

## WORKSHOP

### CREDIBILITY MODELS WITH TIME-VARYING TREND COMPONENTS

BY JOHANNES LEDOLTER, STUART KLUGMAN, CHANG-SOO LEE

#### ABSTRACT

Traditional credibility models have treated the process generating the losses as stable over time, perhaps with a deterministic trend imposed. However, there is ample evidence that these processes are not stable over time. What is required is a method that allows for time-varying parameters in the process, yet still provides the shrinkage needed for sound ratemaking. In this paper we use an automobile insurance example to illustrate how this can be accomplished.

#### KEYWORDS

Credibility; Kalman filter; shrinkage estimation; time series; trend components.

#### 1. INTRODUCTION

The goal of any ratemaking process is to estimate future claims on the basis of prior experience. The experience will be available for many groups over several time periods. It has been long known (MOWBRAY, 1914) that both statistical and business optimality is achieved by first estimating a rate for each group and then reducing the large values and increasing the small ones. Traditionally (e.g., BÜHLMANN and STRAUB, 1972) the initial estimates are sample means. Others (e.g. HACHEMEISTER, 1975) have recommended deterministic trend factors. Most all approaches that are currently used assume that the time series observations from a single group vary independently around a stable mean or trend.

Most time series, however, exhibit time-varying levels as well as autocorrelations among adjacent observations. The optimal forecasts for such series do not assign equal weights to all past observations, but discount the information according to their age; older observations get less weight. See BOX and JENKINS (1976) or ABRAHAM and LEDOLTER (1983) for a thorough discussion. Evidence for time-varying parameters was presented for automobile losses by BAILEY and SIMON (1959). A problem with most standard time series approaches, however, is that they are designed for making forecasts based on single series of relatively long lengths. Typical insurance problems contain many (sometimes hundreds) series of short (3-7 years) duration. Because these short series are occurring in a common external environment (e.g., of rising health care costs, automobile safety improvements, etc.) many of the features will be common to all of the series. The importance of both time series and

cross-sectional effects has also been noted in two recent econometrics papers by GARCIA-FERRER et al. (1987) and ZELLNER and HONG (1989) who use shrinkage methods to predict the economic growth rates of several countries.

The purpose of this paper is to bring together a dynamic model for the time-varying aspects of the problem and a shrinkage technique that takes account of the multiple group aspect. In Section 2 we review the credibility model with time-invariant parameters. In Section 3 we discuss univariate structural time-series models with time-varying trend and seasonal coefficients and we apply the shrinkage approach of Section 2 to the coefficients in the structural time series models. The final section illustrates this approach on actuarial data.

## 2. THE STANDARD CREDIBILITY MODEL

In all of the situations discussed in this paper the data consists of observations  $Y_t^{(i)}$ ,  $i = 1, \dots, k$ ,  $t = 1, \dots, n$  where  $k$  is the number of groups under consideration and  $n$  is the number of periods of observation. Typically, each value represents the amount paid in claims, divided by some measure of the size of the group,  $P_t^{(i)}$ . The objective is to forecast the value for a future period,  $Y_{n+1}^{(i)}$ , for each group.

A linear data generating model for the observations specifies

$$(2.1) \quad Y_t^{(i)} = \underline{x}_{it}' \underline{\beta}^{(i)} + e_t^{(i)} \quad e_t^{(i)} \sim N(0, \sigma^2 / P_t^{(i)})$$

where  $e_t^{(i)}$ , for  $t = 1, \dots, n$  and  $i = 1, \dots, k$ , are independent and  $\underline{x}_{it}$  are  $p \times 1$  known design vectors, usually functions of  $t$ . Two well-known models take (1)  $p = 1$  and  $x_{it} = 1$  and (2)  $p = 2$  and  $\underline{x}_{it} = (1, t)'$ . The data generating model in (1) is part of a special case of the Bühlmann-Straub model (BÜHLMANN and STRAUB, 1972); the linear trend in (2) is part of the Hachemeister model (HACHEMEISTER, 1975). The factor  $P_t^{(i)}$  in (2.1) is a measure of the amount of data that produces the observation  $Y_t^{(i)}$ , which in most actuarial situations is an average of many observations. The forecast of  $Y_{n+1}^{(i)}$ , the observation at a future time period, is provided by the estimate of the mean  $E(Y_{n+1}^{(i)}) = \underline{x}_{i, n+1}' \underline{\beta}^{(i)}$ .

The standard credibility model also assumes that the coefficients  $\underline{\beta}^{(i)}$ , for  $i = 1, \dots, k$ , are independent realizations from a common distribution. That is,

$$(2.2) \quad \underline{\beta}^{(i)} = \underline{b} + \underline{a}^{(i)} \quad \text{where} \quad \underline{a}^{(i)} \sim N(\underline{0}, \sigma^2 \underline{B}).$$

Treating this second level distribution as a prior distribution, the Bayes shrinkage estimate of  $\underline{\beta}^{(i)}$  is given by

$$(2.3) \quad \hat{\underline{\beta}}^{(i)} = \underline{Z}_i \hat{\underline{\beta}}^{(i)} + (\underline{I} - \underline{Z}_i) \underline{b}$$

where

$$(2.4) \quad \hat{\underline{\beta}}^{(i)} = \left( \sum_t P_t^{(i)} \underline{x}_{it} \underline{x}_{it}' \right)^{-1} \sum_t P_t^{(i)} \underline{x}_{it} Y_t^{(i)}$$

is the weighted least squares estimate in group  $i$ ,

$$(2.5) \quad Z_i = B(B + V_i)^{-1},$$

and

$$(2.6) \quad V_i = \left( \sum_t P_t^{(i)} x_{it} x'_{it} \right)^{-1}.$$

A problem with this solution is that estimates of the quantities  $B$  and  $\hat{b}$  must be obtained. A commonly accepted approach is to use the method of moments estimates that have been developed in variance components analysis (see SWAMY, 1971). However, there are a number of drawbacks with this approach. The estimates of  $B$  and  $Z_i$  are biased and, furthermore, the moment estimate of the scaled covariance matrix  $B$  need not be non-negative definite. These drawbacks can be overcome, in part, by either using the iterative estimation approach of DEVYLDER (1981, 1984), or a true Bayes approach instead of an empirical Bayes approach. The details of the Bayes analysis can be found in KLUGMAN (1987). DeVyllder proposes estimators  $\hat{B}$  and  $\hat{b}$  of  $B$  and  $b$  which depend via  $Z_i = B(B + V_i)^{-1}$  on the parameter  $B$  to be estimated. He suggests an iterative procedure where

$$(2.7) \quad \hat{b} = \left( \sum_i Z_i \right)^{-1} \sum_i Z_i \hat{\beta}_i^{(i)},$$

$$H = \sum_i Z_i (\hat{\beta}_i^{(i)} - \hat{b}) (\hat{\beta}_i^{(i)} - \hat{b})' / (k - 1),$$

$$\hat{B} = (H + H') / 2 \hat{\sigma}^2,$$

and

$$\hat{\sigma}^2 = \sum_i \sum_t P_t^{(i)} (Y_t^{(i)} - x'_{it} \hat{\beta}_i^{(i)})^2 / k(n - p).$$

The iterative procedure starts from an initial arbitrary non-negative definite symmetric matrix  $\hat{B}_0$ . It stops if, from one iteration to the next, the elements in  $\hat{B}$  do not change by more than a specified small quantity.

**Remark.** We can think of credibility models as consisting of two components: The first one in equation (2.1) models, for each group separately, the generation of the observations for given values of the coefficients  $\beta^{(i)}$ ; we refer to this as the *data generating model*. The second component in equation (2.2) relates the parameters  $\beta^{(i)}$  in the data generating model across the  $k$  groups; we refer to this as the *shrinkage component*. As mentioned above, a shortcoming of the traditional credibility model in equations (2.1) and (2.2) is that it does not

allow for time-varying coefficients. As a consequence the age of the observation does not enter into the analysis.

### 3. SHRINKAGE ESTIMATION IN MODELS WITH TIME-VARYING COEFFICIENTS

#### 3.1. Analysis of a single series

The following discussion concentrates on a single series (group) and, in order to simplify the presentation, we have omitted the group index  $i$ . In this paper we use structural time series models to incorporate time-varying coefficients into the data generating model. These models (see HARVEY and TODD, 1983; HARVEY, 1984) are of the form

$$(3.1) \quad Y_t = \underline{x}_t' \underline{\beta}_t + e_t, \quad e_t \sim N(0, \sigma^2/P_t)$$

$$\underline{\beta}_t = T \underline{\beta}_{t-1} + \underline{v}_t, \quad \underline{v}_t \sim N(0, \sigma^2 A).$$

As the notation indicates, the  $e_t$ 's are normal and independent with mean zero and variance  $\sigma^2/P_t$ , and the  $\underline{v}_t$ 's are independent and multivariate normal with mean vector zero and covariance matrix  $\sigma^2 A$ . Furthermore  $e_t$  and  $\underline{v}_t$  are mutually independent. Actuaries have used models of this type before. DE JONG and ZEHNWIRTH (1983), for example, use these models in the credibility context and show that the data generating equation of traditional credibility models can be formulated in this form. NEUHAUS (1987) applied this type of model to the prediction of number of policies, claim frequency and mean severity, and he discussed how to select the appropriate model and how to estimate its parameters. A recent application of these models in an insurance context is described by HARVEY and FERNANDEZ (1989) who combine a structural time series model for the size of claims with a model for the number of claims.

The simplest special case of the model in (3.1) assumes that  $p = 1$ ,  $x_t = 1$  and  $T = 1$ . This model allows the mean level  $\beta_t$  of the series to change over time according to a random walk,  $\beta_t = \beta_{t-1} + v_t$ . The exponentially weighted moving average forecasts that arise from this model (see ABRAHAM and LEDOLTER (1986), for example) are a special case of the recursive credibility model discussed by GERBER and JONES (1975) and its generalization by SUNDT (1981). If  $\text{Var}(v_t) = 0$ , implying that the coefficients  $\beta_t = \beta$  are time-invariant, then this model simplifies to the data generating equation of the Bühlmann-Straub model.

Another special case of interest is the model with a time-varying linear trend component where

$$(3.2) \quad \underline{x}_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \underline{\beta}_t = \begin{bmatrix} \beta_{0t} \\ \beta_{1t} \end{bmatrix}, \quad T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

This model allows the slope  $\beta_{1t} = \beta_{1,t-1} + v_{2t}$  and the intercept  $\beta_{0t} = \beta_{0,t-1} + \beta_{1,t-1} + v_{1t}$  to change over time. With  $\lambda_1 = \lambda_2 = 0$  the model in (3.1) reduces to the data generating equation of the Hachemeister model.

If quarterly or monthly data are analyzed, it may be necessary to incorporate a seasonal component. A model with

$$(3.3) \quad \underline{x}_t = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \underline{\beta}_t = \begin{bmatrix} \beta_{0t} \\ \beta_{1t} \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \end{bmatrix}, \quad T = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\underline{v}_t = \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

can be used for quarterly data. The first two components in  $\underline{\beta}_t$  correspond to level and slope at time  $t$ . The last three components of  $\underline{\beta}_t$  correspond to additive seasonal factors. If the  $\lambda$ 's are zero, the model reduces to the Hachemeister linear trend model with seasonal indicators.

The inference in structural time series models (3.1) is discussed in HARVEY and TODD (1983). The standard Kalman filter updating equations (see, for example, JAZWINSKI, 1970; MEINHOLD and SINGPURWALLA, 1983) are used to obtain  $\hat{\beta}_{n|n}$ , the estimate of  $\beta_n$  that is based on the observations  $Y_1, Y_2, \dots, Y_n$ . Furthermore, one can obtain its covariance matrix  $\sigma^2 G_{n|n}$ , predict future coefficients  $\beta_{n+l}$  from  $\hat{\beta}_{n+l|n} = T' \hat{\beta}_{n|n}$ , and future observations  $Y_{n+l}$  from  $\hat{Y}_n(l) = \underline{x}'_{n+l} T' \hat{\beta}_{n|n}$ .

Starting values are needed to initialize the Kalman filter recursions

$$(3.4) \quad \begin{aligned} \hat{\beta}_{t|t-1} &= T \hat{\beta}_{t-1|t-1} \\ \hat{\beta}_{t|t} &= \hat{\beta}_{t|t-1} + k_t (Y_t - \underline{x}'_t \hat{\beta}_{t|t-1}) \\ G_{t|t-1} &= T G_{t-1|t-1} T' + A \\ G_{t|t} &= G_{t|t-1} - k_t \underline{x}'_t G_{t|t-1} \\ k_t &= G_{t|t-1} \underline{x}_t (\underline{x}'_t G_{t|t-1} \underline{x}_t + P_t^{-1})^{-1}. \end{aligned}$$

For a single series in (3.1) we start these recursions with a  $p \times 1$  vector of zeros for  $\hat{\beta}_{0|0}$  and a diagonal matrix with very large diagonal elements for  $G_{0|0}$ . This

non-informative initialization reflects our ignorance about starting values in the absence of prior data. Other initialization approaches are possible (ANSLEY and KOHN, 1985; KOHN and ANSLEY, 1986; DEJONG, 1988), and their relationships are discussed in LEDOLTER, KLUGMAN and LEE (1989).

With a non-informative prior distribution the Kalman filter estimate  $\hat{\beta}_{n|n}$  is an unbiased estimator of the coefficient at time  $n$ ,  $\beta_n$ . The estimate is a weighted average of the  $n$  past observations. In general, older observations receive less weight if there is evidence that the coefficients are time-changing.

The Kalman filter updating equations, and therefore the estimate  $\hat{\beta}_{n|n}$  and the forecast  $\hat{Y}_n(l)$ , depend on the variance ratios  $A$  in equation (3.1). These parameters are estimated by maximum likelihood. The likelihood function of  $\sigma^2$  and  $A$  is obtained from the prediction error decomposition (SCHWEPPE, 1965). Assuming a non-informative initialization the log-likelihood function can be written as

$$(3.5) \quad l(\sigma^2, A; \text{data}) = c - \frac{n-p}{2} \log \sigma^2 - \frac{1}{2} \sum_{t=p+1}^n \log f_t \\ - \frac{1}{2\sigma^2} \sum_{t=p+1}^n (Y_t - \mathbf{x}'_t \hat{\beta}_{t|t-1})^2 / f_t,$$

where  $Y_t - \mathbf{x}'_t \hat{\beta}_{t|t-1}$  is the one-step-ahead prediction error at time  $t$ , and  $\sigma^2 f_t$  is its variance;  $\hat{\beta}_{t|t-1}$  and  $f_t = P_t^{-1} + \mathbf{x}'_t G_{t|t-1} \mathbf{x}_t$  can be obtained from the Kalman filter recursions. The maximization is simplified by the fact that one can concentrate the log-likelihood function with respect to  $\sigma^2$ ; the numerical maximization of the concentrated log-likelihood  $l_c(A; \text{data})$  needs to be carried out for elements in  $A$  only.

### 3.2. Analysis of multiple series and the introduction of shrinkage

So far we have discussed the analysis of a single series with time-varying coefficients. In insurance applications we not only have a single series, but we have  $n$  observations from  $k$  groups, and the estimation of  $A$  can be improved by incorporating information from the other groups. Here we assume that the  $A$  in the  $k$  groups are the same. As the value of  $n$  is usually small relative to  $k$ , it is not possible to estimate separate variance ratios for each series. Assuming independence across the  $k$  groups we can add the log-likelihood functions in (3.5) for the  $k$  groups and obtain estimates of a common  $A$  via numerical optimization. An estimate of the variance  $\sigma^2$  is obtained from

$$(3.6) \quad \hat{\sigma}^2 = \frac{1}{(n-p)k} \sum_{i=1}^k \sum_{t=p+1}^n (Y_t^{(i)} - \mathbf{x}'_{it} \hat{\beta}_{t|t-1}^{(i)})^2 / f_t^{(i)}.$$

The estimate of  $A$  is used to carry out the Kalman filter recursions. This is done for each group separately, using a non-informative initialization. The resulting coefficient estimate  $\hat{\beta}_{n|n}^{(i)}$  provides us with an estimate of the parameter

at time  $n$ ,  $\beta_n^{(i)}$ ; its covariance matrix is given by  $\sigma^2 G_{n|n}^{(i)}$ . The estimate is a weighted average of the  $n$  observations. The estimate of  $A$  determines the weights in this average. Positive variance ratios in  $A$  imply that the importance of each observation in determining the estimate depends on its age. If the variance ratios are zero, then the Kalman filter estimates simplify to the usual regression estimates  $\hat{\beta}^{(i)}$  in equation (2.4).

So far there has been no shrinkage, as we have ignored the cross-sectional correlations. In order to effect shrinkage we introduce a second equation,

$$(3.7) \quad \beta_n^{(i)} = \underline{b}_n + q_n^{(i)} \quad \text{where} \quad q_n^{(i)} \sim N(0, \sigma^2 B_n).$$

This equation specifies that at time  $n$  the coefficient vectors in the structural time series model for the  $k$  groups vary independently around a common value  $\underline{b}_n$ . We combine this equation with the results from the  $k$  separate Kalman filters,

$$(3.8) \quad \hat{\beta}_{n|n}^{(i)} = \underline{b}_n + w_n^{(i)} \quad \text{where} \quad w_n^{(i)} \sim N(0, \sigma^2 G_{n|n}^{(i)})$$

are independent across groups. These two equations yield the standard two-stage credibility model in Section 2. The shrinkage estimate based on (3.7) and (3.8) is given by

$$(3.9) \quad \tilde{\beta}_n^{(i)} = Z_i \hat{\beta}_{n|n}^{(i)} + (I - Z_i) \underline{b}_n,$$

where  $Z_i = B_n (B_n + V_i)^{-1}$  and  $V_i = G_{n|n}^{(i)}$ . The results in Section 2 can be used to estimate  $\underline{b}_n$  and  $B_n$ . In our examples we have used deVylder's iterative approach discussed in Section 2.

### 3.3. Discussion

Adding this second equation to induce shrinkage is somewhat heuristic, but is needed as by itself the model in equation (3.1) does not incorporate cross-sectional correlations.

In theory, a cross-correlation structure can be introduced by specifying a certain covariance structure for the error terms in a multivariate version of the model in (3.1). However, it is usually quite difficult to identify the exact form of the cross-correlation structure, especially for the short time series which are typical with insurance data. We have avoided these modelling issues by introducing a heuristic shrinkage equation at the last available observation period.

Model-based approaches to shrinkage are clearly possible. One alternative to the above heuristic shrinkage approach is a model that introduces a shrinkage equation for the coefficient vector at the initial time period zero. That is, one assumes that  $\beta_0^{(i)} = \underline{b}_0 + q_0^{(i)}$ , where the  $q_0^{(i)}$ , for  $i = 1, \dots, k$ , are independent realizations from a normal distribution with mean vector zero and covariance matrix  $\sigma^2 B_0$ . This implies that at the initial time period the standard actuarial shrinkage model is valid. If the elements in  $A$  are zero, implying that the coefficients in the data generating model are time-invariant, this model and the

traditional credibility model are identical. For time-varying coefficient models we start from the standard actuarial shrinkage model at time zero, but assume that the coefficients for subsequent periods are subject to stochastic change. For the inference in this model one initializes the Kalman filter in each group by the same  $\hat{\beta}_{0|0}^{(i)} = b_0$  and  $G_{0|0}^{(i)} = B_0$ , treats  $b_0$  and  $B_0$  as unknown parameters, and simultaneously obtains estimates of  $\lambda$ ,  $b_0$  and  $B_0$ . This results in shrinkage of the Kalman filter estimates  $\hat{\beta}_{n|n}^{(i)}$  at time  $n$  towards the common initial mean  $b_0$ . But even for modest positive values of  $\lambda$  this shrinkage effect disappears very quickly as  $n$  increases, and for moderate  $n$  there is hardly any shrinkage. It is for these reasons that we have rejected this alternative approach and have concentrated our discussion on the former, somewhat heuristic procedure.

Another model that introduces cross-sectional correlations is one that assumes that the  $k$  coefficients at time  $t$ ,  $\beta_t^{(i)}$ , for  $i = 1, \dots, k$ , vary independently around a common trend component  $b_t$  which itself follows a structural time series model. LEE (1991) studies these common-trend type models in detail, and we hope to report on this work in a future paper.

The advantage of our admittedly heuristic method is that it is more general than the traditional credibility approach. It recognizes the fact that most time series exhibit changing levels, trends and seasonality, and it discounts previous observations when it determines their estimates. The difference between the two approaches is shown best in the case of the Bühlmann-Straub model. The traditional approach shrinks the sample means towards a common average, whereas our new approach shrinks exponentially weighted averages. Furthermore, it can be shown that for  $\lambda = 0$  our approach coincides with the solution in Section 2.

#### 4. EXAMPLES

In this section two examples are given, with the second one being analyzed in detail. These examples provide illustrations of situations in which models that combine time-varying and shrinkage aspects are likely to improve the results.

##### 4.1. Worker's compensation

MEYERS (1984) studies yearly loss ratios under Worker's compensation insurance for 319 classes (occupation groups) and three years. A model without trend component is appropriate since these data are already adjusted for inflation. Meyers uses the Bühlmann-Straub model in his analysis. However, MEYERS and SCHENKER (1983) provide evidence that the loss ratios are not constant, but vary independently from year to year around a common mean. In the notation of our present paper

$$(4.1) \quad x_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \beta_t = \begin{bmatrix} \beta_{0t} \\ \beta_{1t} \end{bmatrix}, \quad T = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & 0 \end{bmatrix},$$



where  $\beta_{1t}$  is an unchanging long-term average and  $\beta_{0t}$  is the level in year  $t$ . An approach that combines this state-space model with shrinkage can be expected to improve the forecasts for future losses, as many of the 319 classes have very small sample sizes.

#### 4.2. Automobile bodily injury

The data for the second example are taken from the automobile insurance industry. Quarterly data on the amount (not adjusted for inflation) paid under the bodily injury component of automobile insurance policies (LOSS) and the number of cars covered by these policies (EXPOSURE) were obtained from 31 states. Only states without no-fault laws were included, as under no-fault

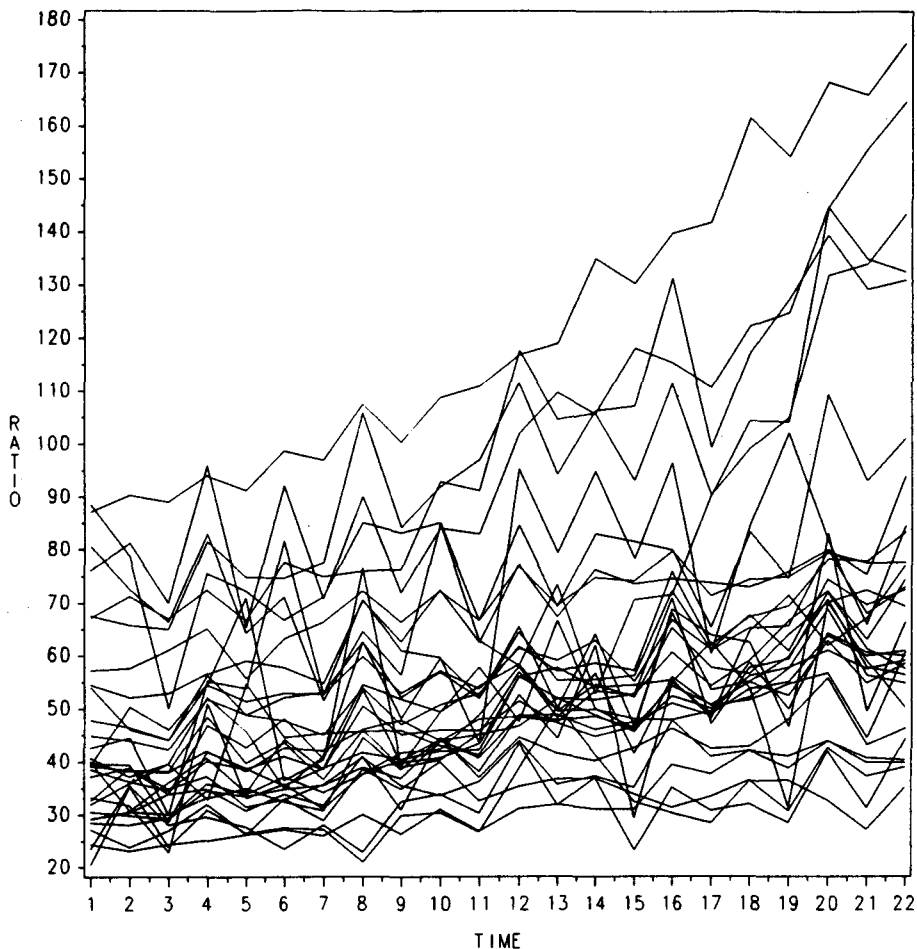


FIGURE 1. Multiple time series plot of the ratio  $R = \text{LOSS}/\text{EXPOSURE}$ .

laws many claims that would otherwise be covered by the liability portion of the insurance are paid under the bodily injury component. Data from the first quarter of 1983 to the second quarter of 1988 ( $n = 22$ ) are used in our analysis.

The ratio  $R_t^{(i)} = \text{LOSS}_t^{(i)} / \text{EXPOSURE}_t^{(i)}$ , where  $t = 1, \dots, 22$  (quarters) and  $i = 1, \dots, 31$  (states) is our dependent variable that needs to be predicted. The multiple time series plot of the ratios  $R_t^{(i)}$  in Figure 1 shows presence of seasonality and a need for a logarithmic transformation. The presence of seasonality is seen more clearly in Figure 2 where we have plotted estimates of the multiplicative seasonal indices for the 31 states. We use the following procedure to obtain the seasonal indices: For each univariate series we calculate centered yearly moving averages to estimate the trend component; we then obtain, for each time period  $t$ , an estimate of the seasonal factor from the ratio of the observation and the corresponding centered moving average; next, we average the seasonal factors for each quarter to obtain seasonal indices for the four quarters; finally, we normalize these indices so that they sum to four. The dot plot of these normalized seasonal indices in Figure 2 shows a seasonal pattern; in the fourth quarter the ratios  $R_t^{(i)}$  tend to be highest.

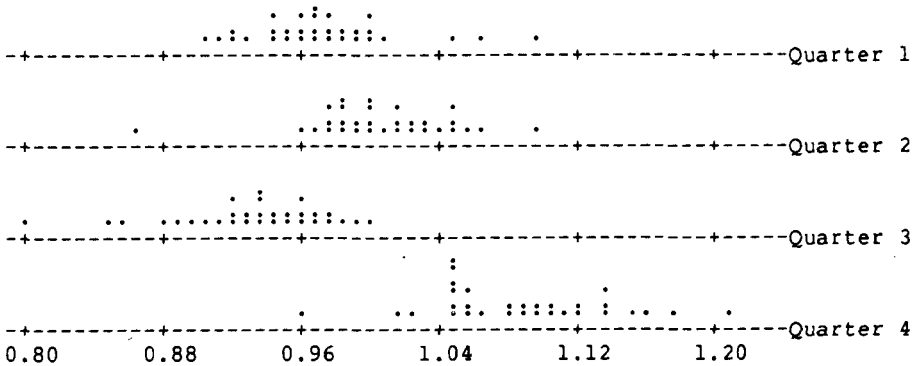


FIGURE 2. Dot plot of the seasonal indices for the 31 states.

A multiple time series plot of the transformed observations,  $Y_t^{(i)} = \log R_t^{(i)}$ , is given in Figure 3. This plot indicates that a linear trend model with additive seasonal components provides a good description of the transformed observations.

In the standard actuarial model it is usually assumed that the variance of the error component is related to the exposure  $P_t^{(i)}$ ; that is,  $\text{Var}(e_t^{(i)}) = \sigma^2 / P_t^{(i)}$ . We now want to check whether this is a reasonable assumption. Since the exposures  $P_t^{(i)}$  do not change much over time, we calculate an average exposure  $\bar{P}^{(i)}$  for each state. Due to size differences among the states, these averages are quite different. Next, we adjust each time series  $Y_t^{(i)}$  for trend and seasonality and calculate an estimate of its variance. The residuals from a regression of  $Y_t^{(i)}$  on time  $t$  and additive seasonal indicators are used to calculate the variance

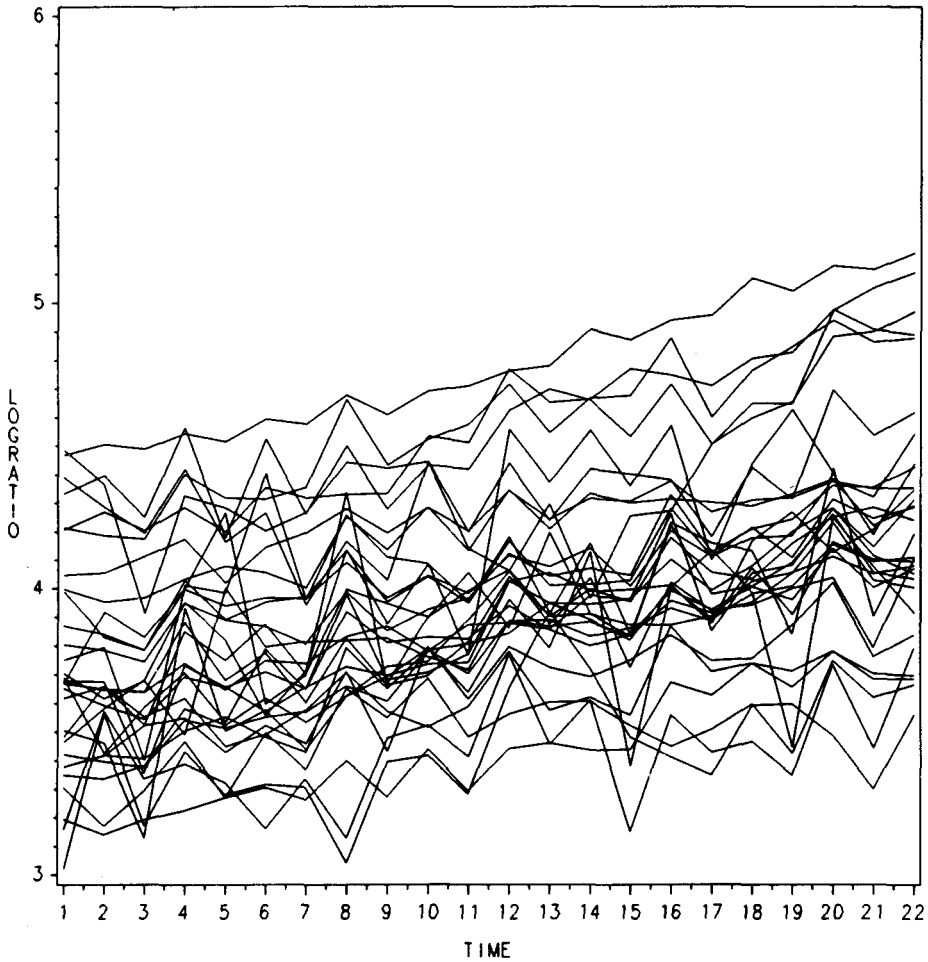


FIGURE 3. Multiple time series plot of  $Y = \log (\text{LOSS/EXPOSURE})$ .

estimate. In Figure 4 we plot the resulting mean square errors against the reciprocal of the average exposures. The linear relationship confirms that  $\text{Var}(e_t^{(i)}) = \sigma^2/P_t^{(i)}$  is a reasonable assumption.

Based on this preliminary analysis we are led to consider the structural time series model with a linear trend and additive seasonal components,

$$(4.2) \quad \begin{aligned} Y_t^{(i)} &= \underline{x}_t' \beta_t^{(i)} + e_t^{(i)} & e_t^{(i)} &\sim N(0, \sigma^2/P_t^{(i)}) \\ \beta_t^{(i)} &= T\beta_{t-1}^{(i)} + \nu_t^{(i)} & \nu_t^{(i)} &\sim N(0, \sigma^2 A) \end{aligned}$$

where  $\underline{x}_t$ ,  $T$  and  $A$  are given in equation (3.3). Our model allows for time-varying coefficients and reduces to a linear trend regression model with quarterly indicators if  $\lambda_1 = \lambda_2 = \lambda_3 = 0$ .

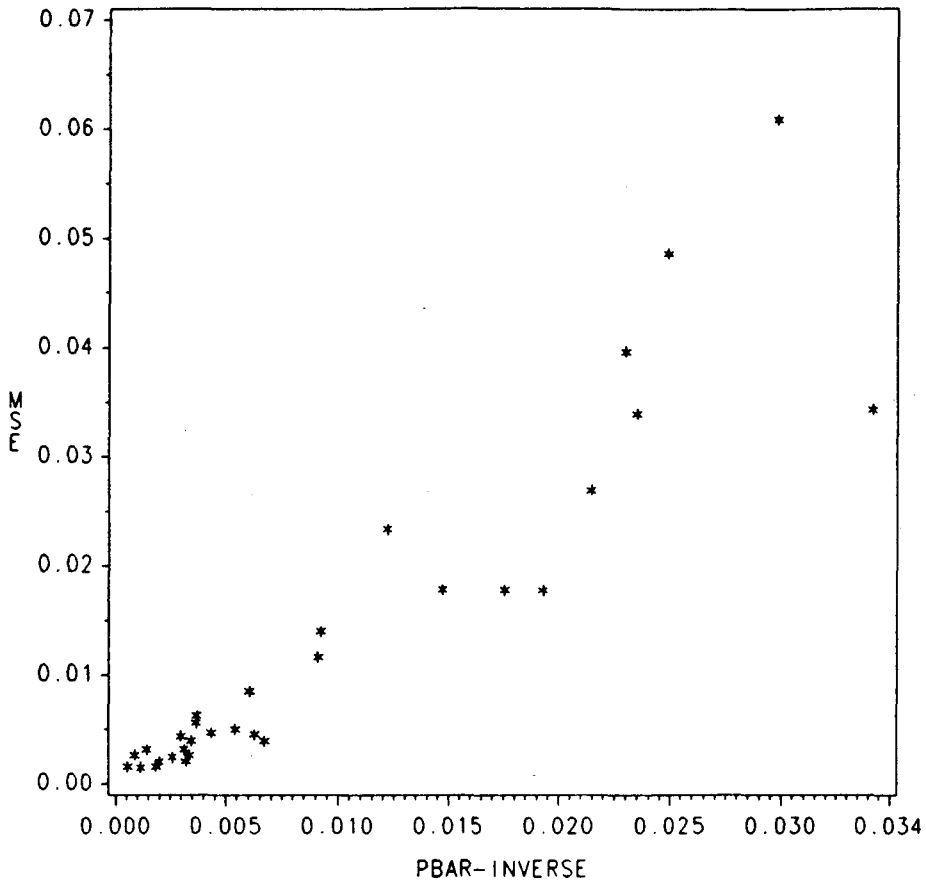


FIGURE 4. Plot of the mean square error from the linear trend regression with seasonal indicators against the reciprocal of average exposure;  $k = 31$  states.

The maximum likelihood estimation approach in Section 3 is applied and, assuming independent groups, estimates of  $\sigma^2$  and the three variance ratios are obtained. It is found that  $\hat{\sigma}^2 = 3.8089 \cdot 10^{-3}$ ,  $\hat{\lambda}_1 = 0.0495$ ,  $\hat{\lambda}_2 = 0.0044$  and  $\hat{\lambda}_3 = 0.00008$ . The estimate  $\hat{\lambda}_3$  is close to zero and the log-likelihood deficiency (ratio),  $l_c(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3) - l_c(\hat{\lambda}_1, \hat{\lambda}_2, 0)$ , is quite small. This implies that the seasonal coefficients do not change much over time. Contours of the log-likelihood function of  $\lambda_1$  and  $\lambda_2$ , for  $\hat{\lambda}_3 = 0.00008$ , are plotted in Figure 5. This plot, as well as the large log-likelihood deficiency  $l_c(\hat{\lambda}_1, \hat{\lambda}_2, 0) - l_c(0, 0, 0) = 19.16$ , shows that a standard least squares approach which assumes time constant intercept and slope coefficients would be inappropriate.

In order to check the adequacy of the structural time series model in equation (4.2) we calculated the standardized one-step-ahead forecast errors for periods 6 through 22. Standardization of the forecast error by its standard error  $\hat{\sigma}f_t^{1/2}$  assures that its variance does not depend on time. We found that

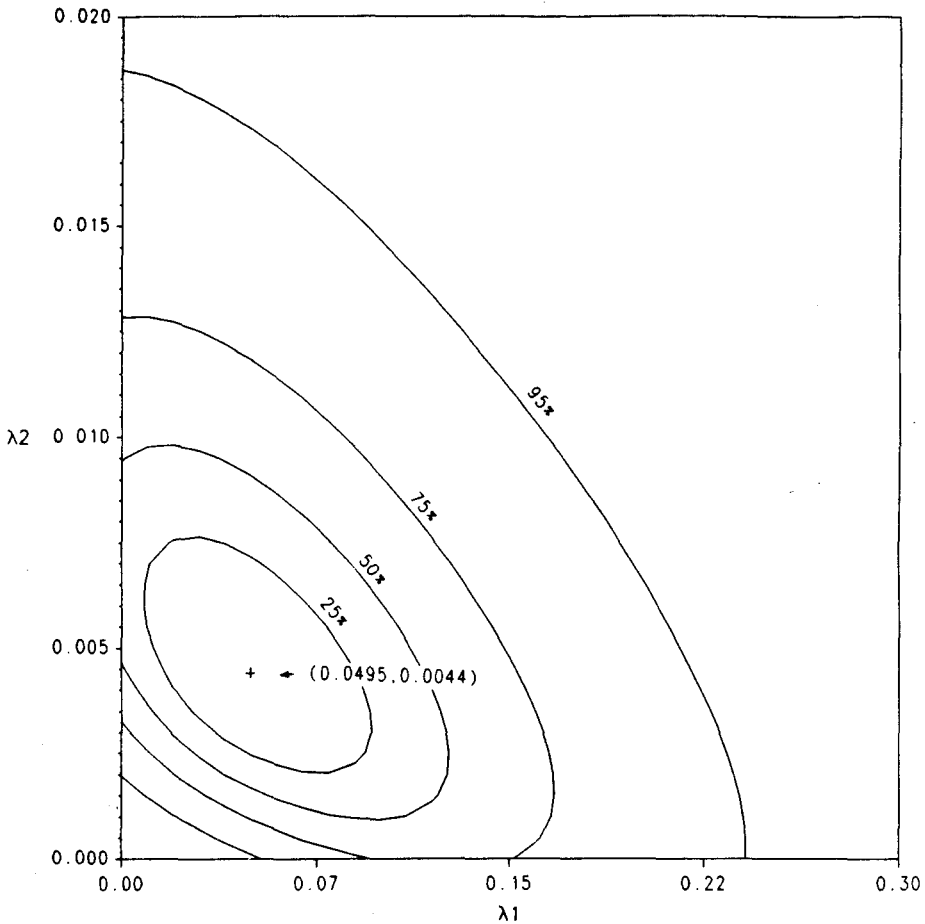


FIGURE 5. Contour plot of the concentrated log-likelihood function. The variance ratio that determines the variability in the seasonal component is set at its estimate  $\hat{\lambda}_3 = 0.00008$ . Probability coverage of the contours is obtained from the Chi-square approximation.

the standardized one-step-ahead forecast errors were serially uncorrelated for essentially all 31 groups.

The estimates of  $\hat{\lambda}_1$ ,  $\hat{\lambda}_2$  and  $\hat{\lambda}_3$  are used to calculate the estimates  $\hat{\beta}_{n/n}^{(i)}$ , for  $n = 22$  (the last available time period) and  $i = 1, \dots, 31$  (states). Dot diagrams of the  $k = 31$  estimates of intercept, slope and seasonal coefficients (only the first one is shown), together with their standard errors, are given in Figure 6. The standard errors are obtained from the diagonal elements in  $\hat{\sigma}^2 G_{n/n}^{(i)}$ .

We notice considerable variability among the  $k = 31$  intercept estimates. Furthermore, we find that the between group variability is much larger than the uncertainty that is associated with each estimate (that is, the within group variability as measured by the standard error of the estimate). This result indicates that there should be no or little benefit to shrink the intercept

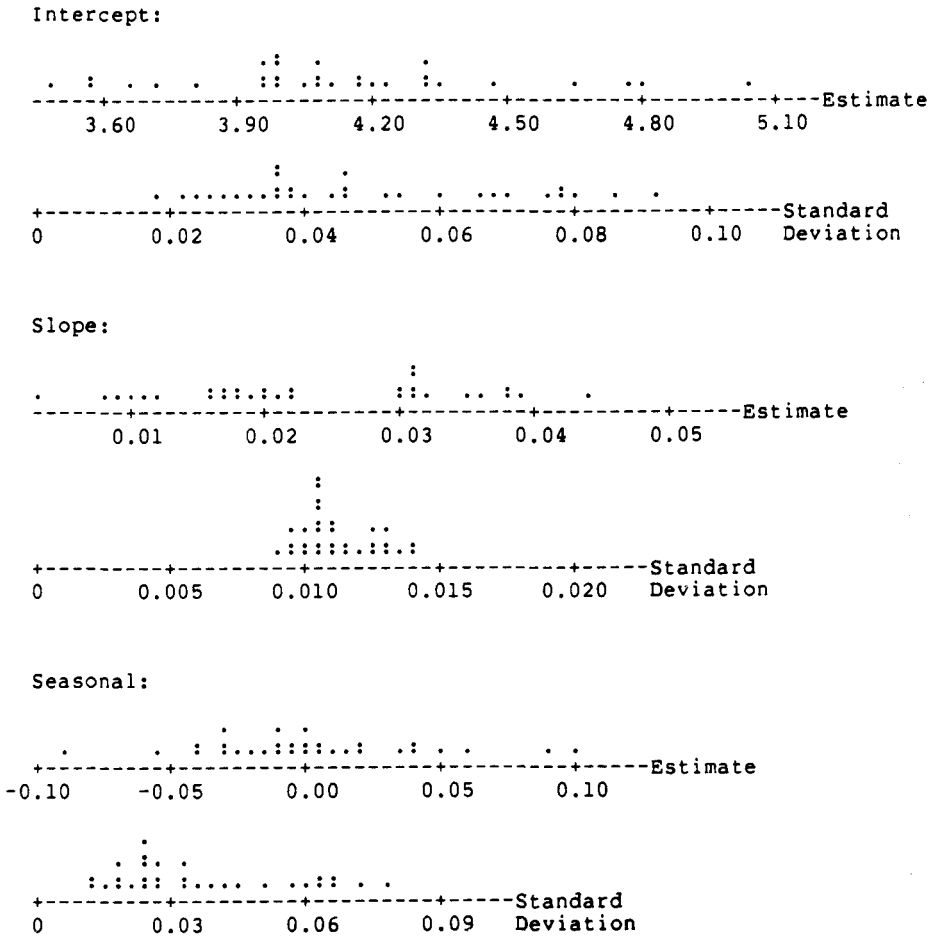


FIGURE 6. Dot plots of the estimates and their standard errors for the intercept, slope, and seasonal coefficients in model (4.2);  $k = 31$  states.

estimates. The dot plots of the slope estimates and their standard errors show a different picture; the within group variability is quite large when compared with the variability between the slope estimates. These pictures suggest that shrinkage procedures should pool the slope estimates towards a common value. The same conclusion is reached for the seasonal factors (the third, fourth and fifth component of the beta vector). They, too, should be shrunk towards common means.

Next, we apply shrinkage and calculate the shrinkage estimate discussed in equation (3.9) of Section 3. That is, we compute

$$(4.3) \quad \tilde{\beta}_n^{(i)} = Z_i \hat{\beta}_{n|n}^{(i)} + (I - Z_i) b_n$$

where  $Z_i = B_n(B_n + V_i)^{-1}$  and  $V_i = G_{n|n}^{(i)}$ . DeVylder's modification in (2.7) is used to estimate  $\hat{b}_n$  and  $B_n$ . The only minor difference is that we are using the maximum likelihood estimate  $\hat{\sigma}^2 = 3.8089 * 10^{-3}$  from the Kalman filter as the estimate of  $\sigma^2$ . In Figure 7 we compare the estimates before and after shrinkage. The graphs confirm what we had anticipated from the results in Figure 6. The slopes and seasonal components are shrunk towards their respective means, whereas the intercepts are essentially unchanged.

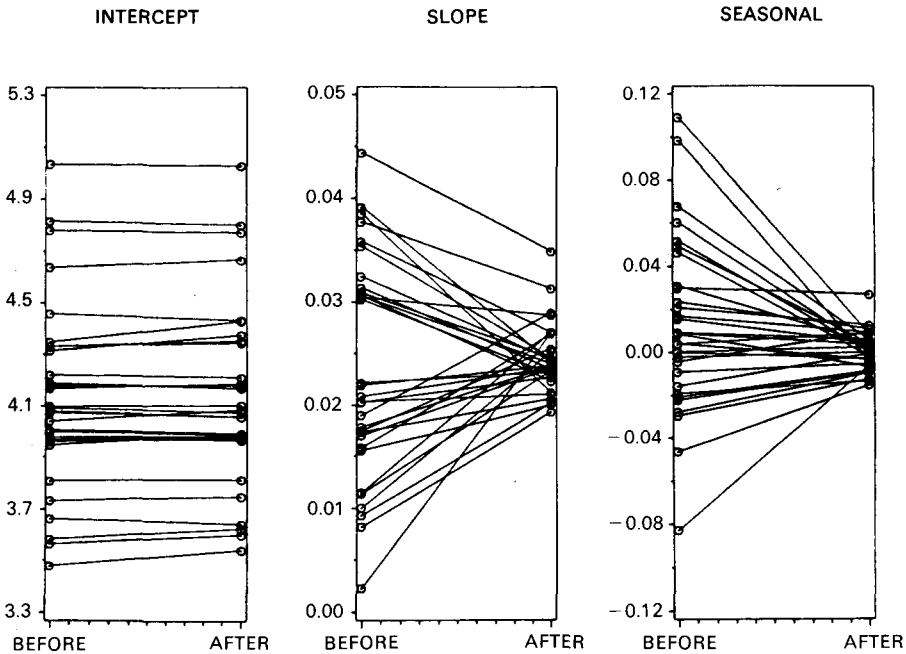


FIGURE 7. Intercept, slope and first seasonal coefficient estimate in model (4.2) before and after shrinkage;  $k = 31$  states.

**Forecast comparisons**

The prediction of future values is a major reason for fitting models to data. We must now investigate whether the proposed new approach leads to forecast improvements. In particular, we address the following two questions:

- (1) Has shrinkage of the coefficients improved the forecasting performance of our time-varying trend component model? To address this issue we compare forecasts that are calculated from the shrinkage estimates  $\tilde{\beta}_n^{(i)}$  in (4.3) [method 1] and forecasts that are calculated from the standard Kalman filter estimates  $\hat{\beta}_{n|n}^{(i)}$  [method 2].

- (2) Has our generalization of incorporating time-varying trend components helped the forecasting? To investigate this question we compare the forecasts that use the shrinkage estimates  $\hat{\beta}_n^{(i)}$  in (4.3) [method 1] with forecasts that are calculated from the shrinkage estimates in the standard regression model with constant linear trend and seasonal indicators [Hachemeister, method 3].

A true test of the forecast performance of a model is obtained by an out-of-sample comparison of forecasts and actual observations. Here we use the last four observations  $R_{19}^{(i)}$  through  $R_{22}^{(i)}$ , for  $i = 1, \dots, 31$ , as our hold-out sample. This is a reasonable choice as actuarial practice bases predictions of future premiums on about four to five years of past data. For each state we calculate four one-step-ahead forecast errors  $R_t - \hat{R}_{t-1}(1)$ ,  $t = 19, \dots, 22$ , where  $\hat{R}_t(1) = \exp[\hat{Y}_t(1)]$  is obtained by applying the inverse transformation to the forecast of the logarithmically transformed data. For each state separately, we then compute the mean square error MSE, the mean absolute deviation (error) MAD, and the mean absolute percent error MAPE. For each measure (MSE, MAD, MAPE) and for each method (methods 1 through 3) we calculate a weighted average that combines the information from the 31 states. The average exposures  $\bar{P}^{(i)}$ ,  $i = 1, \dots, 31$ , are used as weights. The results are given in Table 1. Table 1 also shows the results of a further refinement of method 1 (Kalman filter with shrinkage). In method 1R we shrink the last 4 components of the 5-dimensional coefficient vectors, but leave the first components (intercepts) unchanged.

TABLE 1  
WEIGHTED AVERAGES OF ACCURACY MEASURES. AVERAGE EXPOSURES ARE USED TO COMBINE  
THE INFORMATION FROM  $k = 31$  STATES

	Method 1 Kalman filter model (4.2) with shrinkage		Method 2 Kalman filter model (4.2) without shrinkage	Method 3 Hachemeister constant linear trend & seasonal indicator model with shrinkage
	1	1R		
MSE	32.28	31.88	39.24	38.02
MAD	3.75	3.70	4.20	4.32
MAPE	5.12	4.99	5.35	5.40

In addition to the comparison of the aggregate measures, we compare the measures for each state separately. We assign a score of 1 if in state  $i$  the first method leads to a lower MSE (MAD, MAPE) than the second. The proportion of states where method 1 outperforms method 2 (method 3) is given in Table 2.



TABLE 2  
PROPORTION OF STATES WHERE ONE METHOD OUTPERFORMS THE OTHER

Comparison	MSE	MAD	MAPE
Method 1 vs. Method 2	58	55	58
Method 1R vs. Method 2	71	65	61
Method 1 vs. Method 3	61	58	61
Method 1R vs. Method 3	65	55	55
Method 2 vs. Method 3	55	52	55
Method 1R vs. Method 1	52	48	52

**Comments.** (i) For shrinkage methods we calculate the forecasts  $\hat{Y}_t(l)$  after shrinking the estimates that are obtained at time  $t$ . We carry out a new shrinkage if we go to another forecast origin. (ii) The Kalman filter methods 1 and 2 require estimates of the variance ratios  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ . In order to avoid the numerical maximization of the log-likelihood for each forecast origin  $t$ , we use the estimates that are obtained from the complete data set ( $n = 22$ ). (iii) The transformation  $\hat{R}_t(l) = \exp[\hat{Y}_t(l)]$  results in the median of the predictive distribution of  $R_{t+l}$ . The mean of the predictive distribution can be obtained by incorporating the variance of the predictive distribution into the inverse transformation (see GRANGER and NEWBOLD, 1976). Because differences are usually relatively minor and because it is not obvious whether the mean of the posterior distribution is preferable to the median we have not pursued this adjustment.

### Interpretation of results

Table 1 shows that we can improve the one-step-ahead forecast performance if we allow the trend and the seasonal components to change over time. Comparing the results of the two shrinkage methods (methods 1 and 3) we find that the structural time series model in (4.2) leads to a 15.1. (16.1), 13.2 (14.4), and 5.2 (7.6) percent reduction in MSE, MAD, and MAPE, when it is compared to the Hachemeister model with fixed trend and seasonal components. The numbers in parentheses reflect the improvements if shrinkage is not applied to the intercepts in the structural time series model. Table 2 leads to a similar conclusion. The one-step-ahead forecasts from the structural time series model with shrinkage outperform the forecasts from the Hachemeister model in roughly 60 percent of the states (the proportion varies from 55 to 65 percent, depending on the accuracy measure that is used in the comparison).

Tables 1 and 2 also show that shrinkage of the coefficients improves the forecasts in the structural time series model (4.2). The size of the improvements that are due to shrinkage (method 1 vs method 2) is roughly the same as the one we obtain by allowing the trend and seasonal coefficients in the two

shrinkage methods to change over time (method 1 vs method 3). There is very little difference between the forecasts from the structural time series model without shrinkage and the Hachemeister shrinkage model with fixed trend and seasonal coefficients (method 2 vs method 3).

This example shows the feasibility of an approach that applies shrinkage to the coefficient estimates in structural time series models and illustrates its potential for forecast improvements. GARCIA-FERRER et al. (1987) and ZELLNER and HONG (1989) reach a similar conclusion in their analysis of macroeconomic data. They find that individual country growth rate forecasts are improved by shrinking the forecasts to a common average. However, their shrinkage methods are somewhat different from the ones considered in this paper. Furthermore, they apply shrinkage primarily to forecasts and not to estimates in time-varying coefficient models.

#### ACKNOWLEDGMENTS

STUART KLUGMAN'S research was in part supported by the Huebner Foundation. The authors wish to thank GLENN MEYERS and the Insurance Services Office for the data in Section 4. They also thank WILLIAM BELL, ANDREW HARVEY, WILLIAM WECKER and the referees for helpful comments.

#### REFERENCES

- ABRAHAM, B. and LEDOLTER, J. (1983) *Statistical Methods for Forecasting*. New York: Wiley.
- ABRAHAM, B. and LEDOLTER, J. (1986) Forecast Functions Implied by ARIMA Models and Other Related Forecast Procedures. *International Statistical Review* **54**, 51–66.
- ANSLEY, C. and KOHN, R. (1985) Estimation, Filtering, and Smoothing in State Space Models with Incompletely Specified Initial Conditions. *Annals of Statistics* **13**, 1286–1316.
- BAILEY, R.A. and SIMON, L.J. (1959) An Actuarial Note on the Credibility of Experience of a Single Private Passenger Car. *Proceedings of the Casualty Actuarial Society* **46**, 159–164.
- BOX, G. and JENKINS, G. (1976) *Time Series Analysis, Forecasting, and Control* (2nd ed.). San Francisco: Holden-Day.
- BÜHLMANN, H. and STRAUB, E. (1972) Credibility for Loss Ratios. *Actuarial Research Clearing House*, Number 2.
- DE JONG, P. (1988) The Likelihood for a State Space Model. *Biometrika* **75**, 165–169.
- DE JONG, P. and ZEHNWIRTH, B. (1983) Credibility Theory and the Kalman Filter. *Insurance: Mathematics and Economics* **2**, 281–286.
- DEVYLDER, F. (1981) Practical Credibility Theory with Emphasis on Optimal Parameter Estimation. *ASTIN Bulletin* **12**, 115–131.
- DEVYLDER, F. (1984) Practical Models in Credibility Theory, Including Parameter Estimation. In: *Premium Calculation in Insurance*, DEVYLDER, F., GOOVAERTS, M. and HAEZENDONCK, J., editors, Reidel.
- GARCIA-FERRER, A., HIGHFIELD, R.A., PALM, F. and ZELLNER, A. (1987) Macroeconomic Forecasting using Pooled International Data. *Journal of Business and Economic Statistics* **5**, 53–67.
- GERBER, H. and JONES, D. (1975) Credibility Formulas of the Updating Type. *Transactions of the Society of Actuaries* **27**, 31–46.
- GRANGER, C.W.J. and NEWBOLD, P. (1976) Forecasting Transformed Series. *Journal of the Royal Statistical Society, Series B* **38**, 189–203.
- HACHEMEISTER, C. (1975) Credibility for Regression Models with Applications to Trend. In *Credibility: Theory and Applications*, P. KAHN, ed., New York: Academic Press.

- HARVEY, A. (1984) A Unified View of Statistical Forecasting Procedures. *Journal of Forecasting* **3**, 245–283.
- HARVEY, A. and TODD, P. (1983) Forecasting Economic Time Series with Structural and Box-Jenkins Models: A Case Study. *Journal of Business and Economic Statistics* **1**, 299–315.
- HARVEY, A. and FERNANDES, C. (1989) Time Series Models for Insurance Claims. *Journal of the Institute of Actuaries* **116**, Part 3, 513–528.
- JAZWINSKI, A. (1970) *Stochastic Processes and Filtering Theory*. New York: Academic Press.
- KLUGMAN, S. (1987) Credibility for Classification Ratemaking via the Hierarchical Linear Model. *Proceedings of the Casualty Actuarial Society* **74**, 272–321.
- KOHN, R. and ANSLEY, C. (1986) Estimation, Prediction, and Interpolation for ARIMA Models with Missing Data. *Journal of the American Statistical Association* **81**, 751–761.
- LEDOLTER, J., KLUGMAN, S. and LEE, C. S. (1989) Credibility Models with Time-Varying Trend Components, Technical Report 159, Department of Statistics and Actuarial Science, University of Iowa.
- LEE, C. S. (1991) Time Series Models for the Credibility Estimation of Insurance Premiums, unpublished Ph.D. dissertation (forthcoming), Department of Statistics and Actuarial Science, University of Iowa.
- MEINHOLD, R. and SINGPURWALLA, N. (1983) Understanding the Kalman Filter. *The American Statistician* **37**, 123–127.
- MEYERS, G. (1984) Empirical Bayesian Credibility for Workers' Compensation Classification Ratemaking. *Proceedings of the Casualty Actuarial Society* **71**, 96–121.
- MEYERS, G. and SCHENKER, N. (1983) Parameter Uncertainty in the Collective Risk Model. *Proceedings of the Casualty Actuarial Society* **70**, 111–143.
- MOWBRAY, A. (1914) How Extensive a Payroll is Necessary to Give a Dependable Pure Premium? *Proceedings of the Casualty Actuarial Society* **1**, 24–30.
- NEUHAUS, W. (1987) Early Warning. *Scandinavian Actuarial Journal*, 128–156.
- SCHWEPPE, F. (1965) Evaluation of Likelihood Functions for Gaussian Signals. *IEEE Trans. Inf. Theory* **11**, 61–70.
- SUNDT, B. (1981) Recursive Credibility Estimation. *Scandinavian Actuarial Journal*, 3–21.
- SWAMY, P. A. V. B. (1971) *Statistical Inference in Random Coefficient Regression Models*. New York: Springer.
- ZELLNER, A. and HONG, C. (1989) Forecasting International Growth Rates Using Bayesian Shrinkage and other Procedures. *Journal of Econometrics* **40**, 183–202.

JOHANNES LEDOLTER

*Department of Statistics and Actuarial Science*

*Department of Management Sciences*

*The University of Iowa*

*Iowa City, IA 52242*

STUART KLUGMAN

*College of Business and Public Administration*

*Drake University*

*Des Moines, IA 50311*

CHANG-SOO LEE

*Department of Statistics and Actuarial Science*

*The University of Iowa*

*Iowa City, IA 52242*