case prior opinion may be incorporated informally, subsequent to the analysis. In any case, the distribution assigned to β and θ should be regarded as part of the model.

5. The posterior distribution of w (i.e., the conditional distribution of w given y) may suggest suitable point and interval predictors. With the possible exception of the term posterior distribution, which might be used in referring to any distribution that is conditional on y, the use of Bayesian jargon should be avoided.

Comment: The Kalman Filter and BLUP

James C. Spall

1. INTRODUCTION

Professor Robinson has given a wide-ranging account of best linear unbiased prediction with an impressive array of examples and applications. In this discussion, however, I will restrict my attention to issues regarding the Kalman filter and BLUP.

For ease of discussion, let us restate the random effects model in state-space form as given in Robinson, Section 6. The unobservable random effects (state) vector, u_t, evolves according to

$$u_t = G_t u_{t-1} + w_t,\ u_0 = 0,\ t = 1, 2, \ldots, n, \tag{1.1a}$$

where w_t is a noise term with mean 0 and covariance matrix W_t, and G_t is the state transition matrix. The second equation in the model relates the state vector to the vector of observables y_t:

$$y_t = F_t u_t + v_t, \tag{1.1b}$$

where v_t is a noise term with mean 0 and covariance matrix V_t, and F_t is the measurement matrix. Equations (1.1a, b) can be expressed in the random effects model form of Robinson by writing

$$y = Zu + e,$$

where $y = (y_1^T, y_2^T, \ldots, y_n^T)^T$, $Z = \text{block diag}[F_1, F_2, \ldots, F_n]$, $u = (u_1^T, u_2^T, \ldots, u_n^T)^T$, and $e = (v_1^T, v_2^T, \ldots, v_n^T)^T$. The covariance matrix for u, G in the notation of Robinson, is a function of G_t and W_t, $t = 1, 2, \ldots, n$. The structure of this covariance matrix allows for recursive algorithms of the Kalman filter/smoother form to be used to form BLUP estimates for the components of u. Incidentally, a slightly confusing point in Robinson, Subsection 6.4, is that it is a Kalman smoother, not filter, that produces the BLUP estimate of u based on data y. What Robinson had in mind, I presume, is the common problem where one is interested in an estimate of u_t based only on data through time t (not through some later time); the Kalman *filter*, of course, is used for this problem. For the remainder of this discussion, I will assume that the filtering problem is the one of interest (although virtually all of the ideas would also apply in the smoothing problem).

A couple of other points are worth noting here. First, Sallas and Harville (1988) address a slightly broader problem than that considered above and by Robinson: namely the estimation of random *and* fixed effects via Kalman filter techniques. Second, as noted by Robinson, the Kalman filter is not entirely due to Kalman. The filter equations were essentially derived by others prior to Kalman, but it was Kalman who crystallized much of the thinking in the area and discovered several key relationships to certain systems-theoretic concepts (see Spall, 1988, for further discussion of this).

In the next two sections, I will discuss two problems that were given fairly light treatment in the Robinson paper, but that are important from the point of view of a practitioner. Section 2 describes some problems associated with constructing uncertainty bounds for the filter estimation error $\hat{u}_n - u_n$ when the noise terms have an unknown distribution (as in the general setting of Robinson, equation 1.1). Section 3 elaborates on the brief discussion of Robinson regarding uncertainty in the model parameters θ.

2. UNCERTAINTY BOUNDS FOR $\hat{u}_n - u_n$ IN DISTRIBUTION-FREE SETTINGS

Robinson presents the formula for the covariance matrix of the BLUP estimation error in Section 1 of his paper, and it is well known that this covariance

James C. Spall is a member of the Senior Professional Staff at The Johns Hopkins University, Applied Physics Laboratory, Laurel, Maryland 20723-6099.

matrix can be computed recursively as part of the Kalman filter (or smoother, as appropriate). However, Robinson does not present any discussion as to how this (or other) information can be used to construct probability bounds or uncertainty regions for the estimation error $\hat{u}_n - u_n$ in the distribution-free setting that is the basis for most of the paper (in the case where w_t, v_t are normally distributed for all t, the estimation error is also normally distributed for all t, which makes it relatively straightforward to derive these quantities). We will outline below an approach to allow one to construct probability bounds (and hence uncertainty regions) in the distribution-free setting.

Before presenting the approach, it is worth noting that unlike many other estimators, the Kalman filter estimation error is not asymptotically normally distributed. This follows because central limit effects do not hold (as a result of the disproportionate weight given to more recent terms in the sum), as shown below. Thus standard asymptotic procedures for uncertainty region calculation (such as those that are used in maximum likelihood parameter estimation) do not apply. To establish this asymptotic nonnormality, let us consider models that are in standard uniformly completely controllable and observable (UCC and UCO) form (see, e.g., Jazwinski 1970, pages 232–234; Spall and Wall, 1984; or Anderson and Moore 1979, pages 68–82). Recall from the Cramér–Lévy theorem (e.g., Feller, 1971, pages 525–526), that the sum of two nondegenerate independent random vectors is normally distributed only if both of the random vectors are normally distributed. Straightforward algebra shows that

$$(2.1)\quad \hat{u}_n - u_n = s_{n-1} + (K_n F_n - I)w_n + K_n v_n,$$

where s_{n-1} is the weighted sum of $\{w_t, v_t\}_{t=1}^{n-1}$ and K_n is the Kalman gain matrix. Since UCC and UCO imply that $\alpha_0 I \le P_n \equiv \text{cov}(\hat{u}_n - u_n) \le \alpha_1 I$ with $0 < \alpha_0 \le \alpha_1 < \infty$, we know that $\|K_n F_n - I\| \ge \alpha_2$ and $\|K_n\| \ge \alpha_2$ for some $\alpha_2 > 0$ and any matrix norm (Jazwinski, 1970, pages 234–237; Deyst and Price, 1968; or Deyst, 1973). Thus by (2.1) the contributions of the w_n and v_n terms to $\hat{u}_n - u_n$ will be nonnegligible as $n \to \infty$. Then by the Cramér–Lévy theorem $\hat{u}_n - u_n$ is not asymptotically normal. The implications of the above have been examined in several simulation experiments with uniformly distributed noise terms. Applications of the Kolmogorov–Smirnov goodness-of-fit test (with a null hypothesis that $\hat{u}_n - u_n \sim N(0, P_n)$) yielded p values of less than 10^{-6} for a variety of state-space parameter values, confirming that it is dangerous to assume that $\hat{u}_n - u_n$ is even approximately normal.

The approach to characterizing the uncertainty in $\hat{u}_n - u_n$ is to bound the probability $P(\hat{u}_n - u_n \in E_n)$ for rejection regions E_n such that the complementary region E_n^c is symmetric and convex. For example, E_n might represent the points outside a q-dimensional spheroid, i.e., the values of $\hat{u}_n - u_n$ such that $\|\hat{u}_n - u_n\| \ge c$ for Euclidean norm $\|\cdot\|$ and some $c > 0$. Note that one easy bound is that given by Chebyshev's inequality, e.g., $P(\|\hat{u}_n - u_n\| \ge c) \le E\|\hat{u}_n - u_n\|^2/c^2 = \text{trace} P_n/c^2$. We seek a bound that has the potential to be more precise than the Chebyshev bound. Anderson's inequality (Anderson, 1955; Tong, 1980, page 55) provides a means to this goal.

Let us write

$$\hat{u}_n - u_n = A_n \bar{w}_n + B_n \bar{v}_n$$

where $\bar{w}_n = (w_1^T, \dots, w_n^T)^T$, $\bar{v}_n = (v_1^T, \dots, v_n^T)^T$, $A_n = (A_{n1}, A_{n2}, \dots, A_{nn})$, $B_n = (B_{n1}, B_{n2}, \dots, B_{nn})$, and A_{nt}, B_{nt} are weighting matrices as derived from the Kalman filter and the state equation (see Spall and Wall, 1984, equation 2.2, for B_{nt} and the filter contribution to A_{nt}). Suppose that either the $\{w_t\}$ or $\{v_t\}$ process is normally distributed and that the other sequence has an unknown symmetric, unimodal distribution (the partial assumption of normality is stronger than required for the technique, but is made here for ease of discussion). We now create a surrogate expression that will be used to form a probability that bounds $P(\hat{u}_n - u_n \in E_n)$. By the fact that UCO and UCC imply that the filter is exponentially stable (Jazwinski, 1970, pages 240–242), we have that $\|A_{nt}\| = O(e^{-c_0(n-t)})$ and $\|B_{nt}\| = O(e^{-c_1(n-t)})$ for some $c_0, c_1 > 0$ as $n - t \to \infty$. To apply Anderson's theorem, we leave the weighting matrix sequence associated with the normally distributed noise process unchanged, but modify the other sequence so that all of those weighting matrices have (at least approximately) magnitude equal to the largest magnitude matrix in the sequence. For convenience, suppose that $\{v_t\}$ is the Gaussian sequence. We then create a modified sequence $A_{nt}^* = a_{nt} A_{nt}$ where $|a_{nt}| \ge 1$ and $\|A_{nt}^*\| = O(1)$. Then $A_n^* \bar{w}_n$ is approximately normally distributed by the Lindeberg–Feller form of the central limit theorem and so $A_n^* \bar{w}_n + B_n \bar{v}_n$ is approximately normally distributed with mean 0 and covariance matrix $A_n^* \text{block diag}[W_1, W_2, \dots, W_n] A_n^{*T} + B_n \text{block diag}[V_1, V_2, \dots, V_n] B_n^T$. Since $|a_{nt}| \ge 1$, an iterative application of Anderson's theorem to each w_t term implies

$$\begin{aligned} P(\hat{u}_n - u_n \in E_n) &= 1 - P(A_n \bar{w}_n + B_n \bar{v}_n \in E_n^c) \\ (2.2\text{a})\qquad &\le 1 - P(A_n^* \bar{w}_n + B_n \bar{v}_n \in E_n^c) \\ &= P(A_n^* \bar{w}_n + B_n \bar{v}_n \in E_n). \end{aligned}$$

TABLE 1
Probability values for $\{|\hat{u}_9 - u_9| \geq c\}$

c	Normal	Chebyshev	Bound (2.2a)
$\sqrt{P_9}$	.32	1.0	.56
$2\sqrt{P_9}$	.05	.25	.24
$3\sqrt{P_9}$	.003	.11	.08
$4\sqrt{P_9}$	0^+	.06	.02

Similarly if $\{w_t\}$ has a normal distribution and $\{v_t\}$ has an unknown distribution, we have

$$P(\hat{u}_n - u_n \in E_n) \leq P(A_n\bar{w}_n + B_n^*\bar{v}_n \in E_n), \tag{2.2b}$$

where the $\{B_{nt}^*\}$ are chosen in the same manner as $\{A_{nt}^*\}$ above, and $A_n\bar{w}_n + B_n^*\bar{v}_n$ is approximately normally distributed with mean 0 and the obvious covariance matrix.

For bounded E_n^c (the usual case) and large n, the probability bounds in (2.2a, b) may not be satisfactory since they will approach unity due to the fact that $A_n^*\bar{w}_n$ and $B_n^*\bar{v}_n$ are of order n. However, for shorter realizations (but long enough to achieve practical central limit theorem effects) the bounds may represent an improvement over the Chebyshev inequality, as illustrated below.

For a scalar u_t and y_t setting, Table 1 compares the bound of (2.2a) (using $A_{nt}^* = \max_t|A_{nt}|\ \forall t$) with the probability values resulting from an assumption of normality for all noise terms (so $\hat{u}_n - u_n$ is normal) and from an application of the Chebyshev inequality. All state-space parameters (G_t, W_t, F_t, and V_t) were taken to be unity and, as with Robinson's first-lactation example, $n = 9$. As expected, the probability values for bound (2.2a) lie between those of the Chebyshev inequality and those resulting from the normal distribution assumption for $\hat{u}_9 - u_9$.

3. ESTIMATION OF MODEL PARAMETERS

As noted by Robinson, the problem of estimating variance parameters in linear models has received considerable attention. In the context of the state-space model (1.1a, b) this "identification" problem has been treated extensively in the statistics and (especially) control systems literature, including several special issues of the *IEEE Transactions on Automatic Control* and *Automatica* (most recently the January 1990 issue of *Automatica*).

One of the beauties of the state-space formulation of the random effects model is that under the assumption of normally distributed noise terms the Kalman filter can be used for calculating the log-likelihood function and its derivatives, in addition to its usual application to state (u_t) estimation. This is achieved by writing the log-likelihood in terms of the independent "innovations" sequence $\{y_t - F_tG_t\hat{u}_{t-1}\}_{t-1}^n$. This process has been described in many places (e.g., Goodwin and Payne, 1977, pages 158–159; Sallas and Harville, 1988). As part of this process, Kalman filter type recursions yield the Fisher information matrix, which, when inverted, can serve as an approximation to the covariance matrix of the parameter estimation error.

As suggested by Robinson, the uncertainty in model parameters should be reflected in the uncertainty of the BLUP estimate. This may be viewed as a problem in nuisance parameter analysis, i.e., the estimation of quantities of interest in the presence of uncertainty in other terms within the model. Several authors have considered this problem in the context of state-space models. Ansley and Kohn (1986) present an expression for the asymptotic mean square error for the Kalman filter state estimate $\hat{u}_t$ in the presence of uncertain model parameters. Spall and Garner (1990) consider how uncertainty in certain model parameters (e.g., uncertainty in a scale factor such as σ^2 in the model of (1.1) in Robinson) will affect the precisions for estimates of other nonrandom parameters (e.g., estimates of θ or the fixed effects β in the Robinson Model). Both the Ansley–Kohn and Spall–Garner approaches rely on Kalman filter type recursions, the former on differentiated state estimates and the latter on a differentiated log-likelihood function and score vector.

4. CONCLUDING REMARKS

Although largely in the domain of the control and aerospace engineering communities until the early 1970s, the Kalman filter has now been embraced by the statistical community as the method of choice for a wide range of time series problems. The Kalman filter/state-space approach has been successfully applied in countless problems drawn from perhaps every major branch of the physical and social sciences. For these reasons, I think that the Kalman filter approach is, in the sense of Robinson, "a good thing."