A HIERARCHICAL KALMAN FILTER

Greg Taylor

Taylor Fry Consulting Actuaries
Level 8, 30 Clarence Street
Sydney NSW 2000
Australia

Professorial Associate,
Centre for Actuarial Studies
Faculty of Economics and Commerce
University of Melbourne
Parkville VIC 3052
Australia

Phone: 61 2 9249 2901
Fax: 61 2 9249 2999
greg.taylor@taylorfry.com.au

March 2009

Table of Contents

Summary

Sundt's hierarchical credibility model is generalised to a dynamic form, ie a form in which parameters are assumed to evolve over time. This is done by superimposing a Kalman filter on Sundt's model.

In the process it is shown that both Sundt's model and the Kalman filter may be derived as direct consequences of Hachemeister's credibility regression model.

A numerical example in Section 7 illustrates the application of the hierarchical Kalman filter to a hierarchy of occupational groups. The example shows how parameter estimates produced by the filter track the true values of parameters better than the estimates from a static model when those parameters evolve over time.

Keywords: credibility, Hachemeister model, hierarchical credibility, Kalman filter.

1. Introduction

Hierarchical credibility was introduced by Taylor (1979). The parameters were scalars, as were the observations on them.

This framework was generalised by Sundt (1979, 1980) to include vector parameters and observations. In Sundt's framework a linear regression model is placed at each level of the hierarchy. The model is **static** in the sense that all parameters relate to just a single epoch or time interval.

If, in fact, data were collected over multiple time periods, they would be pooled on the assumption that the parameters remained constant over time. Such pooling of data renders the model sluggish in its response to evolving parameters.

One structure that recognises parameter evolution explicitly is the **Kalman filter** (Kalman, 1960; Jazwinski, 1970). The purpose of the present paper is to adapt Sundt's hierarchical model to a dynamic form by superimposing a Kalman filter on it. It is anticipated that the parameter estimates yielded by the dynamic model will track the evolving parameters better than their static counterparts.

The Kalman filter is derived in Section 4 and Sudnt's results in Section 5. it is shown in the process that both can be viewed as consequences of **Hachemeister's (1975) multi-dimensional credibility model**.

The hierarchical form of the Kalman filter is derived in Section 6. Its use is illustrated in Section 7, where it is applied to the original context of hierarchical credibility (Taylor, 1979). In this case the hierarchy is a tree of **occupational groups** in a **workers compensation** setting.

2. Credibility estimation

Let m be an unknown vector random variable, and let x be an observable vector random variable. A **linear credibility estimator** (or just credibility estimator) of m (based on x) is the linear function $L(x)$ of x which minimises $E\left[L(x)-m\right]^2$.

The linear credibility estimator $L(x)$ is unbiased for m.

3. Standard (multi-dimensional) credibility

Let θ be an unknown random parameter and x an observable random vector. Assume the following:

(3.i) $x = Yb(\theta)+\xi$ for some vector-valued function $b(.)$, deterministic matrix Y, and centred random vector ξ;

(3.ii) θ and ξ are independent.

Let

$$\beta = E\left[b(\theta)\right] \tag{3.1}$$

$$\Lambda = Var\left[b(\theta)\right] \tag{3.2}$$

$$\Phi = Var\left[x \mid \theta\right] = Var\left[\xi\right] \tag{3.3}$$

This is the Hachemeister (1975) framework, and the linear credibility estimator of $b(\theta)$ is

$$\tilde{b} = (1-Z)\beta + Z\hat{b} \tag{3.4}$$

where

$$Z = \Lambda Y^T \Phi^{-1} Y \left(1 + \Lambda Y^T \Phi^{-1} Y\right)^{-1} \tag{3.5}$$

and $\hat{b}$ is the classical regression estimator

$$\hat{b} = \left(Y^T \Phi^{-1} Y\right)^{-1} Y^T \Phi^{-1} x \tag{3.6}$$

Here the upper T denotes matrix transposition.

Extension. Relax the condition that ξ be centred. Then results (3.4) and (3.5) continue to hold but with (3.6) replaced as follows:

$$\hat{b} = \left(Y^T \Phi^{-1} Y\right)^{-1} Y^T \Phi^{-1} \left[x - E\left[\xi\right]\right] \tag{3.6a}$$

Proof. Write

$$u = x - E\left[\xi\right]$$

Then, by Assumption (3.i),

$$u = Yb(\theta) + \eta$$

where is the centred random vector

$$\eta = \xi - E\left[\xi\right]$$

Then results (3.4) – (3.6) hold with *x* replaced by *u*.

□

4. Kalman filter

4.1 Standard Kalman filter

Let b^t, t = 2, 3, etc be an unknown vector parameter that evolves randomly over time, and let x^t be an observable random vector, where b^t and x^t satisfy the following relations.

(4.i) $b^{t+1} = A^{t+1}b^t + \varepsilon^{t+1}$, t = 1, 2, etc

where A^{t+1} is a deterministic square matrix and ε^{t+1} a centred random vector, and the system is initiated by a deterministic vector b^1.

(4.ii) $x^t = Y^t b^t + \xi^t$, t = 1, 2, etc

where Y^t is a deterministic matrix and ξ^t a centred random vector.

(4.iii) Each of the sets $\{\varepsilon^t\}$ and $\{\xi^t\}$ is mutually stochastically independent, and further the two sets are stochastically independent of each other.

Let $\tilde{b}^{t|s}$ be the (unbiased) credibility estimator of b^t based on $\{x^1, x^2, ... x^s\}$. Specifically, on account of Assumption (4.i), define

$$\tilde{b}^{t+1|t} = A^{t+1}\tilde{b}^{t|t} \qquad (4.1)$$

It is straightforward to show that $\tilde{b}^{t+1|t}$ is the credibility estimator of b^{t+1} based on $\{x^1, ... x^t\}$.

Let

$$\Phi^t = Var\left[x^t \mid b^t\right] = Var\left[\xi^t\right] \qquad (4.2)$$

$$\Psi^t = Var\left[\varepsilon^t\right] \qquad (4.3)$$

$$\Lambda^t = E\left[b^t - \tilde{b}^{t|t-1}\right]^2 \qquad (4.4)$$

The credibility estimator $\tilde{b}^{t|t-1}$ is the linear function of x^1, …, x^{t-1} that minimises Λ^t.

By Assumption (4.i) and (4.1),

$$\begin{aligned}\Lambda^t &= E\left[\left(A^t b^{t-1} + \varepsilon^t\right) - A^t \tilde{b}^{t-1|t-1}\right]^2 \\ &= A^t\, Var\left[\tilde{b}^{t-1|t-1}\right]\left(A^t\right)^T + \Psi^t\end{aligned} \qquad (4.5)$$

where use has been made of Assumption (4.iii) and the fact that $\tilde{b}^{t-1|t-1}$ does not depend on ε^t.

The mathematics of finding the credibility estimator $\tilde{b}^{t|t}$ is the same as in Section 3 but with the following replacements:

$x \leftarrow x^t$
$Y \leftarrow Y^t$
$b(\theta) \leftarrow b^t$
$\xi \leftarrow \xi^t$
$\beta \leftarrow \tilde{b}^{t|t-1}$
$\tilde{b} \leftarrow \tilde{b}^{t|t}$

The parallel of Assumption (3.ii) here is independence of b^t and ξ^t, which is guaranteed by Assumption (4.iii).

It then follows that

$\Lambda \leftarrow \Lambda^t$ [from (3.2) and (4.4)]
$\Phi \leftarrow \Phi^t$ [from (3.3) and (4.2)]

Substitution of these replacements into (3.4) and (3.5) yields

$$\tilde{b}^{t|t} = \left(1 - K^t\right)\tilde{b}^{t|t-1} + K^t \hat{b}^t \tag{4.6}$$

where

$$K^t = L^t\left(1 + L^t\right)^{-1} = \left(1 + L^t\right)^{-1} L^t \tag{4.7}$$

$$L^t = \Lambda^t \left(Y^t\right)^T \left(\Phi^t\right)^{-1} Y^t \tag{4.8}$$

$$\hat{b}^t = \left[\left(Y^t\right)^T \left(\Phi^t\right)^{-1} Y^t\right]^{-1} \left(Y^t\right)^T \left(\Phi^t\right)^{-1} x^t \tag{4.9}$$

Since $\hat{b}^t$ depends on just x^t and $\tilde{b}^{t|t-1}$ does not, (4.6) gives

$$\begin{aligned} Var\left[\tilde{b}^{t|t}\right] &= \left(1-K^t\right) Var\left[\tilde{b}^{t|t-1}\right]\left(1-K^t\right)^T + K^t\, Var\left[\hat{b}^t\right]\left(K^t\right)^T \\ &= \left(1-K^t\right)\Lambda^t\left(1-K^t\right)^T + K^t\left[\left(Y^t\right)^T\left(\Phi^t\right)^{-1} Y^t\right]^{-1}\left(K^t\right)^T \end{aligned} \tag{4.10}$$

To simplify this, temporarily suppress the upper *t*. Substitute (4.7) and (4.8) into (4.10), noting that $1 - K = (1 + L)^{-1}$, to obtain

$$Var\left[\tilde{b}^{t|t}\right] = (1+L)^{-1}\left\{\Lambda + L\left[Y^T\Phi^{-1}Y\right]^{-1}L^T\right\}\left(1+L^T\right)^{-1}$$
$$= (1+L)^{-1}\Lambda$$

With the upper *t* restored

$$Var\left[\tilde{b}^{t|t}\right] = \left(1-K^t\right)\Lambda^t \qquad (4.11)$$

Equations (4.1), (4.5) – (4.9) and (4.11) collectively show how to proceed from $\tilde{b}^{t-1|t-1}$ to $\tilde{b}^{t|t}$ and, in doing so, constitute a single iteration of the **Kalman filter** (Kalman, 1960; Jazwinski, 1970).

4.2 Extension

Relax the requirement in Assumption (4.i) that ε^{t+1} be centred. Then the results of Section 4.1 continue to hold with the single change that (4.1) is replaced by the following:

$$\tilde{b}^{t+1|t} = A^{t+1}\tilde{b}^{t|t} + E\left[\varepsilon^{t+1}\right] \qquad (4.1a)$$

Proof. Define the sequence a^t, t = 1, 2, etc by the recursion

$$a^{t+1} = A^{t+1}a^t + E\left[\varepsilon^{t+1}\right]$$
$$a^1 = 0 \qquad (4.12)$$

Then define

$$c^t = b^t - a^t \qquad (4.13)$$

By (4.13) and Assumption (4.i)

$$c^{t+1} = A^{t+1}c^t + \eta^{t+1} \qquad (4.14)$$

where η^{t+1} is the centred random vector

$$\eta^{t+1} = \varepsilon^{t+1} - E\left[\varepsilon^{t+1}\right] \qquad (4.15)$$

By (4.13) and Assumption (4.ii),

$$x^t = Y^t c^t + \zeta^t \qquad (4.16)$$

with

$$\zeta^t = Y^t a^t + \xi^t \qquad (4.17)$$

Note that (4.14) and (4.16) are of the same form as Assumptions (4.i) and (4.ii) with the following replacements:

$$b^t \leftarrow c^t$$
$$\varepsilon^t \leftarrow \eta^t$$
$$\xi^t \leftarrow \zeta^t$$

and where ζ^t is **not** centred.

The derivation of the Kalman filter for this system is by exactly the same argument as given above in relation to Assumptions (4.i) and (4.ii) except that appeal is now made to the extension in Section 3 in order to accommodate the non-centredness of ζ^t.

Note that a^t is deterministic, and so the variance quantities $\Phi^t, \Psi^t, \Lambda^t$ are all left unchanged by the above replacements. The only change required to the Kalman filter is the replacement of (4.1) by (4.1a).

5. Hierarchical credibility

Consider the hierarchical credibility model of Sundt (1979, 1980). Here the unknown parameter and observable random vector x of Section 3 are replaced by unknown random parameters θ_i and observable random vectors x_i, $i = 1,2,\ldots,n$, with i denoting the level within a multi-level hierarchy. The model is defined by the following assumptions.

(5.i) For each $i = 1, \ldots, n$,

$$x_i \mid \theta_1,\ldots,\theta_i = Y_i\, b_i\left(\theta_1,\ldots,\theta_i\right) + \xi_i\left(\theta_1,\ldots,\theta_i\right)$$

for some vector-valued function $b_i(.)$, deterministic matrix Y_i, and random vector $\xi_i\left(\theta_1,\ldots,\theta_i\right)$ satisfying

$$E\left[\xi_i\left(\theta_1,\ldots,\theta_i\right) \mid \theta_1,\ldots,\theta_{i-1}\right] = 0 .$$

(5.ii) For each $i = 2, \ldots, n$,

$b_i\left(\theta_1\ldots,\theta_i\right) = W_{i-1}\, b_{i-1}\left(\theta_1\ldots,\theta_{i-1}\right) + \eta_i\left(\theta_1\ldots,\theta_i\right)$ for some deterministic matrix W_{i-1} and random vector $\eta_i\left(\theta_1\ldots,\theta_i\right)$ satisfying

$$E\left[\eta_i\left(\theta_1,\ldots,\theta_i\right) \mid \theta_1,\ldots,\theta_{i-1}\right] = 0 .$$

For the case $i = 1$,

$$b_1\left(\theta_1\right) = \beta + \eta_1\left(\theta_1\right)$$

with β some known parameter and $E\left[\eta_1\left(\theta_1\right)\right] = 0$.

(5.iii) $\xi_i\left(\theta_1,\ldots,\theta_i\right)$ and $\eta_i\left(\theta_1,\ldots,\theta_i\right)$ do not depend on θ_j, $j > i$.

(5.iv) (a) $\xi_i(\theta_1,...,\theta_i)$ and $\xi_j(\theta_1,...,\theta_j)$ are conditionally independent for $j > i$, given $\theta_1,...,\theta_i$

(b) $\eta_i(\theta_1,...,\theta_i)$ and $\eta_j(\theta_1,...,\theta_j)$ are conditionally independent for $j > i$, given $\theta_1,...,\theta_i$

(c) $\xi_i(\theta_1,...,\theta_i)$ and $\eta_j(\theta_1,...,\theta_j)$ are independent for all *i*,*j*.

Remark 5.1 It follows from these assumptions that

(5.v) x_i and θ_j are independent for $j > i$

(5.vi) For each $i = 1,...,n-1$, $\left(x_1^T,...,x_i^T\right)^T$ and $\left(x_{i+1}^T,...,x_n^T\right)$ are conditionally independent given $\theta_1,...,\theta_i$.

The assumptions (5.i) – (5.iv) are in fact stronger than those of Sundt, who assumed (5.v) and (5.vi) in place of (5.iii) and (5.iv).

Let

$$\Lambda_i = E\,Var\left[b_i(\theta_1,...,\theta_i) \mid \theta_1,...,\theta_{i-1}\right], i = 1,2,...,n \tag{5.1}$$

$$\Phi_i = E\,Var\left[x_i \mid \theta_1,...,\theta_i\right], i = 1,2,...,n \tag{5.2}$$

Now define

$$x_{(i)} = \left(x_i^T,...,x_n^T\right)^T, i = 1,2,...,n \tag{5.3}$$

and define the vectors $\xi_{(i)}, \eta_{(i)}$ similarly. When these vectors are written as conditioned by θ, each component will be understood to be conditioned by all relevant information, ie in $x_{(i)}$, x_i will be conditioned by $\theta_1,...,\theta_i$; x_{i+1} by $\theta_1,...,\theta_{i+1}$, etc. Equivalently, each component is conditioned by $\theta_1,...,\theta_n$.

By Assumptions (5.i) and (5.ii), particularly the linearity of the relations there,

$$x_{(i)} \mid \theta = Y_{(i)} b_i(\theta_1,...,\theta_i) + N_i \eta_{(i)} \mid \theta + \xi_{(i)} \mid \theta \tag{5.4}$$

for some matrices $Y_{(i)}$ and N_i.

To calculate these matrices, note that

$$\begin{aligned} x_{(i+1)} \mid \theta &= Y_{(i+1)} b_{i+1}(\theta_1,...,\theta_{i+1}) + N_{i+1}\eta_{(i+1)} \mid \theta + \xi_{(i+1)} \mid \theta \\ &= Y_{(i+1)} W_i b_i(\theta_1,...,\theta_i) + \left[Y_{(i+1)}, N_{i+1}\right]\eta_{(i)} \mid \theta + \xi_{(i+1)} \mid \theta \end{aligned} \tag{5.5}$$

by Assumption (5.ii).

Combining Assumption (5.i) with (5.5) and equating the result to (5.4) yields

$$Y_{(i)} = \begin{bmatrix} Y_i \\ Y_{(i+1)} W_i \end{bmatrix}, i = 1, 2, ..., n-1 \tag{5.6}$$

$$N_i = \begin{bmatrix} 0 & 0 \\ Y_{(i+1)} & N_{i+1} \end{bmatrix}, i = 1, 2, ..., n-1 \tag{5.7}$$

where the zero sub-matrices have the same numbers of rows as x_i .

The recursion (5.6) and (5.7) is initiated with

$$Y_{(n)} = Y_n, \qquad N_n = 0 \tag{5.8}$$

obtained from (5.4) with *i* = *n*.

Denote

$$\Phi_{(i)} = E\,Var\left[x_{(i)} \mid \theta_1, ..., \theta_i \right] \tag{5.9}$$

$$\Pi_{(i)} = E\,Var\left[x_{(i)} \mid \theta_1, ..., \theta_{i-1} \right] \tag{5.10}$$

Now

$$\begin{aligned} \Phi_{(i)} &= \begin{bmatrix} E\,Var\left[x_i \mid \theta_1, ..., \theta_i \right] & 0 \\ 0 & E\left[x_{(i+1)} \mid \theta_1, ..., \theta_i \right] \end{bmatrix} \\ &= \begin{bmatrix} \Phi_i & 0 \\ 0 & \Pi_{(i+1)} \end{bmatrix}, i = 1, 2, ..., n-1 \end{aligned} \tag{5.11}$$

where the independence assumption (5.iv) has been used in the first equality.

The recursion (5.11) is initiated with

$$\Phi_{(n)} = \Phi_n \tag{5.12}$$

The matrices $\Pi_{(i)}$ may be evaluated by reference to the first equality in (5.5) with *i* + 1 replaced by *i*, whence

$$\Pi_{(i)} = E_{\theta_1, ..., \theta_{i-1}} E_{\theta_i} \left\{ \left[x_{(i)} - Y_{(i)}\, b_i\left(\theta_1, ..., \theta_i \right) \right]^2 \mid \theta_1, ..., \theta_{i-1} \right\}$$

where it is convenient for the time being to show explicitly the variates with respect to which expectations are taken.

The last relation may be expanded as follows:

$$\begin{aligned}\Pi_{(i)} &= E_{\theta_1,\ldots,\theta_{i-1}} E_{\theta_i}\Big[\Big\{\Big\{x_{(i)} - E\big[Y_{(i)}\, b_i(\theta_1,\ldots,\theta_i) \mid \theta_1,\ldots,\theta_i\big]\Big\}\\ &\quad + Y_{(i)}\Big\{E\big[b_i(\theta_1,\ldots,\theta_i) \mid \theta_1,\ldots,\theta_i\big] - E\big[b_i(\theta_1,\ldots,\theta_i) \mid \theta_1,\ldots,\theta_{i-1}\big]\Big\}\Big\}^2 \mid \theta_1,\ldots,\theta_{i-1}\Big]\\ &= \Phi_{(i)} + Y_{(i)}\, \Lambda_i\, Y_{(i)}^T\end{aligned} \tag{5.13}$$

by (5.1) and (5.9).

Let $\tilde{b}_i$ denote the credibility estimator of $b_i(\theta_1,\ldots,\theta_i)$ based on $x_{(1)}$. This means that $\tilde{b}_i$ is the linear function of $x_{(1)}$ that minimises

$$E_{\theta_1\ldots\theta_i} E\Big\{\big[\tilde{b}_i - b_i(\theta_1,\ldots,\theta_i)\big]^2 \mid \theta_1,\ldots,\theta_i\Big\}$$

Sundt (1980) shows that, by Assumption (5.ii), $W_{i-1}\,\tilde{b}_{i-1}$ is the credibility estimator of $E\big[b_i(\theta_1,\ldots,\theta_i) \mid \theta_1,\ldots,\theta_i\big]$, and that, by Assumption (5.v), $\tilde{b}_i$ depends on just the $x_{(i)}$ part of $x_{(1)}$.

The mathematics of finding this estimator is once again the same as in Section 3 but with the following replacements:

$$\begin{aligned}&x \leftarrow x_{(i)}\\ &Y \leftarrow Y_{(i)}\\ &b(\theta) \leftarrow b_i(\theta_1,\ldots,\theta_i)\\ &\xi \leftarrow \xi_{(i)}\\ &\beta \leftarrow W_{i-1}\,\tilde{b}_{i-1}\\ &\tilde{b} \leftarrow \tilde{b}_i\end{aligned}$$

The expectation in the minimisation criterion is also replaced. The objective function $E\big[L(x) - m\big]^2$ is now replaced by

$$E_{\theta_1\ldots\theta_{i-1}} E_{\theta_i}\Big\{\big[\tilde{b}_i - b_i(\theta_1,\ldots,\theta_i)\big]^2 \mid \theta_1,\ldots,\theta_{i-1}\Big\}$$

All expectations on which the credibility estimator depends are replaced correspondingly. As a result

$$\Lambda \leftarrow E_{\theta_1,\ldots,\theta_{i-1}} Var_{\theta_i}\left[b_i\left(\theta_1,\ldots,\theta_i\right) \mid \theta_1,\ldots,\theta_{i-1}\right] = \Lambda_i$$
$$\Phi \leftarrow \Phi_{(i)}$$
$$\hat{b}_i = \left\{\left[Y_{(i)}\right]^T \Phi_{(i)}^{-1} Y_{(i)}\right\}^{-1} \left[Y_{(i)}\right]^T \Phi_{(i)}^{-1} x_{(i)} \tag{5.14}$$

With these replacements, (3.4) and (3.5) become

$$\tilde{b}_i = \left(1 - Z_i\right) W_{i-1} \tilde{b}_{i-1} + Z_i \hat{b}_i \tag{5.15}$$

and

$$Z_i = \Lambda_i \left[Y_{(i)}\right]^T \Phi_{(i)}^{-1} Y_{(i)} \left[1 + \Lambda_i \left[Y_{(i)}\right]^T \Phi_{(i)}^{-1} Y_{(i)}\right]^{-1} \tag{5.16}$$

6. A hierarchical Kalman filter

Section 5 defined the hierarchy of parameters $\left\{b_i\left(\theta_1,\ldots,\theta_i\right) : i = 1,\ldots,n\right\}$ and observations $\left\{x_i : i = 1,\ldots,n\right\}$. Section 4 defined the evolution of a parameter vector over time.

The present section combines the two concepts into an **evolutionary hierarchy** of parameter vectors $\left\{b_i^t : i = 1,\ldots,n; t = 1,2,\ldots\right\}$ and observation vectors $\left\{x_i^t : i = 1,\ldots,n; t = 1,2,\ldots\right\}$. The hierarchy is subject to the following assumptions.

(6.i) For each t = 1, 2, … and each i = 1, …, n, $x_i^t \mid b_1^t,\ldots,b_i^t = Y_i^t b_i^t + \xi_i^t$ for some parameter vector b_i^t, deterministic matrix Y_i^t, and centred random vector ξ_i^t.

(6.ii) For each t = 1, 2, … and each i = 2, …, n, define $c_i^t = b_i^t - W_{i-1}^t b_{i-1}^t$ for some deterministic matrix W_{i-1}^t. For the case i = 1, define $c_1^t = b_1^t - W_0^t \beta^t$ where $\beta^t = E\left[b_1^t\right]$ and W_0^t is some deterministic matrix.

The vectors c_i^t and β^t evolve according to

$$c_i^{t+1} = A_i^{t+1} c_i^t + \varepsilon_i^{t+1}, \quad i = 1,\ldots,n;\ t = 1,2,\ldots$$
$$\beta^{t+1} = A_0^{t+1} \beta^t + \varepsilon_0^{t+1}, \quad t = 1,2,\ldots$$

for deterministic matrices A_i^{t+1} and centred random vectors ε_i^{t+1}.

(6.iii) The random vectors ξ_i^t, ε_j^s are mutually stochastically independent.

Remark 6.1. The random vectors b_i^t are no longer expressed as functions of abstract latent parameters θ_i as in Section 5. Instead, the explicit relations between these vectors are stipulated in Assumption (6.ii).

Remark 6.2. By Assumption (6.ii), the evolution of the b_i^t is according to

$$b_i^{t+1} = W_{i-1}^{t+1} b_{i-1}^{t+1} + A_i^{t+1} b_i^t - A_i^{t+1} W_{i-1}^t b_{i-1}^t + \varepsilon_i^{t+1}, \quad i = 2,...,n;\ t = 1,2,... \tag{6.1}$$

For the case i = 1, this relation continues to hold if b_0^t, b_0^{t+1} are replaced by β^t, β^{t+1}.

The following will use a notation parallel to that of Section 5 but augmented by the superscript representing time. For example, $x_{(i)}^t$ will denote the vector $x_{(i)}$, defined by (5.4), as it occurs in time period *t*.

The following will also use the *t*|*s* superscript notation from Section 4 with the same meaning as there. For example, $\tilde{b}_i^{t|s}$ will denote the linearised estimator of b_i^t based on data $x_{(1)}^1, ..., x_{(1)}^s$.

The objective is to find estimators $\tilde{b}_i^{t|t}$ for i = 1, 2, …, n; t = 2, 3, … . This is done by application of the Kalman filter to the vectors $x_{(1)}^t$.

Begin by noting that the evolution of the b_i^t may be expressed in the form

$$b_i^{t+1} = \sum_{j=0}^{i} {}^iM_j^{t+1} b_j^t + \sum_{j=0}^{i} {}^iN_j^{t+1} \varepsilon_j^{t+1} \tag{6.2}$$

for matrices ${}^iM_j^{t+1}$, ${}^iN_j^{t+1}$ defined below, and where b_0^t in interpreted as β^t.

This relation may be proved by induction i. It is true in the case i = 0, by Assumption (6.ii). Here

$${}^0M_0^{t+1} = A_0^{t+1},\ {}^0N_0^{t+1} = 1 \tag{6.3}$$

For the case i > 0, use (6.2) to substitute for b_{i-1}^{t+1} in (6.1). This yields

$$\begin{aligned} b_i^{t+1} = {} & W_{i-1}^{t+1} \left[\sum_{j=0}^{i-1} {}^{i-1}M_j^{t+1} b_j^t + \sum_{j=0}^{i-1} {}^{i-1}N_j^{t+1} \varepsilon_j^{t+1} \right] \\ & + A_i^{t+1} b_i^t - A_i^{t+1} W_{i-1}^t b_{i-1}^t + \varepsilon_i^{t+1} \end{aligned} \tag{6.4}$$

Equality of (6.4) with (6.2) yields the recursion

$$
\begin{aligned}
&{}^{i}M_{i}^{t+1} = A_{i}^{t+1} \\
&{}^{i}M_{i-1}^{t+1} = W_{i-1}^{t+1}\,{}^{i-1}M_{i-1}^{t+1} - A_{i}^{t+1}W_{i-1}^{t} \\
&{}^{i}M_{j}^{t+1} = W_{i-1}^{t+1}\,{}^{i-1}M_{j}^{t+1}, \quad j = 0,1,\ldots,i-2 \\
&{}^{i}N_{i}^{t+1} = 1 \\
&{}^{i}N_{j}^{t+1} = W_{i-1}^{t+1}\,{}^{i-1}N_{j}^{t+1}, \quad j = 0,1,\ldots,i-1
\end{aligned}
\tag{6.5}
$$

The relation (6.2) for $i = 0, 1, \ldots, n$ may be summarised in the form

$$b_{(0)}^{t+1} = M_{(0)}^{t+1} b_{(0)}^{t} + N_{(0)}^{t+1}\varepsilon_{(0)}^{t+1} \tag{6.6}$$

where the matrix $M_{(0)}^{t+1}$ takes the block form

$$M_{(0)}^{t+1} = \begin{bmatrix} {}^{0}M_{0}^{t+1} & & & \\ {}^{1}M_{0}^{t+1} & {}^{1}M_{1}^{t+1} & & \\ \vdots & \vdots & \ddots & \\ {}^{n}M_{0}^{t+1} & {}^{n}M_{1}^{t+1} & \ldots & {}^{n}M_{n}^{t+1} \end{bmatrix} \tag{6.7}$$

and $N_{(0)}^{t+1}$ is similarly constructed from the blocks ${}^{i}N_{j}^{t+1}$.

Now define $Y_{[i]}^{t}$ (distinct from $Y_{(i)}^{t}$) as the block diagonal matrix with blocks $Y_{i}^{t},\ldots,Y_{n}^{t}$. As in Section 5, where the vectors $x_{(i)}^{t}$ are written without as conditioned by b, each component will be understood to be conditioned by all relevant information, ie in $x_{(i)}^{t}$, x_{i}^{t} will be conditioned by $b_{1}^{t},\ldots,b_{i}^{t}$; x_{i+1}^{t} by $b_{1}^{t},\ldots,b_{i+1}^{t}$, etc. Equivalently, each component is conditioned by $b_{1}^{t},\ldots,b_{n}^{t}$. By Assumption (6.i),

$$x_{(i)}^{t} \mid b^{t} = Y_{[i]} b_{(i)}^{t} + \xi_{(i)}^{t} \tag{6.8}$$

Note that (6.6) and (6.8) with $i = 0$ are of the same form as Assumptions (4.i) and (4.ii) and so the Kalman filter, consisting of (4.1), (4.6) – (4.9) and (4.11), may be applied with the following replacements:

$$
\begin{aligned}
&b^{t} \leftarrow b_{(0)}^{t} \\
&A^{t} \leftarrow M_{(0)}^{t} \\
&x^{t} \leftarrow x_{(0)}^{t} \\
&Y^{t} \leftarrow Y_{[0]}^{t} \\
&\varepsilon^{t} \leftarrow N_{(0)}^{t}\varepsilon_{(0)}^{t} \\
&\xi^{t} \leftarrow \xi_{(0)}^{t}
\end{aligned}
$$

Assumption (6.iii) ensures that the sets $\left\{N_{(0)}^t \varepsilon_{(0)}^t\right\}$ and $\left\{\xi_{(0)}^t\right\}$ are mutually stochastically independent, as required by Assumption (4.iii).

It will be necessary for this filter to map an estimator of $b_{(0)}^{t-1}$ to an estimator of $b_{(1)}^t$, rather than $b_{(0)}^t$. This is because there are no direct observations Y on b_0^t and so no estimator $\hat{b}_0^t$ for use in (6.17).

Define $M_{(1:0)}^t$ to be the matrix obtained by deleting the first row of $M_{(0)}^t$. Define $N_{(1:0)}^t$ similarly. Then (6.6) gives

$$b_{(1)}^t = M_{(1:0)}^t \, b_{(0)}^{t-1} + N_{(1:0)}^t \, \varepsilon_{(0)}^t \tag{6.9}$$

In correspondence with (4.2) – (4.4), define

$$\Phi_{(i)}^t = Var\left[x_{(i)}^t \,|\, b^t \right] = Var\left[\xi_{(i)}^t \right] \tag{6.10}$$

$$\Psi_{(i)}^t = Var\left[\varepsilon_{(i)}^t \right] \tag{6.11}$$

$$\Lambda_{(i)}^t = E\left[b_{(i)}^t - \tilde{b}_{(i)}^{t|t-1} \right]^2 \tag{6.12}$$

Note that the mutual independence of the ξ_i^t implies that $\Phi_{(i)}^t$ has block diagonal structure.

Also define

$$\Gamma_{(i)}^t = Var\left[\tilde{b}_{(i)}^{t|t} \right] \tag{6.13}$$

and note that, in correspondence with (4.5),

$$\Lambda_{(1)}^t = E\left\{ \left[M_{(1:0)}^t \, b_{(0)}^{t-1} + N_{(1:0)}^t \, \varepsilon_{(0)} \right] - M_{(1:0)}^t \, \tilde{b}_{(0)}^{t-1|t-1} \right\}^2 \tag{6.14}$$

$$= M_{(1:0)}^t \, \Gamma_{(0)}^{t-1} \left[M_{(1:0)}^t \right]^T + N_{(1:0)}^t \Psi_{(0)}^t \left[N_{(1:0)}^t \right]^T \tag{6.15}$$

where, in (6.14), use has been made of (6.9) and (6.16) has been anticipated. Use has also been made of Assumption (6.iii).

Then the Kalman filter, with equations now written in operational order, is as follows:

$$\tilde{b}_{(1)}^{t|t-1} = M_{(1:0)}^t \, \tilde{b}_{(0)}^{t-1|t-1} \tag{6.16}$$

$$\hat{b}^t_{(1)} = \left[\left(Y^t_{[1]} \right)^t \left(\Phi^t_{(1)} \right)^{-1} Y^t_{[1]} \right]^{-1} \left(Y^t_{[1]} \right)^T \left(\Phi^t_{(1)} \right)^{-1} x^t_{(1)} \tag{6.17}$$

$\Lambda^t_{(1)}$ according to (6.15)

$$L^t_{(1)} = \Lambda^t_{(1)} \left(Y^t_{[1]} \right)^T \left(\Phi^t_{(1)} \right)^{-1} Y^t_{[1]} \tag{6.18}$$

$$K^t_{(1)} = L^t_{(1)} \left(1 + L^t_{(1)} \right)^{-1} \tag{6.19}$$

$$\tilde{b}^{t|t}_{(1)} = \left(1 - K^t_{(1)} \right) \tilde{b}^{t|t-1}_{(1)} + K^t_{(1)} \hat{b}^t_{(1)} \tag{6.20}$$

$$\Gamma^t_{(1)} = \left(1 - K^t_{(1)} \right) \Lambda^t_{(1)} \tag{6.21}$$

It remains to calculate $\tilde{b}^{t|t}_0$ and hence $\tilde{b}^{t|t}_{(0)}$, and then $\Gamma^t_{(0)}$.

Begin by noting, from (6.1), that

$$b^t_1 = W^t_0 b^t_0 + A^t_1 \left(b^{t-1}_1 - W^{t-1}_0 b^{t-1}_0 \right) + \varepsilon^t_1 \tag{6.22}$$

Now the matrix W^t_0 may not be square. It will be assumed here to have at least as many rows as columns and to be of full rank, this being the more common case in practice (eg see example in Section 7).

As shown in the appendix, it is always possible to transform W^t_0 as follows:

$$R^t W^t_0 = U^t \tag{6.23}$$

where R^t is defined in the appendix, and U^t takes the block diagonal form

$$U^t = \begin{bmatrix} u_1 & & \\ & u_2 & \\ & & \ddots \end{bmatrix} \tag{6.24}$$

with each u_j a column vector with all components unity.

By (6.22) and (6.23),

$$R^t b^t_1 = U^t b^t_0 + R^t A^t_1 \left(b^{t-1}_1 - W^{t-1}_0 b^{t-1}_0 \right) + R^t \varepsilon^t_1$$

Pre-multiply both sides of this relation by the block diagonal matrix

$$\Omega^t = \begin{bmatrix} \omega_1^T & & \\ & \omega_2^T & \\ & & \ddots \end{bmatrix} \tag{6.25}$$

where each ω_j^T is a row vector and

$$\omega_j^T u_j = 1 \tag{6.26}$$

This gives

$$b_o^t = \Omega^t R^t \left[b_1^t - A_1^t \left(b_1^{t-1} - W_0^{t-1} b_0^{t-1} \right) - \varepsilon_1^t \right]$$

It follows from the unbiasedness of ε_1^t that an unbiased estimate of b_0^t is

$$\tilde{b}_0^{t|t} = \Omega^t R^t \left[\tilde{b}_1^{t|t} - A_1^t \left(\tilde{b}_1^{t-1|t-1} - W_0^{t-1} \tilde{b}_0^{t-1|t-1} \right) \right] \tag{6.27}$$

The matrix Ω^t has been constrained by (6.25) and (6.26) but is otherwise still free. It may be chosen so that $\tilde{b}_0^{t|t}$ is minimum variance as well as unbiased. This may be done component of $\tilde{b}_0^{t|t}$ by component. By (6.25), each of these components takes the form

$$\tilde{b} = \omega^T \alpha \tag{6.28}$$

where the subscript on ω has been suppressed and α denotes $R^t[\ldots]$ in (6.27) and, by (6.26), ω is subject to the constraint

$$\omega^T u = 1 \tag{6.29}$$

By (6.28),

$$Var\left[\tilde{b}\right] = \omega^T V \omega \tag{6.30}$$

where V denotes $Var[\alpha]$. It is straightforward to show that the choice of ω minimising (6.30) subject to (6.29) is

$$\omega = \left(u^T V^{-1} u \right)^{-1} V^{-1} u \tag{6.31}$$

This may be evaluated if V is known. Now V is a submatrix of the covariance matrix

$$\begin{aligned}
&R^t\, Var\left[\tilde{b}_1^{t|t}\right]\left(R^t\right)^T + R^t A_1^t\, Var\left[\tilde{b}_1^{t-1|t-1}\right]\left(A_1^t\right)^T\left(R^t\right)^T \\
&+R^t\, A_1^t\, W_0^{t-1} Var\left[\tilde{b}_0^{t-1|t-1}\right]\left[W_0^{t-1}\right]^T\left(A_1^t\right)^T\left(R^t\right)^T \\
&-\left\langle R^t\, A_1^t\, Cov\left[\tilde{b}_1^{t-1|t-1},\tilde{b}_0^{t-1|t-1}\right]\left(W_0^{t-1}\right)^T\left(A_1^t\right)^T\left(R^t\right)^T\right\rangle \\
&-\left\langle R^t\, Cov\left[\tilde{b}_1^{t|t},\tilde{b}_1^{t-1|t-1}\right]\left(A_1^t\right)^T\left(R^t\right)^T\right\rangle \\
&+\left\langle R^t\, Cov\left[\tilde{b}_1^{t|t},\tilde{b}_0^{t-1|t-1}\right]\left(W_0^{t-1}\right)^T\left(A_1^t\right)^T\left(R^t\right)^T\right\rangle
\end{aligned} \qquad (6.32)$$

where *Cov* [*b*,*c*] denotes $E\left\{\left[b-E[b]\right]\left[c-E[c]\right]^T\right\}$ and <.> denotes the symmetrisation of the argument, ie $\left\langle A\,Cov[b,c]\,B^T\right\rangle = A\,Cov[b,c]\,B^T + B\,Cov[c,b]\,A^T$.

Of the six *Var* and *Cov* terms on the right, the first is obtainable from $\Gamma_{(1)}^t$ in (6.21), and the next three from $\Gamma_{(0)}^{t-1}$. The last two may be obtained by calculating $Cov\left[\tilde{b}_{(1)}^{t|t},\tilde{b}_{(0)}^{t-1|t-1}\right]$.

By (6.16) and (6.20),

$$\tilde{b}_{(1)}^{t|t} = \left(1-K_{(1)}^t\right)M_{(1:0)}^t\tilde{b}_{(0)}^{t-1|t-1} + K_{(1)}^t\,\hat{b}_{(1)}^t \qquad (6.33)$$

Note that $\hat{b}_{(1)}^t$ depends on data relating to only time *t* whereas $\tilde{b}_{(0)}^{t-1|t-1}$ depends on data relating to only prior periods. The two are therefore stochastically independent, and so (6.33) yields

$$Cov\left[\tilde{b}_{(1)}^{t|t},\tilde{b}_{(0)}^{t-1|t-1}\right] = \left(1-K_{(1)}^t\right)M_{(1:0)}^t\,\Gamma_{(0)}^{t-1} \qquad (6.34)$$

Remark 6.3. The transformation (6.23) of W_0^t into the form (6.24), while generally applicable, has been selected here because it is convenient for the example in Section 7. In fact, other transformations could have been chosen, eg (6.23) with *m* x *n* matrix U^t taking the block form

$$U^t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

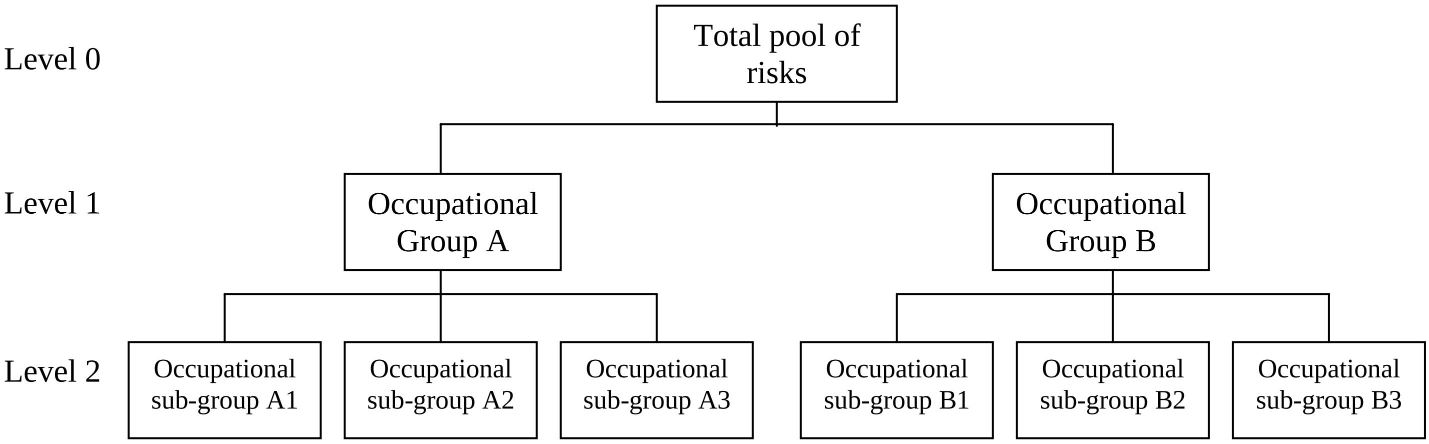
An alternative choice of U^t would lead of course to a different R^t in (6.23) and a different Ω^t satisfying (6.26). The remainder of the reasoning (6.22) – (6.34) would be unchanged.

7. Numerical example

7.1 Data

The following example considers workers compensation claim costs within the hierarchical occupational structure illustrated in Figure 7.1.

Figure 7.1 – Occupational structure



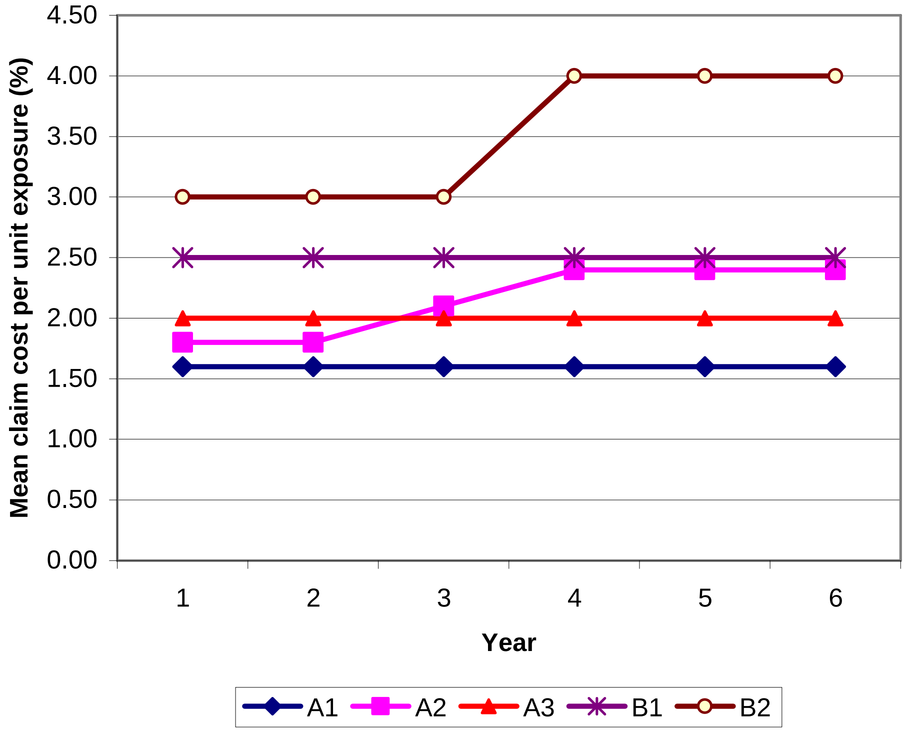
Claims experience has been simulated over 6 years for each of the occupational sub-groups. Each observation is a realisation of a gamma, distribution with parameters as set out in Table 7.1. "CoV" denotes coefficient of variation in the table.

Table 7.1 – Claim cost parameters

Segment	Exposure	Year											
		1		2		3		4		5		6	
		Mean	CoV	Mean	CoV	Mean	CoV	Mean	CoV	Mean	CoV	Mean	CoV
	$M	%	%	%	%	%	%	%	%	%	%	%	%
Pool	275	1.99	4.3	1.99	4.3	2.10	4.3	2.30	4.4	2.30	4.4	2.30	4.4
A	225	1.82	4.7	1.82	4.7	1.96	4.7	2.09	4.8	2.09	4.8	2.09	4.8
B	50	2.75	10.0	2.75	10.0	2.75	10.0	3.25	10.3	3.25	10.3	3.25	10.3
A1	50	1.60	10.0	1.60	10.0	1.60	10.0	1.60	10.0	1.60	10.0	1.60	10.0
A2	100	1.80	7.1	1.80	7.1	2.10	7.1	2.40	7.1	2.40	7.1	2.40	7.1
A3	75	2.00	8.2	2.00	8.2	2.00	8.2	2.00	8.2	2.00	8.2	2.00	8.2
B1	25	2.50	14.1	2.50	14.1	2.50	14.1	2.50	14.1	2.50	14.1	2.50	14.1
B2	25	3.00	14.1	3.00	14.1	3.00	14.1	4.00	14.1	4.00	14.1	4.00	14.1

Figure 7.2 plots the mean claim cost per unit exposure for the different occupational groups. Note that the underlying claim cost of B2 undergoes a step increase in Year 4, while that of A2 ramps up over Years 3 and 4.

Figure 7.2 – Mean claim cost per unit exposure



The simulated claims costs are set out in Table 7.2.

Table 7.2 – Claims data

Occupational group	Claim cost per unit exposure observed in year					
	1	2	3	4	5	6
	%	%	%	%	%	%
A	1.86	1.62	1.86	2.04	2.00	2.09
B	2.10	2.27	2.52	3.65	3.25	3.91
A1	1.66	1.53	1.65	1.36	1.69	1.42
A2	1.96	1.58	1.99	2.32	2.50	2.55
A3	1.86	1.73	1.84	2.13	1.55	1.93
B1	2.27	1.78	2.58	2.76	3.15	3.32
B2	1.94	2.76	2.46	4.54	3.34	4.50

7.2 Static hierarchical credibility rating

For comparison, the occupational groups are first rated by means of Sundt's (1980) hierarchical credibility, as described in Section 5.

This is a static credibility system in the sense that the underlying claim cost parameters are assumed constant over time. When they are in fact variable, as in Table 7.1, this credibility system cannot be expected to track the changes well.

Since the system is static, its natural application at the end of t years is to the **accumulated** claims experience over those t years. Observations on this basis are set out in Table 7.3.

Table 7.3 – Accumulating claims costs per unit exposure

Occupational group	Average claim cost per unit exposure observed over years					
	1	1 to 2	1 to 3	1 to 4	1 to 5	1 to 6
	%	%	%	%	%	%
Pool	1.90	1.82	1.88	1.99	2.04	2.10
A	1.86	1.74	1.78	1.85	1.88	1.91
B	2.10	2.19	2.30	2.64	2.76	2.95
A1	1.66	1.60	1.61	1.55	1.58	1.55
A2	1.96	1.77	1.84	1.96	2.07	2.15
A3	1.86	1.79	1.81	1.89	1.82	1.84
B1	2.27	2.02	2.21	2.35	2.51	2.64
B2	1.94	2.35	2.39	2.92	3.01	3.26

The parametric structure assumed for the occupational hierarchy is as follows.

Let b_i denote the vector of mean claim costs per unit exposure at Level i of the hierarchy. These vectors will be of dimensions 1, 2 and 5 for i=0, 1 and 2 respectively. Let x_2 denote the 5-vector of observations at Level 2.

The hierarchical structure, in the notation of Section 5, is as follows:

$$W_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad W_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \qquad Y_2 = 5 \times 5 \text{ unit matrix}$$

Y_1 is null as there are no observations at Level 1. At least no direct observations, those shown at Levels 0 and 1 in Table 7.3 being merely weighted averages of the observations at Level 2.

The assumed parameters are as set out in Table 7.4.

Table 7.4 – Static hierarchical credibility parameters

Parameter	Value
β	2.0%
Λ_1	$(1.0\%)^2$ x unit matrix (2x2)
Λ_2	$(0.5\%)^2$ x unit matrix (5x5)
Φ_2	$(0.25\%)^2$ x weight matrix (5x5)

The weight matrix referred to in Table 7.4 is the diagonal matrix whose (k,k) element is equal to e_1 / e_k where e_k is the exposure of the k-th occupational group at Level 2. That is, the variances in Φ_2 are inversely proportional to the occupational group exposures in Table 7.1.

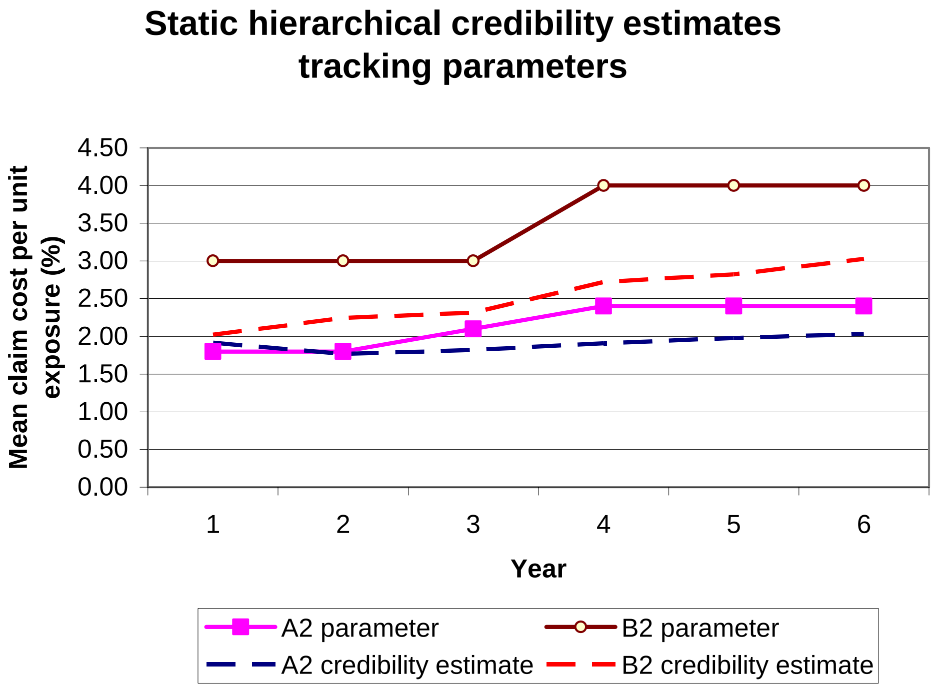
The credibility system (5.15) and (5.16) is applied to the observations, yielding the results set out in Table 7.5.

Table 7.5 – Static hierarchical credibility rating

Occupational group	Static hierarchical credibility rating					
	1	2	3	4	5	6
	%	%	%	%	%	%
A	1.87	1.76	1.80	1.86	1.89	1.92
B	2.09	2.16	2.26	2.55	2.66	2.83
A1	1.78	1.69	1.71	1.71	1.74	1.75
A2	1.91	1.77	1.82	1.91	1.98	2.03
A3	1.87	1.78	1.81	1.87	1.86	1.88
B1	2.16	2.10	2.24	2.46	2.59	2.74
B2	2.02	2.24	2.32	2.72	2.82	3.03

As seen in Figure 7.2, occupational groups A2 and B2 are subject to dramatic change in risk over the 6-year period. Figure 7.3 plots the credibility estimates of mean claim cost per unit exposure against the true parameters. As expected, the static nature of the model produces poor tracking of changing risk parameters.

Figure 7.3



7.3 Dynamic hierarchical credibility rating

In this sub-section the hierarchical Kalman filter developed in Section 6 is applied to the observations set out in Table 7.2.

The hierarchical structure is as in Section 7.2, with W_i^t here taking the same form for each t as W_i in that earlier sub-section. Each A_i^t is a unit matrix, meaning that the parameters evolve according to independent random walks. Further, Y_2^t takes the same form as Y_2 in Section 7.2 and R^t from (6.23) is the 5×5 unit matrix.

The assumed dispersion parameters are set out in Table 7.6.

Table 7.6 – Dispersion parameters for hierarchical Kalman filters

Parameter	Value
Φ_2^t	As for Φ_2 in Section 7.2
Ψ_0^t	$(0.1\%)^2$
Ψ_1^t	$(0.15\%)^2$ x unit matrix (2x2)
Ψ_2^t	$(0.25\%)^2$ x unit matrix (5x5)

The hierarchical filter (6.16) – (6.21) is applied to the observations. It needs to be adapted slightly to the present hierarchy in order to apply to suffix (2) rather than (1). The result (6.27) requires similar adaptation to produce $\tilde{b}_1^{t|t}$ and $\tilde{b}_0^{t|t}$ rather than just the latter.

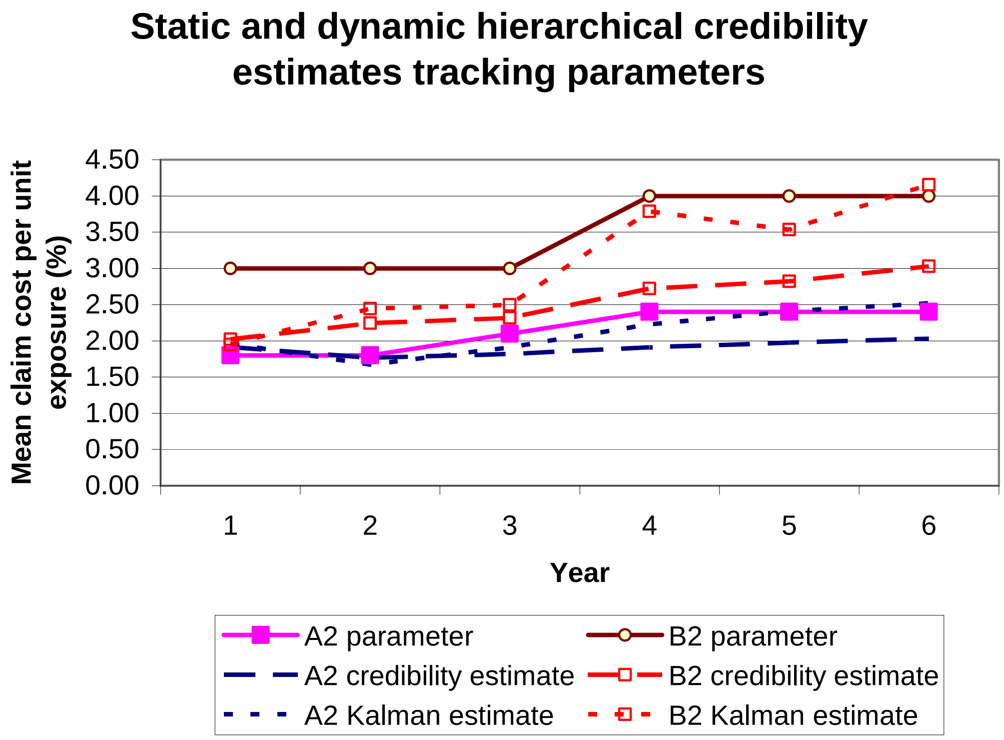
The results produced by the filter are set out in Table 7.7, which also reproduces, for comparison, true parameters, observations, and static hierarchical credibility estimates.

Table 7.7 – Hierarchical Kalman filter credibility rating

Occupational group	Observations and credibility ratings at end of year					
	1	2	3	4	5	6
	%	%	%	%	%	%
			Observations			
A	1.86	1.62	1.86	2.04	2.00	2.09
B	2.10	2.27	2.52	3.65	3.25	3.91
A1	1.66	1.53	1.65	1.36	1.69	1.42
A2	1.96	1.58	1.99	2.32	2.50	2.55
A3	1.86	1.73	1.84	2.13	1.55	1.93
B1	2.27	1.78	2.58	2.76	3.15	3.32
B2	1.94	2.76	2.46	4.54	3.34	4.50
			Static hierarchical credibility rating			
A	1.87	1.76	1.80	1.86	1.89	1.92
B	2.09	2.16	2.26	2.55	2.66	2.83
A1	1.78	1.69	1.71	1.71	1.74	1.75
A2	1.91	1.77	1.82	1.91	1.98	2.03
A3	1.87	1.78	1.81	1.87	1.86	1.88
B1	2.16	2.10	2.24	2.46	2.59	2.74
B2	2.02	2.24	2.32	2.72	2.82	3.03
			Hierarchical Kalman filter rating			
A	1.85	1.66	1.80	1.95	1.94	2.00
B	2.08	2.20	2.42	3.26	3.25	3.70
A1	1.71	1.57	1.65	1.52	1.62	1.52
A2	1.96	1.67	1.91	2.23	2.41	2.52
A3	1.87	1.75	1.84	2.07	1.73	1.88
B1	2.21	1.96	2.35	2.73	2.96	3.25
B2	1.95	2.44	2.50	3.79	3.53	4.15
			True parameter			
A	1.82	1.82	1.96	2.09	2.09	2.09
B	2.75	2.75	2.75	3.25	3.25	3.25
A1	1.60	1.60	1.60	1.60	1.60	1.60
A2	1.80	1.80	2.10	2.40	2.40	2.40
A3	2.00	2.00	2.00	2.00	2.00	2.00
B1	2.50	2.50	2.50	2.50	2.50	2.50
B2	3.00	3.00	3.00	4.00	4.00	4.00

Figure 7.4 provides a graphical comparison of true parameters with static and dynamic credibility estimates for the two occupational groups, A2 and B2, whose parameters have changed over time. The dynamic estimates (dotted lines) are seen to track the parameters (solid lines) far better than the static estimates (dashed lines).

Figure 7.4



8. Conclusion

When the parameters of a hierarchy evolve through time, the conventional (static) hierarchical credibility estimates of Sundt (1979, 1980) (Section 5) are likely to track poorly.

The dynamic equivalent of Sundt's model explicitly recognises parameter evolution. This is done by superimposing a Kalman filter on the hierarchical framework (Section 6). This is likely to lead to much better tracking of the evolving parameters by the credibility estimates. This is illustrated numerically in Section 7.

Appendix. Proof of (6.23).

Suppose W is of dimension m x n with $m \geq n$. Since W is of full rank, it is possible to identify n linearly independent rows. A row permutation, represented by m x n matrix P, will position these in the first n rows, ie

$$PW = \begin{bmatrix} X \\ Y \end{bmatrix}$$

where X is the n x n matrix consisting of the linearly independent rows, and is therefore non-singular. Then

$$APW = \begin{bmatrix} 1 \\ Y \end{bmatrix}$$

where

$$A = \begin{bmatrix} X^{-1} & 0 \\ 0 & 1 \end{bmatrix}$$

A sequence of elementary row operations, represented by the matrix M, may be applied so that Y is replaced by a matrix, each of whose rows consists of a single unit element and the rest zero.

Finally, a permutation of rows, represented by m x m matrix Q gives $QMAPW$ of the form (6.24). Then $R^t = QMAP$.

□

References

Hachemeister, C.A. (1975). Credibility for regression models with application to trend. Appears in Kahn (1975, 129-163).

Jazwinski, A.H. (1970). **Stochastic processes and filtering theory**. Academic Press, New York, NY.

Kalman, R.E. (1960). A new approach to linear filtering and prediction systems. **Journal of Basic Engineering**, 82, 340-345.

Sundt, B. (1979). A hierarchical credibility regression model. **Scandinavian Actuarial Journal**, 107-114.

Sundt, B. (1980). A multi-level hierarchical credibility regression model. **Scandinavian Actuarial Journal**, 25-32.

Taylor, G.C. (1979). Credibility analysis of a general hierarchical model. **Scandinavian Actuarial Journal**, 1-12.