separately.

4. SEVERAL GROUPS

Suppose now that there are G groups, and that the number of claims on the gth group at time t is N_{gt}. The total number of claims at time t, that is the sum of the N_{gt}s, will be denoted by N_t. The problem now is to predict the total value of claims in each group,

$$S_{g,T+1} = \sum_{j=1}^{N_{g,T+1}} Y_{gj,T+1}.$$

The approach adopted is to let the model for the size of claims in each group contain a stochastic trend component which is common to all groups. The stochastic model for the number of claims described earlier is then generalized in a parallel fashion.

It is assumed that data on individual claims are available in each group.

4.1 *Size of Claims*

The model for the size of claims

$$y_{jgt} = \mu_t + \mu_g + \varepsilon_{jt}, \qquad \varepsilon_{jt} \sim \text{NID}(0, \sigma_\varepsilon^2) \tag{4.2}$$

where μ_t is a random walk as in (3.2b) and the μ_gs, $g = 2, \ldots, G$ are fixed parameters. These parameters are to be interpreted as multiplicative factors for the common trend component, as it enters into the model for the size of claims for a given group. In other words the level of claims in different groups is not the same, but the long-run movements they exhibit follow the same pattern, reflecting changes in common influences stemming from changes in unmeasured economic and social variables.

In matrix terms (4.2) is

$$\boldsymbol{\Psi}_t = \boldsymbol{i}\,\mu_t + \boldsymbol{\Psi}_t \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \text{NID}(0, \sigma_\varepsilon^2 \boldsymbol{I}) \tag{4.3}$$

where y_t, $\boldsymbol{i}$ and $\boldsymbol{\varepsilon}_t$ are $N_t \times 1$, such that $\boldsymbol{\mu}$ is a $(G-1) \times 1$ vector and $\boldsymbol{\Psi}_t$ is an $N_t \times (G-1)$ matrix

$$\boldsymbol{\Psi}_t = \begin{bmatrix} \mathbf{0} & & \mathbf{0} \\ \boldsymbol{i}_2 & \ddots & \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{i}_G \end{bmatrix} \tag{4.4}$$

where $\boldsymbol{i}_g$ is an N_{gt} vector of ones. The model is a special case of the dynamic factor model described in Fernández Macho *et al.* (1987),[4] and various generalizations

within this framework might prove to be useful. However, if T is small, the scope for estimating a richer model than the one above may be limited.

The model may be estimated by applying the Kalman filter described in § 3.1. The only modification is that this filter is applied to $\boldsymbol{y}_t$ and to the columns of $\boldsymbol{\Psi}_t$. This enables an estimator of $\boldsymbol{\mu}$ to be computed by carrying out a multivariate regression of the innovations from $\boldsymbol{y}_t$ on the innovations from each of the columns of $\boldsymbol{\Psi}_t$. In this way $\boldsymbol{\mu}$ may be concentrated out of the likelihood function; see Kohn and Ansley (1985)[11] or Harvey (1989),[6] Sect. 3.4. A starting value for $\boldsymbol{\mu}_0$ is obtained from one of the observations in the first group. This is possible because the mean of the first group is μ_t. An alternative way of computing the likelihood function is by adapting the more general algorithm given by De Jong (1988).[3]

A somewhat simpler approach enables one to estimate the μ_ts independently of the hyperparameter, q. The mean in each group satisfies

$$\bar{y}_{gt} = \mu_t + \mu_g + \bar{\varepsilon}_{gt}, \tag{4.5}$$

and so aggregating over time gives

$$\bar{y}_g = \sum_{t=1}^{T} \bar{y}_{gt}/T = \sum_{t=1}^{T} \mu_t/T + \mu_g + \sum_{t=1}^{T} \varepsilon_{gt}/T, \qquad g = 1, \ldots, G. \tag{4.6}$$

Since $\mu_g = 0$ for $g = 1$,

$$\bar{y}_g - \bar{y}_1 = \mu_g + \Sigma\, \bar{\varepsilon}_{gt}/T - \Sigma\, \bar{\varepsilon}_{1t}/T. \tag{4.7}$$

Thus the term involving the underlying stochastic level has disappeared, suggesting the estimator

$$\hat{\mu}_g = \bar{y}_g - \bar{y}_1, \qquad g = 2, \ldots, G. \tag{4.8}$$

The variance of this estimator is

$$\text{Var}\,(\hat{\mu}_g) = \sigma_\varepsilon^2 \left[\sum_{t=1}^{T} (1/N_{gt}) + \sum_{t=1}^{T} (1/N_{1t}) \right] / T^2. \tag{4.9}$$

These estimates of the group effects may be subtracted from the observations in the corresponding groups. The parameters σ_ε^2 and q are then estimated using the algorithm of § 3.1.

4.2 *Number of Claims*

As with the size of claims, it seems reasonable to assume that the stochastic movements in the level are common to all the groups. Thus the number of claims in each of the g-th groups follow independent Poisson distributions with mean

$$\lambda_{gt}^{+} = \lambda_t \exp(\boldsymbol{x}_t'\boldsymbol{\delta}) \exp(\boldsymbol{x}_{gt}'\boldsymbol{\delta}_g), \qquad g = 1, \ldots, G \tag{4.10}$$

where $\boldsymbol{x}_{gt}$ is a vector of explanatory variables for the g-th group, $\boldsymbol{x}_t$ is a vector of explanatory variables entering into the common trend component, and $\boldsymbol{\delta}_g$ and $\boldsymbol{\delta}$

are corresponding vectors of unknown parameters. In the simplest case $x_{gt} = 1$ for all g, in which case $\exp(\delta_g)$ is the factor by which the common trend must be multiplied for the j-th group. The δ_gs play a similar rôle to the μ_gs in (4.2), and, as there, some kind of normalization is required. The most appropriate normalization is to let the sum of the $\exp(\delta_g)$s be G, since then the mean of the gth group is $\lambda \exp(x_t'\delta)$ if $\delta_g = 0$. More generally

$$\sum_{g=1}^{G} \exp(x_{gt}'\delta_g) = G. \tag{4.11}$$

The prediction equations for λ_t are given by (2.6), just as in a univariate model. The G independent observations can now be used to update $a_{t|t-1}$ and $b_{t|t-1}$ by bringing them in one at a time. The net result is two composite updating equations,

$$a_t = a_{t|t-1} + \sum_g N_{gt} \tag{4.12a}$$

$$b_t = b_{t|t-1} + \exp(x_t'\delta) \sum_g \exp(x_{gt}'\delta) = b_{t|t-1} + G \exp(x_t'\delta). \tag{4.12b}$$

The joint density function of the numbers of claims, conditional on the numbers in previous time periods, N_{t-1}, is

$$p(N_{1t}, \ldots, N_{Gt}|N_{t-1}) = \prod_{g=1}^{G} p(N_{gt}|N_{g-1,t}, \ldots, N_{1t}, N_{t-1}) \tag{4.13}$$

Each of these G conditional distributions is negative binomial with parameters $(a_{g-1,t}, b^+_{g-1,t})$ where

$$b^+_{gt} = b_{gt} \exp(-x_t'\delta) \exp(-x_{gt}'\delta) \tag{4.14}$$

and a_{gt} and b_{gt} are given by the recursions

$$a_{gt} = a_{g-1,t} + N_{gt} \tag{4.15a}$$

$$b_{gt} = b_{g-1,t} + \exp(x_t'\delta)\exp(x_{gt}'\delta) \tag{4.15b}$$

with starting values $a_{0,t} = a_{t|t-1}$ and $b_{0,t} = b_{t|t-1}$. The net effect of (4.15) is of course the single set of recursions in (4.12), with $a_t = a_{Gt}$ and $b_t = b_{Gt}$. The full likelihood function is obtained by summing the conditional densities in (4.13) over all t from $t = 1$ to T.

The main drawback with this model is that estimation must be carried out nonlinearly with respect to $\delta_1, \ldots, \delta_G$ as well as ω. However, a preliminary estimator of ω can be constructed by summing the observations in each time period and treating them as a univariate series from a Poisson distribution with parameter

$$\sum_g \lambda^+_{gt} = \sum_g \lambda_t \exp(x_{gt}'\delta_g) = G\,\lambda_t \exp(x_t'\delta) \tag{4.16}$$

The updating recursions are precisely those given in (4.15), except that G is replaced by unity in (4.15b). However, there is no reason why (4.15b) should not be used. It is interesting to observe that $\tilde{\lambda}_t$ is an EWMA of the average number of claims per group in each time period.

If $\boldsymbol{x}_{gt}=(1 \log P_{gt})'$, and if the coefficients of $\log P_{gt}$ are unity and each P_{gt} is roughly constant over time, i.e. $P_{gt} \simeq P_g$, preliminary estimates of $\delta_1, \ldots, \delta_g$ may be obtained by summing the number of claims in each group over time. If the λ_ts are treated as though they were fixed, the total number of claims per group over the full time period, N_g, would have a Poisson distribution with parameter

$$\sum_{t=1}^{T} \lambda_t^+ = e^{\delta g} P_g \sum_{t=1}^{T} \lambda_t \exp(\boldsymbol{x}_t'\boldsymbol{\delta}), \qquad g = 1, \ldots, G. \tag{4.17}$$

In view of the normalization rule, the distribution of N, the sum of claims over all groups, is Poisson with mean $\Sigma\lambda_t \exp(\boldsymbol{x}_t'\boldsymbol{\delta})$. This suggests the following preliminary estimator:

$$\hat{\delta}_g = \log (N_g/N P_g), \qquad g=1, \ldots, G. \tag{4.18}$$

5. CONCLUSION AND EXTENSIONS

This article has set out time series models which can be used for modelling claims. The statistical specification of these models invites estimation by maximum likelihood and allows tests of model specification to be constructed. Such tests, which are described in Harvey and Durbin (1986)[8] and Harvey (1989),[6] could be extended to cover some of the more complex multivariate situations described here.

One practical problem which arises is that the values of individual claims may not be available. In the absence of such information predictions of the expected value of claims could still be made from the aggregate model, (3.3), provided that $\bar{y}_t$, $t=1, \ldots, T$, is observed. The fact that this is the geometric, rather than the arithmetic, mean of the individual claims makes its availability unlikely. Rather than attempt to work with some kind of approximation, it may be worth considering the use of a gamma distribution for the size of claims S_t, since the sum of identically distributed gamma variables is still gamma. We plan to describe the implementation of a dynamic model based on gamma distributions in a later article.

ACKNOWLEDGEMENTS

We would like to thank Piet de Jong, Johannes Ledolter, Keith Ord, Leigh Roberts and Neil Shephard for helpful comments. However, we are solely responsible for the views expressed here and for any errors. We would also like to thank CNPq and the ESRC for financial support, the latter in conjunction with the project 'Multivariate Aspects of Structural Time Series Models'.

REFERENCES

(1) BEARD, R. E., PENTIKAINEN, T. & PESONEN, E. (1984). *Risk Theory*, 3rd Ed. London: Methuen.

(2) BOX, G. E. P. & JENKINS, G. M. (1976). *Time Series Analysis: Forecasting and Control*, revised edition. San Francisco: Holden-Day.

(3) DE JONG, P. (1988). The Diffuse Kalman Filter, Unpublished paper, University of British Columbia.

(4) FERNÁNDEZ MACHO, F. J., HARVEY, A. C. & STOCK, J. H. (1987). Forecasting and Interpolation Using Vector Autoregressions with Common Trends. *Annales d'Economie et de Statistique*, **6/7**, 279–287.

(5) GERBER, H. U. (1979). *An Introduction to Mathematical Risk Theory*. Huebner Foundation Monograph, No. 8. Homewood, Illinois: Richard D. Irwin, Inc.

(6) HARVEY, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

(7) HARVEY, A. C. & FERNANDES, C. (1989). Time Series Models for Count Data or Qualitative Observations. *Journal of Business and Economic Statistics* (to appear).

(8) HARVEY, A.C. & DURBIN, J. (1986). The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling. *Journal of the Royal Statistical Society*, Series A, **149**, 187–227.

(9) HACHEMEISTER, C. (1975). 'Credibility for Regression Models with Application to Trend', in *Credibility: Theory and Applications*, P. Kahn (ed.). New York: Academic Press.

(10) KITAGAWA, G. & GERSCH, W. (1984). A Smoothness Priors–State Space Modelling of Time Series with Trend and Seasonality. *Journal of the American Statistical Association*, **79**, 378–389.

(11) KOHN, R. & ANSLEY, C. F. (1985). Efficient Estimation and Prediction in Time Series Regression Models. *Biometrika*, **72**, 694–697.

(12) LEDOLTER, J., KLUGMAN, S. & LEE, C. (1989). Credibility Models with Time-Varying Trend Components, Technical Report No. 159, Department of Statistics, University of Iowa.

(13) MCCULLAGH, P. & NELDER, J. A. (1983). *Generalised Linear Models*. London: Chapman and Hall.

(14) ROSENBERG, B. (1973). Random Coefficient Models: the Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression. *Annals of Economic and Social Measurement*, **2**, 399–428.

(15) SMITH, J. Q. (1979). A Generalization of the Bayesian Steady Forecasting Model. *Journal of the Royal Statistical Society*, **B 41**, 375–387.

(16) SMITH, R. L. & MILLER, J. E. (1986). A Non-Gaussian State Space Model and Application to Prediction of Records. *Journal of the Royal Statistical Society*, **B 48**, 79–88.