Dynamic Linear Models With Switching

R. H. SHUMWAY and D. S. STOFFER*

The problem of modeling change in a vector time series is studied using a dynamic linear model with measurement matrices that switch according to a time-varying independent random process. We derive filtered estimators for the usual state vectors and also for the state occupancy probabilities of the underlying nonstationary measurement process. A maximum likelihood estimation procedure is given that uses a pseudo-expectation-maximization algorithm in the initial stages and nonlinear optimization. We relate the models to those considered previously in the literature and give an application involving the tracking of multiple targets.

KEY WORDS: Change points; Expectation-maximization algorithm; Nonlinear models; State-space; Target tracking.

1. INTRODUCTION

One way of modeling change in an evolving time series is by assuming that the dynamics of some underlying model changes discontinuously at certain undetermined points in time. In this article we will be concerned primarily with modeling change in the dynamic linear model, a general form that includes autoregressive integrated moving average (ARIMA) and classical regression models as special cases.

In a stable dynamic linear model, one assumes that some underlying $q \times 1$ observation vectors $\mathbf{y}_t$ are connected to unobserved $p \times 1$ signal vectors of interest $\mathbf{x}_t$ through the observation equation

$$\mathbf{y}_t = A_t\mathbf{x}_t + \mathbf{v}_t \tag{1}$$

for the time points $t = 1, \ldots, n$, where A_t are $q \times p$ measurement matrices that convert the unobserved signal measurements into the data vectors $\mathbf{y}_t$. The vectors $\mathbf{v}_t$ are independent zero-mean white-noise vectors with common $q \times q$ covariance matrix R. The measurement or design matrices A_t are usually regarded as specified and may be used to model situations involving structured multiple signals or where there are missing observations (see, for example, Shumway 1988, sec. 3.3). The description of the model is completed by noting that the signal process $\mathbf{x}_t$ was generated from a starting point $\mathbf{x}_0$ with mean $\boldsymbol{\mu}$ and covariance matrix Σ using the state equations

$$\mathbf{x}_t = \Phi\mathbf{x}_{t-1} + \mathbf{w}_t, \tag{2}$$

where Φ is a $p \times p$ transition matrix that models the evolution of the signal vector $\mathbf{x}_t$ through time and $\mathbf{w}_t$ is a $p \times 1$ white-noise process, assumed to be independent of $\mathbf{v}_t$, with covariance matrix Q. The parameters of the model are the initial mean and covariance μ and Σ plus the state transition matrix Φ and covariance Q along with the measurement covariance matrix R. The problems of estimating the parameters by maximum likelihood and the signal vector $\mathbf{x}_t$ by Kalman filtering and smoothing techniques have been exhaustively treated in the literature [see Shumway (1988) for a review of some of them]. For purposes of maximum likelihood estimation it is convenient to assume that the initial vector $\mathbf{x}_0$ along with the errors $\mathbf{v}_t$ and $\mathbf{w}_t$ have multivariate normal distributions. It should be recognized that the model defined in (1) and (2) usually is applied with a considerably reduced parameter space. For example, structural models are often applied that represent the observed series as the sum of unobserved trend and seasonal components (see, for example, Kitagawa and Gersch 1984; Harvey and Todd 1983). The two examples considered here also involve specializing (1) and (2) by reducing the number of nonzero parameters considered.

Generalizations of the preceding model to include the possibility of changes occurring over time have been approached by allowing changes in the error covariances (see Harrison and Stevens 1976 or Gordon and Smith 1988, 1990) or by assigning mixture distributions to the observation errors $\mathbf{v}_t$ (see Peña and Guttman 1988). Approximations to filtering were derived in all of the articles just cited. An application to monitoring renal transplants was described in Smith and West (1983) and in Gordon and Smith (1990). Changes can also be modeled in the classical regression case by allowing switches in the design matrices, as in Quandt (1972). Applications of the switching approach to modeling changes in econometric time series have been reviewed by Tsurimi (1988). Switching via a stationary Markov chain with independent observations has been developed by Lindgren (1978) and Goldfeld and Quandt (1973). Markov switching for dependent data has been applied by Hamilton (1989) to detect changes between positive and negative growth periods in the economy. Applications to speech recognition have been considered by Juang and Rabiner (1985); hidden Markov models are summarized by Rabiner and Juang (1986). An application of the idea of switching to the tracking of multiple targets has been considered in Bar-Shalom and Tse (1975) and in Bar-Shalom (1978) who obtained approximations to Kalman filtering in terms of weighted averages of the innovations. They

* R. H. Shumway is Professor, Division of Statistics, University of California, Davis, CA 95616. D. S. Stoffer is Associate Professor, Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260. R. H. Shumway's time was supported in part by the Science and Technology Division, Institute for Defense Analysis. D. S. Stoffer's time was supported in part by National Science Foundation Grant #DMS-9000522 and by the Centers for Disease Control through a cooperative agreement with the Association for Schools of Public Health (ASPH). We are indebted to an anonymous referee who pointed out a flaw in our original attempt at treating the case where the A_t's are Markov dependent and to Gabriel Frankel of the Institute for Defense Analysis and Mohsen Pouramadhi of Northern Illinois University for a number of helpful suggestions.

© 1991 American Statistical Association
Journal of the American Statistical Association
September 1991, Vol. 86, No. 415, Theory and Methods

call their approximations "probabilistic data association filters."

2. DYNAMIC LINEAR MODELS WITH SWITCHING

Our approach is motivated primarily by the problem of tracking a large number of moving targets using a vector $\mathbf{y}_t$ of sensors. In this problem one does not know at any given time point which target any given sensor has detected. Hence it is the structure of the measurement matrix A_t in (1) that is changing and not the dynamics of the underlying signal $\mathbf{x}_t$ or the noises $\mathbf{v}_t$ or $\mathbf{w}_t$. To illustrate, assume a 3×1 vector of satellite measurements $\mathbf{y}_t = (y_{t_1}, y_{t_2}, y_{t_3})'$ is observing some combination of tracks or signals $\mathbf{x}_t' = (x_{t_1}, x_{t_2}, x_{t_3})$. For the measurement matrix

$$A_t = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

in the defining model (1), it is clear that all sensors are observing target x_{t_1}, whereas for the measurement matrix

$$A_t = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

the first sensor y_{t_1} observes the second target x_{t_2}, the second sensor y_{t_2} observes target x_{t_1}, and the third sensor y_{t_3} observes the third target x_{t_3}. All possible detection configurations will define a set of possible values for A_t, say $M_1, M_2, \ldots, M_m$, as a collection of plausible measurement matrices for $p = 3$ sensors tracking one, two, or three targets.

As a second example of the switching model to be considered here, consider the case where the dynamics of the linear model changes suddenly over the history of a given single realization. For example, Lam (1990) has given the following generalization of Hamilton's (1989) model for detecting positive and negative growth periods in the economy. Assume the representation

$$y_t = z_t + n_t, \qquad (3)$$

where z_t is an autoregressive series and n_t is a random walk with a drift that switches between two values α_0 and $\alpha_0 + \alpha_1$. That is,

$$n_t = n_{t-1} + \alpha_0 + \alpha_1 S_t$$

with $S_t = 0$ or 1 depending on whether we are in state 1 or state 2. Suppose, for purposes of illustration, that

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + w_t$$

is a second-order autoregressive series with $\operatorname{var}(w_t) = \sigma_w^2$. Lam (1990) wrote Equation (3) in differenced form

$$\nabla y_t = z_t - z_{t-1} + \alpha_0 + \alpha_1 S_t \qquad (4)$$

that we may take as the observation equation (1) with state vector

$$\mathbf{x}_t = (z_t, z_{t-1}, \alpha_{0t}, \alpha_{1t})'$$

and $M_1 = (1, -1, 1, 0)$ and $M_2 = (1, -1, 1, 1)$ determining the two possible economic conditions. The state equation (2) is of the form

$$\begin{bmatrix} z_t \\ z_{t-1} \\ \alpha_{0t} \\ \alpha_{1t} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_{t-1} \\ z_{t-2} \\ \alpha_{0,\,t-1} \\ \alpha_{1,\,t-1} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad (5)$$

which models the constant drift in terms of two constant processes, α_{0t} and α_{1t}. In this case a rising economy might be characterized by drift $\alpha_{0t} + \alpha_{1t}$ and a falling economy by α_{0t}.

To incorporate a reasonable switching structure for the measurement matrix into the dynamic linear model (1) and (2) that is compatible with both the practical situations described above, we assume that the m possible configurations are states in a nonstationary independent process defined by the time-varying probabilities

$$\pi_j(t) = \Pr\{A_t = M_j\}, \qquad (6)$$

for $j = 1, \ldots, m$ and $t = 1, 2, \ldots, n$, independent of past measurement matrices $A_1, \ldots, A_{t-1}$ and of past data $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{t-1}$. Important information about the current state of the measurement process is given by the filtered probabilities for being in state j, defined as the conditional probabilities

$$\pi_j(t \mid t) = \Pr\{A_t = M_j \mid Y_t\}, \qquad (7)$$

which also vary as a function of time. We will use the notation $Y_s = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_s\}$ $(s = 1, 2, \ldots, n)$ to denote the space spanned by the observations $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_s$. The filtered probabilities (7) give time-varying estimates of the probability of being in state j given the past and present Y_t. The most important estimators for purposes of tracking the process are the filtered estimators for the state vectors $\mathbf{x}_t$, say

$$\mathbf{x}_t^t = E(\mathbf{x}_t \mid Y_t), \qquad (8)$$

and the filter covariances defined by

$$P_t^t = \operatorname{cov}(\mathbf{x}_t \mid Y_t) \qquad (9)$$

since these are the best current estimators of position.

It is clear that if the parameters of the dynamic linear model are known, we would like to be able to compute quickly both an estimator for the smoothed trajectory and its variance given by Equations (8) and (9) and an estimator for the probabilities of each of the current measurement configurations given by Equation (7). In many cases we will not know the parameters of the dynamic linear model, given by the initial means and by the transition and covariance matrices mentioned in the discussion following (1) and (2). In these cases a procedure for estimating such parameters by maximum likelihood will be of interest. A Bayesian direction can also be taken that assigns prior distributions to the unknown parameters (see Gordon and Smith 1988, 1990, or Peña and Guttman 1988). Our approach here will be in terms of classical maximum likelihood estimation using an approximation to the Expectation-Maximization (EM) algorithm (Dempster, Laird, and Rubin 1977) in combination with a standard nonlinear optimization procedure.

To summarize we derive exact equations for the filtered probabilities and state vectors in Section 3. The estimators given for the state vectors do not involve mixtures of the normal distribution as in Gordon and Smith (1988, 1990) or Peña and Guttman (1987) so that no approximations are needed to avoid the geometric increase in computations. In Section 3 we derive a procedure for estimating the parameters using a pseudo-EM algorithm to get close to the final values before changing to a classical nonlinear optimization algorithm.

3. FILTERING

In this section, we establish the recursions for the filters associated with the state process $\mathbf{x}_t$ and the switching process A_t. The filters in this section are an essential part of the maximum likelihood estimation procedure which will be derived in Section 4.

First, consider the derivation of filters for the measurement matrices A_t. Let $f_j(t \mid t-1)$ denote the conditional density of y_t, given $A_t = M_j$ and the past; we know immediately that, under the multivariate normal assumptions, the conditional density $f_j(t \mid t-1)$ is that of a normal with mean $M_j\mathbf{x}_t^{t-1}$ and covariance matrix

$$\Sigma_{tj} = M_j P_t^{t-1} M_j' + R, \tag{10}$$

where

$$\mathbf{x}_t^{t-1} = E(\mathbf{x}_t \mid Y_{t-1}) \tag{11}$$

is the one-step filtered estimator for $\mathbf{x}_t$. Then the updating equation to get $\Pr\{A_t = M_j \mid Y_t\}$ is given by

$$\pi_j(t \mid t) = \frac{\pi_j(t)\, f_j(t \mid t-1)}{\sum_{k=1}^{m} \pi_k(t)\, f_k(t \mid t-1)}, \tag{12}$$

where we assume that the distribution $\pi_j(t)$ $(j = 1, \ldots, m)$ has been specified prior to observing $\mathbf{y}_1, \ldots, \mathbf{y}_t$.

A potential weakness of the model is the need to specify the time-varying prior probabilities $\pi_j(t)$. If the investigator has no reason to prefer one state over another at time t, he or she might choose uniform probabilities $\pi_j(t) = m^{-1}$. Smoothness can be introduced by letting

$$\pi_i(t) = \sum_{j=1}^{m} \pi_{ij}\pi_j(t-1 \mid t-1), \tag{13}$$

where the nonnegative weights are chosen so that $\sum_j \pi_{ij} = 1$. If the process A_t were Markov with transition probabilities π_{ij}, then (13) would be the update for the filter probability, as has been shown in Lindgren (1978) or Kitagawa (1987) (see, also, Rabiner and Juang 1986). The difficulty in extending the approach here to the Markov-dependent case is the dependence in $\mathbf{y}_t$ which makes it necessary to enumerate over the possible histories to derive Equation (12). Equation (13) has $\pi_i(t)$ as a function of the past observations Y_{t-1} and hence is inconsistent with model assumption (10). Nevertheless, this seems to be a reasonable compromise that allows the data to modify the probabilities $\pi_j(t)$.

The filtered estimators $\mathbf{x}_t^{t-1}$ and $\mathbf{x}_t^t = E(x_t \mid Y_t)$ are given by

$$\mathbf{x}_t^{t-1} = \Phi \mathbf{x}_{t-1}^{t-1}, \tag{14}$$

$$P_t^{t-1} = \Phi P_{t-1}^{t-1} \Phi' + Q, \tag{15}$$

$$\mathbf{x}_t^t = \mathbf{x}_t^{t-1} + \sum_{j=1}^{m} \pi_j(t \mid t)\, K_{tj}(\mathbf{y}_t - M_j \mathbf{x}_t^{t-1}), \tag{16}$$

and

$$P_t^t = \sum_{j=1}^{m} \pi_j(t \mid t)(I - K_{tj}M_j)P_t^{t-1}, \tag{17}$$

where

$$K_{tj} = P_t^{t-1} M_j' \Sigma_{tj}^{-1} \tag{18}$$

with Σ_{tj} defined as in Equation (10). This exhibits the updated filters $\mathbf{x}_t^t$ and filtered covariances as weighted combinations of the m scaled innovations and covariances, respectively, corresponding to each of the possible measurement matrices. The equations are somewhat similar to the approximations introduced by Bar-Shalom and Tse (1975) as "probabilistic data association filters" [see, also, the approximations of Gordon and Smith (1988) and Peña and Guttman (1988)].

To verify (16), let $I(A_t = M_j)$ be the indicator function of the set $A_t = M_j$ and note that

$$\begin{aligned}
\mathbf{x}_t^t &= E(\mathbf{x}_t \mid Y_t) = E[E(\mathbf{x}_t \mid Y_t, A_t) \mid Y_t] \\
&= E\left\{\sum_{j=1}^{m} E(\mathbf{x}_t \mid Y_t, A_t = M_j)\, I(A_t = M_j) \mid Y_t\right\} \\
&= E\left\{\sum_{j=1}^{m} [\mathbf{x}_t^{t-1} + K_{tj}(y_t - M_j x_t^{t-1})]\, I(A_t = M_j) \mid Y_t\right\} \\
&= \sum_{j=1}^{m} \pi_j(t \mid t)[\mathbf{x}_t^{t-1} + K_{tj}(\mathbf{y}_t - M_j \mathbf{x}_t^{t-1})],
\end{aligned}$$

where K_{tj} is given by (18). Equation (17) is derived in a similar fashion.

To review, we have shown how to extend the classical Kalman filtering recursions to the case where the measurement matrices are switching (or not) in accordance with a nonstationary independent measurement process. In this model we have derived estimators for the probability of being in state j at time t given the past and present Y_t in Equation (12). The modified results for the filtered state estimators (16) show that the filtered estimators involve weighted combinations of the gain-adjusted innovations. The filtered covariances P_t^t again involve weighted combinations of the conventional estimators.

4. MAXIMUM LIKELIHOOD ESTIMATION

To develop a procedure for maximum likelihood estimation, note first of all that the innovations form of the log-likelihood function is proportional to

$$\ln L'(\boldsymbol{\theta}) = \sum_{t=1}^{n} \ln\left(\sum_{j=1}^{m} \pi_j(t)\, f_j(t \mid t-1)\right), \tag{19}$$

where $f_j(t \mid t-1)$ are the multivariate normal densities defined earlier with means $M_j\mathbf{x}_t^{t-1}$ and covariance matrices given in (10). We may consider maximizing (19) directly as a function of the parameters $\Theta = (\boldsymbol{\mu}, \Phi, Q, R)$, or we may consider applying the EM algorithm to the complete-data log-likelihood. In general, the EM algorithm converges nicely in the initial stages and more slowly in the final stages, where it is advantageous to switch to a standard nonlinear optimization procedure.

To apply the EM algorithm, note in the Appendix that the log-likelihood of the complete data, $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n, A_1, A_2, \ldots, A_n$, and $\mathbf{y}_1, \ldots, \mathbf{y}_n$, can be written as in Shumway and Stoffer (1982) or Shumway (1988) with an additional term corresponding to the unknown probabilities $\pi_j(t)$. This leads to the same regression equations for the transition matrix Φ and the error covariance Q expressed in terms of the same smoothers $\mathbf{x}_t^n = E(\mathbf{x}_t \mid Y_n)$ and smoother covariances P_t^n and $P_{t,t-1}^n$. We note only the modification (A.16) for the last smoothed covariance. The equation for updating R given in (A.8) involves smoothed probabilities $\pi_j(t \mid n)$ that have not been computed. The backwards recursions for the smoothed probabilities involve integrating over mixtures of normal distributions and are excessively complicated. Monte Carlo integration techniques such as the Gibbs sampler (Carlin, Polson, and Stoffer 1990) may be useful, but here we apply Equation (A.8) in the EM algorithm assuming that the smoothed probabilities can be approximated by the filtered probabilities $\pi_j(t \mid t)$. The resulting "pseudo-EM" algorithm works quite well as is evident from the example considered in the following section.

Since the pseudo-EM algorithm is quite slow in later iterations and does not necessarily increase the incomplete-data log-likelihood or converge to maximizing values for the parameters, a variable-metric (Fletcher–Powell–Davidon) nonlinear optimization procedure (see Nash and Walker-Smith 1987) was applied to finish the maximization of Equation (19) in the latter stages.

5. AN EXAMPLE

To illustrate the procedures of the preceding sections, we return to the problem of tracking multiple targets, as introduced in Section 1. The example given here uses a set of contrived data that simulates the behavior of three sensors observing various configurations of one to three targets over $n = 100$ time points. The difficulty is that it is not possible to identify which configuration of targets is being observed at any given time point. This introduces a whole collection of plausible measurement matrices, $M_1, M_2, \ldots, M_m$ at each time point to serve as possible models for the true measurement matrix A_t. Two of these matrices, corresponding to observing only one track on all three sensors or observing one track on each of the three sensors, were given in Section 1. The complete collection of measurement matrices investigated at each point as plausible explanations for the data forms the set of possible states for the hidden Markov chain in this example. A listing of the $m = 10$ configurations assumed to be possible at each time point is shown in Table 1. Notice that the two measurement matrices shown in Section 1 can be identified as M_1 and M_7 in Table 1.

The underlying tracks were generated according to a model with a transition matrix of the form $\Phi = \text{diag}(1.005, .990, 1.000)$ which corresponds to three tracks, a random walk, and two tracks that increase and decrease slightly over time.

Table 1. Definition of Sensor Target Associations Used to Determine m = 10 Possible Measurement Matrices in Example

Measurement matrix	Measurements and observed tracks		
	y_{t1}	y_{t2}	y_{t3}
M_1	x_{t1}	x_{t1}	x_{t1}
M_2	x_{t1}	x_{t1}	x_{t2}
M_3	x_{t1}	x_{t2}	x_{t1}
M_4	x_{t2}	x_{t1}	x_{t1}
M_5	x_{t1}	x_{t3}	x_{t2}
M_6	x_{t1}	x_{t2}	x_{t3}
M_7	x_{t2}	x_{t1}	x_{t3}
M_8	x_{t2}	x_{t3}	x_{t1}
M_9	x_{t3}	x_{t2}	x_{t1}
M_{10}	x_{t3}	x_{t1}	x_{t2}

The matrices Q and R were taken to be diag(.0025, .0025, .0025) and diag(.0625, .0625, .0625), respectively, and the initial mean was $\mu = (5, 5, 5)'$ with $n = 100$ points. The measurement matrices were switched three times during the first 100 points, leading to the observed data shown in Figure 1. Since all tracks started at the same mean, the observed data corresponds roughly to three one-dimensional targets originating from a common launch point. To sim-

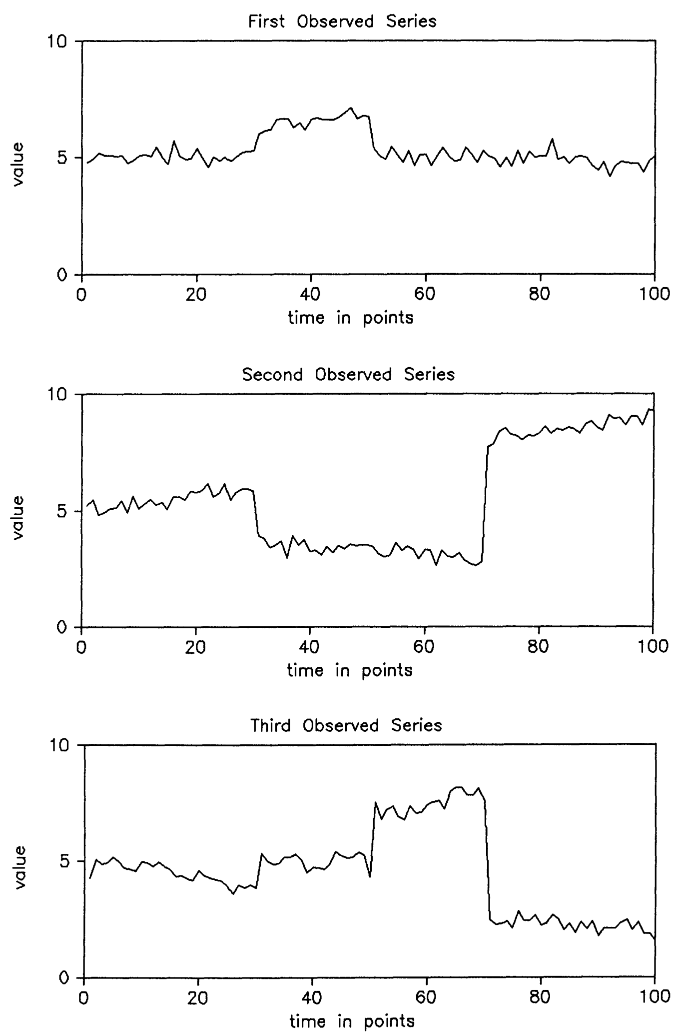

Figure 1. Original Measurements for Tracking Example.

plify further we have not exhibited the nonlinear dynamics of the measurement matrices, which would certainly be present in any realistic application. It is fairly clear from the observed tracks where the transition points lie and which series (staying level, decreasing, increasing) that we are observing. We have chosen a fairly straightforward set of realizations to test the method.

There is the question in any application as to whether time varying probabilities or the parameters in the dynamic linear model are more important or whether the two are equally important. Often the dynamics of the tracks are assumed known so that those parameters are known. In that case adjustment of the initial probabilities might be very useful for effectively pruning highly improbable measurement matrices. In other cases the dynamics of the signal or track are unknown, and the initial probabilities simply may not matter that much. In this case parameter estimation might proceed for some fixed reasonable a priori set of probabilities, such as $\pi_j(t) = m^{-1}$. The filtered probabilities (12) can serve as the important indicator of whether or not we are in state j at time t. In the present example there are 10 initial probabilities. Because of the large number of states that are common in examples involving tracking, we chose to hold all of these probabilities fixed at reasonable starting values and identify the tracks by varying the parameters in the state-space model. The $m = 10$ initial probabilities were all taken as .1, corresponding to assuming that each configuration is equally likely.

A summary of parameter estimates using 50 iterations of the EM algorithm and several varible-metric (Fletcher–Powell–Davidon) iterations is shown in Table 2. It is clear that the log-likelihood (19) increases at each step and that the transition matrix parameters converge rather quickly, say after 5 iterations. Estimating variances requires somewhat more time, and these may be assumed known to increase processing speed. One iteration requires running the filters and smoothers for both the probabilities and the signal tracks, as detailed in Section 2. Time requirements were 6 min., 3 min. and 1 min. respectively for 8086-, 80286-, and 80386-based microcomputers. It is clear that tremendous reductions in computing time are available through parallel processing, which can be done in Equations (12), (16), and (17) using parallel paths for each possible measurement matrix. Hence the multiple path filters would require no more time than a single Kalman filter and smoother.

Figure 2 shows 4 of the 10 filtered probabilities, computed using the forward recursions (12) and (14)–(18). These are states exhibiting nonzero probabilities over a significant portion of the range. It is clear that after an initial period of indecisiveness ($t < 20$), the filter indicates M_{10} for $20 \le t \le 29$, M_6 for $30 \le t \le 49$, M_9 for $50 \le t \le 69$, and M_{10} again for $70 \le t \le 100$.

Figure 3 shows the signal tracks, estimated using Equations (14)–(18), and we note that they compare well to the input series mixed to determine the observed realizations in Figure 1. In fact, we did not plot the original tracks because they would have overlayed the estimated tracks almost exactly.

6. DISCUSSION

We have developed an approach to modeling change in a vector time series that uses a model for switching that is different than considered by Gordon and Smith (1988, 1990) or by Peña and Guttman (1988). In general restricting the switching to the measurement matrices simplifies the Kalman filtering recursions considerably since the equations are exact and it is not necessary to approximate mixture of normals as in the Gordon and Smith approach.

The model is still general enough to include most structural models useful in modeling and monitoring changes of regime in vector time series. Even changes of the kind considered by Gordon and Smith can be accommodated as long as one is willing to assume that they occur as abrupt changes in the configuration of elements observed in the state vec-

Table 2. Iterations for Parameters in Tracking Example

Iteration no.	ϕ_{11}	ϕ_{22}	ϕ_{33}	q_{11}	q_{22}	q_{33}	r_{11}	r_{22}	r_{33}	ln L'
					Pseudo-EM					
0	1.000	1.000	1.000	.1000	.1000	.1000	.1000	.1000	.1000	−316.67
5	1.006	.990	.999	.0272	.0283	.0332	.0658	.0653	.0744	−252.50
10	1.006	.990	.999	.0118	.0129	.0160	.0516	.0533	.0643	−232.53
15	1.006	.991	.999	.0070	.0081	.0102	.0509	.0522	.0664	−226.84
20	1.006	.991	.999	.0047	.0059	.0073	.0512	.0519	.0682	−223.93
50	1.006	.991	.999	.0016	.0026	.0025	.0534	.0516	.0724	−219.79
					Nonlinear optimization (Variable Metric)					
51	1.006	.991	.999	.0013	.0025	.0023	.0533	.0516	.0732	−219.70
52	1.006	.991	.999	.0010	.0029	.0019	.0533	.0516	.0723	−219.51
53	1.006	.991	.999	.0011	.0022	.0015	.0533	.0515	.0723	−219.41
54	1.006	.991	.999	.0011	.0021	.0015	.0533	.0515	.0722	−219.38
55	1.006	.991	.999	.0011	.0020	.0015	.0533	.0514	.0722	−219.38
56	1.006	.991	.999	.0011	.0023	.0012	.0536	.0466	.0711	−219.36
57	1.006	.991	.999	.0011	.0024	.0011	.0536	.0461	.0709	−219.33
58	1.006	.991	.999	.0011	.0024	.0012	.0533	.0460	.0704	−219.32
*	1.005	.990	1.000	.0025	.0025	.0025	.0625	.0625	.0625	

*denotes true parameter values.

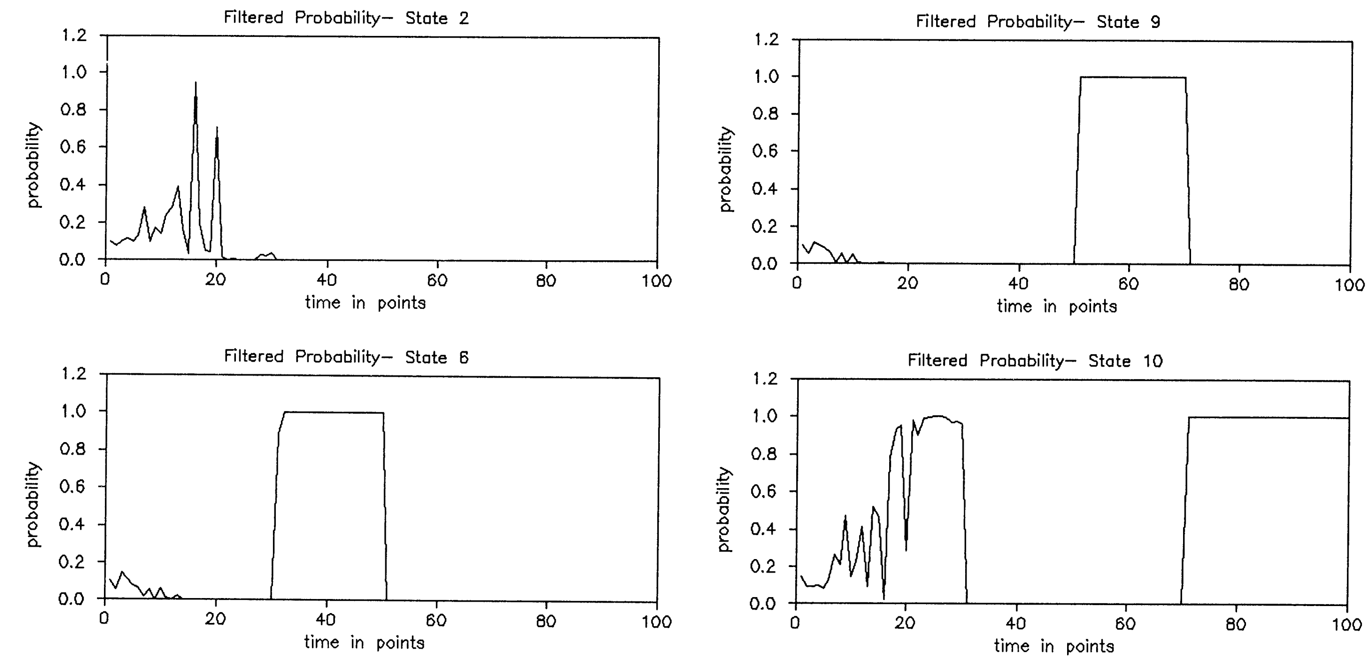

Figure 2. Filtered Probabilities for Most Probable Measurement Matrices.

tor. Modeling of abrupt changes in the observation error, however, is precluded under our formulation, making our switching structure of relatively little use in robust detection of outliers as considered by Penã and Guttman.

We mention also that our approach allows for simultaneous estimation of parameters in the dynamic linear model by maximum likelihood. This means that the sequential model selection approach and parameter estimation can be accomplished simultaneously.

It should be noted in closing that more general models can be considered that introduce dependence in the measurement matrices as they evolve over time. For example, the measurement matrices may form a stationary Markov chain with constant initial and transition probabilities. The cost of increasing the level of generality to the dependent case seems to be high in computational effort since the possible dependent paths must be laboriously traced back through the chain, as in Hamilton (1989) who analyzed a simple two-state Markov structure. In addition one will have to assume reasonable values for the transition probabilities or treat them as additional parameters. Thus it seems that the simple extensions to dependent switching transitions that obtained for independent data (see Lindgren 1978) do not carry over easily to the dependent data case given here.

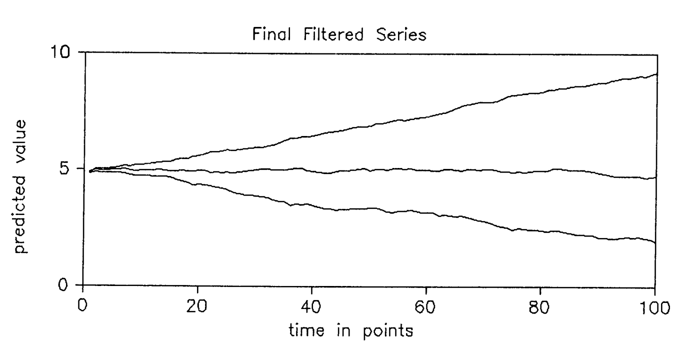

Figure 3. Filtered Estimators for Tracks.

APPENDIX: PSEUDO–EM ALGORITHM

We give details for the pseudo-EM algorithm used in the first stages of the procedure for maximizing the incomplete-data log-likelihood (16). If we imagine the unobserved and observed components of the model (1) and (2) generated in the order $\mathbf{x}_0$, $\mathbf{x}_1$, A_1, $\mathbf{y}_1$, $\mathbf{x}_2$, A_2, $\mathbf{y}_2$, ..., the complete-data log-likelihood can be written as

$$\begin{aligned}\ln L(\boldsymbol{\theta}) = & -\frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x_0 - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_0 - \boldsymbol{\mu}) - \frac{n}{2}\ln|Q| \\ & - \frac{1}{2}\sum_{t=1}^{n}(\mathbf{x}_t - \Phi\mathbf{x}_{t-1})'Q^{-1}(\mathbf{x}_t - \Phi\mathbf{x}_{t-1}) \\ & + \sum_{t=1}^{n}\sum_{j=1}^{m} I(A_t = M_j)\ln\pi_j(t) - \frac{n}{2}\ln|R| \\ & - \frac{1}{2}\sum_{t=1}^{n}\sum_{j=1}^{m} I(A_t = M_j)(\mathbf{y}_t - A_t\mathbf{x}_t)'R^{-1}(\mathbf{y}_t - A_t\mathbf{x}_t).\end{aligned} \tag{A.1}$$

The EM algorithm requires that we maximize the conditional expectation

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = E_{\theta_0}[\ln L(\boldsymbol{\theta}) \mid Y_n] \tag{A.2}$$

with respect to $\boldsymbol{\theta}$ at each step, where $\boldsymbol{\theta}_0$ is the parameter value at the previous iteration.

Now, taking conditional expectations in (A.1), noting that

$$\pi_j(t \mid n) = E[I(A_t = M_j) \mid \mathrm{Y}_n], \tag{A.3}$$

leads to

$$\hat{\pi}_j(t) = \pi_j(t \mid n), \tag{A.4}$$

$$\hat{\boldsymbol{\mu}} = \mathbf{x}_0^n, \tag{A.5}$$

$$\hat{\Phi} = BA^{-1}, \tag{A.6}$$

$$\hat{Q} = n^{-1}(C - B\Phi' - \Phi B' + \Phi A \Phi'), \tag{A.7}$$

and

$$\hat{R} = n^{-1} \sum_{t=1}^{n} \sum_{j=1}^{m} \pi_j(t \mid n)[(\mathbf{y}_t - M_j\mathbf{x}_t^n)(\mathbf{y}_t - M_j\mathbf{x}_t^n)' + M_j P_t^n M_j']. \tag{A.8}$$

The covariance matrix Σ is held fixed at some reasonable value. The matrices A, B, and C are defined as

$$A = \sum_{t=1}^{n} (P_{t-1}^n + \mathbf{x}_{t-1}^n \mathbf{x}_{t-1}^{n\prime}), \tag{A.9}$$

$$B = \sum_{t=1}^{n} (P_{t,t-1}^n + \mathbf{x}_t^n \mathbf{x}_{t-1}^{n\prime}), \tag{A.10}$$

and

$$C = \sum_{t=1}^{n} (P_t^n + \mathbf{x}_t^n \mathbf{x}_t^{n\prime}). \tag{A.11}$$

These matrices involve the usual Kalman smoothers $\mathbf{x}_t^n = E(\mathbf{x}_t \mid Y_n)$ and their covariances

$$P_t^n = E\{(\mathbf{x}_t - \mathbf{x}_t^n)(\mathbf{x}_t - \mathbf{x}_t^n)' \mid Y_n\}$$

and

$$P_{t,t-1}^n = E\{t\mathbf{x}_t - \mathbf{x}_{t-1}^n)(\mathbf{x}_{t-1} - \mathbf{x}_t^n)' \mid Y_n\}.$$

The smoothers are derived under the assumption that the A_t are stochastic in a manner analogous to that used for deriving the filters in Section 3. We obtain for $t = n, n - 1, \ldots,$

$$\mathbf{x}_{t-1}^n = \mathbf{x}_{t-1}^{t-1} + J_{t-1}(\mathbf{x}_t^n - \mathbf{x}_t^{t-1}), \tag{A.12}$$

where

$$J_t = P_{t-1}^{t-1}\Phi'(P_t^{t-1})^{-1}. \tag{A.13}$$

The smoother covariances satisfy

$$P_{t-1}^n = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1})J_{t-1}' \tag{A.14}$$

for $t = n, n - 1, \ldots, 1$ and

$$P_{t-1,t-2}^n = P_{t-1}^{t-1}J_{t-2}' + J_{t-1}(P_{t,t-1}^n - \Phi P_{t-1}^{t-1})J_{t-2}' \tag{A.15}$$

for $t = n, n - 1, \ldots, 2$ subject to

$$P_{n,n-1}^n = \sum_{j=1}^{m} \pi_j(n \mid n)(I - K_{nj}M_j)\Phi P_{n-1}^{n-1}, \tag{A.16}$$

where K_{nj} is defined in (15).

As discussed in the text, the difficulties encountered in computing $\pi_j(t \mid n)$ led us to replace it with $\pi_j(t \mid t)$ in (A.8). The advantages of using the EM algorithm are (a) its stability under the switching structure, and (b) the simple regression computations involved at each step as exhibited in (A.6)–(A.8).

[Received February 1990. Revised January 1991.]

REFERENCES

Bar-Shalom, Y. (1978), "Tracking Methods in a Multi-Target Environment," *IEEE Transactions on Automatic Control,* 23, 618–626.

Bar-Shalom, Y., and Tse, E. (1975), "Tracking in a Cluttered Environment With Probabilistic Data Association," *Automatica,* 11, 4451–4460.

Carlin, B. P., Polson, N. G., and Stoffer, D. S., (1990), "A Monte Carlo Approach to Nonnormal and Nonlinear State Space Modeling," preprint.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society,* Ser. B, 35, 1–38.

Goldfeld, S. M., and Quandt, R. E. (1973), "A Markov Model for Switching Regression," *Journal of Econometrics,* 1, 3–16.

Gordon, K., and Smith, A. F. M. (1988), "Modeling and Monitoring Discontinuous Changes in Time Series," in *Bayesian Analysis of Time Series and Dynamic Linear Models,* ed. J. C. Spall, New York: Marcel Dekker, pp. 359–392.

——— (1990), "Modeling and Monitoring Biomedical Time Series," *Journal of the American Statistical Association,* 85, 328–337.

Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica,* 57, 357–384.

Harrison, P. J., and Stevens, C. F. (1976), "Bayesian Forecasting" (with discussion), *Journal of the Royal Statistical Society,* Ser. B, 38, 205–247.

Harvey, A. C., and Todd, P. H. J. (1983), "Forecasting Economic Time Series With Structural and Box–Jenkins Models," *Journal of Business and Economic Statistics,* 1, 299–315.

Juang, B. H. and Rabiner, L. R. (1985), "Mixture Autoregressive Hidden Markov Models for Speech Signals," *Transactions of the IEEE Acoustics, Speech, and Signal Processing,* 33, 1404–1413.

Kitagawa, G. (1987), "Non-Gaussian State-Space Modeling of Nonstationary Time Series," *Journal of the American Statistical Association,* 82, 1032–1041.

Kitagawa, G., and Gersch, W. (1984), "A Smoothness Priors Modeling of Time Series With Trend and Seasonality," *Journal of the American Statistical Association,* 79, 378–389.

Lam, P. S. (1990), "The Hamilton Model With a General Autoregressive Component: Estimation and Comparison With Other Models of Economic Time Series," preprint.

Lindgren, G. (1978), "Markov Regime Models for Mixed Distributions and Switching Regressions," *Scandinavian Journal of Statistics,* 5, 81–91.

Nash, J. C., and Walker-Smith, M. (1987), *Nonlinear Parameter Estimation,* New York: Marcel Dekker.

Peña, D., and Guttman, I. (1988), "Bayesian Approach to Robustifying the Kalman Filter," in *Bayesian Analysis of Time Series and Dynamic Linear Models,* ed. J. C. Spall, New York: Marcel Dekker, pp. 227–254.

Quandt, R. E. (1972), "A New Approach to Estimating Switching Regressions," *Journal of the American Statistical Association,* 67, 306–310.

Rabiner, L. R., and Juang, B. H. (1986), "An Introduction to Hidden Markov Models," *IEEE Acoustics, Speech, and Signal Processing Magazine,* Jan., 4–16.

Shumway, R. H. (1988), *Applied Statistical Time Series Analysis,* New York: Prentice-Hall.

Shumway, R. H., and Stoffer, D. S. (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis,* 3, 253–264.

Smith, A. F. M., and West, M. (1983), "Monitoring Renal Transplants: An Application of The Multiprocess Kalman Filter," *Biometrics,* 39, 867–878.

Tsurimi, H. (1988), "Survey of Bayesian and Non-Bayesian Testing of Model Stability in Econometrics," in *Bayesian Analysis of Time Series and Dynamic Linear Models,* ed. J. C. Spall, New York: Marcel Dekker, pp. 75–100.

CORRECTIONS

R. H. Shumway and D. S. Stoffer, "Dynamic Linear Models With Switching," 86, No. 415 (September 1991), 763–769.

Scott Vander Wiel of AT&T Bell Laboratories has pointed out the following oversight in our development of the filtering equations established in Section 3, p. 765. The conditional densities $f_j(t|t-1) = f\{\mathbf{y}_t | Y_{t-1}, A_t = M_j\}$ $(j = 1, \ldots, m)$ were needed to establish the recursions for the filter associated with the state process. The conditional mean vectors and covariance matrices are correct as stated; however, the densities are mixtures of normals.

Let $A(l)$ denote the set $\{A_1 = M_{j_1}, \ldots, A_{t-1} = M_{j_{t-1}}\}$, where l refers to a particular $(t-1)$-tuple, $(j_1, \ldots, j_{t-1})$, $l, \ldots m^{t-1}$. Then, from standard Kalman filter results with fixed design matrices, the density $f\{\mathbf{y}_t | Y_{t-1}, A_t = M_j, A(l)\}$ is multivariate normal with mean vector $M_j x_t^{t-1}(l)$ and covariance matrix $M_j P_t^{t-1}(l) M_j' + R$, where the argument l denotes filtering with respect to the particular set of design matrices specified by $A(l)$. Of course, $x_t^{t-1}(l)$ and $P_t^{t-1}(l)$ can be obtained from the standard Kalman filter recursions. It is clear that the density $f_j(t|t-1)$ is a mixture of the normal densities with mixing probabilities $\pi(l|t-1) = \Pr\{A(l) | Y_{t-1}\}$, $l = 1, \ldots, m^{t-1}$; that is,

$$f_j(t|t-1) = \sum_l \pi(l|t-1) f\{\mathbf{y}_t | Y_{t-1}, A_t = M_j, A(l)\}, \qquad j = 1, \ldots, m.$$

These densities involve computation of $x_t^{t-1}(l)$ and $P_t^{t-1}(l)$, and calculation of these quantities involves iterating over an exponentially increasing number of possible histories. Because our objective was able to compute quickly both an estimator for the smoothed trajectory and its variance, we wanted to avoid lengthy calculations where it was necessary to filter and smooth over all possible histories. The normal densities with mean vectors $M_j x_t^{t-1}$ and covariance matrices $M_j P_t^{t-1} M_j' + R$, as given in Section 3, can be considered approximations to the densities $f_j(t|t-1)$; these approximations allow for the quick updating scheme specified in the article. Moreover, these single normal densities minimize the Kullback–Leibler distance for the normal mixtures described earlier. (Similar approximations were used in Gordon and Smith 1990 and Peña and Guttman 1988, to mention a few.) As seen from the example in Section 5, the performance and tractability of the scheme is impressive.

REFERENCES

Gordon, K., and Smith, A. F. M. (1990), "Modeling and Monitoring Biomedical Time Series," *Journal of the American Statistical Association,* 85, 328–337.

Peña, D., and Guttman, I. (1988), "Bayesian Approach to Robystifying the Kalman Filter," in *Bayesian Analysis of Time Series and Dynamic Linear Models,* ed. J. C. Spall, New York: Marcel Dekker, pp. 227–254.