KM 38 (1991)
pg.45-52

Optimality and Initialisation of the Kalman Filter.

Comments on "A note on the calculation of latent trajectories in the quasi Markov simplex model by means of the regression method and the discrete Kalman filter" by C.V. Dolan & P.C.M. Molenaar.

J. Molenaar

and

J.H.L. Oud

1. Introduction.

In the following we refer to Dolan and Molenaar's paper mentioned above by DM. In DM a comparison is made between two methods to estimate latent trajectories: the discrete Kalman filter and the regression estimator. This comparison is interesting because it brings together people from the field of control theory on the one hand and people mainly acquainted with factor analytic and regression techniques on the other. However, the presentation in DM has some confusing aspects culminating in the statement that the regression estimator is superior compared to the Kalman filter. This statement is remarkable, since the Kalman filter is designed to be optimal in control theory and so, according to DM, the regression method should be something more optimal than optimal. To clarify the most important issues we shall focus below on two points. In §2 we point out the characteristics of estimation theory in a dynamical system. In §3 we show the consequences of specific choices for the initialisation of the Kalman filter. Our conclusions are summarized in §4. In Appendix A it is pointed out under which conditions the Kalman filter reduces to the regression resp. Bartlett estimator.
For ease of comparison, our comments will follow, as far as possible, the notation used in DM, although this is not commonly accepted in the literature. We shall notationally distinguish the estimator from the estimate by underlining the estimator.
For readers not familiar with the terminology used in the social sciences we explain two frequently used terms: an estimator is called 'cross-sectional', if it is based on data gathered at

Institute for Mathematics Consulting
Faculty of Mathematics & Computing Science
Eindhoven University of Technology (TUE)
P.O. Box 513
5600 MB EINDHOVEN
The Netherlands

Instituut voor Orthopedagogiek
University of Nijmegen (KUN)
P.O. Box 9103
6500 HD NIJMEGEN
The Netherlands

a single point in time, whereas it is called 'longitudinal', if it makes use of data, gathered at distinct, successive points in time. Part of the confusion in DM stems from the fact that the regression estimator is originally designed as a cross-sectional estimator (just as the Bartlett estimator; see [3] but that it is applied longitudinally in DM. How the regression method relates to the Kalman filter and to the Kalman smoother, when the former is applied longitudinally, will be addressed below.

2. **Dynamical versus statical estimation in the longitudinal case.**

By a dynamical system we understand a system which evolves in time. We assume that the system is characterized by a vector $\eta(t)$. In control theory $\eta(t)$ is called 'state', whereas in factor analysis the terms 'factors' or 'latent variables' are used. The vector $\eta(t)$ is in general not directly measurable, but it has to be estimated from measurements on related properties of the system. We assume that data $\mathbf{y}(t)$ is acquired on successive, discrete times t_i, $i=1,2,\ldots$. It is the great merit of Kalman [2] to have shown that under some assumptions the estimation of $\eta(t_j)$ from data $\{\mathbf{y}(t_1),\ldots,\mathbf{y}(t_j)\}$ can be performed by a recursive algorithm. It is therefore necessary that the states $\eta(t_{i+1})$ and $\eta(t_i)$ at consecutive times are linearly related (eq. (6) in DM). Also, a linear relationship between $\eta(t_i)$ and the data $\mathbf{y}(t_i)$ at all times t_i has to be assumed (eq.(7) in DM). Furthermore, the matrices $B(t_{i+1},t_i)$ $\Lambda(t_i)$, $\Psi(t_i)$ and $\Theta(t_i)$, as defined in DM, have to be known in advance. In practice, the determination of these matrices, i.e. the modelling and estimation of the system, is a separate problem. It might be solved, for example, by means of the LISREL program as mentioned in DM. If data become available continuously in time instead of at discrete times, the so-called Kalman-Bucy filter could be applied. The mathematical derivation of the Kalman-Bucy filter is considerably more complex than that of the discrete time Kalman filter [11].

The algorithm by Kalman yields the estimator $\underline{\eta}(t_i|t_j)$ for the state $\eta(t_i)$ in terms of the data $\{\mathbf{y}(t_1),\ldots,\mathbf{y}(t_j)\}$. The cases $t_i > t_j$, $t_i = t_j$ and $t_i < t_j$ are referred to as prediction, filtering and smoothing, respectively. The recursive formulae for filtering and (fixed interval) smoothing are given in DM. The estimator $\underline{\eta}(t_i|t_j)$ is designed to be unbiased and to have minimum variance. Whether it really has these properties depends on the initialisation of the recursive scheme. We shall deal with this point further in §3.

It is important to realize that the Kalman filter has originally been derived to cope with the problem of estimating the state of a dynamical system in real time: as soon as new data $\mathbf{y}(t_{i+1})$ becomes available, the current estimate $\eta(t_i|t_i)$ can be updated (via formulae (8)-(11) in DM), resulting in the new estimate $\eta(t_{i+1}|t_{i+1})$. Because of its recursiveness, the procedure is efficient both in time and storage. As counterpart of dynamical estimation, or filtering at consecutive times, we like to introduce the term statical estimation for the case that all data have been recorded before the estimation starts. If the underlying system model has a recursive character, as is e.g. embodied in eq.(6) in DM, and the estimation is statical, application of the Kalman fixed interval smoother $\underline{\eta}(t_i|T)$ comes into consideration with T the latest time point. Because the Kalman smoother $\underline{\eta}(t_i|T)$ is based on more information, viz. $\{\mathbf{y}(t_1),...,\mathbf{y}(T)\}$, than the Kalman filter $\underline{\eta}(t_i|t_i)$, which merely exploits the data $\{\mathbf{y}(t_1),...,\mathbf{y}(t_i)\}$, it is quite clear that the smoothed estimate is more reliable than the filtered one, except for $t_i = T$. This manifests itself in the variances: the diagonal elements of the filter covariance matrix $\mathbf{V}(t_i)$ in (11) of DM are always larger than or equal to those of the smoothed covariance matrix $\mathbf{V}(t_i|T)$ in (21) of DM.

In statical estimation other estimators are, of course, also applicable, e.g. the regression or the Bartlett estimators, though these originally cross-sectionally designed methods do not take into account any recursive character of the underlying model. In DM it is argued that for statical estimation the Kalman smoother and the regression estimator yield the same results. It should be noted that this central theme in DM is proven for T=2 and with respect to a special model, the Markov simplex model (see appendices in DM). It is illustrated for T=10 with this special model by means of a simulation study. However, because both estimators are designed to have minimum variance, it is trivially true for the general case as well under certain initialisation conditions for the Kalman smoother. These conditions will be pointed out in §3. This does certainly not mean that both methods are equally valuable in practice. Because the Kalman smoother makes use of the recursiveness of the model, the matrices involved have the sizes q x q, q x m and m x m with q and m the dimensions of η and $\mathbf{y}$ respectively. These dimensions are independent of the length T of the time series. Application of the regression estimator requires matrices, which are at least a factor of T larger, so that this method quickly becomes intractable for increasing T. This aspect has also been stated in DM. However, also for short time series the Kalman smoother probably deserves preference because of its greater efficiency. Also the calculation of the smoother covariance matrix $\mathbf{V}(t_i|T)$ can be performed easily on the basis of small sized matrices. In fact, there is little in favor of using the regression method instead of the Kalman smoother in case of statical estimation. Moreover, in the social sciences and in control theory estimation is often dynamical. The Kalman filter has many potentialities as a device for monitoring the development of individuals and it is currently used in that way in education [5]. For that purpose, it clearly makes no sense to wait with the data processing until the end of development is reached. On the contrary, the monitoring is usually meant to make timely intervention possible.

3. Initialisation of the Kalman filter

Because the Kalman filter is recursive, one has to start it at some (virtual) time, say $t_0 < t_1$. In § 3.3 of [9] several possibilities are listed. The initialisation of the Kalman filter and the consequences for its optimality are seldom dealt with systematically in the literature. This is not surprising, because it may be shown, see e.g. [1], that under certain conditions the influence of the initialisation damps out if time elapses. This aspect has been demonstrated in [12] via a simple example. By plotting the diagonal elements of $\mathbf{V}(t_i)$ it is directly established from visual

inspection when the influence of initialisation gets lost. So, apart from a transient regime, the filter results are independent of the initialisation. This implies that in the long run the filter (and also the predictor and smoother) is unbiased and has minimum variance.
In DM, as often in social sciences, the emphasis is on short time series, which are still in the transient regime. Let us therefore examine this aspect more carefully. As dealt with below, the Kalman filter has its optimality properties of unbiasedness and minimum variance at one time if it had these properties already at some time earlier in its history. So, formally, one should have at one's disposal at the (virtual) time t_0 an estimate obtained from an estimator with the optimality properties. Let us denote this estimate and its associated covariance matrix by η_0 and $\mathbf{V}_0$. We also have to know the transition matrix $B(t_1,t_0)$ in (6) of DM and the noise covariance matrix $\Psi(t_0)$. If started with initial estimates $\eta(t_0|t_0) = \eta_0$ and $\mathbf{V}(t_0) = \mathbf{V}_0$, the filter would remain unbiased and keep minimum variance for all t_i, i= 1,2,... . This holds for filtering as well as for prediction and smoothing. As for the minimum variance property this is inherent to the derivation of the filter, see e.g. [1,2,4]. As for the unbiasedness this can be shown as follows. Let us introduce the estimation error $e(t_i)$ by

$$e(t_i) = \underline{\eta}(t_i|t_i) - \eta(t_i)$$

From (6)-(8) in DM one may deduce the recursive relation

$$e(t_{i+1}) = (\mathbf{I} - \mathbf{K}(t_{i+1})\Lambda(t_{i+1}))B(t_{i+1},t_i)e(t_i) + (\mathbf{I} - \mathbf{K}(t_{i+1})\Lambda(t_{i+1}))\ \zeta(t_i) + \mathbf{K}(t_{i+1})\ \varepsilon(t_{i+1})$$

If we take the conditional mean of both sides, given the initial state $\eta(t_0)$, we find that

$$E(e(t_{i+1}) \mid \eta(t_0)) = 0 \quad \text{iff} \quad E(e(t_i) \mid \eta(t_0)) = 0.$$

This implies that, conditionally on $\eta(t_0)$, $\underline{\eta}(t_{i+1}|t_{i+1})$ is unbiased if and only if $\underline{\eta}(t_i|t_i)$ is unbiased. So, if $\underline{\eta}(t_0|t_0)$ is unbiased, this also holds for all $\underline{\eta}(t_i|t_i)$, i = 1,2,... .

In practice, however, one has at time t_0 little or no information about the system. Then, two strategies come into consideration:

a. One takes more or less reliable guesses for the prediction estimates $\eta(t_1|t_0)$ and $\mathbf{V}(t_1|t_0)$ and starts the filter by applying (8) and (11) in DM. The consequence is that in the transient regime the filter will in general neither be unbiased nor have minimum variance.

b. One chooses $\eta(t_1|t_0)$ arbitrarily and for $\mathbf{V}(t_1|t_0)$ a diagonal matrix with very big diagonal elements. The filter interprets this choice as if information is transferred from t_0 to t_1, but this information is completely unreliable and therefore ignored. As pointed out in Appendix A, and also in [6,8], the Kalman filter estimator $\underline{\eta}(t_1|t_1)$ reduces in this case to the cross-sectional Bartlett estimator at time t_1 which is unbiased, i.e. $E(e(t_1)|\eta(t_1)) = 0$ and also $E(e(t_1)|\eta(t_0)) = 0$. Because the filtered estimator at t_1 is thus unbiased, all filtered, predicted and smoothed estimators at the following times are also unbiased.

In DM approach a. is approximately followed. From the Appendix in DM it may be deduced that one sets rather arbitrarily $\eta(t_1|t_0) = 0$ and $\mathbf{V}(t_1|t_0) = E[\eta(t_1)\eta'(t_1)]$. In Appendix A it is pointed out that for this specific choice the Kalman estimator $\underline{\eta}(t_1|t_1)$ reduces to the cross-sectional regression estimator at time t_1. This estimator is biased, $E(e(t_1)|\eta(t_1)) \neq 0$ and $E(e(t_1)|\eta(t_0)) \neq 0$, although it has less variance than the (minimum variance unbiased) Bartlett estimator. For $t_i \geq t_1$ it leads to Kalman filter estimators $\underline{\eta}(t_i|t_j)$, which are biased. The corresponding variances, i.e. the diagonal elements of $\mathbf{V}(t_i)$, are in general and especially in the transient regime smaller than for the initialisation approach b. It is, however, by no means clear that this is an advantage in view of the clear bias. Altogether, there does not seem much to recommend the regression estimator when one is interested in estimating particular subjects' latent values or trajectory. In our view method b. using the Bartlett estimator, is the preferred way of initialisation of the Kalman filter in most practical situations. It is a cautious procedure, because it allows the estimator to have some more variance for getting minimum variance unbiasedness in return.

4. Conclusions

We summarize our conclusions:

i. For dynamical estimation the recursive Kalman filter is preferable to the longitudinal application of the regression estimator. The implementation of the Kalman filter is simple and the matrices involved are independent of the length of the time series. Also, the covariance matrix estimates can easily be calculated in terms of the same matrices before any measurement is done.

ii. For statical estimation the initialisation of the Kalman filter in DM is chosen such that the (longitudinally applied) regression estimator based on the data $\{\mathbf{y}(t_1),...,\mathbf{y}(T)\}$, becomes equivalent to the Kalman smoothed estimator $\underline{\eta}(t_i|T)$, so that in Table 2 the equivalence of columns 2 and 3 and of colums 5 and 6 is trivial. For the same reasons as given under i., the Kalman smoother is always to be preferred to the regression estimator.

iii. Comparison of colums 1 and 2 and colums 4 and 5 in Table 2 in DM is not fair, because colums 1 and 4 are based on less information than used for colums 2,3,5 and 6 (except for t = 10). It gives the false impression that the regression estimator as developed in factor analysis would be more optimal than the Kalman filter.

iv. For short time series the results of the Kalman filter (and also of the predictor and smoother) are quite sensitive to the way the filter is initialized. In most practical applications method b in the preceding section, using the Bartlett estimator, is the preferred procedure to start the filter.

References.

1. Jazwinski, A.C., 1970, Stochastic processes and filtering theory, Academic Press (New York).
2. Kalman, R.E., 1960, A new approach to linear filtering and prediction theory, Trans.Am.Soc.Met.Eng., J. Basic Eng. D 82, pp. 35-43.
3. Lawley, D.N., & Maxwell, A.E. (1971), Factor analysis as a statistical method. London: Butterworths.
4. Molenaar, J. and Visser, H., February 1989, Fundamental aspects of the Kalman filter with examples regarding load forecasting and acid rain, KEMA Scientific & Technical Reports, ISSN 0167-8590.
5. Oud, J.H.L. (199), Leerlingen volgen met LISKAL. In C. Aarnoutse & M. Voeten (Eds.), Gaat en onderwijst: Liber amicorum voor dr. M.J.C. Mommers (pp. 99-120). Tilburg: Zwijsen.
6. Oud, J.H., v.d. Bercken, J.H. and Essers, R.J., 1990, Longitudinal factor score estimation using the Kalman filter, Appl. Psychological Measurement, Vol. 14, No. 4, pp. 395-418.
7. Otter, P.W. (1986), Dynamic structural systems under indirect observation: identifiability and estimation aspects from a system theoretic perspective, Psychometrika, 51, 415-428.
8. Oud, J.H. and Molenaar, J., 1988, The Kalman filter as a factor score estimator compared with the regression and Bartlett estimators, Proceedings of the SMABS-88 conference, Groningen, pp. 210-222.
9. Poulisse, H.N.J., 1983, State- en parameter estimation as chemometrical concepts, thesis, University of Nijmegen.
10. Priestly, M.B. & Subba Rao, T. (1975), The estimation of factor scores and Kalman filtering for discrete stationary processes. International Journal of Control, 21, 971-975.
11. Ruymgaart, P.A. and Soong, T.T., 1988, Mathematics of Kalman-Bucy filtering, Springer Verlag, Berlin etc.
12. Visser, H. and Molenaar, J., December 1988, Kalman filter analysis in dendroclimatology, Biometrics, Vol. 44, pp. 929-940.

Appendix A.

Here, we point out under which conditions the Kalman filter reduces to the regression resp. Bartlett estimator. The latter estimator is also known as the generalized least squares estimator. The insights presented here are in particular important for the discussion concerning the initialisation of the Kalman filter.

Applying eq. (8) in DM at $t_0 = 0$, the Kalman filter estimator $\eta(1|1)$ with $t_1=1$ can be written as

$$\underline{\eta}(1|1) = [\mathbf{I} - \mathbf{K}(1)\Lambda(1)]\eta(1|0) + \mathbf{K}(1)\mathbf{y}(1) \tag{A.1}$$

with $\mathbf{I}$ the $q \times q$ unity matrix. The corresponding covariance matrix $\mathbf{V}(1)$ follows from (11) in DM:

$$\mathbf{V}(1) = [\mathbf{I} - \mathbf{K}(1)\Lambda(1)]\mathbf{V}(1|0) \tag{A.2}$$

The gain matrix $\mathbf{K}$ is given, according to (9) in DM, by

$$\mathbf{K}(1) = \mathbf{V}(1|0)\Lambda'(1)[\Lambda(1)\mathbf{V}(1|0)\Lambda'(1) + \Theta(1)]^{-1} \tag{A.3}$$

For our purpose it is convenient to rewrite these expressions using the matrix equalities 7.B.1-6 presented in Appendix 7B of [1]. We assume the matrices $\mathbf{V}(1|0)$ and $\Theta(1)$ to be positive definite. Then, we may write

$$\mathbf{K}(1) = [\mathbf{V}^{-1}(1|0) + \Lambda'(1)\Theta^{-1}(1)\Lambda(1)]^{-1}\Lambda'(1)\Theta^{-1}(1) \tag{A.4}$$

$$\begin{aligned} \mathbf{I} - \mathbf{K}(1)\Lambda(1) &= [\mathbf{I} + \mathbf{V}(1|0)\Lambda'(1)\Theta^{-1}(1)\Lambda(1)]^{-1} \\ &= [\mathbf{V}^{-1}(1|0) + \Lambda'(1)\Theta^{-1}(1)\Lambda(1)]^{-1}\mathbf{V}^{-1}(1|0) \end{aligned} \tag{A.5}$$

Substitution of these relations into (A.1) and (A.2) yields:

$$\begin{aligned} \underline{\eta}(1|1) = {} & [\mathbf{I} + \mathbf{V}(1|0)\Lambda'(1)\Theta^{-1}(1)\Lambda(1)]^{-1}\eta(1|0) + \\ & [\mathbf{V}^{-1}(1|0) + \Lambda'(1)\Theta^{-1}(1)\Lambda(1)]^{-1}\Lambda'(1)\Theta^{-1}(1)\mathbf{y}(1) \end{aligned} \tag{A.6}$$

$$\mathbf{V}(1) = [\mathbf{V}^{-1}(1|0) + \Lambda'(1)\Theta^{-1}(1)\Lambda(1)]^{-1} \tag{A.7}$$

Let us now consider two special cases:

(i) We take $\eta(1|0)$ arbitrarily and $\mathbf{V}(1|0)$ of the form $\mathbf{V}(1|0) = \lambda\mathbf{I}$ with λ a constant. Taking λ bigger and bigger implies that the predictor $\underline{\eta}(1|0)$ is considered to be more and more unreliable. In the limit $\lambda \rightarrow \infty$ we then have:

(A.8) $\underline{\eta}(1|1) = [\Lambda'(1)\Theta^{-1}(1)\Lambda(1)]^{-1}\Lambda'(1)\Theta^{-1}(1)\mathbf{y}(1)$

(A.9) $\mathbf{V}(1) = [\Lambda'(1)\Theta^{-1}(1)\Lambda(1)]^{-1}$

If we compare these expressions with §8.4 in [3] we conclude that in this limit the Kalman filter estimator at $t = 1$ reduces to the cross-sectional Bartlett estimator at $t = 1$.

(ii) We take $\underline{\eta}(1|0) = 0$ and $\mathbf{V}(1|0) = \Phi(1) \equiv E(\eta(1)\eta'(1))$, the covariance matrix of the factors at $t = 1$. This matrix is positive definite and assumed to be known. In practice, this latter point usually leads to extra assumptions, as discussed in §3. In this case we arrive at

(A.10) $\underline{\eta}(1|1) = [\Phi^{-1}(1) + \Lambda'(1)\Theta^{-1}(1)\Lambda(1)]^{-1}\Lambda'(1)\Theta^{-1}(1)\mathbf{y}(1)$

(A.11) $\mathbf{V}(1) = [\Phi^{-1}(1) + \Lambda'(1)\Theta^{-1}(1)\Lambda(1)]^{-1}$

We find in this case the well-known regression estimator as presented in eq.(3)-(5) in DM and §8.3 of [3].