Understanding the Kalman Filter

RICHARD J. MEINHOLD and NOZER D. SINGPURWALLA*

This is an expository article. Here we show how the successfully used Kalman filter, popular with control engineers and other scientists, can be easily understood by statisticians if we use a Bayesian formulation and some well-known results in multivariate statistics. We also give a simple example illustrating the use of the Kalman filter for quality control work.

KEY WORDS: Bayesian inference; Box-Jenkins models; Forecasting; Exponential smoothing; Multivariate normal distribution; Time series.

1. INTRODUCTION

The Kalman filter (KF) commonly employed by control engineers and other physical scientists has been successfully used in such diverse areas as the processing of signals in aerospace tracking and underwater sonar, and the statistical control of quality. More recently, it has also been used in some nonengineering applications such as short-term forecasting and the analysis of life lengths from dose-response experiments. Unfortunately, much of the published literature on the KF is in the engineering journals (including the original development, in Kalman 1960 and Kalman and Bucy 1961), and uses a language, notation, and style that is alien to statisticians. Consequently, many practitioners of statistics are not aware of the simplicity of this useful methodology. However, the model, the notions, and the techniques of Kalman filtering are potentially of great interest to statisticians owing to their similarity to linear models of regression and time series analysis, and because of their great utility in applications.

In actuality, the KF may be easily understood by the statistician if it is cast as a problem in Bayesian inference and we employ some well-known elementary results in multivariate statistics. This feature was evidently first published by Harrison and Stevens (1971, 1976), who were primarily interested in Bayesian forecasting. However, the particular result presented by them is in a nontutorial manner, with emphasis placed on the implementation of the KF. Our aim, on the other hand, is to provide an exposition of the key notions of the approach in a single source, laying out its derivation in a few easy steps, filling in some clarifying technical details, giving an example, and giving an interpretation of results. A more mathematical discussion of the KF emphasizing the stochastic differential equation approach is given by Wegman (1982). We feel that once it is demystified, the KF will be used more often by applied statisticians.

*Richard J. Meinhold is a graduate student, and Nozer D. Singpurwalla is Professor of Operations Research and Statistics at George Washington University, Washington, D.C. 20052. The work of the second author was supported in part by the Office of Naval Research Contract N00014-77-C-0263 and by the U.S. Army Research Office under Grant DAAG-29-80-C-0067 with George Washington University.

2. THE KALMAN FILTER MODEL: MOTIVATION AND APPLICATIONS

Let $Y_t, Y_{t-1}, \ldots, Y_1$, the data (which may be either scalars or vectors), denote the observed values of a variable of interest at times $t, t-1, \ldots, 1$. We assume that Y_t depends on an unobservable quantity θ_t, known as the *state of nature*. Our aim is to make inferences about θ_t, which may be either a scalar or a vector and whose dimension is independent of the dimension of Y_t. The relationship between Y_t and θ_t is linear and is specified by the *observation equation*

$$Y_t = F_t\theta_t + v_t, \tag{2.1}$$

where F_t is a known quantity. The *observation error* v_t is assumed to be normally distributed with mean zero and a known variance V_t, denoted as $v_t \sim N(0, V_t)$.

The essential difference between the KF and the conventional linear model representation is that in the former, the state of nature—analogous to the regression coefficients of the latter—is not assumed to be a constant but may change with time. This dynamic feature is incorporated via the *system equation,* wherein

$$\theta_t = G_t\theta_{t-1} + w_t, \tag{2.2}$$

G_t being a known quantity, and the *system equation error* $w_t \sim N(0, W_t)$, with W_t known. Since there are many physical systems for which the state of nature θ_t changes over time according to a relationship prescribed by engineering or scientific principles, the ability to include a knowledge of the system behavior in the statistical model is an apparent source of attractiveness of the KF. Note that the relationships (2.1) and (2.2) specified through F_t and G_t may or may not change with time, as is also true of the variances V_t and W_t; we have subscripted these here for the sake of generality.

In addition to the usual linear model assumptions regarding the error terms, we also postulate that v_t is independent of w_t; while extension to the case of dependency is straightforward, there is no need in this article to do so.

2.1 Some Applications

To look at how the KF model might be employed in practice, we consider a simplified version of the frequently referenced example of tracking a satellite's orbit around the earth. The unknown state of nature θ_t

could be the position and speed of the satellite at time t, with respect to a spherical coordinate system with origin at the center of the earth. These quantities cannot be measured directly. Instead, from tracking stations around the earth, we may obtain measurements of distance to the satellite and the accompanying angles of measurement; these are the Y_t's. The principles of geometry, mapping Y_t into θ_t, would be incorporated in F_t, while v_t would reflect the measurement error; G_t would prescribe how the position and speed change in time according to the physical laws governing orbiting bodies, while w_t would allow for deviations from these laws owing to such factors as nonuniformity of the earth's gravitational field, and so on.

A less complicated situation is considered by Phadke (1981) in the context of statistical quality control. Here the observation Y_t is a simple (approximately normal) transform of the number of defectives observed in a sample obtained at time t, while $\theta_{1,t}$ and $\theta_{2,t}$ represent, respectively, the true defective index of the process and the drift of this index. We then have as the observation equation

$$Y_t = \theta_{1,t} + v_t,$$

and as the system equations

$$\theta_{1,t} = \theta_{2,t} + w_{1,t}$$

$$\theta_{2,t} = \theta_{2,t-1} + w_{2,t}.$$

In vector notation, this system of equations becomes

$$\boldsymbol{\theta}_t = G\boldsymbol{\theta}_{t-1} + \mathbf{U}_t,$$

where

$$\boldsymbol{\theta}_t = \begin{bmatrix}\theta_{1,t}\\ \theta_{2,t}\end{bmatrix} \quad \text{and} \quad \mathbf{U}_t = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}w_{1,t}\\ w_{2,t}\end{bmatrix};$$

$$G \equiv \begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix}$$

does not change with time.

If we examine $Y_t - Y_{t-1}$ for this model, we observe that, under the assumption of constant variances, namely, $V_t \equiv V$ and $W_t \equiv W$, the autocorrelation structure of this difference is identical to that of an *ARIMA* (0,1,1) process in the sense of Box and Jenkins (1970). Although such a correspondence is sometimes easily discernible, we should in general not, because of the discrepancies in the philosophies and methodologies involved, consider the two approaches to be equivalent.

3. THE RECURSIVE ESTIMATION PROCEDURE

The term "Kalman filter" or "Kalman filtering" refers to a recursive procedure for inference about the state of nature θ_t. The key notion here is that given the data $\mathbf{Y}_t = (Y_t, \ldots, Y_1)$, inference about θ_t can be carried out through a direct application of Bayes's theorem:

$$\text{Prob}\{\text{State of Nature} \mid \text{Data}\}$$

$$\propto \text{Prob}\{\text{Data} \mid \text{State of Nature}\}$$

$$\times \text{Prob}\{\text{State of Nature}\}, \qquad (3.1)$$

which can also be written as

$$P(\theta_t \mid \mathbf{Y}_t) \propto P(Y_t \mid \theta_t, \mathbf{Y}_{t-1}) \times P(\theta_t \mid \mathbf{Y}_{t-1}), \qquad (3.2)$$

where the notation $P(A \mid B)$ denotes the probability of occurrence of event A given that (or conditional on) event B has occurred. Note that the expression on the left side of (3.2) denotes the *posterior distribution* for θ at time t, whereas the first and second expressions on the right side denote the *likelihood* and the *prior distribution* for θ, respectively.

The recursive procedure can best be explained if we focus attention on time point $t-1$, $t = 1,2,\ldots,$ and the observed data until then, $\mathbf{Y}_{t-1} = (Y_{t-1}, Y_{t-2}, \ldots, Y_1)$. In what follows, we use matrix manipulations in allowing for Y and/or θ to be vectors, without explicitly noting them as such.

At $t-1$, our state of knowledge about θ_{t-1} is embodied in the following probability statement for θ_{t-1}:

$$(\theta_{t-1} \mid \mathbf{Y}_{t-1}) \sim N(\hat{\theta}_{t-1}, \Sigma_{t-1}), \qquad (3.3)$$

where $\hat{\theta}_{t-1}$ and Σ_{t-1} are the expectation and the variance of $(\theta_{t-1} \mid \mathbf{Y}_{t-1})$. In effect, (3.3) represents the posterior distribution of θ_{t-1}; its evolution will become clear in the subsequent text.

It is helpful to remark here that the recursive procedure is started off at time 0 by choosing $\hat{\theta}_0$ and Σ_0 to be our best guesses about the mean and the variance of θ_0, respectively.

We now look forward to time t, but in two stages:

1. prior to observing Y_t, and
2. after observing Y_t.

Stage 1. Prior to observing Y_t, our best choice for θ_t is governed by the system equation (2.2) and is given as $G_t\theta_{t-1} + w_t$. Since θ_{t-1} is described by (3.3), our state of knowledge about θ_t is embodied in the probability statement

$$(\theta_t \mid \mathbf{Y}_{t-1}) \sim N(G_t\hat{\theta}_{t-1}, R_t = G_t\Sigma_{t-1}G_t' + W_t); \qquad (3.4)$$

this is our prior distribution.

In obtaining (3.4), which represents our prior for θ_t in the next cycle of (3.2), we used the well-known result that for any constant C

$$X \sim N(\mu, \Sigma) \Rightarrow CX \sim N(C\mu, C\Sigma C'),$$

where C' denotes the transpose of C.

Stage 2. On observing Y_t, our goal is to compute the posterior of θ_t using (3.2). However, to do this, we need to know the likelihood $\mathcal{L}(\theta_t \mid \mathbf{Y}_t)$, or equivalently $P(Y_t \mid \theta_t, \mathbf{Y}_{t-1})$, the determination of which is undertaken via the following arguments.

Let e_t denote the *error* in predicting Y_t from the point $t-1$; thus

$$e_t = Y_t - \hat{Y}_t = Y_t - F_tG_t\hat{\theta}_{t-1}. \qquad (3.5)$$

Since F_t, G_t, and $\hat{\theta}_{t-1}$ are all known, observing Y_t is equivalent to observing e_t. Thus (3.2) can be rewritten as

$$P(\theta_t \mid Y_t, \mathbf{Y}_{t-1}) = P(\theta_t \mid e_t, \mathbf{Y}_{t-1}) \propto P(e_t \mid \theta_t, \mathbf{Y}_{t-1}) \times P(\theta_t \mid \mathbf{Y}_{t-1}), \quad (3.6)$$

with $P(e_t \mid \theta_t, \mathbf{Y}_{t-1})$ being the likelihood.

Using the fact that $Y_t = F_t\theta_t + v_t$, (3.5) can be written as $e_t = F_t(\theta_t - G_t\hat{\theta}_{t-1}) + v_t$, so that $E(e_t \mid \theta_t, \mathbf{Y}_{t-1}) = F_t(\theta_t - G_t\hat{\theta}_{t-1})$.

Since $v_t \sim N(0, V_t)$, it follows that the likelihood is described by

$$(e_t \mid \theta_t, \mathbf{Y}_{t-1}) \sim N(F_t(\theta_t - G_t\hat{\theta}_{t-1}), V_t). \quad (3.7)$$

We can now use Bayes's theorem (Eq. (3.6)) to obtain

$$P(\theta_t \mid Y_t, \mathbf{Y}_{t-1}) = \frac{P(e_t \mid \theta_t, \mathbf{Y}_{t-1}) \times P(\theta_t \mid \mathbf{Y}_{t-1})}{\int_{\text{all } \theta_t} P(e_t, \theta_t \mid \mathbf{Y}_{t-1}) d\theta_t}, \quad (3.8)$$

and this best describes our state of knowledge about θ_t at time t. Once $P(\theta_t \mid Y_t, \mathbf{Y}_{t-1})$ is computed, we can go back to (3.3) for the next cycle of the recursive procedure. In the next section, we show that the posterior distribution of (3.8) is of the form presented in (3.3).

4. DETERMINATION OF THE POSTERIOR DISTRIBUTION

The tedious effort required to obtain $P(\theta_t \mid \mathbf{Y}_t)$ using (3.8) can be avoided if we make use of the following well-known result in multivariate statistics (Anderson 1958, pp. 28–29), and some standard properties of the normal distribution.

Let X_1 and X_2 have a bivariate normal distribution with means μ_1 and μ_2, respectively, and a covariance matrix

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix};$$

we denote this by

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right]. \quad (4.1)$$

When (4.1) holds, the conditional distribution of X_1 given X_2 is described by

$$(X_1 \mid X_2 = x_2) \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}). \quad (4.2)$$

The quantity $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ is called the *regression function,* and $\Sigma_{12}\Sigma_{22}^{-1}$ is referred to as the *coefficient of the least squares regression of X_1 on x_2.*

As a converse to the relationship (4.1) implies (4.2), we have the result that whenever (4.2) holds, and when $X_2 \sim N(\mu_2, \Sigma_{22})$, then (4.1) will hold; we will use this converse relationship.

For our situation, we suppress the conditioning variables $\mathbf{Y}_{t-1}$ and let X_1 correspond to e_t, and X_2 correspond to θ_t; we denote this correspondence by $X_1 \Leftrightarrow e_t$ and $X_2 \Leftrightarrow \theta_t$. Since $(\theta_t \mid \mathbf{Y}_{t-1}) \sim N(G_t\hat{\theta}_{t-1}, R_t)$ (see (3.4)), we note that

$$\mu_2 \Leftrightarrow G_t\hat{\theta}_{t-1}$$

and

$$\Sigma_{22} \Leftrightarrow R_t.$$

If in (4.2) we replace X_1, X_2, μ_2, and Σ_{22} by e_t, θ_t, $G_t\hat{\theta}_{t-1}$, and R_t, respectively, and recall the result that $(e_t \mid \theta_t, \mathbf{Y}_{t-1}) \sim N(F_t(\theta_t - G_t\hat{\theta}_{t-1}), V_t)$ (Eq. (3.7)), then

$$\mu_1 + \Sigma_{12}R_t^{-1}(\theta_t - G_t\hat{\theta}_{t-1}) \Leftrightarrow F_t(\theta_t - G_t\hat{\theta}_{t-1}),$$

so that $\mu_1 \Leftrightarrow 0$ and $\Sigma_{12} \Leftrightarrow F_tR_t$; similarly,

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11} - F_tR_tF_t' \Leftrightarrow V_t,$$

so that $\Sigma_{11} \Leftrightarrow V_t + F_tR_tF_t'$.

We now invoke the converse relation mentioned previously to conclude that the joint distribution of θ_t and e_t, given $\mathbf{Y}_{t-1}$, can be described as

$$\left[\begin{pmatrix} \theta_t \\ e_t \end{pmatrix} \middle| \, Y_{t-1} \right] \sim N\left[\begin{pmatrix} G_t\hat{\theta}_{t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} R_t & R_tF_t' \\ F_tR_t & V_t + F_tR_tF_t' \end{pmatrix} \right]. \quad (4.3)$$

Making e_t the conditioning variable and identifying (4.3) with (4.1), we obtain via (4.2) the result that

$$(\theta_t \mid e_t, \mathbf{Y}_{t-1}) \sim N[G_t\hat{\theta}_{t-1} + R_tF_t'(V_t + F_tR_tF_t')^{-1}e_t, \; R_t - R_tF_t'(V_t + F_tR_tF_t')^{-1}F_tR_t]. \quad (4.4)$$

This is the desired posterior distribution. We now summarize to highlight the elements of the recursive procedure.

After time $t-1$, we had a posterior distribution for θ_{t-1} with mean $\hat{\theta}_{t-1}$ and variance Σ_{t-1} (Eq. (3.3)). Forming a prior for θ_t with mean $G_t\hat{\theta}_{t-1}$ and variance $R_t = G_t\Sigma_{t-1}G_t' + W_t$ (Eq. (3.4)) and evaluating a likelihood given $e_t = Y_t - F_tG_t\hat{\theta}_{t-1}$ (Eq. (3.5)), we arrive at the posterior density for θ_t; this has mean

$$\hat{\theta}_t = G_t\hat{\theta}_{t-1} + R_tF_t'(V_t + F_tR_tF_t')^{-1}e_t \quad (4.5)$$

and variance

$$\Sigma_t = R_t - R_tF_t'(V_t + F_tR_tF_t')^{-1}F_tR_t. \quad (4.6)$$

We now continue through the next cycle of the process.

5. INTERPRETATION OF RESULTS AND CONCLUDING REMARKS

If we look at (4.4) for obtaining some additional insight into the workings of the Kalman filter, we note that the mean of the posterior distribution of $(\theta_t \mid e_t, \mathbf{Y}_{t-1})$ is indeed the regression function of θ_t on e_t. The mean (regression function) is the sum of two quantities $G_t\hat{\theta}_{t-1}$, and a multiple of the *one step ahead forecast error* e_t.

We first remark that $G_t\hat{\theta}_{t-1}$ is the mean of the prior distribution of θ_t (see (3.4)), and by comparing (4.3) and (4.4) to (4.1) and (4.2) we verify that the multiplier of e_t, $R_tF_t'(V_t + F_tR_tF_t')^{-1}$, is the coefficient of the least squares regression of θ_t on e_t (conditional on $\mathbf{Y}_{t-1}$). Thus one way to view Kalman filtering is to think of it as an updating procedure that consists of forming a preliminary (prior) guess about the state of nature and then adding a correction to this guess, the correction being

determined by how well the guess has performed in predicting the next observation.

Second, we should clarify the meaning of regressing θ_t on e_t since this pair constitutes but a single observation and the regression relationship is not estimated in the familiar way. Rather, we recall the usual framework of sequential Bayesian estimation, wherein a new posterior distribution arises with each successive piece of data. At time zero, the regression of θ_1 on e_1 is determined entirely by our prior specifications. On receiving the first observation, the value of e_1 is mapped into $\hat{\theta}_1$ through this function, which is then replaced by a new regression relation based on $\hat{\theta}_1$, F_1, G_1, V_1, and W_1. This in turn is used to map e_2 into $\hat{\theta}_2$, and so on as the process continues in the usual Bayesian prior/posterior iterative manner; see Figure 1. Thus Kalman filtering can also be viewed as the evolution of a series of regression functions of θ_t on e_t, at times $0,1,\ldots,t-1,t$, each having a potentially different intercept and regression coefficient; the evolution stems from a learning process involving all the data.

The original development of the Kalman filter approach was motivated by the updating feature just described, and its derivation followed via the least squares estimation theory. The Bayesian formulation described here yields the same result in an elegant manner and additionally provides the attractive feature of enabling inference about θ_t through a probability distribution rather than just a point estimate.

6. ILLUSTRATIVE EXAMPLES

6.1 The Steady Model

We consider two examples to illustrate the preceding mechanism and its performance.

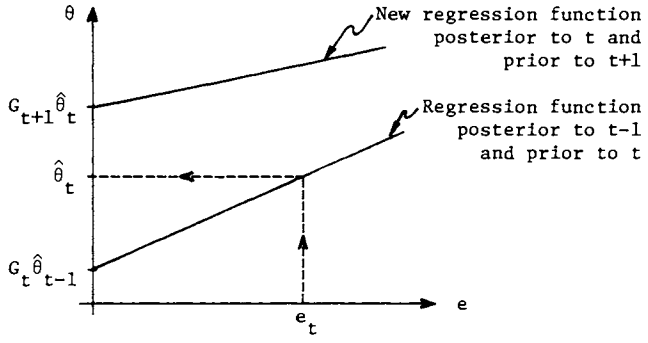

Figure 1. Regression of θ_t on e_t

We first return to the quality control model of Section 2.1, simplified by the removal of the drift parameter. This yields

$$Y_t = \theta_t + v_t \text{ (Obs. Eqn.)}$$

and (6.1)

$$\theta_t = \theta_{t-1} + w_t \text{ (Sys. Eqn.)}.$$

This is a simplest possible nontrivial KF model (sometimes referred to in the forecasting literature as the steady model); it also corresponds, in the sense of possessing the same autocorrelation structure (assuming constant variances), to a class of *ARIMA* (0, 1, 1) models of Box and Jenkins (1970). In this situation, $F_t \equiv G_t \equiv 1$; if we further specified that $\Sigma_0 = 1$, $V_t \equiv 2$, $W_t \equiv 1$, we can easily demonstrate inductively that $R_t = G_t\Sigma_{t-1}G_t' + W_t \equiv 2$, and from (4.6), $\Sigma_t \equiv 1$. In (4.5), then, our recursive relationship becomes

$$\begin{aligned}\hat{\theta}_t &= \hat{\theta}_{t-1} + \tfrac{1}{2}(Y_t - \hat{\theta}_{t-1}) \\ &= \tfrac{1}{2}(Y_t + \hat{\theta}_{t-1}) \\ &= \sum_{j=0}^{t-1} (\tfrac{1}{2})^{j+1} Y_{t-j} + (\tfrac{1}{2})^t \hat{\theta}_0. \end{aligned} \quad (6.2)$$

Table 1. A Simulation of the Process Described in Section (6.2)

t	$v_t \simeq N(0, 2)$	$w_t \simeq N(0, 1)$	F_t	$\theta_t = G_t\theta_{t-1} + w_t$	$Y_t = F_t\theta_t + v_t$	$\hat{\theta}_t$	Σ_t
0				− .353		4.183	1
1	− .376	.887	1.3	1.063	1.007	− .619	.608
2	.023	−1.021	.8	− .489	− .368	− .350	.842
3	− .898	−1.207	.9	− .962	−1.764	− .527	.812
4	1.645	.150	1.1	− .331	1.281	.338	.696
5	− .701	− .329	1.2	− .163	− .897	− .434	.636
6	− .257	.448	1.0	.366	.109	− .097	.734
7	−1.766	.403	1.1	.220	−1.524	− .550	.690
8	−1.551	−1.069	.9	− .959	−2.414	−1.050	.795
9	.443	.186	.9	.666	1.042	.732	.807
10	.300	− .267	1.0	.066	.366	.366	.751
11	− .745	.406	1.2	.373	− .297	− .213	.640
12	−1.175	− .789	.8	− .602	−1.657	− .638	.846
13	1.281	.386	1.1	.687	2.037	.967	.699
14	−1.017	− .753	.7	− .409	−1.304	− .041	.912
15	− .991	− .120	.9	.085	− .915	− .324	.820
16	.861	.524	1.0	.566	1.427	.436	.752
17	−1.047	.224	1.3	− .059	−1.124	− .542	.593
18	.076	− .356	1.1	− .386	− .348	− .290	.678
19	2.498	− .907	1.2	− .714	1.641	.704	.635
20	1.482	− .881	.9	−1.238	.368	.370	.789
21	−1.683	.023	.7	.642	−1.234	− .543	.926
22	1.072	.632	.6	.953	1.644	.275	1.008
23	− .684	− .314	1.1	− .791	−1.554	− .687	.712
24	− .580	− .217	1.0	− .612	−1.192	− .658	.741
25	−1.894	1.927	.9	2.233	.116	.264	.801

We see then that in this simple situation the KF estimator of θ_t, and thus Y_{t+1}, is actually equivalent to that derived from a form of exponential smoothing.

6.2 A Numerical Example

We present in Table 1 a numerical example involving a simulation of the (scalar-dimensional) general model of (2.1) and (2.2). We continue to specify $\Sigma_0 = 1$, $V_t \equiv 2$, $W_t \equiv 1$, but incorporate cyclical behavior in θ_t by setting

$$G_t = \frac{1}{2}\sin\left(\frac{\pi}{2}[2t+1]\right) = (-1)^t/2,$$

while F_t is in the nature of the familiar independent variable of ordinary regression. This situation clearly cannot be contained in any class of the *ARIMA* family; instead it is analogous, if not equivalent, to the transfer function model approach of Box and Jenkins (1970).

Starting with a value for θ_0, the disturbances v_t and w_t were generated from a table of random normal variates and used in turn to produce, via the system and observation equations, the processes $\{\theta_t\}$ and $\{Y_t\}$, of which only the latter would ordinarily be visible. A "bad guess" value of $\hat{\theta}_0$ was chosen; as can be seen in Figure 2, where the actual values of θ_t and their estimates $\hat{\theta}_t$ are plotted, the effect of this error is short-lived. The reader may find it conducive to a better understanding of the model to work through several iterations of the recursive procedure.

[*Received October 1981. Revised July 1982.*]

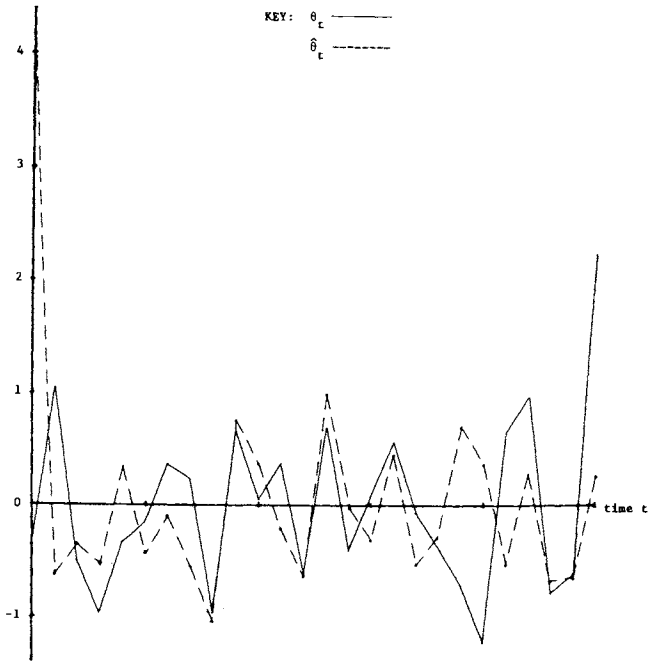

Figure 2. A Plot of the Simulated Values of θ_t, the State of Nature at Time t, and Their Estimated Values $\hat{\theta}_t$ Via the Kalman Filter

REFERENCES

ANDERSON, T.W. (1958), *An Introduction to Multivariate Statistical Analysis*, New York: John Wiley.

BOX, G.E.P., and JENKINS, G.M. (1970), *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day.

HARRISON, P.J., and STEVENS, C.F. (1971), "A Bayesian Approach to Short-Term Forecasting," *Operations Research Quarterly*, 22, 341–362.

——— (1976), "Bayesian Forecasting (with discussion)," *Journal of the Royal Statistical Society*, Ser. B, 38, 205–247.

KALMAN, R.E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, 82, 34–45.

KALMAN, R.E., and BUCY, R.S. (1961), "New Results in Linear Filtering and Prediction Theory," *Journal of Basic Engineering*, 83, 95–108.

PHADKE, M.S. (1981), "Quality Audit Using Adaptive Kalman Filtering," *ASQC Quality Congress Transactions—San Francisco*, 1045–1052.

WEGMAN, E.J. (1982), "Kalman Filtering," in *Encyclopedia of Statistics*, eds. Norman Johnson and Samuel Kotz, New York: John Wiley.