

LINEAR FILTERING AND RECURSIVE CREDIBILITY ESTIMATION

BY BEN ZEHNWIRTH

Macquarie University, North Ryde, Australia

ABSTRACT

Recursive credibility estimation is discussed from the viewpoint of linear filtering theory. A conjunction of geometric interpretation and the innovation approach leads to general algorithms not developed before. Moreover, covariance characterizations considered by other researchers drop out elegantly as a result of geometric considerations. Examples are presented of Kalman type filters valid for non-Gaussian measurements

KEYWORDS

Credibility, filtering theory, linear Bayesian theory, geometry, Kalman filter, prospective ratemaking, Gram-Schmidt, Fourier series.

1. INTRODUCTION AND SUMMARY

There have appeared a number of papers, fairly tightly connected, concerned with recursive credibility formulae. An early paper that occupies a somewhat central position is that of GERBER and JONES (1975), which develops credibility formulae of the updating type, valid if and only if the covariance structure (5.12) holds. The other papers notably, JEWELL (1976), SUNDT (1981, 1983) and KREMER (1982) develop recursive formulae for a variety of other evolutionary type models, the last emphasizing the relationships with modern models of time series. Last, but not least, the paper of DE JONG and ZEHNWIRTH (1983) relates some credibility models to the Kalman filter, perhaps, the most important algorithm in linear stochastic system theory.

The basic purpose of the present paper is to unify many existing results in recursive credibility theory and moreover develop more general ones. To achieve this, we adopt a geometric interpretation of recursive linear least squares estimation theory in the spirit of GERBER and JONES (1975) and DE VYLDER (1976). There is also a side benefit to be had by adopting a geometric approach—it reduces both the conceptual and algebraic burdens. The practical importance to actuaries of the present paper lies in the fact that once a model for premium rate-making is postulated, the estimators of parameters, premium forecast and associated errors may be derived quite readily using the general results contained herein. Moreover, the recursive nature of the formulae affords economy of computing space and time.

The main results here are established with the aid of KAILATH's (1974) innovation technique which has found fruitful applications in linear filtering theory. It

is intimately related to the well known Gram-Schmidt orthogonalization scheme and Fourier series.

Suppose \hat{Y} is a forecast of the random quantity Y with associated mean-square error C , based on some past measurements. Given a new measurement X we wish to update our forecast of Y and its associated mean-square error C . Let \hat{X} represent the forecast of X based on the past measurements. The innovation, $e = X - \hat{X}$, represents what is "new" in the new measurement X . The updated forecast of Y is

$$(1.1) \quad \hat{Y} + Ke$$

where the weight K is given by

$$(1.2) \quad K = E[Ye]\{E[e^2]\}^{-1}.$$

The mean-square error of the updated forecast (1.1) is

$$(1.3) \quad C - KE[Ye].$$

The foregoing results are treated in elaborate detail in Sections 3 and 4. In Section 5 we consider a general prospective ratemaking framework and indicate how covariance structures considered by GERBER and JONES (1975), JEWELL (1976) and SUNDT (1981) drop out elegantly as a result of the geometric interpretation of the problem. Finally, in Section 6 Kalman type filters are derived for two different models using results developed earlier in the paper. The filters are related to the work of SUNDT (1981, 1983) and DE JONG and ZEHNWIRTH (1983).

2. HILBERT SPACE OF SQUARE-INTEGRABLE RANDOM VARIABLES

For the purposes of the present paper it is convenient to formulate some definitions and terminology and to state two classical projection theorems.

Consider a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The Hilbert space $\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathbb{P})$ is the linear space of measurable functions from Ω into \mathbb{R} whose second moment exist. We identify with the element $X \in \mathcal{H}$, the equivalence class $\{\tilde{X}: \tilde{X} = X \text{ a.e.}\}$. The inner product $\langle X, Y \rangle$ for any two elements X and Y in \mathcal{H} is defined by

$$\langle X, Y \rangle = E[XY].$$

Accordingly, the corresponding $\|\cdot\|$ is defined by

$$\|X\| = (E[X^2])^{1/2}$$

It is beneficial to extend the definition of the inner product $\langle \cdot, \cdot \rangle$ to random vectors. Suppose $\mathbf{X} = (X_1, \dots, X_n)$ and $\mathbf{Y} = (Y_1, \dots, Y_m)$ where $\mathbf{X} \in \mathcal{H}^n$ and $\mathbf{Y} \in \mathcal{H}^m$. Define $\langle \mathbf{X}, \mathbf{Y} \rangle$ by

$$\langle \mathbf{X}, \mathbf{Y} \rangle = E[\mathbf{X}\mathbf{Y}'].$$

This is not an inner product in the true sense—it is a matrix. However, if we ignore this deficiency, the projection theorem can be used as a quick mnemonic way of obtaining the approximate optimal estimators (theorem 3.2)

The following properties of the bilinear functional $\langle \cdot, \cdot \rangle$ are noted.

$$(2.1) \quad \langle \mathbf{A}\mathbf{X}, \mathbf{B}\mathbf{Y} \rangle = \mathbf{A} \langle \mathbf{X}, \mathbf{Y} \rangle \mathbf{B}'$$

for any two matrices \mathbf{A} and \mathbf{B} of appropriate dimensions.

$$(2.2) \quad \|\mathbf{A}\mathbf{X}\|^2 = \mathbf{A} \|\mathbf{X}\|^2 \mathbf{A}'$$

and

$$(2.3) \quad \langle \mathbf{X}, \mathbf{Y} \rangle' = \langle \mathbf{Y}, \mathbf{X} \rangle.$$

We state two classical projection theorems applicable to any Hilbert space (borrowed from LUENBERGER (1969)).

THEOREM 2.1. *Let \mathcal{H} be a Hilbert space and \mathcal{L} a closed subspace of \mathcal{H} . Corresponding to any vector $\mathbf{Y} \in \mathcal{H}$, \exists a unique $\mathbf{X}^* \in \mathcal{L}$ such that*

$$\|\mathbf{Y} - \mathbf{X}^*\|_* = \inf_{\mathbf{X} \in \mathcal{L}} \|\mathbf{Y} - \mathbf{X}\|_*$$

where $\|\cdot\|_*$ is the norm defined on \mathcal{H} .

Furthermore, a necessary and sufficient condition that $\mathbf{X}^* \in \mathcal{L}$ be the unique minimization vector is that $\mathbf{Y} - \mathbf{X}^*$ be orthogonal (\perp) to \mathcal{L} .

In what follows denote by $\mathcal{P}(\mathbf{Y}|\mathcal{L})$ the projection of \mathbf{Y} onto \mathcal{L} , that is $\mathcal{P}(\mathbf{Y}|\mathcal{L}) = \mathbf{X}^*$.

THEOREM 2.2. *Let \mathcal{L} be a closed subspace of a Hilbert space \mathcal{H} . Suppose N is a closed subspace of \mathcal{L} so that $\mathcal{L} = N \oplus N^\perp$ where N^\perp is the orthogonal complement of N in \mathcal{L} . If $\mathbf{Y} \in \mathcal{H}$ then*

$$\mathcal{P}(\mathbf{Y}|\mathcal{L}) = \mathcal{P}(\mathbf{Y}|N) + \mathcal{P}(\mathbf{Y}|N^\perp).$$

3. LINEAR ESTIMATION OF A RISK PARAMETER

One of the key problems in credibility theory is the estimation of a risk parameter. Suppose $Y \in \mathcal{H}$ is a (non-observable) risk parameter and X_0, X_1, \dots, X_n are (observable) measurements in \mathcal{H} . A linear estimator of Y based on X_0, X_1, \dots, X_n is any linear combination

$$Y^* = \sum_{i=1}^n a_i X_i, \quad (a_i \in \mathbb{R})$$

with mean-square error

$$\|Y - Y^*\|^2.$$

Denote by $\mathcal{L}_k = \mathcal{L}(X_0, X_1, \dots, X_k)$ the closed linear subspace spanned by the elements X_0, X_1, \dots, X_k . Also for notational simplification denote by $\mathcal{P}_k(\tilde{X})$ the projection $\mathcal{P}(\tilde{X}|\mathcal{L}_k)$, of \tilde{X} onto \mathcal{L}_k where $\tilde{X} \in \mathcal{H}$.

The following fundamental result is based on the projection theorem in \mathcal{H} . It is discussed in LUENBERGER (1969) and appears under various guises in NORBERG (1979) and references therein. It is included here for the sake of completeness.

THEOREM 3.1. *Suppose $\mathbf{X} = (X_0, X_1, \dots, X_n)' \in \mathcal{H}^{n+1}$ and $\mathcal{P}_n(Y) = \hat{\mathbf{a}}' \mathbf{X}$ where $\hat{\mathbf{a}} = (\hat{a}_0, \dots, \hat{a}_n)'$.*

Then,

$$\hat{\mathbf{a}}' = \langle Y, \mathbf{X} \rangle \| \mathbf{X} \|^2$$

and the mean-square error $\| Y - \mathcal{P}_n(Y) \|^2$ is

$$\| Y - \mathcal{P}_n(Y) \|^2 = \| Y \|^2 - \langle Y, \mathbf{X} \rangle \| \mathbf{X} \|^2 \langle \mathbf{X}, Y \rangle.$$

PROOF. The projection theorem 2.1 gives

$$Y - \mathcal{P}_n(Y) \perp X_i; \quad i = 0, 1, \dots, n$$

whence,

$$\hat{\mathbf{a}}' \langle \mathbf{X}, X_i \rangle = \langle Y, X_i \rangle; \quad i = 0, 1, \dots, n.$$

The expression for $\hat{\mathbf{a}}'$ follows from the last set of equalities whereas the expression concerning the mean-square error follows by noting that $Y - \mathcal{P}_n(Y) \perp \mathcal{P}_n(Y)$. We remark that the matrix $\mathbf{G} = \| \mathbf{X} \|^2$ is called the *Gram* matrix.

COROLLARY. *If $X_0 \equiv 1$ then $\mathcal{P}_n(Y)$ is the inhomogeneous linear Bayes rule which may be written*

$$\langle Y, 1 \rangle + C[Y, \mathbf{X}^*] C^{-1}[\mathbf{X}^*](\mathbf{X}^* - \langle \mathbf{X}^*, 1 \rangle)$$

with associated mean-square error (Bayes risk)

$$C[Y] - C[Y, \mathbf{X}^*] C^{-1}[\mathbf{X}^*] C[\mathbf{X}^*, Y]$$

where the vector $\mathbf{X}^ = (X_1, \dots, X_n)$ and the covariances $C[\cdot, \cdot]$ and $C[\cdot]$ are defined as follows:*

For any two vectors $\mathbf{U} \in \mathcal{H}^n$ and $\mathbf{V} \in \mathcal{H}^m$

$$C[\mathbf{U}, \mathbf{V}] = \langle \mathbf{U}, \mathbf{V} \rangle - \langle \mathbf{U}, 1 \rangle \langle 1, \mathbf{V} \rangle$$

and

$$C[\mathbf{U}] = C[\mathbf{U}, \mathbf{U}].$$

We now discuss straightforward extensions of the abovementioned results to vector parameters.

Suppose $\mathbf{Y} = (Y_1, \dots, Y_m)' \in \mathcal{H}^m$ is a vector risk parameter to be estimated on the basis of the measurement vector $\mathbf{X} = (X_0, X_1, \dots, X_n) \in \mathcal{H}^{n+1}$. We restrict attention to linear estimators, namely $\sum a_j X_j$, of each component Y_i of the vector

Y. Write, $\mathbf{A} = (a_{ij})$, an $m \times n$ matrix. The optimal linear estimator minimizes

$$\sum_{i=1}^m \left\| Y_i - \sum_{j=1}^n a_{ij} X_j \right\|^2$$

over all matrices \mathbf{A} of dimension $m \times n$.

THEOREM 3.2 (Luenberger). *If $\hat{\mathbf{A}}\mathbf{X}$ is the optimal linear estimator of \mathbf{Y} then*

$$\hat{\mathbf{A}} = \langle \mathbf{Y}, \mathbf{X} \rangle \| \mathbf{X} \|^{-2}$$

and the error covariance matrix of $\hat{\mathbf{A}}\mathbf{X}$ is given by

$$\| \mathbf{Y} - \hat{\mathbf{A}}\mathbf{X} \|^2 = \| \mathbf{Y} \|^2 - \langle \mathbf{Y}, \mathbf{X} \rangle \| \mathbf{X} \|^{-2} \langle \mathbf{X}, \mathbf{Y} \rangle.$$

PROOF. The results follow from the observation that the optimization decomposes into a separate problem for each component Y_i of the risk parameter vector \mathbf{Y} . The i th subproblem is simply that of finding $\mathcal{P}_n(Y_i)$. That is

$$\begin{aligned} \hat{\mathbf{A}}\mathbf{X} &= (\mathcal{P}_n(Y_1), \dots, \mathcal{P}_n(Y_m))'. \\ &= \mathcal{P}_n(\mathbf{Y}), \text{ say.} \end{aligned}$$

We remark that trace $\| \mathbf{Y} - \hat{\mathbf{A}}\mathbf{X} \|^2$ represents the mean-square error of $\mathcal{P}_n(\mathbf{Y})$. It is also known as the Bayes risk of $\mathcal{P}_n(\mathbf{Y})$ relative to squared error loss function.

COROLLARY 1. *If \mathbf{T} is a fixed $r \times m$ matrix then the optimal linear estimator of $\mathbf{T}\mathbf{Y}$ is $\mathbf{T}\mathcal{P}_n(\mathbf{Y})$ with error covariance $\mathbf{T} \| \mathbf{Y} - \mathcal{P}_n(\mathbf{Y}) \|^2 \mathbf{T}'$.*

COROLLARY 2. *If $X_0 \equiv 1$ then $\mathcal{P}_n(\mathbf{Y})$ is the inhomogeneous linear Bayes rule for \mathbf{Y} , which may be written*

$$\langle \mathbf{Y}, \mathbf{1} \rangle + C[\mathbf{Y}, \mathbf{X}^*] C^{-1}[\mathbf{X}^*] (\mathbf{X}^* - \langle \mathbf{X}^*, \mathbf{1} \rangle)$$

with error covariance matrix,

$$C[\mathbf{Y}] - C[\mathbf{Y}, \mathbf{X}^*] C^{-1}[\mathbf{X}^*] C[\mathbf{X}^*, \mathbf{Y}].$$

All the foregoing results are well known to both linear filtering theorists and credibility theorists.

4. THE GEOMETRY OF RECURSIVE RISK PARAMETER ESTIMATION

In many practical situations the elements X_0, X_1, X_2, \dots represent measurements taken sequentially in time. The optimal linear estimator of a risk parameter Y based on the measurements to time n , viz., X_0, X_1, \dots, X_n is $\hat{Y}_n = \mathcal{P}_n(Y)$ with mean-square error

$$C_n = \| Y - \hat{Y}_n \|^2.$$

If X_{n+1} is the next measurement then its best linear estimator based on \mathcal{L}_n is $\mathcal{P}_n(X_{n+1})$. Accordingly, the innovation of the new information acquired at time $n+1$ is

$$e_{n+1} = X_{n+1} - \mathcal{P}_n(X_{n+1}).$$

Put $e_0 = X_0$ and write $e_0^* = e_0 / \|e_0\|$, then by virtue of theorem 3.1

$$\mathcal{P}_0(X_1) = \langle X_1, e_0^* \rangle e_0^*$$

whence,

$$e_1 = X_1 - \langle X_1, e_0^* \rangle e_0^*.$$

By virtue of the projection theorem 2.1, $e_0 \perp e_1$ and $\mathcal{L}_1 = \mathcal{L}(e_0, e_1)$. It follows that

$$e_2 = X_2 - \langle X_2, e_0^* \rangle e_0^* - \langle X_2, e_1^* \rangle e_1^*$$

where $e_1^* = e_1 / \|e_1\|$.

Subsequently,

$$(4.1) \quad e_{n+1} = X_{n+1} - \sum_{j=0}^n \langle X_{n+1}, e_j^* \rangle e_j^*$$

where

$$e_j^* = e_j / \|e_j\|; \quad j = 0, 1, 2, \dots$$

We observe that the normalized innovations $\{e_j^*\}$ represent the orthonormal system obtained by the well-known Gram-Schmidt orthogonalization process. It follows, trivially, that the innovation sequence $\{e_j\}$ is orthogonal.

The closed linear subspace \mathcal{L}_{n+1} may be decomposed

$$(4.2) \quad \mathcal{L}_{n+1} = \mathcal{L}_n \oplus \mathcal{L}(e_{n+1}).$$

In view of the projection theorem 2.2,

$$(4.3) \quad \hat{Y}_{n+1} = \hat{Y}_n + \mathcal{P}(Y|e_{n+1})$$

where application of theorem 3.1 yields,

$$(4.4) \quad \mathcal{P}(Y|e_{n+1}) = \langle Y, e_{n+1} \rangle \|e_{n+1}\|^{-2} e_{n+1}.$$

Alternatively decompose \mathcal{L}_{n+1} thus:

$$(4.5) \quad \mathcal{L}_{n+1} = \mathcal{L}(e_0) \oplus \dots \oplus \mathcal{L}(e_{n+1}).$$

The Fourier series of Y based on \mathcal{L}_{n+1} is

$$(4.6) \quad \hat{Y}_{n+1} = \sum_{j=0}^{n+1} \langle Y, e_j^* \rangle e_j^*$$

whereas the Fourier series based on \mathcal{L}_n is

$$(4.7) \quad \hat{Y}_n = \sum_{j=0}^n \langle Y, e_j^* \rangle e_j^*.$$

The difference between expression (4.6) and (4.7) yields expression (4.3)

We note that the key element in the foregoing analysis is the orthogonality property of the innovation sequence $\{e_j\}$.

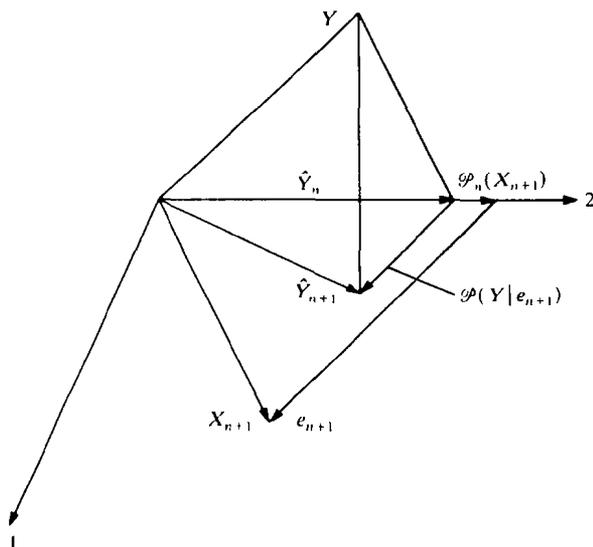


FIGURE 4.1 The geometry of recursive risk parameter estimation

Figure 4.1 shows the geometry of recursive risk parameter estimation. The co-ordinate axis labelled 2 represents \mathcal{L}_n and the 1-2 plane represents \mathcal{L}_{n+1} . Observe that $Y - \hat{Y}_n \perp \mathcal{L}_n$, $Y - \hat{Y}_{n+1} \perp \mathcal{L}_{n+1}$, $e_{n+1} \perp \mathcal{L}_n$ and $e_{n+1} \perp \hat{Y}_n$.

We point out that if $X_0 \equiv 1$ then $1 \in \mathcal{L}_n$ whence we have the unbiasedness properties,

$$\langle Y - \hat{Y}_n, 1 \rangle = 0$$

(4.8) and

$$\langle e_{n+1}, 1 \rangle = 0.$$

Denote by C_n the mean-square error,

$$\|Y - \hat{Y}_n\|^2.$$

Examination of fig. 4.1 leads to

$$\begin{aligned} (4.9) \quad C_{n+1} &= C_n - \|\mathcal{P}(Y|e_{n+1})\|^2 \\ &= C_n - \langle Y, e_{n+1} \rangle^2 \|e_{n+1}\|^{-2}. \end{aligned}$$

Write,

$$(4.10) \quad K_{n+1} = \langle Y, e_{n+1} \rangle \|e_{n+1}\|^{-2},$$

then equations (4.3) and (4.9) may be recast

$$(4.11) \quad \hat{Y}_{n+1} = \hat{Y}_n + K_{n+1}e_{n+1}$$

and

$$(4.12) \quad C_{n+1} = C_n - K_{n+1}\langle Y, e_{n+1} \rangle$$

respectively.

The preceding analysis also applies to the estimation of a vector risk parameter $Y \in \mathcal{H}^m$. Recall that,

$$\begin{aligned} \hat{Y}_n &= \mathcal{P}_n(Y) \\ &= (\mathcal{P}_n(Y_1), \dots, \mathcal{P}_n(Y_m))'. \end{aligned}$$

Let C_n represent the error covariance matrix of \hat{Y}_n . The following recursions are obtained.

$$(4.13) \quad \hat{Y}_{n+1} = \hat{Y}_n + K_{n+1}e_{n+1}$$

and

$$(4.14) \quad C_{n+1} = C_n - K_{n+1}\langle e_{n+1}, Y \rangle$$

where

$$(4.15) \quad K_{n+1} = \langle Y, e_{n+1} \rangle \|e_{n+1}\|^{-2}.$$

Finally, we remark that the preceding recursions also carry over to vector valued measurements X_0, X_1, \dots .

5. THE GEOMETRY OF RECURSIVE PROSPECTIVE RATEMAKING

In the present section we adopt the general prospective rate-making formulation of GERBER and JONES (1975).

Let X_i represent the claims cost (or loss ratio, etc.) in the i th period. The premium forecast for period $n+1$ based on the measurements $X_0 (\equiv 1), X_1, \dots, X_n$ is denoted by P_{n+1} . This premium is the optimal affine estimator (inhomogeneous linear Bayes rule) of X_{n+1} based on the measurements X_1, X_2, \dots, X_n .

That is,

$$P_{n+1} = \mathcal{P}_n(X_{n+1}).$$

The innovation in the measurement X_n is

$$e_n = X_n - \mathcal{P}_{n-1}(X_n).$$

Since,

$$\mathcal{L}_n = \mathcal{L}_{n-1} \oplus \mathcal{L}(e_n)$$

we have,

$$(5.1) \quad \mathcal{P}_n(X_{n+1}) = \mathcal{P}_{n-1}(X_{n+1}) + \mathcal{P}(X_{n+1}|e_n)$$

where

$$(5.2) \quad \mathcal{P}(X_{n+1}|e_n) = \langle X_{n+1}, e_n \rangle \|e_n\|^{-2} e_n.$$

In keeping with GERBER and JONES (1975) write

$$(5.3) \quad Z_n = \langle X_{n+1}, e_n \rangle \|e_n\|^{-2}$$

whence,

$$(5.4) \quad P_{n+1} = \mathcal{P}_{n-1}(X_{n+1}) + Z_n(X_n - P_n).$$

We emphasize that the last formula holds true in general.

We now focus on formula (4) of SUNDT (1981) which examines the situation where there exist constants b_n , c_n and d_n such that

$$P_{n+1} = b_n + c_n P_n + d_n X_n.$$

Combining this with formula (5.4) above yields

$$\mathcal{P}_{n-1}(X_{n+1}) = (c_n + Z_n)P_n + (d_n - Z_n)X_n + b_n.$$

As $\mathcal{P}_{n-1}(X_{n+1})$ should not depend on X_n , we must have $d_n = Z_n$, and as $P_n = \mathcal{P}_{n-1}(X_n)$ we obtain

$$(5.5) \quad \mathcal{P}_{n-1}(X_{n+1}) = a_n \mathcal{P}_{n-1}(X_n) + b_n$$

with $a_n = c_n + Z_n$. That is, the premium forecast for period $n+1$ based on \mathcal{L}_{n-1} is an affine function of the premium forecast for period n also based on \mathcal{L}_{n-1} . Since the innovations $\{e_j\}$ are orthogonal, $\mathcal{P}_{n-1}(X_{n+1})$ and $\mathcal{P}_{n-1}(X_n) (= P_n)$ have the Fourier series representations

$$(5.6) \quad \mathcal{P}_{n-1}(X_{n+1}) = \sum_{i=0}^{n-1} \langle X_{n+1}, e_i^* \rangle e_i^*$$

and

$$(5.7) \quad \mathcal{P}_{n-1}(X_n) = \sum_{i=0}^{n-1} \langle X_n, e_i^* \rangle e_i^*$$

where we recall that the sequence $\{e_i^*\}$ represents the orthonormal innovations.

Substituting (5.6) and (5.7) into (5.5) gives

$$(5.8) \quad \langle X_{n+1}, e_i^* \rangle = a_n \langle X_n, e_i^* \rangle; \quad i = 1, \dots, n-1$$

and

$$(5.9) \quad \langle X_{n+1}, e_0^* \rangle = a_n \langle X_n, e_0^* \rangle + b_n.$$

Also let

$$(5.10) \quad b_n = E[X_{n+1}] - a_n E[X_n].$$

We are now in a position to derive the covariance characterization (5) of SUNDT (1981, p. 5).

Equation (5.8) can be written as

$$\langle X_{n+1}, e_i \rangle = a_n \langle X_n, e_i \rangle$$

and insertion of (4.1) gives

$$\left\langle X_{n+1}, X_i - \sum_{j=0}^{i-1} \langle X_j, e_j^* \rangle e_j^* \right\rangle = a_n \left\langle X_n, X_i - \sum_{j=0}^{i-1} \langle X_j, e_j^* \rangle e_j^* \right\rangle$$

that is,

$$\langle X_{n+1}, X_i \rangle - \sum_{j=0}^{i-1} \langle X_n, e_j^* \rangle \langle X_{n+1}, e_j^* \rangle = a_n \left[\langle X_n, X_i \rangle - \sum_{j=0}^{i-1} \langle X_n, e_j^* \rangle \langle X_n, e_j^* \rangle \right].$$

Combining the last equation with (5.8) we obtain

$$\langle X_{n+1}, X_i \rangle - \langle X_n, e_0^* \rangle \langle X_{n+1}, e_0^* \rangle = a_n [\langle X_n, X_i \rangle - \langle X_n, e_0^* \rangle \langle X_n, e_0^* \rangle],$$

that is,

$$(5.11) \quad C[X_{n+1}, X_i] = a_n C[X_n, X_i]; \quad i = 1, \dots, n-1.$$

The converse is straightforward.

The case for which $a_n \equiv 1$ and $b_n \equiv 0$ in (5.5) makes (5.4) a credibility formula of the updating type in the spirit of GERBER and JONES (1975). Equation (5.11) now reduces to the covariance structure.

$$(5.12) \quad C[X_i, X_j] = \begin{cases} V_i + W_i; & i = j \\ V_i; & i < j \end{cases}$$

in agreement with GERBER and JONES (1975).

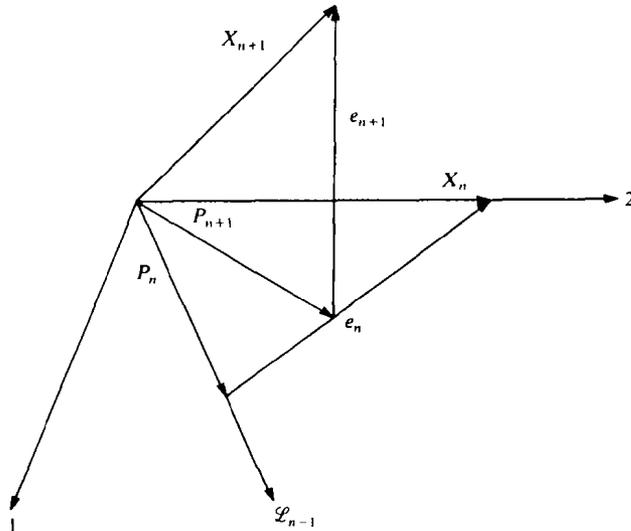


FIGURE 5.1 The geometry of credibility formulae of the updating type in the spirit of GERBER and JONES (1975)

Figure 5.1 shows the geometry of credibility formulae of the updating type. The co-ordinate axis labelled 2 represents X_n and the 1-2 plane represents \mathcal{L}_n . Let E_n represent the mean-square error of P_n , that is

$$E_n = \|e_n\|^2.$$

Figure 5.1 depicts the following orthogonality relations: $e_n \perp e_{n+1}$, $e_n \perp P_n$ and $e_{n+1} \perp \mathcal{L}_n$.

These may be used to obtain a number of expressions connecting Z_n and second-order moments of X_n , e_n , P_n etc. In particular

$$(5.13) \quad E_{n+1} = \|X_{n+1} - X_n\|^2 - (1 - Z_n)^2 E_n$$

assuming (5.12) holds.

We can also demonstrate (5.13) mathematically thus: From expression preceding (5.1)

$$X_{n+1} - X_n = P_{n+1} - P_n + e_{n+1} - e_n.$$

Substituting (5.4) with $\mathcal{P}_{n-1}(X_{n+1}) = P_n$ into the last equation gives

$$X_{n+1} - X_n = (1 - e_n)Z_n + e_{n+1}.$$

Recognizing the fact that $e_n \perp e_{n+1}$ now yields (5.13).

GERBER and JONES (1975) also derive the relations,

$$(5.14) \quad Z_1 = W_1(W_1 + V_1)^{-1}$$

$$(5.15) \quad Z_n = (W_n - W_{n-1} + Z_{n-1}V_{n-1})(W_n - W_{n-1} + Z_{n-1}V_{n-1} + V_n)^{-1}$$

which will be alluded to in the next section.

6. KALMAN TYPE FILTERS

In the present section we examine some applications of the algorithms developed earlier to two special models and relate them to the classical Kalman filter for which both measurement and system noises are Gaussian (JAZWINSKI (1969)).

6.1. Bühlmann Model

Consider a risk characterized by a parameter Y . Associated with this risk are measurements X_1, X_2, \dots . The following assumptions are made:

ASSUMPTION 1. *Conditional on Y fixed, the measurements X_1, X_2, \dots are independent.*

ASSUMPTION 2. *Conditional on Y fixed, the mean and variance of each X_i can be written $E[X_i|Y] = \mu(Y)$ and $C[X_i|Y] = \sigma^2(Y)$ respectively.*

Without loss of generality assume $\mu(Y) = Y$ and for notational convenience write, $y_0 = E[Y]$, $\sigma_0^2 = E[\sigma^2(Y)]$ and $v = C[Y]$

Adopting the notation and terminology of the preceding sections, the recursion for the inhomogeneous linear Bayes rule for Y is,

$$(6.1.1) \quad \hat{Y}_{n+1} = \hat{Y}_n + K_{n+1} e_{n+1}.$$

In view of assumptions 1 and 2 and the unbiasedness conditions (4.8) it follows that

$$\langle X_{n+1}, e_j \rangle = \langle Y, e_j \rangle; \quad j = 0, 1, \dots, n.$$

This means,

$$\mathcal{P}_n(X_{n+1}) = \mathcal{P}_n(Y) (= P_{n+1})$$

whence,

$$e_{n+1} = X_{n+1} - \hat{Y}_n.$$

Consider now the inner product,

$$\begin{aligned} \langle Y, e_{n+1} \rangle &= \langle Y - \hat{Y}_n + \hat{Y}_n, e_{n+1} \rangle \\ &= \langle Y - \hat{Y}_n, e_{n+1} \rangle, \end{aligned}$$

the latter equality following from $\hat{Y}_n \perp e_{n+1}$. Further,

$$\begin{aligned} \langle Y - \hat{Y}_n, e_{n+1} \rangle &= \langle Y - \hat{Y}_n, Y - \hat{Y}_n + X_{n+1} - Y \rangle \\ &= C_n \end{aligned}$$

the latter equality following from the orthogonality condition

$$Y - \hat{Y}_n \perp X_{n+1}$$

as a result of (4.8) and assumptions 1 and 2.

Consequently,

$$(6.1.2) \quad \langle Y, e_{n+1} \rangle = C_n.$$

Moreover,

$$\begin{aligned} (6.1.3) \quad \|e_{n+1}\|^2 &= \|X_{n+1} - \hat{Y}_n\|^2 \\ &= \|Y - \hat{Y}_n\|^2 + \|X_{n+1} - Y\|^2 \\ &= C_n + E[C[X_{n+1}|Y]] \\ &= C_n + \sigma_0^2 \end{aligned}$$

where the second equality follows from

$$Y - \hat{Y}_n \perp X_{n+1} - Y.$$

Substituting (6.1.2) and (6.1.3) into (4.10) yields,

$$(6.1.4) \quad K_{n+1} = C_n (C_n + \sigma_0^2)^{-1}.$$

For continuity write (4.12) again,

$$(6.1.5) \quad C_{n+1} = C_n - K_{n+1} C_n.$$

Substituting (6.1.4) into (6.1.5) gives

$$(6.1.6) \quad C_{n+1}^{-1} = C_n^{-1} + \sigma_0^{-2}.$$

In summary, we have developed the Kalman filter

$$(6.1.7) \quad \hat{Y}_{n+1} = \hat{Y}_n + K_{n+1}(X_{n+1} - \hat{Y}_n)$$

$$(6.1.8) \quad K_{n+1} = C_n(C_n + \sigma_0^2)^{-1}$$

$$(6.1.9) \quad C_{n+1}^{-1} = C_n^{-1} + \sigma_0^{-2}$$

with initial conditions

$$\hat{Y}_0 = y_0 \quad \text{and} \quad C_0 = v.$$

We point out that if Y has a Gamma distribution and $X_1|Y$ is Poisson with mean Y (implying that $\sigma^2(Y) = Y$ and $y_0 = \sigma_0^2$) then

$$\hat{Y}_n = E[Y|X_1, \dots, X_n]$$

and

$$C_n = E[C[Y|X_1, \dots, X_n]].$$

Moreover, by virtue of a fundamental result in linear Bayes theory (HARTIGAN (1969)), the same classical Kalman filter (6.1.7) to (6.1.9) is obtained if we assume instead that

$$Y \sim \text{Normal}(y_0, v)$$

and

$$X_n|Y \sim \text{Normal}(Y, \sigma_0^2).$$

See DE JONG and ZEHNWIRTH (1983) for more details.

In passing we also note that since $\mathcal{P}_n(X_{n+1}) = \mathcal{P}_n(Y)$ it follows that

$$K_{n+1} = Z_{n+1}.$$

Combining (6.1.4) and (6.1.5) gives

$$Z_{n+1} = Z_n(Z_n + 1)^{-1},$$

which is also a consequence of expression (5.15).

Moreover,

$$(6.1.10) \quad P_{n+1} = P_n + Z_n(X_n - P_n)$$

and

$$(6.1.11) \quad E_{n+1} = [(E_n - \sigma_0^2)^{-1} + \sigma_0^{-2}]^{-1} + \sigma_0^2.$$

The last expression also follows from (5.13).

6.2. Evolutionary Risk Parameter Model

In the present sub-section we imagine that we have a sequence of risk parameters Y_1, Y_2, \dots and corresponding measurements X_1, X_2, \dots . The measurement

equations are given by

ASSUMPTION 1

$$X_n | Y_n \sim \text{Poisson}(Y_n)$$

and

$$C[X_i, X_j | Y_i, Y_j] = 0, \quad i \neq j$$

The system equations (that is, the equations indicating how the parameters evolve over time) are given by

ASSUMPTION 2

$$E[Y_n | Y_{n-1}] = Y_{n-1}$$

and

$$C[Y_n | Y_{n-1}] = \nu_n, \quad (\nu_n \in \mathbb{R}).$$

We also assume independence between the measurement and system “noises”. That is,

ASSUMPTION 3

$$C[X_n, Y_{n+1} | Y_n] = 0.$$

Now, put $E[Y_1] = y_0$, a constant, and write

$$C_{n+1|n} = \|Y_{n+1} - \hat{Y}_n\|^2$$

and

$$C_{n+1} = \|Y_{n+1} - \hat{Y}_{n+1}\|^2$$

where in the present context,

$$\hat{Y}_n = \mathcal{P}_n(Y_n).$$

Applying the projection theorem to the decomposition (4.2) gives

$$\mathcal{P}_{n+1}(Y_{n+1}) = \mathcal{P}_n(Y_{n+1}) + K_{n+1}e_{n+1}$$

where now

$$K_{n+1} = \langle Y_{n+1}, e_{n+1} \rangle \|e_{n+1}\|^{-2}.$$

In view of assumptions 2 and 3

$$\langle Y_{n+1}, e_j \rangle = \langle Y_n, e_j \rangle; \quad j = 0, 1, \dots, n$$

whence,

$$\mathcal{P}_n(Y_{n+1}) = \mathcal{P}_n(Y_n).$$

Similarly, in view of assumption 1

$$\mathcal{P}_n(X_{n+1}) = \mathcal{P}_n(Y_{n+1}).$$

It follows that,

$$(6.2.1) \quad \hat{Y}_{n+1} = \hat{Y}_n + K_{n+1}e_{n+1}$$

where

$$e_{n+1} = X_{n+1} - \hat{Y}_n.$$

Consider the inner product

$$(6.2.2) \quad \begin{aligned} \langle Y_{n+1}, e_{n+1} \rangle &= \langle Y_{n+1} - \hat{Y}_n + \hat{Y}_n, e_{n+1} \rangle \\ &= \langle Y_{n+1} - \hat{Y}_n, e_{n+1} \rangle \\ &= \langle Y_{n+1} - \hat{Y}_n, Y_{n+1} - \hat{Y}_n + X_{n+1} - \hat{Y}_{n+1} \rangle \\ &= C_{n+1|n}, \end{aligned}$$

the second equality follows by noting that

$$\hat{Y}_n \perp X_{n+1} - \hat{Y}_n$$

and the last equality follows by noting that

$$Y_{n+1} - \hat{Y}_n \perp X_{n+1} - \hat{Y}_{n+1}.$$

Now,

$$\begin{aligned} C_{n+1|n} &= \|Y_{n+1} - Y_n + Y_n - \hat{Y}_n\|^2 \\ &= C + \|Y_{n+1} - Y_n\|^2, \end{aligned}$$

since

$$Y_{n+1} - Y_n \perp Y_n - \hat{Y}_n.$$

Hence,

$$(6.2.3) \quad C_{n+1|n} = C_n + \nu_n.$$

Turning now to the computation of $\|e_{n+1}\|^2$ we have

$$(6.2.4) \quad \begin{aligned} \|e_{n+1}\|^2 &= \|Y_{n+1} - \hat{Y}_n\|^2 + \|X_{n+1} - Y_{n+1}\|^2 \\ &= C_{n+1|n} + E[Y_{n+1}] \\ &= C_{n+1|n} + y_0, \end{aligned}$$

the second equality following from assumptions 1 and 2.

Application of (4.9) with Y_{n+1} playing the role of Y gives

$$(6.2.5) \quad C_{n+1} = C_{n+1|n} - K_{n+1} \langle Y_{n+1}, e_{n+1} \rangle.$$

Combining equations (6.2.1) to (6.2.5) yields the Kalman filter

$$(6.2.6) \quad \hat{Y}_{n+1} = \hat{Y}_n + K_{n+1}(X_{n+1} - \hat{Y}_n)$$

$$(6.2.7) \quad K_{n+1} = C_{n+1|n}(C_{n+1|n} + y_0)^{-1}$$

$$(6.2.8) \quad C_{n+1}^{-1} = C_{n+1|n}^{-1} + y_0^{-1}$$

and

$$(6.2.9) \quad C_{n+1|n} = C_n + \nu_n$$

with initial conditions

$$\hat{Y}_0 = y_0 \quad \text{and} \quad C_{1|0} = \nu_0.$$

We point out again the connection with the classical Kalman filter. That is, if instead of assumption 1 we have:

ASSUMPTION 1

$$X_n | Y_n \sim \text{Normal}(Y_n, y_0).$$

In addition to assumptions 2 and 3 we also assume

ASSUMPTION 4

$$Y_n | Y_{n-1} \sim \text{Normal}(Y_{n-1}, \nu_n).$$

The same Kalman filter (6.2.6) to (6.2.9) is obtained.

The prospective rating algorithm is given by

$$(6.2.10) \quad P_{n+1} = P_n + Z_n(X_n - P_n)$$

and

$$(6.2.11) \quad E_{n+1} = \{(E_n - y_0)^{-1} + y_0^{-1}\}^{-1} + y_0$$

where again

$$Z_n = K_n$$

Although the two preceding models satisfy (5.12) we conclude by emphasizing that the general algorithms presented in Sections 4 and 5 may be applied to any model and in particular the models considered by SUNDT (1981, 1983) satisfying the more general structure (5.10) and (5.11).

REFERENCES

- DE JONG, P and ZEHNWIRTH, B (1983) Credibility Theory and the Kalman Filter *Insurance Mathematics and Economics* 2, 281-286
- DE VYLDER, F (1976) Geometrical credibility *Scandinavian Actuarial Journal* 121-149
- GERBER, H U and JONES, D A (1975) Credibility formulas of the Updating Type In *Credibility Theory and applications*, edited by P M Kahn Academic Press New York, pp 89-105
- HARTIGAN, J A (1969) Linear Bayesian Methods *Journal Royal Statistical Society Series B* 31, 446-454
- JAZWINSKI, A H (1970) *Stochastic Processes and Filtering Theory* Academic Press New York
- JEWELL, W S (1976) Two Classes of Covariance Matrices Giving Simple Linear Forecasts *Scandinavian Actuarial Journal*, 15-29
- KAILATH, T (1974) A View of Three Decades of Linear Filtering Theory *IEEE Trans Inform Thy* 19, 750-760
- KREMER, E (1982) Exponential Smoothing and Credibility Theory *Insurance Mathematics and Economics* 1, 213-217
- LUENBERGER, D G (1969) *Optimization by vector space methods* John Wiley New York
- NORBERG, R (1979) The Credibility Approach to Experience Rating *Scandinavian Actuarial Journal*, 181-221
- SUNDT, B. (1981) Recursive Credibility Estimation *Scandinavian Actuarial Journal*, 3-21
- SUNDT, B (1983) Finite Credibility Formulae in Evolutionary Models *Scandinavian Actuarial Journal*, 106-116

