MANAGEMENT SCIENCE
Vol. 23, No. 7, March, 1977
Printed in U.S.A.

KALMAN FILTERING APPLIED TO STATISTICAL FORECASTING*

G. W. MORRISON† AND D. H. PIKE‡

This paper describes the use of the Kalman Filter in a certain class of forecasting problems. The time series is assumed to be modeled as a time varying mean with additive noise. The mean of the time series is assumed to be a linear combination of known functions. The coefficients appearing in the linear combination are unknown. Under such assumptions, the time series can be described as a linear system with the state vector of the system being the unknown parameters and present value of the mean of the process. The Kalman Filter can be used under these circumstances to obtain an "optimal" estimate of the state vector. One of the distinct advantages of the Kalman Filter is that time varying coefficients can be permitted in the model. Examples using the Kalman Filter in forecasting are presented.

1. Introduction

The Kalman Filter has been used as a forecasting tool in several special cases (see [1], [2], and [8]). There has not, however, been any attempt to develop a general methodology for using the Kalman Filter in statistical forecasting. This paper presents a general class of forecasting models to which Kalman Filtering can be applied. It is shown that the Kalman Filter model can be regarded as a generalization of the least squares model. The similarities and differences between the two models are discussed. Examples of the use of the Kalman Filter in statistical forecasting are also presented.

2. Model Assumptions

In the discussion which follows, it will be assumed that a time series can be represented by:

$$Y(t) = \mu(t) + \varepsilon(t), \qquad t \in \{1, 2, \ldots\}, \tag{1}$$

where $\mu(t)$ is the mean value function of the time series at time t. The variable $\varepsilon(t)$ appearing in (1) is assumed to be a zero mean random variable with known variance $R(t)$. It is also assumed that the random variables $\varepsilon(t)$ and $\varepsilon(u)$ are uncorrelated for $t \neq u$. That is,

$$E(\varepsilon(t)) = 0, \tag{2}$$

$$E(\varepsilon(t)\varepsilon(u)) = R(t)\delta_{tu}, \tag{3}$$

where δ_{tu} is the Kronecker delta.

The mean value function, $\mu(t)$, is assumed to be defined by the recurrence relation:

$$\mu(t+1) = \mu(t) + (\mathbf{g}(t))'\mathbf{a}(t), \qquad t \in \{0, 1, \ldots\}, \tag{4}$$

where $\mathbf{g}(t)$ and $\mathbf{a}(t)$ are k-dimensional column vectors and the superscript on $\mathbf{g}(t)$ denotes transposition. The th components of $\mathbf{g}(t)$ and $\mathbf{a}(t)$ are respectively $g_i(t)$ and $a_i(t)$. It is assumed that $\mathbf{g}(t)$ is known for all $t \in \{0, 1, \ldots\}$ and that $\mu(0)$ is known.

If the "state," $\mathbf{X}(t)$, of the time series described by (1) and (4) is defined by:

$$\mathbf{X}'(t) = (\mu(t), a_1(t), \ldots, a_k(t)), \tag{5}$$

(4) may be augmented by a series of linear equations which describe the dynamic

* Accepted by Ambar G. Rao; received March 3, 1975. This paper has been with the authors 8 months, for 2 revisions.

† Computer Sciences Division, Oak Ridge National Laboratory.
‡ University of Tennessee.

behavior of $\mathbf{a}(t)$. The most general linear model of the state of the system is:

$$\mathbf{X}(t+1) = \mathbf{A}(t)\mathbf{X}(t) + \mathbf{U}(t) + \boldsymbol{\eta}(t), \tag{6}$$

where $\mathbf{A}(t)$ is a known transition matrix, $\mathbf{U}(t)$ is a known vector function and $\boldsymbol{\eta}(t)$ is a zero mean random vector with known covariance matrix $\mathbf{Q}(t)$. As with $\varepsilon(t)$, $\boldsymbol{\eta}(t)$ is assumed to have no serial correlation.

To illustrate the use of (6) in model building, several examples are given:

(a) CONSTANT COEFFICIENT CASE. The case where $\mathbf{a}(t) = \mathbf{a}$ for all t is modeled by choosing $\mathbf{U}(t)$ and $\mathbf{Q}(t)$ to be null and

$$\mathbf{A}(t) = \begin{bmatrix} 1 & g_1(t) & g_2(t) & \cdots & g_{k-1}(t) & g_k(t) \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}. \tag{7}$$

In this case, (6) reduces to:

$$\mathbf{X}(t+1) = \mathbf{A}(t)\mathbf{X}(t). \tag{6a}$$

The linear model used as the ordinary least squares model falls into this class. The ordinary least squares model is (see [5]):

$$y(t) = (\mathbf{f}(t))'\mathbf{a} + \varepsilon(t), \tag{8}$$

where $\mathbf{f}(t)$ is a known vector function and $\mathbf{a}$ is a vector of unknown coefficients. If we define

$$\mu(0) = 0, \quad \text{and} \tag{9}$$

$$\mathbf{g}(t) = \mathbf{f}(t+1) - \mathbf{f}(t), \tag{10}$$

then (6a) and (7) define the equivalent linear system model. Table 1 shows the transition matrix $\mathbf{A}(t)$, the system state vector $\mathbf{X}(t)$, and the time series model for the special case where $f_{i+1}(t) = t^i/i!$ for $i = 0, 1, 2$.

TABLE 1

Models when $f_{i+1}(t) = t^i/i!$

Model	Time Series	State: $\mathbf{X}(t)$	Transition Matrix: $\mathbf{A}(t)$
Constant	$\mu(t) = a_0$ $y(t) = \mu(t) + \varepsilon(t)$	$(\mu(t))$	(1)
Linear	$\mu(t) = \mu(t-1) + a_1$ $y(t) = \mu(t) + \varepsilon(t)$	$\begin{bmatrix} \mu(t) \\ a_1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$
Quadratic	$\mu(t) = \mu(t-1) + a_1 + \frac{1}{2}a_2(t-\frac{1}{2})$ $y(t) = \mu(t) + \varepsilon(t)$	$\begin{bmatrix} \mu(t) \\ a_1 \\ a_2 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & t+\frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(b) RANDOMLY VARYING PARAMETERS. (6), in general, represents the case where the parameters of the model given in (4) vary randomly in time. Cooley and Prescott [3] as well as Ozkan [7] discuss the use of models of this form in constructing macroeconomic models. These models are of the general form:

$$\mathbf{X}(t+1) = \mathbf{A}(t)\mathbf{X}(t) + \mathbf{U}(t). \tag{6b}$$

For example, Cooley and Prescott formulate a model of this form with $\mathbf{A}(t)$ reflecting constant coefficients (similar to (7)).

(c) MODELING HYPOTHESIZED VARIATIONS IN PARAMETERS. Two frequently encountered situations will serve to illustrate the modeling capability of (6). First, consider the model where

$$\mu(t+1) = \mu(t) + a_1 + V(t), \tag{11}$$

where $V(t)$ is a serially correlated random variable. If $a_2(t)$ is defined to be $V(t)$ and a first order autoregressive model is used to describe $a_2(t)$:

$$a_2(t+1) = \rho a_2(t) + \eta_3(t), \tag{12}$$

then

$$\mathbf{X}(t+1) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & \rho \end{bmatrix} \mathbf{X}(t) + \begin{bmatrix} 0 \\ 0 \\ \eta_3(t) \end{bmatrix}. \tag{13}$$

Another model of interest is

$$\mu(t+1) = a_1(t) + \mu(t), \tag{14}$$

$$a_1(t+1) = \alpha\mu(t), \qquad \alpha \text{ a known parameter;} \tag{15}$$

then

$$\mathbf{X}(t+1) = \begin{bmatrix} 1 & 1 \\ \alpha & 0 \end{bmatrix} \mathbf{X}(t). \tag{16}$$

This model can be thought of as an "exponential growth rate" model. The rate of increase, $a_1(t+1)$ being dependent upon μ. This model is used in a subsequent example.

In summary, the problem is to estimate the state of the system at time t, given a model of a linear system described by (6) and a sequence of noisy observations of one element of the state vector (1). The linear model given by (6) includes the class of problems amendable to solution by ordinary least squares. The model in (6) admits a larger class of problems and may thus be viewed as a generalization of the constant coefficient models normally encountered with ordinary least squares.

3. Kalman Filter

The Kalman Filter is a method for estimating the state vector of a linear dynamic system from noisy observations. The Kalman Filter assumes a model of the form:

$$\mathbf{X}(t+1) = \mathbf{A}(t)\mathbf{X}(t) + \mathbf{U}(t) + \boldsymbol{\eta}(t), \tag{17}$$

$$y(t) = \mathbf{H}(t)\mathbf{X}(t) + \boldsymbol{\varepsilon}(t), \tag{18}$$

for $t \in \{0, 1, \ldots\}$, where $y(t)$ is a noisy observation of $\mathbf{X}(t)$. In the system described by (17) and (18) the matrices $\mathbf{U}(t)$, $\mathbf{A}(t)$, and $\mathbf{H}(t)$ must be known for each instant of time. One must also have knowledge of the noise covariance matrices of $\boldsymbol{\eta}(t)$ and $\boldsymbol{\varepsilon}(t)$ which are denoted respectively as $\mathbf{Q}(t)$ and $\mathbf{R}(t)$. The Kalman Filter produces an estimate of $\mathbf{X}(t)$, $\hat{\mathbf{X}}(t)$, from the observations:

$$\mathbf{Y}(t) = (y(1), y(2), \ldots, y(t)). \tag{19}$$

The estimate is chosen as that $\hat{\mathbf{X}}(t)$ which minimizes:

$$Z(\mathbf{X}(t)) = \big(\mathbf{X}(t) - \hat{\mathbf{X}}(t)\big)'\big(\mathbf{X}(t) - \hat{\mathbf{X}}(t)\big), \tag{20}$$

where $\mathbf{X}(t)$ denotes the true value of $\mathbf{X}(t)$. That is, $\hat{\mathbf{X}}(t)$ minimizes the length of the error vector $\mathbf{X}(t) - \hat{\mathbf{X}}(t)$.

The Kalman Filter is a sequential estimation procedure for producing optimal estimates of $\mathbf{X}(t)$, $\hat{\mathbf{X}}(t)$ as each new observation $y(t)$ becomes available. Figure 1

shows the detailed calculations. Each period the observation $y(t)$ is used to update the estimate of $\hat{\mathbf{X}}(t-1)$, producing a new estimate, $\hat{\mathbf{X}}(t)$. The gain vector, $\mathbf{K}(t)$, in step 4 is applied to the observation error, $\mathbf{Y}(t) - \mathbf{H}(t)[\mathbf{A}(t-1)\hat{\mathbf{X}}(t-1) + \mathbf{U}(t-1)]$, in step 6. The matrix $\mathbf{G}(t)$ is the posterior error covariance matrix (the error in $\hat{\mathbf{X}}(t)$ given observations $\mathbf{Y}(t)$), while $\mathbf{P}(t)$ is the prior error covariance matrix (the error in $\hat{\mathbf{X}}(t)$ given $\mathbf{Y}(t-1)$). The Kalman Filter equations are initialized with estimates of $\mathbf{G}(0)$ (a positive definite matrix) and $\hat{\mathbf{X}}(0)$ (an unbiased estimate of $\mathbf{X}(t)$).

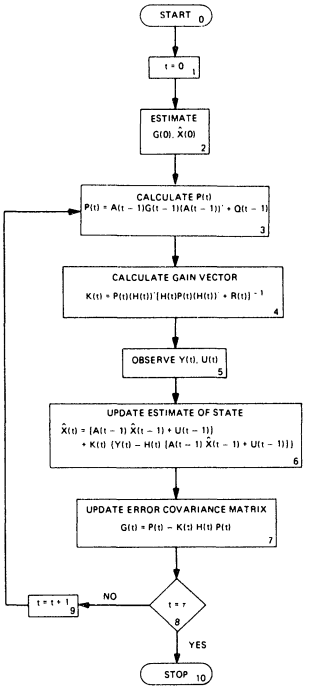

FIGURE 1. Flow Diagram of Kalman Filter Equations.

4. Discussion

The model in (1) and (6) is of the form of (17) and (18) if:

$$\mathbf{H}(t) = \mathbf{H} = (1, 0, \ldots 0), \qquad t \in \{1, 2, \ldots\}. \tag{21}$$

The Kalman Filter can therefore be used to produce optimal estimates of $\mathbf{X}(t)$. Forecasting is achieved by using (17) and (18), recursively. For any positive integer τ:

$$\hat{\mathbf{X}}(t+\tau) = \prod_{i=t}^{t+\tau-1} \mathbf{A}(i)\hat{\mathbf{X}}(t), \tag{22}$$

$$\hat{y}(t+\tau) = \mathbf{H}(t)\hat{\mathbf{X}}(t+\tau) = \mathbf{H}(t) \prod_{i=t}^{t+\tau-1} \mathbf{A}(i)\hat{\mathbf{X}}(t), \tag{23}$$

where $\hat{y}(t+\tau)$ denotes the forecasted value of $y(t+\tau)$ given $\hat{\mathbf{X}}(t)$.

There are several advantages to using the Kalman Filter model in place of ordinary least squares. First, the Kalman Filter has certain computational advantages. Note that the optimal estimates of the coefficients produced by least squares are given by (refer to (8) for the model definition)

$$\hat{\mathbf{a}}(t) = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{Y}(t), \tag{25}$$

where $\mathbf{F}' = [f(0), f(1), \ldots, f(t-1)]$. The matrix being inverted in (25) is a kth order matrix (where k = number of coefficients being estimated), while the matrix being inverted in (18) is a 1×1 matrix. Thus, use of the Kalman Filter to estimate $\mathbf{X}(t)$ avoids the matrix inversion step required in least squares. The Kalman Filter also has the advantage that the matrices appearing in Figure 1 are of fixed dimension. The matrices $\mathbf{F}$ and $\mathbf{Y}(t)$ in (25) increase with t. (25) also requires the entire history of observations, $\mathbf{Y}(t)$, while the Kalman Filter requires only the most recent observation, $y(t)$ and the previous state estimate. Second, referring to Figure 1, it should be noted that by removing step 6, the resulting flow chart gives an algorithm for precomputing the gains. These gains can be stored and used in the step 6 equation to update the state estimates as observations are received.

A third advantage of the Kalman Filter is the fact that the general linear model given by (6) allows one to easily formulate a variety of models with nonstationary coefficients. The fact that $\hat{X}(t)$ depends upon $R(t)$, the measurement noise covariance matrix, is a desirable property. This is clearly indicated where the process model is the constant model of Table 1 with $R(t) = R$. In this case, $\mathbf{X}(t)$, $\mathbf{G}(t)$, $\mathbf{K}(t)$ and $\mathbf{H}(t)$ reduce to scalars (see [3, pp. 113–114]):

$$\hat{X}(t+1) = \hat{X}(t) + \frac{G(0)}{R(t+1)G(0)}\left[y(t+1) - \hat{X}(t)\right]. \tag{26}$$

This is one of the few cases where an analytic relationship for $\mathbf{K}(t)$, the "gain matrix," can be obtained. The least squares estimate of $\mathrm{X}(t)$ is:

$$\hat{X}(t+1) = \hat{X}(t) + (t+1)^{-1}\left[y(t+1) - \hat{X}(t)\right]. \tag{27}$$

Note that the estimate of X in (26) is a function of R. The gain which is applied to the forecast error is inversely proportional to R. That is, errors are not weighted heavily if there is a large amount of noise present (i.e., large R). Note also that the gains obtained by (26) and (27) converge to a common gain as t approaches infinity. The rate of convergence depends upon the relative values of $G(0)$ and R. This same general behavior is exhibited for any general linear system. To see this, note that the equation for the gain $\mathbf{K}(t)$ may be rewritten as

$$\mathbf{K}(t) = \mathbf{G}(t)\mathbf{H}'(t)\mathbf{R}^{-1}(t). \tag{28}$$

Computationally, (28) is not desirable. However, (28) shows the basic behavior of the gain matrix. From (28) it is clear that, in general, the gain is proportional to the uncertainty in the state estimate ($\mathbf{G}(t)$) and inversely proportional to the uncertainty in measurement noise ($\mathbf{R}(t)$).

Using the criterion defined by (16), Kalman [6] has shown that if $\hat{\mathbf{X}}(0)$ is unbiased and $\varepsilon(t)$ is normally distributed, the Kalman Filter is the optimal filter. If either assumption is violated, a linear filter may not be an optimal filter. It should be noted, however, that the Kalman Filter is always the optimal filter if the search for an optimal filter is restricted to the class of linear filters.

5. Examples

The Kalman Filter model was compared to the least squares model in forecasting electrical power demand for the United States. Twenty-seven years of data were available (for the period 1945–1972). The first fourteen years of data were used to initialize the models. The linear and quadratic models in Table 1 were compared to the corresponding least squares linear and quadratic models. The Kalman Filter models were initialized using $\mathbf{G}(0) = I$, $\hat{\mathbf{X}}(0)$ = estimate of $\mathbf{X}(0)$ using the 14-period

least squares estimate of a, $\mathbf{R}(t) = \mathbf{R}$ = variance of residuals from the 14-period least squares model.[1]

Forecasts were made beginning with period 14 for the following period. Table 2 shows the power demand and forecasted demand for periods 15–27. No difference was observed between the Kalman Filter forecasts and the corresponding least squares forecasts. The "SSE" row in the table is the sum-of-squared-errors computed according to:

$$\text{SSE} = \sum_{i=15}^{27} (y(i) - \hat{y}(i))^2, \tag{29}$$

where $\hat{y}(i)$ is the forecasted value of $y(i)$.

TABLE 2
Comparison of Several Forecast Models

Period	Observation	Linear Model Stationary	Quadratic Model Stationary	Linear Model Nonstationary
1	1.91			
2	2.18			
3	2.41			
4	2.49			
5	2.81			
6	3.18		Initialization Period	
7	3.43			
8	3.84			
9	4.11			
10	4.81			
11	5.29			
12	5.58			
13	5.69			
14	6.27			
15	6.83	6.44	6.84	6.84
16	7.21	6.89	7.34	7.37
17	7.76	7.32	7.79	7.89
18	8.31	7.79	8.30	8.46
19	8.90	8.27	8.85	9.06
20	9.53	8.78	9.43	9.69
21	10.40	9.32	10.1	10.4
22	11.10	9.92	10.8	11.1
23	12.0	10.5	11.6	11.9
24	13.1	11.2	12.4	12.8
25	13.9	12.0	13.4	13.8
26	14.7	12.7	14.4	14.8
27	15.8	13.5	15.3	15.9
Sum of Squared Errors		22.225	1.286	0.335

The "exponential growth rate model" discussed previously (16) was constructed to illustrate the distinct advantage of the Kalman Filter over least squares models. In this case, the process model was assumed to be the exponential growth model as defined by (16): The coefficient α was chosen to minimize the sum-of-squared-forecasting errors for periods 5–14. Alpha was restricted to the grid $\{0, 0.001, 0.002, \ldots, 0.1\}$. A minimum was achieved for $\alpha = 0.005$. This value of α was used in forecasting for periods 15-27. $\mathbf{G}(0)$, $\hat{\mathbf{X}}(0)$, and $R(t) = R$ were identical to those values used in the linear Kalman Filter model described above.

[1] In general, $R(t)$ may depend on time. This would be the case if one were updating the estimate of $R(t)$ at each point in time. In this example, $R(t)$ was fixed at the initial estimate.

Table 2 illustrates that this model was superior to either of the models where $a(t)$ was restricted to be non-time varying.

Figure 2 illustrates the results of the various forecasting schemes.

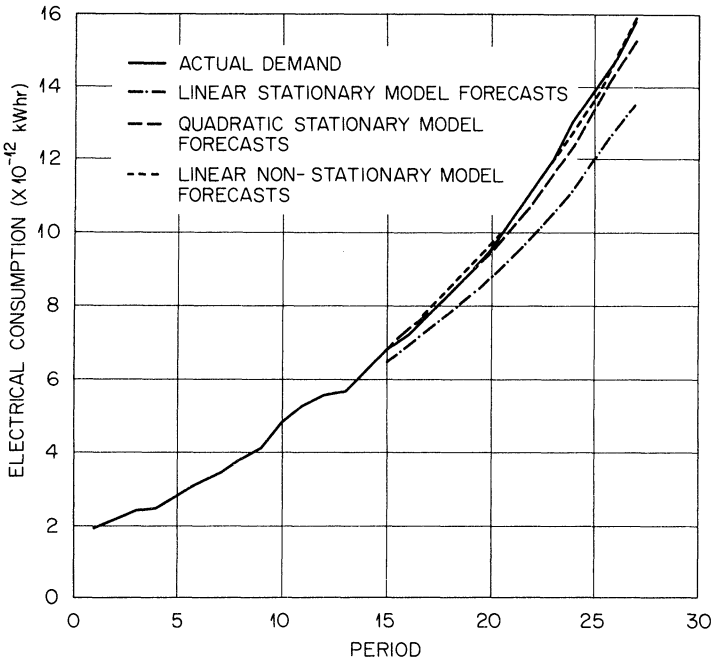

FIGURE 2. Comparison of Forecasting Techniques. All Forecasts are One Period Forecasts.

6. Summary

It has been shown that the Kalman Filter is a viable forecasting technique. It is similar in many respects to the least squares approach to forecasting. There are some distinct computational advantages to be realized using the Kalman Filter. If the coefficients in the Kalman Filter are restricted to be stationary, then the Kalman Filter and the least squares approach to forecasting can be expected to yield similar results. The distinct advantage of the Kalman Filter is that the assumption of stationarity of the model coefficients can be relaxed. If nonstationary linear models are used, results can be superior to the least squares model.

References

1. CAPEHART, B. L. AND BAUER, C. S., "Kalman Filtering Applications in Economic Inventory Systems," 39th Annual ORSA Meeting (May 6, 1971).
2. CHANG, S. H. AND FYFFE, D. E., "Estimation of Forecast Errors for Seasonal-Style Goods Sales," *Management Science*, Vol. 18, No. 2 (October 1971), pp. B89–B96.
3. COOLEY, T. F. AND PRESCOTT, E. C., "An Adaptive Regression Model," *International Economic Review*, Vol. 14, No. 2 (June 1974), pp. 364–371.
4. GELB, ARTHUR, *Applied Optimal Estimation*, MIT Press, Cambridge, Massachusetts, 1974.
5. JOHNSTON, J., *Econometric Methods*, 2nd ed., McGraw-Hill, New York, 1972, pp. 121–175.
6. KALMAN, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, (ASME), Vol. 82D (March 1960), pp. 35–45.
7. OZKAN, T. O., "Application of Kalman Filter to Parameter Estimation of Macroeconomic Models," M.S. Thesis, Massachusetts Institute of Technology (1974).
8. RYAN, H. W. AND KAMAT, S. J., "Application of Kalman Filter to Inventory Control," 41st Annual ORSA Meeting (April 26, 1972).
9. SORENSON, H. W., "Kalman Filtering Techniques," *Advances in Control Systems*, Vol. 3, Academic Press, 1966.