Application of the Kalman Filter

a Markov Switching Model

by

Kazume Nishiyama

A dissertation submitted to the Graduate Faculty in Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1997

UMI Number: 9720122

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

This manuscript has been read and accepted for the Graduate Faculty in Economics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

1/27/97 — S. Neftci (FNE)

Date — Chair of Examining Committee

1/27/97 — Linda N. Edwards

Date — Executive Officer

Prof. Salih Neftci

Prof. Michael Grossman

Prof. Mitchell Kellman

Supervisory committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

APPLICATION OF THE KALMAN FILTER
TO A MARKOV SWITCHING MODEL

by

Kazume Nishiyama

Adviser: Professor Salih N. Neftçi

This paper proposes a linear approximation in the form of Kalman filter for models with Markov regime shifts. James Hamilton (1989) developed the optimal nonlinear algorithm for this type of model in Econometrica. However, the nonlinear filter lacks a closed form solution. It involves highly complicated and time-consuming procedures and is difficult to implement. On the other hand, the optimal linear filter, the Kalman filter, has a closed form solution and is simple and easy to implement. This paper investigates how well the Kalman filter approximates the optimal filter in various situations.

In the first section of this paper, the Kalman filter formulae are derived for a model with hidden states following a first order Markov process. Next, we apply the Kalman filter to US real GNP series to establish business cycle dates. These dates are compared with those of Hamilton and the National Bureau of Economic Research. Finally, simulations are conducted to compare the two filters in more general contexts.

The results indicate that the optimal linear filter approximates the nonlinear filter remarkably well. When the Kalman filter is applied to US real GNP series, it yields business cycle dates which are very similar to those of NBER and Hamilton. Moreover, the results of simulations indicate that the larger the unexpected jumps or shifts in the series, the better the Kalman filter approximates the nonlinear filter. Further, for extremely large unexpected jumps and shifts, the Kalman filter actually outperforms the nonlinear filter.

Acknowledgments

I am indebted to Salhi N. Neftci, my dissertation advisor, and Michael Grossman for many helpful comments and advise for the completion of this thesis. I am truly grateful for their kind support and guidance throughout my years in the Graduate Center.

Table of Contents

Tables

Figures

1 Introduction

It has been of considerable interest to characterize the nature of economic time series. One of the recent approaches to this investigation revolves around the Markov switching models. These models assume that the hidden state of the economy is subject to discrete regime shifts governed by a Markov process.

Hamilton (1988) analyzed the term structure of interest rates using the model in which the mean level of the short term interest rate is subject to a two-state second order Markov process and the disturbance term follows an AR(4) with an error whose variance is also subject to regime changes. The practical difficulties in this type of model arise from their nonlinear behavior which cannot be handled by conventional econometric methods. Hamilton proposed a nonlinear iterative procedure to model shifts in regimes. In his 1989 paper, he applied this method on quarterly US GNP series and was successful in assigning business cycle dates. The Markov switching model owes its recent popularity to Hamilton's algorithm proposed in his 1988 and 1989 papers.

Hamilton's nonlinear filter yields a series of conditional distribution of the hidden state and the observed variable at each t, given the current and past values of the latter. The conditional distribution of the observed variable is used to construct the likelihood function which is maximized to estimate parameters. The inference about the regime can be made using the result from a smoothing technique.

Nonlinear filter is optimal in the sense of minimizing mean squared errors(MSE). However, in general nonlinear filter lacks a closed form solution, moreover involves highly complicated procedures at each stage of iteration, and is time-consuming to implement in practice. Due to its complexity, applications of nonlinear filter has been limited to relatively small systems (Hamilton, 1990)

Efforts to overcome these difficulties have been made. Hamilton (1990) himself proposed a use of an EM algorithm to maximize an ill-behaved likelihood function more easily. McCulloch and Tsay (1992) analyzed a general Markov switching model in Baysian framework using Gibbs sampling technique. This techniques is used to extracting marginal distributions from conditional distributions of variables.

This paper proposes an optimal linear approximation in the form of the Kalman filter for the markov switching model. It is well known that an optimal linear filter for a non- Gaussian process is the optimal linear filter for Gaussian analog of the original non-Gaussian process (Lipster and Shiryayev, 1977).

When the Kalman filter is applied to a Gaussian process, it yields a series of estimators and an associated covariance matrix. These estimators are optimal in the sense of minimizing MSE. If it is applied to non-Gaussian process, it loses its optimality, but it still remains to be optimal within the class of linear estimators. Our basic idea is to approximate the nonlinear filter by the optimal linear filter. If the result of the linear approximation is "close enough" to that of the optimal nonlinear filter, the use of the optimal linear filter on a non-Gaussian process is justified.

Also, if the data contain errors in variables, applying the optimal linear approximation may be more appropriate. Hence, errors in variables can provide another justification for obtaining Gaussian analogs of finite state Markov processes.

This paper is arranged as follows. In section 2, the simplest model with the states governed by a first-order Markov process is presented. In addition, the Kalman filter is derived in a way which shows its optimality among all linear estimators. Further. properties of the Kalman filter applied on a non-Gaussian process is compared to the Kalman filter applied on a Gaussian process. The comparison with the nonlinear filter is also made. In section 3, we run a Kalman filter on the same data set used by Hamilton to compare the performances of the Kalman filter and the nonlinear filter in establishing historical business cycle dates. Section 4 uses simulations to compare the two filters in various situations econometricians face when they attempt to predict the state of the economy with a priori estimates.

2 Model with Markov Regime Changes

2.1 Model

Let $\{y_t\}$ be a stationary process whose behavior is governed by underlying states of the economy, θ_t. Assume that θ_t is a two-state first-order Markov process, with the states being economic upswings (state 1) and economic downturns (state 0).

$$\theta_t = \begin{cases} a_1 & \text{if state 1 occurs} \\ a_0 & \text{otherwise} \end{cases}$$

The probability law for the transition between these two states can be described by the following probabilities.

$$p_{11} = P(\theta_t = a_1 | \theta_{t-1} = a_1)$$

$$p_{10} = P(\theta_t = a_0 | \theta_{t-1} = a_1) = 1 - p_{11}$$

$$p_{00} = P(\theta_t = a_0 | \theta_{t-1} = a_0)$$

$$p_{01} = P(\theta_t = a_1 | \theta_{t-1} = a_0) = 1 - p_{00}$$

There are various possible models which incorporate this Markov switching state. For example, in his 1989 paper, Hamilton presented a mean-switching model with autocorrelated error terms. He also presented a model which assumes that not only the mean but also the variance of error terms is subject to a Markov switching states. One also can formulate models with some exogenous variables whose coefficients are function of the states.

However, for ease of exposition, we formulate the most simple model as follows.

$$y_t = \theta_t + \eta_t \tag{1}$$

where $\eta_t \sim N(0, \sigma_\eta^2)$. In this model, only the mean value of y_t is subject to an occasional discrete shifts between the two regimes. Due to these shifts, the model exhibits non-linear behavior. Further, we assume that the economic agents observe the value of y_t at each t. However, they do not directly observe which state the economy is in. Then, our problem is to make a statistical inference about the state of the economy at each point in time, given all the information in the past. We assume that the information up to and including time t can be summarized by a vector $Y_t = (y_t, y_{t-1}, \cdots, y_0)$.

Hamilton (1989) proposed a non-linear filter which deals with such non-linear models with regime shifts. Our purpose is to present an optimal linear filter for this type of model. To do so, it is useful to represent the model in a state space form. Introduce an indicator vector $X_t = [x_{1t} \;\; x_{0t}]'$ such that:

$$x_{it} = \begin{cases} 1 & \text{if state i occurs} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 0$. Then, θ_t can be written as

$$\theta_t = a_1 x_{it} + a_0 x_{0t} = J' X_t,$$

where $J' = [a_1 \;\; a_0]$. Therefore, equation (1) becomes:

$$y_t = J' X_t + \eta_t. \tag{2}$$

We are now in the position to derive the transition equation for the vector X_t. Consider the expected value of $x_{j,t+1}$ conditional on Y_t:

$$E(x_{j,t+1}|Y_t) = E(x_{j,t+1}|X_t) = p_{1j} x_{1t} + p_{0j} x_{0t},$$

or simply,

$$E(X_{t+1}|Y_t) = \Lambda' X_t,$$

where Λ is the transition matrix of the Markov process, θ_t:

$$\Lambda = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{00} & p_{00} \end{bmatrix}.$$

The left hand side of the above equation is the expected value of the next period's indicator vector conditional on all the information available at time t. From the construction of the indicator vector, we can see $E(x_{j,t+1}|Y_t)$ equals a probability of being in each state conditional on information up to time t with j being the state prevailing at time $t+1$. Hence, $E(x_{j,t+1}|Y_t)$ takes on a value between 0 and 1, while a realization of $x_{j,t+1}$ is either 1 or 0. Therefore, to formulate an equation governing the transition from X_t to X_{t+1}, we introduce a disturbance terms, $\varepsilon_{j,t+1}$ to fill the gap between the actual value taken by $x_{j,t+1}$ and its conditional expectation for $j = 0, 1$. Then we have:

$$x_{j,t+1} = \sum_{i=0}^{1} p_{ij} x_{it} + \varepsilon_{j,t+1},$$

or simply,

$$X_{t+1} = \Lambda' X_t + \varepsilon_{t+1} \tag{3}$$

where $\varepsilon_t = [\varepsilon_{1t} \;\; \varepsilon_{0t}]'$.

The behavior of the disturbance vector ε_{t+1} is worth investigating. From the binary property of $x_{j,t+1}$, and from the construction of the transition equation (3), $\varepsilon_{j,t+1}$ is necessarily binary and can be written as:

$$\begin{aligned} \varepsilon_{j,t+1} &= x_{j,t+1} - E(x_{j,t+1}|X_t) \\ &= x_{j,t+1} - \sum_{i=0}^{1} p_{ij} x_{it}. \end{aligned} \tag{4}$$

Equation (4) implies that $E(\varepsilon_{j,t+1}|X_t) = 0$. Moreover, conditional on the value of X_t,

$$\varepsilon_{j,t+1} = \begin{cases} 1 - \sum_{i=0}^{1} p_{ij}x_{it} & \text{with probability} \quad \sum_{i=0}^{1} p_{ij}x_{it} \\ -\sum_{i=0}^{1} p_{ij}x_{it} & \text{with probability} \quad 1 - \sum_{i=0}^{1} p_{ij}x_{it}. \end{cases}$$

Note that, conditional on X_t, the probability that $x_{j,t+1} = 1$ is $\sum_{i=0}^{1} p_{ij}x_{it}$, and the probability that $x_{j,t+1} = 0$ is $1 - \sum_{i=0}^{1} p_{ij}x_{it}$. The conditional variance of $\varepsilon_{j,t+1}$ is:

$$\begin{aligned} E(\varepsilon_{j,t+1}^2|X_t) &= \left(\sum_{i=0}^{1} p_{ij}x_{it}\right)\left(1 - \sum_{i=0}^{1} p_{ij}x_{it}\right) \\ &= \sum_{i=0}^{1} p_{ij}x_{it} - \sum_{i=0}^{1} p_{ij}^2 x_{it} \\ &= E(x_{j,t+1}|X_t) - \sum_{i=0}^{1} p_{ij}^2 x_{it} \end{aligned} \tag{5}$$

for $j = 1, 0$. The derivation of the conditional variance of the error term relies on the fact that $x_{jt} = 1$ or 0 so that $x_{jt}^2 = x_{jt}$ and $x_{it}x_{jt} = 0$ for $i \neq j$.

Equations (2), (3), and (5) can be further simplified by using the relationship, $x_{0t} = 1 - x_{1t}$, to yield:

$$y_t = (a_1 - a_0)x_t + a_0 + \eta_t, \text{ and} \tag{6}$$

$$x_{t+1} = (p_{11} - p_{01})x_t + p_{01} + \varepsilon_{t+1} \tag{7}$$

$$\sigma_{t+1}^2 = E(x_{t+1}|X_t) - (p_{11}^2 - p_{01}^2)x_t - p_{01}^2, \tag{8}$$

where we replaced x_t for x_{1t} and ε_t for ε_{1t} for notational convenience. Equations (6) and (7) are the state space representation of our model, and (8) describes the evolution of the variance of the binary process, $\{\varepsilon_t\}$.

2.2 Optimal Linear Filter

Let $Y_t = (y_t, y_{t-1}, ..., y_0)$. We seek the optimal linear filter of the model that yields a series of linear estimators for the value of x_{t+1}, given Y_t. The optimal linear estimator.

$E^*(x_{t+1}|Y_t) = A^*Y_t + b^*$, is defined here to minimize the expected mean square error (MSE):

$$E[\| x_t - (AY_{t-1} + b) \|^2]. \tag{9}$$

For our purpose, the following theorems are necessary.

THEOREM 1:

Let the random variable X and Y be jointly distributed random variables such as:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \left(\begin{bmatrix} m_x \\ m_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right).$$

Then,

$$E^*(X|Y) = m_x + \Sigma_{xy}\Sigma_{yy}^{-1}(Y - m_y). \tag{10}$$

THEOREM 2:

Suppose that $X, Y_1, Y_2, \cdots, Y_k$ are jointly distributed, with $Y_1, Y_2, \cdots, Y_k$ mutually orthogonal to one another. Then,

$$E^*(X|Y_1, Y_2, \cdots, Y_k) = E^*(X|Y_1) + E^*(X|Y_2) + \cdots + E^*(X|Y_k) - (k-1)m_x \tag{11}$$

For the proofs of Theorem 1 and 2, please see Anderson and Moore (1979).

Define the following notations:

$E^*(x_{t+1}|Y_t) = \hat{x}_{t+1|t}$,

$E^*(x_t|Y_t) = \hat{x}_{t|t}$, and

$E^*(y_{t+1}|Y_t) = \hat{y}_{t+1|t}$.

For all t, the vector Y_t can be divided into two orthogonal parts: Y_{t-1} and v_t. v_t is the innovation containing the new information which becomes available when the new observation, y_t, arrives. A part of the information carried by y_t is already contained in Y_{t-1} through $E^*(y_t|Y_{t-1})$. Thus, we can express the innovation at time t by $v_t = y_t - E^*(y_t|Y_t)$. By the orthogonality condition of the optimal linear estimator. Y_{t-1} and v_t are mutually uncorrelated. Therefore, using Theorem 2, we have:

$$E^*(x_t|Y_t) = E^*(x_t|Y_{t-1}) + E^*(x_t|v_t) - E(x_t). \tag{12}$$

From Theorem 1, the second term in (12) becomes:

$$\begin{aligned} E^*(x_t|v_t) &= E(x_t) + Cov(x_t, v_t)Var(v_t)^{-1}v_t \\ &= E(x_t) + (a_1 - a_0)\Sigma_{t|t-1}((a_1 - a_0)^2\Sigma_{t|t-1} + \sigma_\eta)^{-1} \\ &\quad \times (y_t - (a_1 - a_0)\hat{x}_{t|t-1} - a_0), \end{aligned} \tag{13}$$

since

$$\begin{aligned} Cov(x_t, v_t) &= Cov(x_t, (a_1 - a_0)(x_t - x_{t|t-1}) + \eta_t) \\ &= (a_1 - a_0)\Sigma_{t|t-1} \end{aligned}$$

and

$$\begin{aligned} Var(v_t) &= Var((a_1 - a_0)(x_t - x_{t|t-1}) + \eta_t) \\ &= (a_1 - a_0)^2\Sigma_{t|t-1} + \sigma_\eta^2. \end{aligned}$$

Substituting (13) into (12) and noting $E^*(x_t|Y_{t-1}) = \hat{x}_{t|t-1}$, we obtain:

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + (a_1 - a_0)\Sigma_{t|t-1}$$

$$\times ((a_1 - a_0)^2 \Sigma_{t|t-1} + \sigma_\eta^2)^{-1}(y_t - (a_1 - a_0)\hat{x}_{t|t-1} - a_0) \tag{14}$$

Therefore, the best linear estimator of x_{t+1}, given Y_t is:

$$\begin{aligned}
\hat{x}_{t+1|t} &= (p_{11} - p_{01})x_{t|t-1} + p_{01} \\
&= (p_{11} - p_{01})\hat{x}_{t|t-1} + p_{01} + (p_{11} - p_{01})(a_1 - a_0)\Sigma_{t|t-1} \\
&\quad \times ((a_1 - a_0)^2 \Sigma_{t|t-1} + \sigma_\eta^2)^{-1}(y_t - (a_1 - a_0)\hat{x}_{t|t-1} - a_0) \\
&= (p_{11} - p_{01})\hat{x}_{t|t-1} + p_{01} + k_t(y_t - (a_1 - a_0)\hat{x}_{t|t-1} - a_0),
\end{aligned} \tag{15}$$

where

$$k_t = (p_{11} - p_{01})(a_1 - a_0)\Sigma_{t|t-1}((a_1 - a_0)^2 \Sigma_{t|t-1} + \sigma_\eta^2)^{-1}. \tag{16}$$

The evolution of the one period ahead prediction error, $\Sigma_{t|t-1}$ is given as:

$$\begin{aligned}
\Sigma_{t+1|t} &= E[(x_{t+1} - \hat{x}_{t+1|t})^2 | X_t] \\
&= (p_{11} - p_{01})^2 \Sigma_{t|t-1} + \sigma_{\varepsilon,t+1}^2 \\
&\quad - (p_{11} - p_{01})^2 (a_1 - a_0)^2 (\sigma_\eta^2 + (a_1 - a_0)^2 \Sigma_{t|t-1})^{-1} \Sigma_{t|t-1}^2 \\
&= (p_{11} - p_{01})^2 \Sigma_{t|t-1} + \sigma_{\varepsilon,t+1}^2 - k_t (p_{11} - p_{01})(a_1 - a_0)\Sigma_{t|t-1}.
\end{aligned} \tag{17}$$

Equation (8) shows that $\sigma_{\varepsilon,t+1}$ depends on $E(x_{t+1}|x_t)$ and x_t. Thus, $\sigma_{\varepsilon,t+1}$ is time-varying and has to be updated at each iteration. The best linear estimator of $\sigma_{\varepsilon,t+1}$ can be obtained by replacing $E(x_{t+1}|x_t)$ and x_{t-1} by $\hat{x}_{t+1|t}$ and $\hat{x}_{t|t}$ in (5). Therefore,

$$\sigma_{\varepsilon,t+1}^2 = \hat{x}_{t+1|t} - (p_{11}^2 - p_{01}^2)\hat{x}_{t|t} - p_{01}^2. \tag{18}$$

Given initial values, x_0, $\Sigma_{1|0}$ and $\sigma_{\varepsilon,0}^2$, equation (14), (15), (17), and (18) reccursively yield a series of the linear optimal estimates given the observation of y available in the previous time period, $\hat{x}_{t|t-1}$, and prior variance, $\Sigma_{t|t-1}$. The filter can be interpreted as follows. At the beginning of time t, economic agents know $\hat{x}_{t|t-1}$ and $\Sigma_{t|t-1}$. When observation y_t becomes available, they update their estimation to yield $\hat{x}_{t|t}$, incorporating the prediction error v_t. Given this best estimate of x_t available at time t, the one period ahead estimation, $\hat{x}_{t+1|t}$, is formed using the trajectory of $\{x_t\}$ given by (7). The variance of the transition error, $\sigma_{\varepsilon,t+1}$, is estimated, preceding the updating of the variance of one-period ahead estimation error, $\Sigma_{t+1|t}$.

The recursive equations (15) and (17) can be obtained also by applying a standard Kalman filter formulae on the state space equations (6) and (7). assuming that x_t. ε_t, are Gaussian with the variance of ε_t same as the non-Gaussian case for all t.

2.3 Properties of the Optimal Linear Filter

The properties of the optimal linear filter applied to a non-Gaussian system can be most efficiently presented in comparison to that applied to a Gaussian system. Anderson and Moore (1979) indicated that if a Kalman filter is applied to a Gaussian process, it yields a series of optimal unbiased estimators in the sense of minimizing MSE among all forms of estimators (not necessarily linear). Moreover, $x_{t|t-1}$ and $\Sigma_{t|t-1}$ are the conditional mean and variance of x_t given Y_{t-1}, respectively. Therefore, when two series $\{x_t\}$ and $\{y_t\}$ are jointly normally distributed, the Kalman filter updates the conditional distribution of the hidden state since a conditional normal distribution, as well as normal distribution, can be completely characterized by its mean and covariance matrix. Note that if x_t and Y_t are jointly normally distributed, $E(x_t|Y_t) = E(x_t) + \Sigma_{xy}\Sigma_{yy}^{-1}(Y_t - E(Y_t))$. Therefore, replacing $E^*(x_{t+1}|Y_t)$ by $E(x_{t+1}|Y_t)$ in the derivation of the Kalman filter in section (2.2) and carrying out the whole derivation will yield the Kalman filter which shares the same form. The above exposition also proves that we can obtain the optimal linear filter obtained in subsection 2.2 by simply applying the standard Kalman filter formulae on the Gaussian analog of a non-Gaussian system[1]. However, if a system fails to be Gaussian, as in our model, the resulting estimator series is no longer optimal among all forms of estimators. This is because the nonlinear system, (6), (7), and (8), cannot be fully described by their first and second moments, whereas the Kalman filter utilizes only the first and second moments of the series.

It is technically possible to obtain a filter for a non-Gaussian system that yields a series of optimal MSE estimators among all forms of estimators. At each iteration, given $P(x_t|Y_{t-1})$ available in the previous time period, calculate the conditional density $P(x_t|Y_{t1})$ by incorporating the new observation y_t. Then, calculate $\hat{x}_{t|t}$ as follows:

$$\hat{x}_{t|t} = \int_{-\infty}^{+\infty} x_t P(x_t|Y_t) dx_t$$

The one-period ahead estimator $x_{t+1|t}$ can be obtained by using equation (7). The conditional mean at each time period can be obtained in the similar manner. This is the approach taken by Hamilton (1989). His basic filter updates the conditional joint density, $P(x_t, x_{t-1}, \cdots, x_{t-r+1}|Y_t)$, for each t, where r is the order of AR error process of the measurement equation specified in his model. Parameters are estimated by maximizing the likelihood function constructed as a byproduct of the filter.

Although the Kalman filter applied on the non-Gaussian system is not universally optimal, it has closed form solution and is easy to implement. Moreover, it is said that the Kalman filter is more robust with respect to errors in variables. In the following two sections, we compare the performances of the linear optimal filter and the nonlinear filter.

[1] In other words, we are considering the processes, $\{\tilde{x}_t\}$, $\{\tilde{y}_t\}$, and $\{\tilde{\varepsilon}_t\}$, whose first and second moments are the same as those of the corresponding original series, $\{x_t\}$, $\{y_t\}$, and $\{\varepsilon\}$. Then, the analog series $\{\tilde{x}_t\}$, $\{\tilde{y}_t\}$, and $\{\tilde{\varepsilon}_t\}$ is also described by the system, (6), (7), and (8).

3 Determination of Peaks and Troughs

In this section, we investigate the performance of the Kalman filter, using the quarterly US real GNP. In order to compare the results with those of Hamilton's, we use the observations from 1951:II to 1984:IV[2]. Our objective is to assign a probability of the economy belonging to a specific state at each time t and to determine peaks and troughs of the GNP series. These results are compared with those of Hamilton and NBER.

We applied the Kalman filter derived in section 2.2 twice: once with the parameter values used by Hamilton (KF(1))[3] and then with crude a priori parameter estimates (KF(2)). The use of the crude parameter estimates makes the procedures simpler and easier. Table 2 presents the parameter values used for each filter.

The Kalman filter generates series of both prior and posterior estimates of probabilities, $P(x_t|Y_{t-1})$ and $P(x_t|Y_t)$, respectively. Note that $\hat{x}_{t|t}$ equals to $P(x_t = 1|Y_t)$ by construction. Hamilton's basic filter generates a series of the optimal estimates of $P(x_t|Y_t)$. To be consistent with Hamilton's basic filter, we use the estimates of $P(x_t = 1|Y_t)$ to establish the historical business cycle dates. Following Hamilton, for each t we determine the state to be in recovery if $P(x_t = 1|Y_t) > 0.5$ and in recession otherwise. The results and the parameter values used are found in Table 1 and Table 2, respectively. Also, figure 2 reports $P(x_t = 0|Y_t)$ estimated by the nonlinear filter, KF(1), and KF(2).

---

[2]The same data set was used by Hamilton (1989).

[3]The estimate of σ_η is not available from Hamilton's results since his model assumes η_t to follow AR(4) process with white noise error term, say v_t. The value for σ_η is estimated by simulation from the standard deviation of v_t.

Detecting business cycle turning points by Kalman filter has four differences compared to Hamilton's nonlinear filter. First, Kalman filter is a linear approximation of the optimal nonlinear filter. Second, Kalman filter utilizes only the information up to time t, whereas Hamilton uses a full-time smoother which generates the series of expected probabilities of each state given all the information up to time T, $P(x_t|Y_T)$. Moreover, the model we assume is simpler than Hamilton's. Our model is described by equations (6) and (7), whereas Hamilton, in addition, assumes that the disturbance term η_t follows AR(4) process[4]. Finally, a priori parameter estimates obtained by crude methods are used for the KF(2).

Despite these simplifications, the dates determined by the Kalman filter both with Hamilton's and crude parameter estimates adhere to those of NBER surprisingly well. Table 1 shows that the dating of peaks and troughs determined by the Kalman filters correspond to those of NBER's except the trough in 1961 and 1975, and the peaks experienced in 1960 and 1969. Given these results, our parsimonious model and the crude but easy a priori estimation of parameters turn out to be advantages to the use of the Kalman filter, enhancing time-efficiency.

[4]It is possible to incorporate AR(4) disturbances and to obtain smoothed estimates in the Kalman filter (See Anderson and Moore (1979)). However, to pursue simplicity, we choose to use the parsimonious model. If US GNP series is truly generated by the system described by Hamilton's model, our model would be a misspecification. The effect of this type of misspecifiction is tested by simulation in the next section.

4 Simulations

In this section, we compare the Kalman filter and Hamilton's basic nonlinear filters in more general contexts. Econometricians face unexpected disturbances or structural changes in a series when they attempt to determine the hidden state of the series using the parameter values estimated from the past data. In such a situation. performances of the Kalman filter and nonlinear filter not only with accurate a priori parameter estimates but also with inaccurate estimates must be taken into account. In this section, performances of the two filters are examined in various situations described as follows. (a) Accurate a priori estimates are available, and there is no model misspecification. (b) Misspecification: the simplest mean-switching model is specified when the true series has serially correlated disturbance term. (c) The series is subject to occasional jumps. (d) The series experiences a sudden and permanent shift in the mean.

In order to compare performances of the Kalman filter and nonlinear filter. 1000 series of size T=134 is generated with index series x_t which is predetermined according to NBER business cycle dates. The parameters (p_{11}, p_{00}, p_1) which is necessary to run the Kalman and nonlinear filters are obtained by a crude method from the index series. The two filters are applied on each series to obtain the following statistics.

$$MSE = E[\| x_t - \hat{x}_{t|t} \|^2]$$

and N, the number of times that $P(x_t = 1|Y_t)$ fails to predict the true state of the economy if the user determine the true state to be 1 if $P(x_t = 1|Y_t) > 0.5$. and 0 otherwise. MSE is calculated after $\hat{x}_{t|t}$ is clipped at 1 and 0. These statistics are collected for 1000 simulations and their means and standard deviations are calculated. This process was done for each of situations, (a),(b),(c), and (d).

4.1 Correct Estimates and Correct Model Specification

In the first set of simulations, we assume that the series is truly generated based on the Markov mean-switching model in equation (6) and (7). We also assume that researchers correctly specify the model and have accurate parameter estimates. To generate the series following equation (6), we used the Hamilton's estimation for the value of parameters $(a_1, a_0, p_{11}, p_{00})$, the crude a priori estimate for p_1, and σ_η estimated by simulation (See Footnote 3). The same parameter values are used to run each filter. Therefore, the simulations are done assuming that researchers have accurate estimates of the parameters a priori. The results indicate that the mean of MSE of Kalman filter is more than that of nonlinear filter. However, the mean of the number of false inference differ only by 1.733 ($\bar{N} = 14.7240$ for the Kalman filter and $\bar{N} = 12.9910$ for the nonlinear filter). The results is listed in Table 3:(a).

4.2 Misspecification

Next, we examine the two filters' performances in the case of misspecification. Suppose that the true series is generated by the model specified by Hamilton (1989):

$$y_t = (a_1 - a_0)x_t + a_0 + \epsilon_t$$

$$\epsilon_t = \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + \phi_3\epsilon_{t-3} + \phi_4\epsilon_{t-4} + v_t,$$

where x_t is a Markov switching process. Assume that researchers erroneously specifies the model ignoring AR(4) disturbance process as:

$$y_t = (a_1 - a_0)x_t + a_0 + \eta_t.$$

This is the model we used to determine the business cycle dates in section 3. Our purpose in this subsection is to determine the effect of this type of misspecification. The results of this simulation has an important implication for the investigation in section 3, since the drastic simplification made in applying the Kalman filter involves some risk of erroneously specifying a model.

In generating the series, we used parameter values $(a_1, a_0, \phi_1, \phi_2, \phi_3, \phi_4, \sigma_v)$ obtained by Hamilton's maximum likelihood estimation. Three filters are run on each of 1000 sample paths created with this parameter set: Hamilton's basic filter with the correctly specified model, Hamilton's basic filter with the erroneously specified model, and the Kalman filter with an erroneously specified model. In running these filters, we assume that an accurate estimate of the variance of η_t is available.

The results are shown in Table 3:(b). The comparison between the nonlinear filter reveals that the misspecification has little effect on the performances of the nonlinear filter with the misspecified model. The mean of MSE and N of the Kalman filter are greater than those of the nonlinear filter by 0.0140 and 1.9130, respectively. These results justify the use of the Kalman filter despite its parsimonious model and suboptimality.

4.3 Occasional Jumps

Next, the effects of occasional jumps on performances of the two filters are investigated. The jump series ω_t is generated as follows.

$$\omega_t = \begin{cases} z_t & \text{if } \mid z_t \mid > \delta \cdot \sigma_z \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

where $z_t \sim N(0, \sigma_z)$. Therefore, the whole disturbance is a sum of η_t and ω_t. We assume that econometricians have an accurate estimate of σ_η^2, but cannot expect the jumps. Thus, they have an inaccurate estimate for the variance of the disturbance term. We run both the nonlinear filter and the Kalman filter for various values of σ_z and δ. We present the results in Figure 1:(c).

As the magnitude and frequency of the jumps becomes larger, the means of MSE and N increase for the both filters. MSE and N of the Kalman filter exceed those of the nonlinear filter for a series with jumps of low magnitude and low frequency. However, the maximum difference is 0.0125 and 1.7230, respectively. Moreover, the difference in performances between the two filters become smaller as jumps become more frequent and larger. Further, the Kalman filter outperforms the nonlinear one for larger and more frequent jumps.

4.4 Sudden Shifts in the Mean

We also investigated effects of an unexpected permanent shift in the mean of the series. Researchers are assumed to possess and use an a priori estimate of a_1 and a_0 which is accurate before this structural change. Being unable to anticipate the structural change, they mistakenly use inaccurate estimates. Figure 1:(d) shows the results. Both the mean of MSE and N of the Kalman filter are larger than those of the nonlinear filter for modest shifts. However, MSE and N of the Kalman filter never exceed those of the nonlinear filter by more than 0.0127 and 2.5330, respectively. The larger the magnitude of shifts, the larger the means of MSE and N become both for the Kalman filter and the nonlinear filter. However, the degree at which the nonlinear filter outperforms the Kalman filter becomes smaller as the shift becomes larger. Especially, in the case of downward shifts, the Kalman filter eventually outperforms the nonlinear filter. This asymmetry is due to the fact that the preassigned index series (according to NBER business cycle dating) contains larger number of 1's than

0's. Out of 134 period, 109 periods are assigned the value of 1 and 25 periods are assigned 0. Simulations for different preassigned index series are done in appendix.

The common characteristics for all types of simulations is that the standard deviations of MSE and N are smaller for the Kalman filter.

5 Conclusions

We have investigated the possibility of the use of the Kalman filter for models with Markov regime shifts. Our objective is to present the Kalman filter as an alternative algorithm to determine the hidden state of a series. The Kalman filter formulae are derived for this type of model, its theoretical properties are examined in comparison to those of the nonlinear filter. The Kalman filter applied to this type of model is only optimal within the class of linear estimators. However, it has a closed form solution and easy to implement. Therefore, if the Kalman filter approximates Hamilton's optimal filter reasonably well, its use can be justified.

To examine its performance, the Kalman filter is applied to US real GNP series, and the results are compared with those of Hamilton and NBER. Despite its theoretical suboptimality, we found the Kalman filter with a reasonably good a priori parameter establishes the business cycle dates as well as NBER's and Hamilton's dating scheme.

Next, we conducted simulations to examine the performances of the Kalman filter in more general contexts. The results of simulations confirm the ability of the Kalman filter to indicate the hidden state of a series. When the model is correctly specified with the accurate a priori estimates, the mean of N of the Kalman filter is greater than that of the nonlinear filter only by 1.733 (out of 134). Performances of both the Kalman and the nonlinear filters are negatively affected by misspesification of the model, unexpected occasional jumps, and permanent shifts in the mean of the series.

However, the mean of MSE or N of the Kalman filter differs from those of the nonlinear filter by less than 2. Moreover, the larger the magnitudes of misspecification, jumps, or shifts, the better approximation the Kalman filter becomes to the nonlinear filter, and eventually outperforms the latter. In sum, the Kalman filter is found to be more robust to shocks (or structural changes) in the economy which are not expected by researchers.

In conclusion, its simplicity and easy application, coupled with its reasonable performance, justifies the use of the Kalman filter in making probabilistic inference about the hidden state of a series.

A Appendix

In section 4.4, we observe that the performances of both the Kalman and nonlinear filters are asymmetric between downward and upward shifts. In the appendix, we investigate the effect of different index series on the performance of both filters in determining the turning points. We prepare four different index series. For each index series and different degrees of shifts, 1000 series of size $T = 100$ are generated. We apply both the Kalman and nonlinear filters on each series and calculate MSE and N. We assume that the unexpected shift occurs at $t = 51$. We present the results in Figure 3.

When the index series contains larger number of the states of higher mean than that of lower mean, both the Kalman filter and nonlinear filter adjust better to the unexpected upward shifts than to downward shifts with the same magnitude, and vice versa. Moreover, the Kalman filter adjust better than the nonlinear filter to the shift of the opposite to the dominant states as the magnitude of the shifts become larger.

Table 1

Determination of Peaks and Troughs

Peaks

NBER	Hamilton	KF(1)	KF(2)
1953:III	1953:III	1953:III	1953:III
1957:III	1957:I	1957:III	1957:III
1960:II	1960:II	1960:III	1960:III
1969:IV	1969:III	1969:III	1969:III
1973:IV	1974:I	1973:IV	1974:I
1980:I	1979:II	1980:I	1980:I
1981:III	1981:II	1981:III	1981:III

Troughs

NBER	Hamilton	KF(1)	KF(2)
1954:II	1954:II	1954:II	1954:II
1958:II	1958:I	1958:II	1958:II
1961:I	1960:IV	1960:IV	1960:IV
1970:IV	1970:IV	1970:IV	1970:IV
1975:I	1975:I	1975:II	1975:II
1980:III	1980:III	1980:III	1980:III
1982:IV	1982:IV	1982:IV	1982:IV

Hamilton's Full-Sample Smoother (Reported in Hamilton's 1989 paper)

Kalman Filter(1) uses Hamilton's parameter values

Kalman Filter(2) uses crude parameter estimates

Table 2

Parameter Values

Parameter	Hamilton's	Crude Estimates
a_1	1.1643	1.1007
a_0	-0.3577	-0.3408
p_{11}	0.9049	0.9394
p_{00}	0.7550	0.8235
p_1	–	0.7463
σ_η	*0.8195	0.8467
ϕ_1	0.0140	–
ϕ_2	-0.0580	–
ϕ_3	-0.2470	–
ϕ_4	-0.2130	–
σ_v	0.7690	–

* The value is obtained by simulation.

Table 3 (a)

Accurate a Priori Estimates

No Model Misspecification

True System/Model: $y_t = (a_1 - a_0)x_t + a_0 + \varepsilon_t$

	MSE		N	
	mean	std	mean	std
Kalman Filter	0.0844	0.0117	14.7240	3.2460
Nonlinear Filter	0.0714	0.0157	12.9910	3.4060

Table 3 (b)

Model Misspecification

True System: $y_t = (a_1 - a_0)x_t + a_0 + \varepsilon_t$

where $\varepsilon_t = \phi_1\varepsilon_{t-1} + \phi_2\varepsilon_{t-2} + \phi_3\varepsilon_{t-3} + \phi_4\varepsilon_{t-4}$

Model: $y_t = (a_1 - a_0)x_t + a_0 + \varepsilon_t$

	MSE		N	
	mean	std	mean	std
Kalman Filter	0.0831	0.0118	14.4640	3.0988
Nonlinear Filter	0.0722	0.0172	13.0640	3.5814
Nonlinear Filter (AR(4))	0.0703	0.0166	12.7110	3.4493

$\phi_1 = 0.014$, $\phi_2 = -0.058$, $\phi_3 = -0.247$, $\phi_4 = -0.213$, $\sigma_v = 0.7690$

Figure 1
(c) Occasional Jumps

True System:

$$y_t = (a_1 - a_0)x_t + a_0 + \varepsilon_t + \omega_t$$

$$\omega_t = \begin{cases} z_t & \text{if } |z_t| > \delta \cdot \sigma^2 \\ 0 & \text{otherwise,} \end{cases} \qquad \text{where } z_t \sim N(0, \sigma_z^2)$$

Model:

$$y_t = (a_1 - a_0)x_t + a_0 + \varepsilon_t$$

——— Kalman Filter

- - - - - Nonlinear Filter

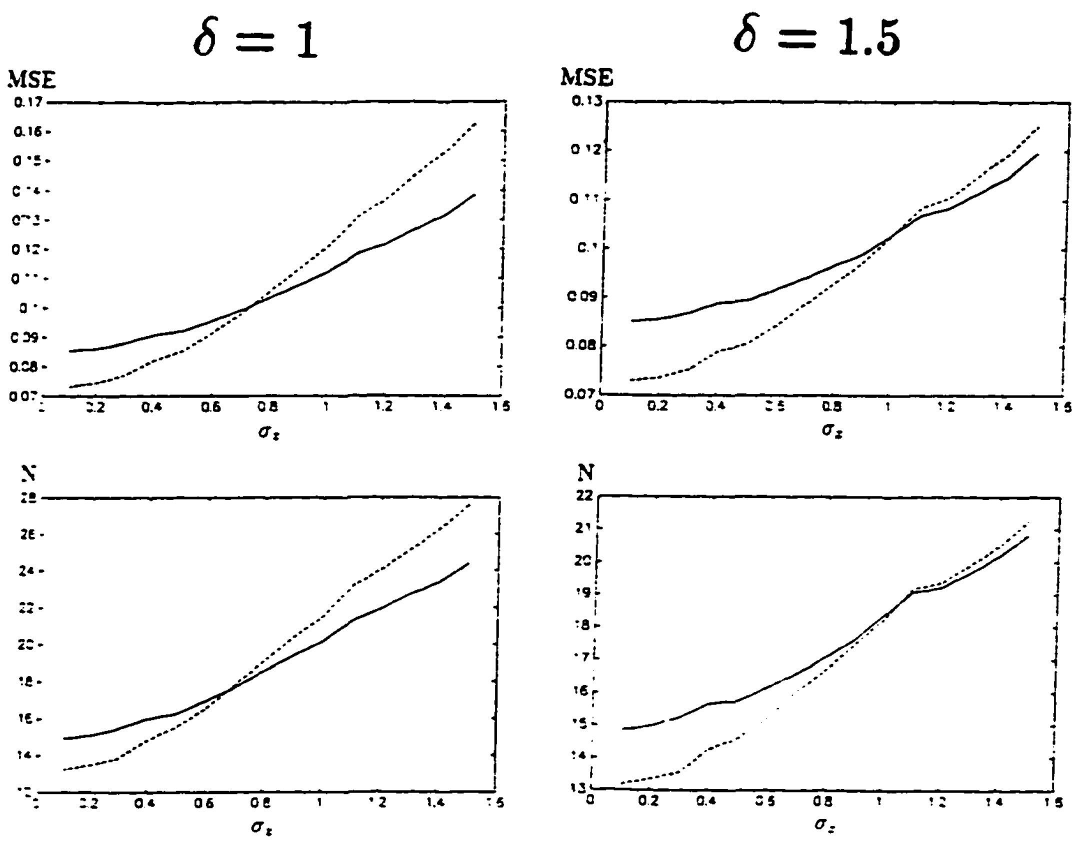


Figure 1 (Continued)
(c) Occasional Jumps

True System:

$$y_t = (a_1 - a_0)x_t + a_0 + \varepsilon_t + \omega_t$$

$$\omega_t = \begin{cases} z_t & \text{if } |z_t| > \delta \cdot \sigma^2 \\ 0 & \text{otherwise,} \qquad \text{where } z_t \sim N(0, \sigma_z^2) \end{cases}$$

Model:

$$y_t = (a_1 - a_0)x_t + a_0 + \varepsilon_t$$

——— **Kalman Filter**

- - - - - **Nonlinear Filter**

$\delta = 2.0$

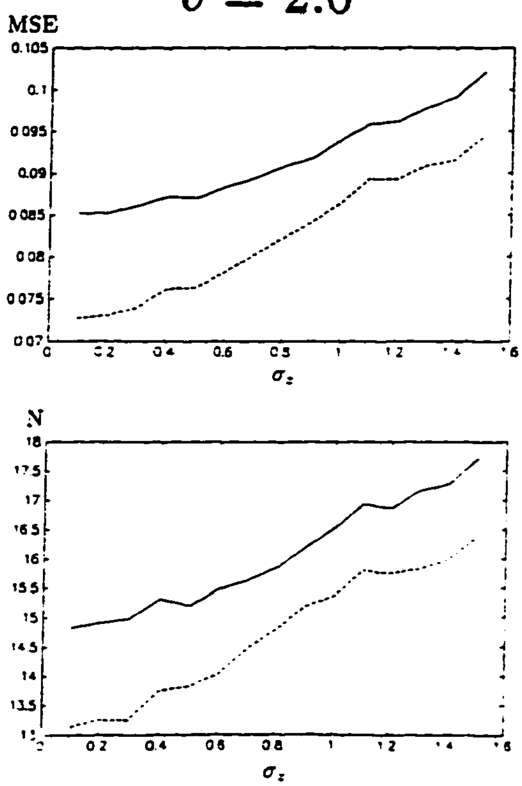


Figure 1
(d) Sudden Shift in the Mean

True System:

$$y_t = (a_1 - a_0)x_t + a_0 + \mu_t + \varepsilon_t$$

$$\mu_t = \begin{cases} \mu & \text{if } t >= \tau \\ 0 & \text{otherwise} \end{cases}$$

Model:

$$y_t = (a_1 - a_0)x_t + a_0 + \varepsilon_t$$

——— **Kalman Filter**

- - - - - **Nonlinear Filter**

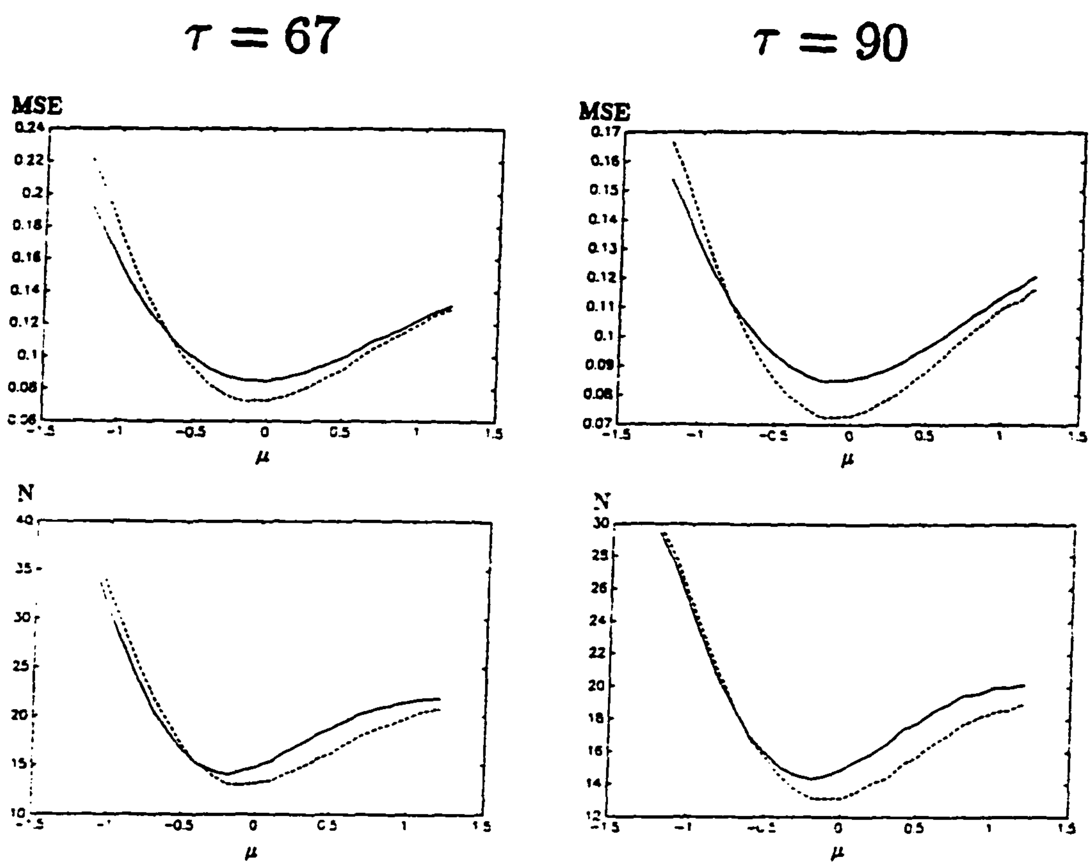


Figure 2
Estimation of $P(x_t = 0|Y_t)$

Nonlinear Filter

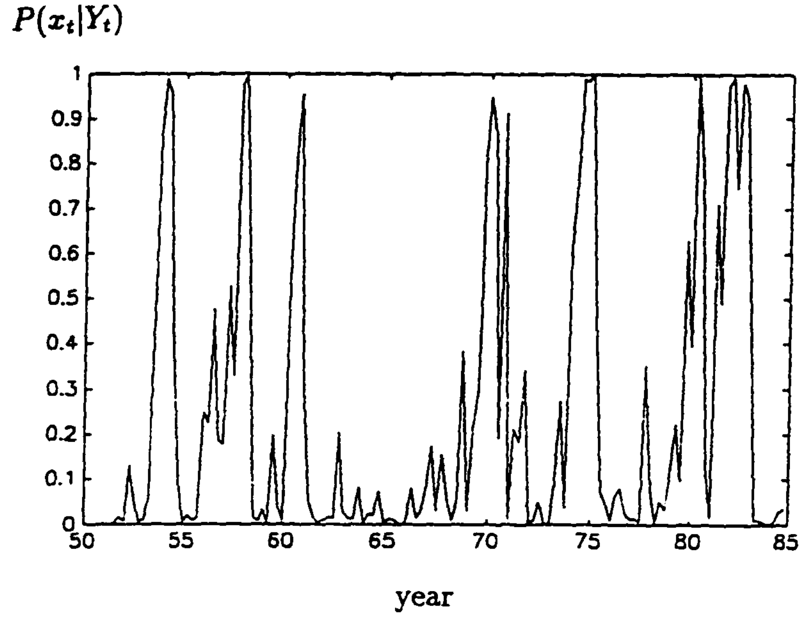
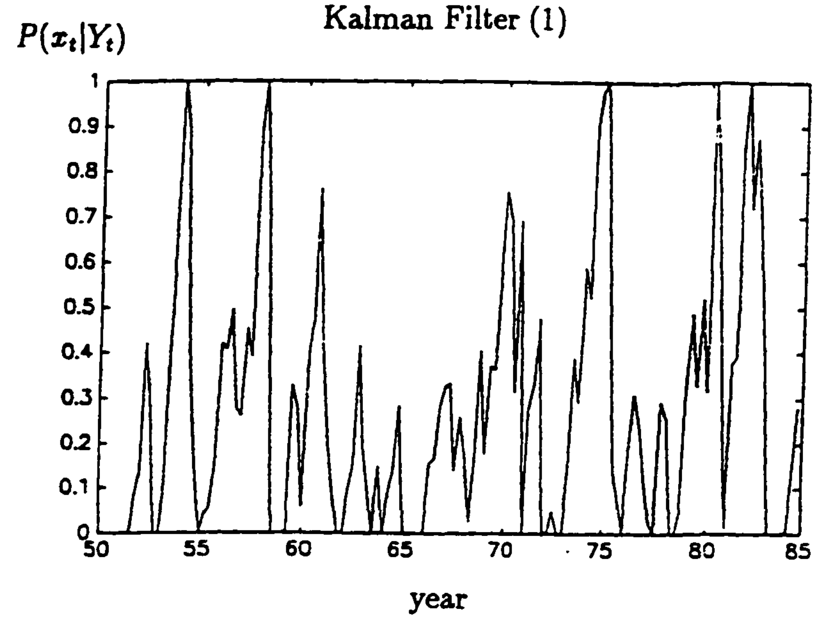


Figure 2 (Continued)
Estimation of $P(x_t = 0|Y_t)$

Kalman Filter (1)



Kalman Filter (2)

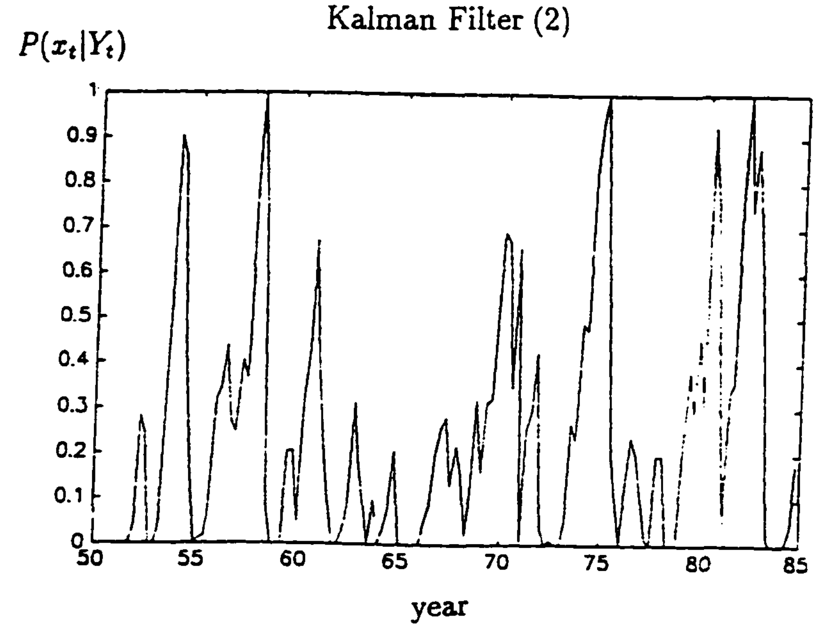
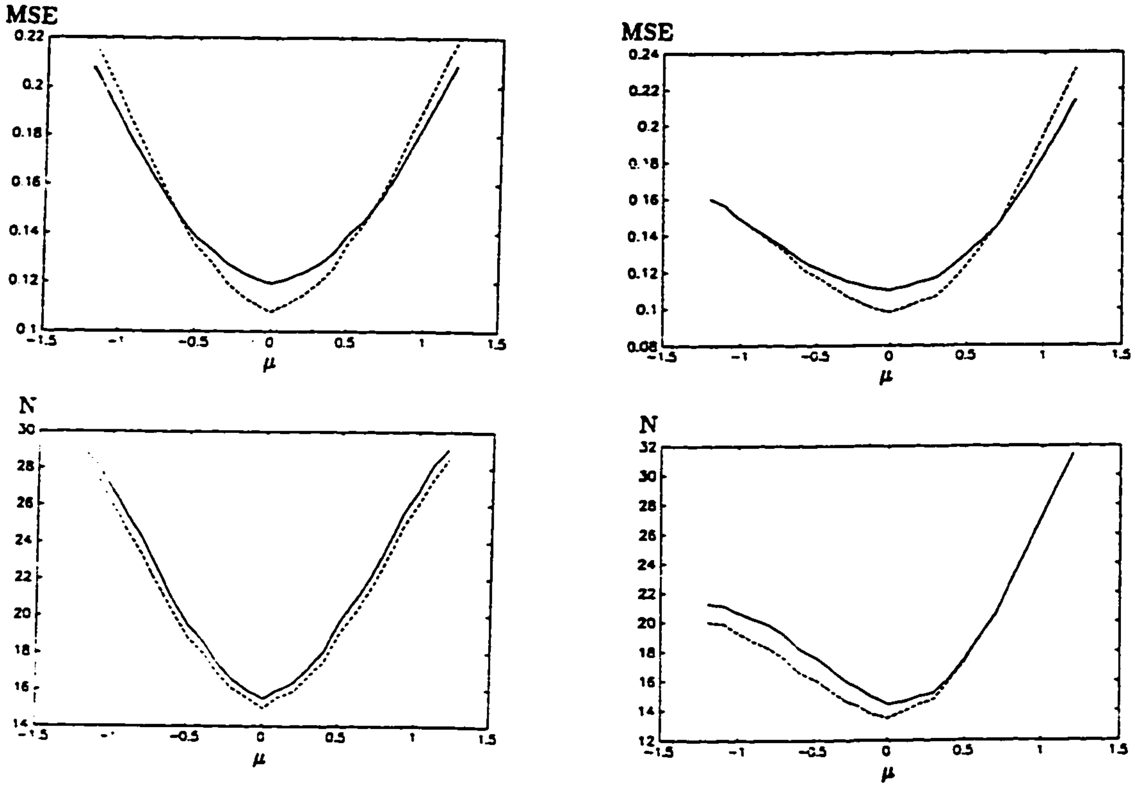


Figure 3

(i)
$p_{11} = 0.8$
$p_{00} = 0.8163$
$p_1 = 0.5$

(ii)
$p_{11} = 0.6667$
$p_{00} = 0.8696$
$p_1 = 0.3$



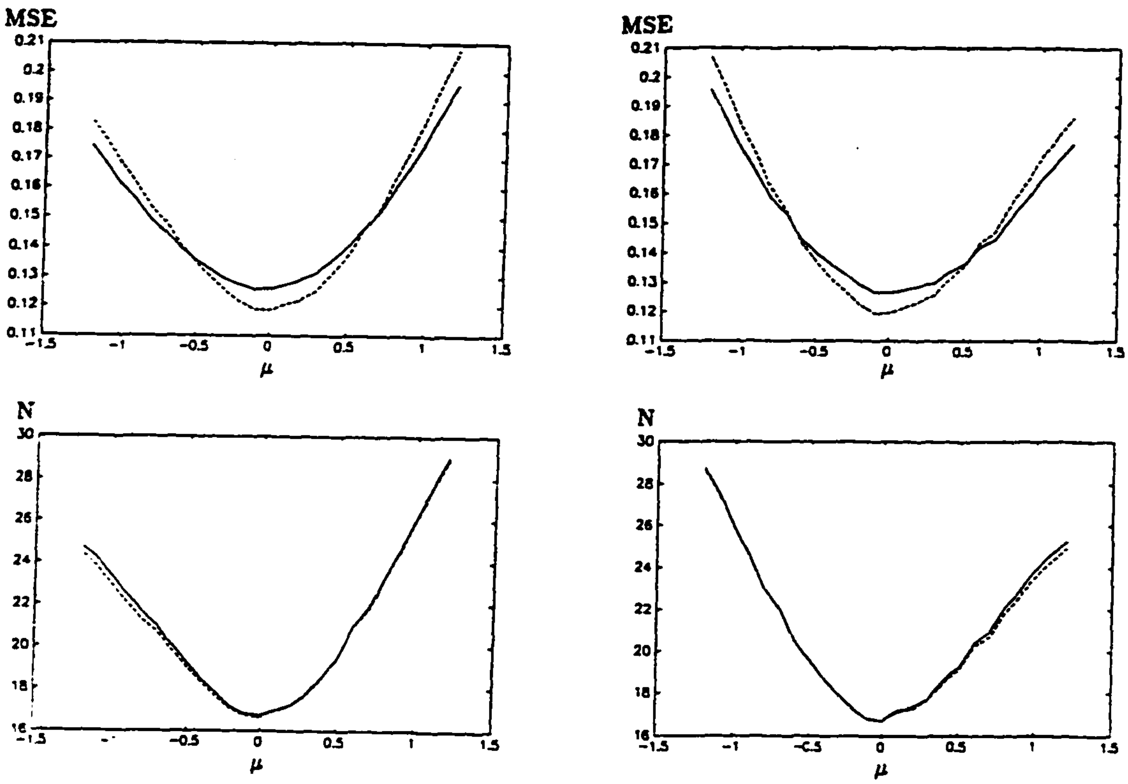
——— Kalman filter

- - - - - Nonlinear Filter

Figure 3 (Continued)

(iii)
$p_{11} = 0.5$
$p_{00} = 0.6780$
$p_1 = 0.4$

(iv)
$p_{11} = 0.6667$
$p_{00} = 0.5128$
$p_1 = 0.6$



——— Kalman Filter

- - - - - Nonlinear Filter

References

[1] James Albert and Siddhartha Chib. Bayes inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts. mimeo, Bowling Green State University and Washington University, 1992.

[2] Brian D. O. Anderson and John B. Moore. *Optimal Filtering.* Prentice Hall. Englewood Cliffs, N.J., 1979.

[3] James D. Hamilton. Rational-expectations econometric analysis of changes in regime. *Journal of Economic Dynamics and Control*, 12:385–423, 1988.

[4] James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384, 1989.

[5] James D. Hamilton. Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45:39–70, 1990.

[6] N. V. Krichagina, R. S. Lipster, and E. Y. Rubinovich. Kalman filter for markov processes. In R. S. Lipster N. V. Krylov and A. A. Novikov, editors, *Steklov Seminar: Statistics and Control of Stochastic Processes*, pages 197–213. Optimization Software, Inc., New York, 1984.

[7] Robert E. McCulloch and Ruey S. Tsay. Statistical analysis of macroeconomic time series via markov switching models. mimeo, University of Chicago, 1992.