Journal of Econometrics 60 (1994) 1–22. North-Holland

Dynamic linear models with Markov-switching

Chang-Jin Kim*

Korea University, Seoul, 136-701, Korea
York University, North York, Ont. M3J 1P3, Canada

Received June 1991, final version received June 1992

In this paper, Hamilton's (1988, 1989) Markov-switching model is extended to a general state–space model. This paper also complements Shumway and Stoffer's (1991) dynamic linear models with switching, by introducing dependence in the switching process, and by allowing switching in both measurement and transition equations. Building upon ideas in Hamilton (1989), Cosslett and Lee (1985), and Harrison and Stevens (1976), a basic filtering and smoothing algorithm is presented. The algorithm and the maximum likelihood estimation procedure is applied in estimating Lam's (1990) generalized Hamilton model with a general autoregressive component. The estimation results show that the approximation employed in this paper performs an excellent job, with a considerable advantage in computation time.

A state–space representation is a very flexible form, and the approach taken in this paper therefore allows a broad class of models to be estimated that could not be handled before. In addition, the algorithm for calculating smoothed inferences on the unobserved states is a vastly more efficient one than that in the literature.

Key words: State–space model; Markov-switching; Basic filtering; Smoothing; Generalized Hamilton model

1. Introduction

Model instability is sometimes defined as a switch in a regression equation from one subsample period (or regime) to another. An F-test proposed by Chow (1960) may be applied in testing for structural changes for the case where the dates that separate subsamples are known.

Correspondence to: Chang-Jin Kim, Department of Economics, Korea University, Anam-dong, Seongbuk-ku, Seoul, 136-701, Korea.

* I have benefited substantially from the comments and suggestions of James Hamilton. I am also indebted to two anonymous referees, Pok-Sang Lam, Myung-Jig Kim, and the participants in seminars at York University and the 1992 North American Winter Meeting of the Econometric Society for their comments on earlier drafts of this paper. Remaining shortcomings and any errors are my responsibility.

In a lot of cases, however, researchers may have little information on the dates at which parameters change, and thus need to make inferences about the turning points as well as on the significance of parameter shifts. Quandt (1958, 1960), Farley and Hinich (1970), Kim and Siegmund (1989), and Chu (1989), for example, considered models in which they permitted at most one switch in the data series with an unknown turning point. Quandt (1972), Goldfeld and Quandt[1] (1973), Brown et al. (1975), Ploberger et al. (1989), and Kim and Maddala (1991) considered models that allow for more than one switch. For other tests of structural change with unknown change point, see Andrews (1990) and the references therein. Also see Wecker (1979), Sclove (1983), and Neftci (1984) for other related models of predicting turning points. An interesting and important aspect of these models in that the time at which a structural change occurs is endogenous to the model.

Recently, Hamilton's (1988, 1989) state-dependent Markov-switching model drew a lot of attention in modeling structural changes in dependent data. His model can be viewed as an extension of Goldfeld and Quandt's (1973) model to the important case of structural changes in the parameters of an autoregressive process. Applications of his model are numerous. For example, Garcia and Perron (1991) applied Hamilton's approach to modeling structural changes in U.S. *ex post* real interest rates and inflation series, Engel and Hamilton (1990) applied it to the exchange rate market, and Cecchetti, Lam, and Mark (1990) applied it to modeling the stock market.

The purpose of this paper is to extend Hamilton's (1988, 1989) Markov-switching model to the state-space representation of a general dynamic linear model, which includes autoregressive integrated moving average (ARIMA) and classical regression models as special cases. [For applications of state-space models in econometrics and time series analysis, see Watson and Engle (1983), Engle and Watson (1985), and Harvey (1985).] So far as the structural changes in the state–space models are concerned, the multi-regime model of Harrison and Stevens (1976) also allows for the regime to switch endogenously, according to a stationary Markov chain. However, in this case it is assumed that the parameters in different regimes are known and that there are known transition probabilities from one regime to the next [Harvey (1991, p. 348)]. One of the most recent applications of the switching approach in state–space models includes Shumway and Stoffer (1991), who considered a dynamic linear model with measurement matrices that switch endogenously according to an independent random process. Their approach was motivated primarily by the problem of tracking a large number of possible moving targets using a vector of sensors. [For surveys of related issues, see Shumway and Stoffer (1991) and Bar-Shalom (1978).] By restricting the switching to the measurement equations and

[1] Quandt's (1972) model assumes that the probability of a switch does not depend upon what regime is in effect, while Goldfeld and Qunadt's (1973) model explicitly allows for such a dependence by introducing a Markov-switching.

assuming regimes to be serially independent, they simplified the Kalman filtering recursions considerably. In economics, however, the shifts in regimes may not be independent. For example, once the economy is in a regime (a recession or a boom), the regime may persist for a while. In this case, modeling switches between regimes according to an independent random process may not be reasonable.

Building upon ideas introduced in Hamilton (1988, 1989), Cosslett and Lee (1985), and Harrison and Stevens (1976), basic filtering and smoothing algorithms for a Markov-switching state–space model are presented. Maximum likelihood estimation of the unknown parameters of the model is also considered. The model and the algorithm introduced are applied to the estimation of the Hamilton model with a general autoregressive component proposed by Lam (1990). The state–space representation of the generalized Hamilton model is a special case of the model presented in this paper.

The state–space representation is a very flexible form, and the approach in this paper allows a broad class of models to be estimated that could not be handled before. For example, with the state–space formulation one can handle regime shifts with moving average (MA) terms, in which the MA coefficients switch according to a Markov process. In addition to developing a more flexible algorithm in dealing with regime shifts, one other contribution of this paper is a vastly more efficient algorithm for calculating smoothed inferences on the unobserved states than that in the literature.

The plan of the paper is as follows. Section 2 introduces the model. Maximum likelihood estimation is described, along with basic filtering and smoothing algorithms. We employ approximations to optimal filtering similar to those proposed by Harrison and Stevens (1976). Section 3 applies the approach to Lam's (1990) generalized Hamilton model, and compares the results to Lam's exact maximum likelihood estimates. Section 4 concludes the paper.

2. The model, filtering, and smoothing

2.1. Specification of the model

Consider the following state–space representation of a dynamic linear model with switching in both measurement and transition equations:

$$y_t = F_{S_t} x_t + \beta_{S_t} z_t + e_t, \tag{2.1}$$

$$x_t = A_{S_t} x_{t-1} + \gamma_{S_t} z_t + G_{S_t} v_t, \tag{2.2}$$

and

$$\begin{pmatrix} e_t \\ v_t \end{pmatrix} \sim \mathrm{N}\left(0, \begin{pmatrix} R & 0 \\ 0 & Q \end{pmatrix}\right), \tag{2.3}$$

where the transition equation (2.2) describes the evolution of a $J \times 1$ vector $\boldsymbol{x}_t$ of characteristics of a physical process in response to a $K \times 1$ vector z_t of weakly exogenous or lagged dependent variables and an $L \times 1$ vector of disturbances $\boldsymbol{v}_t$. The measurement equation (2.1) describes the relation between the unobserved state $\boldsymbol{x}_t$ and an $N \times 1$ vector of measurements y_t. Here, it is assumed that the parameters of the model $F_{S_t}, \beta_{S_t}, A_{S_t}, \gamma_{S_t}$, and G_{S_t} are dependent upon a state variable S_t.

To make the above model tractable, the econometrician must specify a stochastic process for the state variable S_t. Hamilton (1988, 1989) proposes to model S_t as the outcome of an unobserved discrete-time, discrete-state Markov process, building on an original idea by Goldfeld and Quandt (1973). Assuming an M-state, first-order Markov process, we can write the transition probability matrix in the following way:

$$p = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{pmatrix}, \tag{2.4}$$

where $p_{ij} = \Pr[S_t = j | S_{t-1} = i]$ with $\sum_{j=1}^{M} p_{ij} = 1$ for all i. The parameter matrices of the model F_{S_t}, β_{S_t}, A_{S_t}, γ_{S_t}, and G_{S_t} may be known under different regimes or states, but in some circumstances a particular element of a parameter matrix may take on different values which are unknown. Our model incorporates the latter case as well.

When the F_{S_t} matrices, for example, are known under different states, F_m $(m = 1, 2, \ldots, M)$ refers to the known parameter matrix when the state or regime m prevails. When a particular element of the F_{S_t} matrix switches from one state to another, and when the values of that element are unknown under different states, it can be modeled in the following way. Assuming that the state variable S_t can take the values of $1, 2, \ldots, M$, the (i,j)th element of the F_{S_t} can be specified as

$$f_{i,j,S_t} = f_{i,j,1}\, S_{1t} + \cdots + f_{i,j,M}\, S_{Mt}, \tag{2.5}$$

where S_{mt} takes the value 1 when S_t is equal to m and 0 otherwise. The $f_{i,j,m}$'s $(m = 1, 2, \ldots, M)$ are, in principle, part of the parameters to be estimated. As a comparison, notice that Harrison and Stevens (1976) suggested the use of discrete-valued grids, covering a range likely to include plausible values for the variances when the variances of measurement or transition equations are assumed to be heteroskedastic.

2.2. Basic filtering and estimation of the model

Suppose the parameters of the model specified in the previous subsection are known. Let $\psi_{t-1} \equiv (y'_{t-1}, y'_{t-2}, \ldots, y'_1, z'_t, z'_{t-1}, \ldots, z_1)'$ denote the vector of observations received as of time $t-1$. In the usual derivation of the Kalman filter for a fixed-coefficient state-space model, the goal is to form a forecast of the unobserved state vector x_t based on ψ_{t-1}, denoted $x_{t|t-1}$,

$$x_{t|t-1} = \mathrm{E}(x_t \mid \psi_{t-1}).$$

Similarly, in the conventional fixed-coefficient case, the matrix $P_{t|t-1}$ denotes the mean square error of the forecast:

$$P_{t|t-1} = \mathrm{E}[(x_t - x_{t|t-1})(x_t - x_{t|t-1})' \mid \psi_{t-1}].$$

Here the goal will be to form a forecast of x_t based not just on ψ_{t-1} but also conditional on the random variable S_t taking on the value j and on S_{t-1} taking on the value i,

$$x_{t|t-1}^{(i,j)} = \mathrm{E}(x_t \mid \psi_{t-1}, S_t = j, S_{t-1} = i).$$

The proposed algorithm calculates a battery of M^2 such forecasts for each date t, corresponding to every possible value for i and j. Associated with these forecasts are M^2 different mean squared error matrices:

$$P_{t|t-1}^{(i,j)} = \mathrm{E}[(x_t - x_{t|t-1})(x_t - x_{t|t-1})' \mid \psi_{t-1}, S_t = j, S_{t-1} = i].$$

The algorithm is as follows:

$$x_{t|t-1}^{(i,j)} = A_j x_{t-1|t-1}^{i} + \gamma_j z_t, \tag{2.6}$$

$$P_{t|t-1}^{(i,j)} = A_j P_{t-1|t-1}^{i} A'_j + G_j Q G'_j, \tag{2.7}$$

$$\eta_{t|t-1}^{(i,j)} = y_t - F_j x_{t|t-1}^{(i,j)} - \beta_j z_t, \tag{2.8}$$

$$H_t^{(i,j)} = F_j P_{t|t-1}^{(i,j)} F'_j + R, \tag{2.9}$$

$$K_t^{(i,j)} = P_{t|t-1}^{(i,j)} F'_j [H_t^{(i,j)}]^{-1}, \tag{2.10}$$

$$x_{t|t}^{(i,j)} = x_{t|t-1}^{(i,j)} + K_t^{(i,j)} \eta_{t|t-1}^{(i,j)}, \tag{2.11}$$

$$P_{t|t}^{(i,j)} = (I - K_t^{(i,j)} F_j) P_{t|t-1}^{(i,j)}, \tag{2.12}$$

where $x_{t-1|t-1}^{i}$ is an inference about x_{t-1} based on information up to time $t-1$, given $S_{t-1} = i$; $x_{t|t-1}^{(i,j)}$ is an inference about x_t based on information up to time $t-1$, given $S_t = j$ and $S_{t-1} = i$; $P^{(i,j)}_{t|t-1}$ is the variance covariance matrix of $x^{(i,j)}_{t|t-1}$ conditional on $S_t = j$ and $S_{t-1} = i$; $\eta^{(i,j)}_{t|t-1}$ is the conditional forecast error of y_t based on information up to time $t-1$, given $S_{t-1} = i$ and $S_t = j$; $H^{(i,j)}_t$ is the conditional variance of forecast error $\eta^{(i,j)}_{t|t-1}$; and $K^{(i,j)}_t$ is the Kalman gain.

As noted by Gordon and Smith (1988) and Harrison and Stevens (1976), each iteration of the above Kalman filtering produces an M-fold increase in the number of cases to consider. (Even when M, the total number of different regimes, is 2, there would be over 1000 cases to consider by the time $t = 10$.) It is necessary to introduce some approximations to make the above Kalman filtering operable. The key is to collapse terms in the right way at the right time. Therefore, it remains to reduce the $(M \times M)$ posteriors ($x^{(i,j)}_{t|t}$ and $P^{(i,j)}_{t|t}$) into M to complete the above Kalman filtering. We employ the following approximations similar to those proposed by Harrison and Stevens (1976):[2]

$$x^j_{t|t} = \frac{\sum_{i=1}^{M} \Pr[S_{t-1} = i, S_t = j \mid \psi_t]\, x^{(i,j)}_{t|t}}{\Pr[S_t = j \mid \psi_t]}, \tag{2.13}$$

$$P^j_{t|t} = \frac{\sum_{i=1}^{M} \Pr[S_{t-1} = i, S_t = j \mid \psi_t]\, \{P^{(i,j)}_{t|t} + (x^j_{t|t} - x^{(i,j)}_{t|t})(x^j_{t|t} - x^{(i,j)}_{t|t})'\}}{\Pr[S_t = j \mid \psi_t]}, \tag{2.14}$$

where ψ_t refers to information available at time t.

Here, it might be worthwhile to clarify the sense in which the above algorithm involves an approximation. If $x^{(i,j)}_{t|t}$ in eq. (2.11) represented $\mathrm{E}[x_t \mid S_{t-1} = i, S_t = j, \psi_t]$, then it is straightforward to show that (2.13) calculates $\mathrm{E}[x_t \mid S_t = j, \psi_t]$ exactly. Similarly, if $P^{(i,j)}_{t|t}$ in (2.12) represented $\mathrm{E}[(x_t - x^{(i,j)}_{t|t}) \times (x_t - x^{(i,j)}_{t|t})' \mid S_{t-1} = i,\ S_t = j, \psi_t]$, then (2.14) calculates $\mathrm{E}[(x_t - x^j_{t|t}) \times (x_t - x^j_{t|t})' \mid S_t = j, \psi_t]$ exactly. The approximation arises because (2.11) does not calculate $\mathrm{E}[x_t \mid S_{t-1} = i, S_t = j, \psi_t]$ exactly. This formula would give the conditional expectation if, conditional on ψ_{t-1}, and on $S_t = j$ and $S_{t-1} = i$, the distribution of x_t is Normal. However, the distribution of x_t conditional on ψ_{t-1}, $S_t = j$, and $S_{t-1} = i$ is a mixture of Normals for $t > 2$. One can still motivate (2.11) as the linear projection of x_t on y_t and $x^i_{t-1|t-1}$ (taking S_t and S_{t-1} as given). Thus the algorithm is certainly calculating a sensible inference about x_t. Notice, however, that (2.11) is not calculating the linear projection of x_t

[2] In generalizing the Kalman filter and the dynamic linear model to account for a Markov process on the M processes, Highfield (1990) 'collapses' the Kalman filter to a single posterior at each t as in Gordon and Smith (1988). For the quality of various approaches to this 'collapsing', refer to Smith and Markov (1980).

on $y_t, y_{t-1}, \ldots$ since $x^i_{t-1|t-1}$ is a nonlinear function of $y_{t-1}, y_{t-2}, \ldots$. There is plenty of precedent for this; for example, multi-period forecasts of a time-varying coefficient model are doing the same sort of thing.

The last thing that remains to be considered to complete the Kalman filtering is to calculate the $\Pr[S_{t-1} = i,\ S_t = j|\psi_t]$ and other probability terms. The following procedure explains how a complete basic filtering can be performed using the above eqs. (2.6) through (2.14). Notice that the basic filter accepts three inputs and has three outputs. The three inputs are $x^i_{t-1|t-1}, P^i_{t-1|t-1}$, and the joint conditional probability $\Pr[S_{t-2} = i',\ S_{t-1} = i|\psi_{t-1}]$, where ψ_{t-1} refers to information available at time $t-1$. The three outputs are $x^j_{t|t},\ P^i_{t|t}$, and the conditional probability $\Pr[S_{t-1} = i, S_t = j|\psi_t]$. The arguments below follow Hamilton (1989) with slight modifications.

Step 1. Calculate

$$\Pr[S_{t-1} = i, S_t = j|\psi_{t-1}]$$
$$= \Pr[S_t = j|S_{t-1} = i] \times \sum_{i'=1}^{M} \Pr[S_{t-2} = i',\ S_{t-1} = i|\psi_{t-1}], \tag{2.15}$$

where $\Pr[S_t = j|S_{t-1} = i]$ is given by (2.4).

Step 2. Calculate the joint conditional density function of y_t and (S_{t-1}, S_t):

$$f(y_t, S_{t-1} = i, S_t = j|\ \psi_{t-1})$$
$$= f(y_t|S_{t-1} = i, S_t = j, \psi_{t-1}) \times \Pr[S_{t-1} = i, S_t = j|\ \psi_{t-1}], \tag{2.16}$$

where

$$f(y_t|S_{t-1} = i, S_t = j, \psi_{t-1})$$
$$= (2\pi)^{-N/2}|\,H_t^{(i,j)}|^{-1/2} \exp(-\tfrac{1}{2}\eta^{(i,j)\prime}_{t|t-1} H_t^{(i,j)^{-1}}\ \eta^{(i,j)}_{t|t-1}).$$

Step 3. Calculate

$$\Pr[S_{t-1} = i, S_t = j|\psi_t] = \frac{f(y_t, S_{t-1} = i, S_t = j|\psi_{t-1})}{f(y_t|\psi_{t-1})}, \tag{2.17}$$

where

$$f(y_t|\psi_{t-1}) = \sum_{j=1}^{M}\sum_{i=1}^{M} f(y_t, S_{t-1} = i, S_t = j|\psi_{t-1}). \tag{2.16'}$$

Step 4. Then from (2.13), (2.14), and output from step 3, we get $x^j_{t|t}$ and $P^j_{t|t}$. The remaining output $\Pr[S_t = j|\psi_t]$ can be calculated by

$$\Pr[S_t = j|\psi_t] = \sum_{i=1}^{M} \Pr[S_{t-1} = i, S_t = j|\psi_t]. \tag{2.18}$$

As a by-product of running the above Kalman filter, the conditional log-likelihood function can be obtained from step 3. The sample conditional log-likelihood is

$$LL = \log(f(y_T, y_{T-1}, \dots |\psi_0)) = \sum_{t=1}^{T} \log(f(y_t|\psi_{t-1})). \tag{2.19}$$

Actually, (2.19) is an approximation to the log-likelihood function, for the same reasons discussed earlier. The filter above is derived under the assumption that parameters of the underlying model are known. To estimate the parameters of the model, we can maximize the log-likelihood function defined in eq. (2.19) with respect to the underlying unknown parameters, using a nonlinear optimization procedure.

An important note is in order. In our basic filter presented in this section, we tried to derive the distribution of x_t conditional on ψ_{t-1}, $S_t = j$, and $S_{t-1} = i$ $(i, j = 1, 2, \dots, M)$. Instead, we could derive the distribution of x_t conditional on ψ_{t-1}, $S_t = j$, $S_{t-1} = i$, and $S_{t-2} = h$ $(h, i, j = 1, 2, \dots, M)$. In this case, we need to collapse the M^3 posteriors to M^2 at each iteration, and it is straightforward to modify the basic filter in (2.6)–(2.14) appropriately. For example, the superscripts (i, j) and i in eqs. (2.6)–(2.12) will be changed to (h, i, j) and (h, i), respectively, and eqs. (2.13) and (2.14) will be rewritten as

$$x^{(i,j)}_{t|t} = \frac{\sum_{h=1}^{M} \Pr[S_{t-2} = h, S_{t-1} = i, S_t = j|\psi_t]\, x^{(h,i,j)}_{t|t}}{\Pr[S_{t-1} = i, S_t = j|\psi_t]}, \tag{2.13'}$$

$$P^{(i,j)}_{t|t} = \frac{\sum_{h=1}^{M} \Pr[S_{t-2} = h, S_{t-1} = i, S_t = j|\psi_t]\,\{P^{(h,i,j)}_{t|t} + (x^{(i,j)}_{t|t} - x^{(h,i,j)}_{t|t})(x^{(i,j)}_{t|t} - x^{(h,i,j)}_{t|t})'\}}{\Pr[S_{t-1} = i, S_t = j|\psi_t]}. \tag{2.14'}$$

In this way, one can obtain more efficient inferences about the unobserved x_t and S_t. In general, as we carry more states at each iteration, we can get more efficient inferences, but only at the cost of increased computation time and the tractability of the model. When there is no lagged dependent variable in the state–space representation, carrying M^2 states is usually enough. This point will be made clear from an example in section 3. When r ($r > 0$) lagged dependent variables are present in the state–space representation, however, one should carry at least M^{r+1} states at each iteration.

2.3. Smoothing

Once parameters are estimated, we can get an inference about S_t and $\boldsymbol{x}_t$ based on all the information in the sample, $\Pr[S_t = j|\psi_T]$ and $\boldsymbol{x}_{t|T}$ ($t = 1, 2, \ldots, T$).

Consider the following derivation of the joint probability that $S_t = j$ and $S_{t+1} = k$ based on full information:[3]

$$\begin{aligned}
&\Pr[S_t = j, S_{t+1} = k|\psi_T] \\
&= \Pr[S_{t+1} = k|\psi_T] \times \Pr[S_t = j|S_{t+1} = k, \psi_T] \\
&\approx \Pr[S_{t+1} = k|\psi_T] \times \Pr[S_t = j|S_{t+1} = k, \psi_t] \\
&= \frac{\Pr[S_{t+1} = k|\psi_T] \times \Pr[S_t = j, S_{t+1} = k, \psi_t]}{\Pr[S_{t+1} = k|\psi_t]} \\
&= \frac{\Pr[S_{t+1} = k|\psi_T] \times \Pr[S_t = j|\psi_t] \times \Pr[S_{t+1} = k|S_t = j]}{\Pr[S_{t+1} = k|\psi_t]},
\end{aligned} \tag{2.20}$$

$$\Pr[S_t = j|\psi_T] = \sum_{k=1}^{M} \Pr[S_t = j, S_{t+1} = k|\psi_T]. \tag{2.21}$$

Notice that the algorithm in (2.20) involves an approximation as we go from the first line to the second line. To investigate the nature of approximation involved, define $\boldsymbol{h}_{t+1,T} = (y'_{t+1}, y'_{t+2}, \ldots, y'_T, z'_{t+1}, z'_{t+2}, \ldots, z'_T)'$, for $T > t$. That is, $\boldsymbol{h}_{t+1,T}$ is the vector of observations from date $t+1$ to T. Then, we have

$$\begin{aligned}
&\Pr[S_t = j|S_{t+1} = k, \psi_T] \\
&= \Pr[S_t = j|S_{t+1} = k, \boldsymbol{h}_{t+1,T}, \psi_t] \\
&= \frac{f(S_t = j, \boldsymbol{h}_{t+1,T}|S_{t+1} = k, \psi_t)}{f(\boldsymbol{h}_{t+1,T}|S_{t+1} = k, \psi_t)} \\
&= \frac{\Pr(S_t = j|S_{t+1} = k, \psi_t)\, f(\boldsymbol{h}_{t+1,T}|S_{t+1} = k, S_t = j, \psi_t)}{f(\boldsymbol{h}_{t+1,T}|S_{t+1} = k, \psi_t)}.
\end{aligned} \tag{2.22}$$

[3] Here, it is assumed that no lagged dependent variables appear in the model. For the derivation of a similar smoothing algorithm in the context of a general non-Gaussian state–space model, refer to Kitagawa (1989). Also, refer to Hamilton (1991) for related issues.

Provided that

$$f(\boldsymbol{h}_{t+1,T}|S_{t+1}=k,S_t=j,\psi_t)=f(\boldsymbol{h}_{t+1,T}|S_{t+1}=k,\psi_t), \tag{2.23}$$

we have $\Pr[S_t=j|S_{t+1}=k,\psi_T]=\Pr[S_t=j|S_{t+1}=k,\psi_t]$, and (2.20) would be exact. In this case, if S_{t+1} were somewhat known, then y_{t+1} and z_{t+1} would contain no information about S_t beyond that contained in S_{t+1} and ψ_t. For the Hamilton model (1989) with no lagged dependent variables, (2.23) holds exactly and therefore (2.20) is exact. For the state–state model with Markov-switching presented in this paper, however, (2.23) does not hold exactly, and this why an approximation is involved in (2.20).

The smoothing algorithm derived above can be generalized to the models with r ($r>1$) lagged dependent variables. For example, for the Hamilton model (1989) with r lagged dependent variables, the exact smoothing algorithm is written

$$\begin{aligned}
&\Pr[S_{t-r+1},\ldots,S_t,S_{t+1}|\psi_T]\\
&=\frac{\Pr[S_{t-r+2},\ldots,S_t,S_{t+1}|\psi_T]\times\Pr[S_{t-r+1},\ldots,S_t,S_{t+1}|\psi_t]}{\Pr[S_{t-r+2},\ldots,S_t,S_{t+1}|\psi_t]}\\
&=\frac{\Pr[S_{t-r+2},\ldots,S_t,S_{t+1}|\psi_T]\times\Pr[S_{t-r+1},\ldots,S_t|\psi_t]\times\Pr[S_{t+1}|S_t]}{\Pr[S_{t-r+2},\ldots,S_t,S_{t+1}|\psi_t]},
\end{aligned} \tag{2.20$'$}$$

since

$$f(\boldsymbol{h}_{t+1,T}|S_{t-r+1},\ldots,S_t,S_{t+1},\psi_t)=f(\boldsymbol{h}_{t+1,T}|S_{t-r+2},\ldots,S_t,S_{t+1},\psi_t) \tag{2.23$'$}$$

holds exactly. In the case of the state–space model with Markov-switching that includes r lagged dependent variables,[4] the same smoothing algorithm can be used. But in this case, as (2.23′) is not exact, the algorithm involves an approximation. Notice the above algorithms for calculating the smoothed inference about the unobserved states are vastly more efficient than those in Hamilton (1989) and Lam (1990), in terms of its simplicity and the computation time.

Keeping eq. (2.20) in mind, we now turn to the derivation of the smoothing algorithm for the vector $\boldsymbol{x}_t$. Like the basic filtering in subsection 2.2, the

[4] In this case, as mentioned in section 2.2, M^{r+1} states should be carried at each iteration of basic filtering and smoothing.

smoothing algorithm for the vector x_t can be written as follows, given that $S_t = j$ and $S_{t+1} = k$:

$$x_{t|T}^{(j,k)} = x_{t|t}^{j} + \tilde{P}_t^{(j,k)}(x_{t+1|T}^{k} - x_{t+1|t}^{(j,k)}), \tag{2.24}$$

$$P_{t|T}^{(j,k)} = P_{t|t}^{j} + \tilde{P}_t^{(j,k)}(P_{t+1|T}^{k} - P_{t+1|t}^{(j,k)})\,\tilde{P}_t^{(j,k)\prime}, \tag{2.25}$$

where $\tilde{P}_t^{(j,k)} = P_{t|t}^{j} A_k' [P_{t+1|t}^{(j,k)}]^{-1}$; $x_{t|T}^{(j,k)}$ is inference of x_t based on full sample and $P_{t|T}^{(j,k)}$ is the variance–covariance matrix of $x_{t|T}^{(j,k)}$; $x_{t|t}^{j}$ and $P_{t|t}^{j}$ are given by eqs. (2.13) and (2.14).

As $\Pr[S_t = j|\psi_T]$ is not dependent upon $x_{t|T}$, we can first calculate smoothed probabilities, and then these smoothed probabilities can be used to get smoothed values of x_t, $x_{t|T}$. Given the above smoothing algorithms, actual smoothing can be performed by applying approximations similar to those introduced in the basic filtering.

Step 1. Run through the basic filter in the previous subsection for $t = 1, \dots, T$ and store the resulting sequences $x_{t|t-1}^{(i,j)}$, $P_{t|t-1}^{(i,j)}$, $x_{t|t}^{j}$, $P_{t|t}^{j}$, $\Pr[S_t = j|\psi_{t-1}] = \sum_{i=1}^{M} \Pr[S_{t-1} = i, S_t = j|\psi_{t-1}]$, and $\Pr[S_t = j|\psi_t]$ from eqs. (2.6), (2.7), (2.13), (2.14), (2.15), and (2.18), respectively, for $t = 1, 2, \dots, T$.

Step 2. For $t = T-1, T-2, \dots, 1$, get the smoothed joint probability $\Pr[S_t = j, S_{t+1} = k|\psi_T]$ and $\Pr[S_t = j|\psi_T]$ according to (2.20) and (2.21), and save them. Here, $\Pr[S_T = j|\psi_T]$, the starting value for smoothing, is given by the final iteration of the basic filter.

Step 3. Then, we can use the smoothed probabilities from step 2 to collapse the $M \times M$ elements of $x_{t|T}^{(j,k)}$ and $P_{t|T}^{(j,k)}$ into M by taking weighted averages. At each iteration of (2.24) and (2.25), for $t = T-1, T-2, \dots, 1$, collapse the $M \times M$ elements into M in the following way by taking weighted averages over state S_{t+1}:

$$x_{t|T}^{j} = \frac{\sum_{k=1}^{M} \Pr[S_t = j,\ S_{t+1} = k|\psi_T]\, x_{t|T}^{(j,k)}}{\Pr[S_t = j|\psi_T]}, \tag{2.26}$$

$$P_{t|T}^{j} = \frac{\sum_{k=1}^{M} \Pr[S_t = j,\ S_{t+1} = k|\psi_T]\,\{P_{t|T}^{(j,k)} + (x_{t|T}^{j} - x_{t|T}^{(j,k)})(x_{t|T}^{j} - x_{t|T}^{(j,k)})'\}}{\Pr[S_t = j|\psi_T]}. \tag{2.27}$$

Step 4. From step 3, the smoothed value of $x_{t|T}^{(j)}$ is dependent only upon states at time t. By taking a weighted average over the states at time t, we can get $x_{t|T}$ from

$$x_{t|T} = \sum_{j=1}^{M} \Pr[S_t = j | \psi_T] \, x_{t|T}^{j}. \tag{2.28}$$

3. An example: State–space estimation of the Hamilton model with a general autoregressive component

In modeling the time series behavior of U.S. real GNP, Hamilton (1989) considered the case in which real GNP is generated by the sum of two independent unobserved components, one following an autoregressive process with a unit root, and the other following a random walk with a Markov-switching error term. Lam (1990) generalized the Hamilton model to the case in which the autoregressive component need not contain a unit root. The model that he proposed is

$$\tilde{y}_t = n_t + x_t, \tag{3.1}$$

$$n_t = n_{t-1} + \beta_{S_t}, \tag{3.2}$$

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_r x_{t-r} + u_t, \quad u_t \sim \text{i.i.d. N}(0, \sigma^2), \tag{3.3}$$

$$\beta_{S_t} = \delta_0 + \delta_1 S_t, \qquad S_t = 0, 1, \tag{3.4}$$

and

$$\Pr[S_t = 0 | S_{t-1} = 0] = q \quad \text{and} \quad \Pr[S_t = 1 | S_{t-1} = 1] = p,$$

where $\tilde{y}_t$ is the log of real GNP and n_t and x_t are random walk and autoregressive components of the real GNP; β_{S_t} is the error term, which evolves according to a two-state Markov process. [In general, β_{S_t} in (3.4) can be assumed to evolve according to an *M-state* Markov process, whose process may be represented by (2.4) and (2.5). But following Hamilton (1989) and Lam (1990), we will assume a two-state Markov process.] By taking the first difference of $\tilde{y}_t$, eq. (3.1) can be written as

$$y_t = (x_t - x_{t-1}) + \delta_0 + \delta_1 S_t, \tag{3.5}$$

where $y_t = \tilde{y}_t - \tilde{y}_{t-1}$.

Unlike the original Hamilton model with a unit root in (3.3), one difficulty is that the states of an observation include the whole history of the Markov process. This is clear from the following expression for x_t, which is obtained by solving eq. (3.5) backward in time:

$$x_t = \left(\sum_{i=1}^{t} y_i - \delta_0 t - \delta_1 \sum_{i=1}^{t} S_i \right) + x_0. \tag{3.6}$$

Lam (1990) estimated the model by treating the sum of previous states as an additional state variable. This is possible because the sum of previous states is also Markovian. Notice, however, that the above generalized Hamilton model is a special case of the model presented in this paper, and it can be estimated using the algorithm introduced in the previous section. That is, the model can be estimated without introducing an additional state variable, and thus the estimation procedure may be simplified significantly. If we cast the model in the following state–space form by combining eqs. (3.3) and (3.5), it is easy to see that the model is a special case of our general model in (2.1) and (2.2):

$$y_t = [1 \quad -1 \quad 0 \quad \cdots \quad 0] \begin{bmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-r+1} \end{bmatrix} + \beta_{S_t} z_t, \tag{3.7}$$

$$\begin{bmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-r+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_r \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ x_{t-3} \\ \vdots \\ x_{t-r+r} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{3.8}$$

where $z_t = 1$ for all t.

We apply the algorithms introduced in the previous section to estimate Lam's (1990) generalized Hamilton model in (3.7) and (3.8). From a Monte Carlo experiment, Lam (1990) showed that there was substantial bias toward finding the Hamilton model with a general autoregressive process when the data-generating process was the original Hamilton mode with a unit root in (3.3). He reports more evidence for the unit root hypothesis in the autoregressive component of real GNP in the whole post-war sample period of 1947:2–1986:4 than in the sample period 1952:2–1984:4. As we are interested in the estimation of

Table 1

Maximum likelihood estimates of the Hamilton model with general autoregressive component: Comparison of estimates from Lam's (1990) model and our state-space model.[a]

Parameter	Estimate	
	Lam (1990)	State-space
$\hat{p}$	0.957 (0.019)	0.954 (0.022)
$\hat{q}$	0.508 (0.101)	0.465 (0.170)
$\hat{\delta}_0$	− 1.483 (0.151)	− 1.457 (0.420)
$\hat{\delta}_1$	2.447 (0.160)	2.421 (0.424)
$\hat{\sigma}$	0.771 (0.047)	0.773 (0.052)
$\hat{\phi}_1$	1.244 (0.063)	1.246 (0.087)
$\hat{\phi}_2$	− 0.382 (0.064)	− 0.367 (0.086)
$\hat{x}_0$	6.376 (0.127)	5.224 (1.684)
$\hat{x}_{-1}$	—	0.535 (2.699)
Log-likelihood value	− 174.97	− 176.33

[a](1) The model is estimated using 100 times the log-difference in quarterly real GNP. (2) Standard errors are in the parentheses.

the generalized Hamilton model without a unit root in (3.3), we focus our analysis on the sample period 1952:2–1984:4,[5] in which period he found less evidence for a unit root. Also, we assume a second-order autoregressive process for the autoregressive component in (3.3) as in Lam (1990).

Lam treats x_0, the initial value for the autoregressive component, as an additional parameter to be estimated. Following his strategy, we also treat each element of $\boldsymbol{x}_{0|0} = (x_0 \; x_{-1})'$, the initial condition for the state vector ($\boldsymbol{x}_t$) from our state–state representation, as an additional parameter to be estimated.[6] In this case, as we are assuming that elements in $\boldsymbol{x}_{0|0}$ are some unknown constants, each element in $P_{0|0}$, the variance–covariance matrix of $\boldsymbol{x}_{0|0}$, is fixed at zero.

Table 1 reports the estimation results obtained by applying both Lam's algorithm and this paper's algorithm. Estimates of the parameters from the state–space model (using the algorithm proposed in this paper) are close to those from Lam (1990). Except for the initial values of the autoregressive component, all the estimates from the state–space model are within one standard error of those from Lam (1990). [Similar results were obtained from the whole post-war sample period, but the results are not reported here.] Table 2 reports and

[5]As Lam's model includes two lagged dependent variables, his algorithm begins with third observation. In our state–space representation of his model, however, no lagged dependent variable shows up. Our algorithm begins with first observation. Therefore, in order to make the results directly comparable, we employ sample period 1954: 4–1984:4.

[6]In the usual estimation of the state–space models, the initial values for the state vector and its variance–covariance matrix are set to their steady-state values when the transition equation is stationary.

Table 2

Probability that the economy is in the fast growth state based on currently available information and information from the full sample: Comparison of estimates from Lam's (1990) model and our state-space model.

	$\Pr[S_t = 1 \mid \psi_t]$		$\Pr[S_t = 1 \mid \psi_T]$	
Quarter	Lam (1990)	State-space	Lam (1990)	State-space
52.4	0.916814	0.990125	0.998682	0.994438
53.1	0.999975	0.999978	1.000000	0.999988
53.2	0.999973	0.999922	1.000000	0.999953
53.3	0.997743	0.993995	0.999990	0.996281
53.4	0.997653	0.993184	0.999995	0.993728
54.1	0.975372	0.934344	0.999836	0.960409
54.2	0.998244	0.992119	0.999979	0.995564
54.3	0.999990	0.999957	1.000000	0.999976
54.4	0.999963	0.999928	0.999997	0.999960
55.1	0.999995	0.999993	1.000000	0.999996
55.2	0.999565	0.999340	0.999963	0.999629
55.3	0.999955	0.999936	0.999998	0.999964
55.4	0.999779	0.999583	0.999990	0.999733
56.1	0.990158	0.985378	0.999551	0.991679
56.2	0.999543	0.999048	0.999976	0.999438
56.3	0.996999	0.994509	0.999747	0.996910
56.4	0.999928	0.999857	0.999995	0.999919
57.1	0.999376	0.999005	0.999973	0.999294
57.2	0.978538	0.978477	0.999668	0.984490
57.3	0.998226	0.998038	0.999987	0.980551
57.4	0.090938	0.097006	0.011585	0.011182
58.1	0.001468	0.002892	0.001127	0.005039
58.2	0.935936	0.961286	0.997189	0.977910
58.3	0.999949	0.999954	1.000000	0.999974
58.4	0.999977	0.999977	0.999999	0.999987
59.1	0.998540	0.998997	0.999930	0.999437
59.2	0.999975	0.999983	0.999998	0.999970
59.3	0.811593	0.804273	0.994120	0.879021
59.4	0.997334	0.997273	0.999918	0.998468
60.1	0.999982	0.999976	1.000000	0.999968
60.2	0.902201	0.894281	0.999746	0.922835
60.3	0.979368	0.981706	0.999994	0.969037
60.4	0.795111	0.778957	0.999211	0.861931
61.1	0.997989	0.997666	1.000000	0.998684
61.2	0.999594	0.999515	0.999999	0.999727
61.3	0.999638	0.999743	0.999998	0.999856
61.4	0.999982	0.999987	1.000000	0.999993
62.1	0.998882	0.999145	0.999999	0.999518
62.2	0.999049	0.999422	0.999999	0.999674
62.3	0.999292	0.999443	1.000000	0.999596
62.4	0.965012	0.969798	0.999934	0.982811

Table 2 (continued)

	$Pr[S_t = 1 \mid \psi_t]$		$Pr[S_t = 1 \mid \psi_T]$	
Quarter	Lam (1990)	State–space	Lam (1990)	State–space
63.1	0.999750	0.999785	1.000000	0.999879
63.2	0.999836	0.999814	1.000000	0.999896
63.3	0.999945	0.999958	1.000000	0.999976
63.4	0.996689	0.997324	0.999986	0.998498
64.1	0.999995	0.999996	1.000000	0.999997
64.2	0.997999	0.997805	0.999998	0.998766
64.3	0.999690	0.999777	1.000000	0.999872
64.4	0.997926	0.997930	0.999983	0.998838
65.1	0.999998	0.999997	1.000000	0.999998
65.2	0.999870	0.999831	0.999999	0.999905
65.3	0.999965	0.999970	0.999999	0.999983
65.4	0.999998	0.999998	1.000000	0.999999
66.1	0.999993	0.999989	1.000000	0.999994
66.2	0.996920	0.996138	0.999872	0.997830
66.3	0.999976	0.999960	1.000000	0.999977
66.4	0.999758	0.999446	0.999991	0.999689
67.1	0.999859	0.999747	0.999992	0.999858
67.2	0.999853	0.999708	0.999976	0.999836
67.3	0.999993	0.999986	0.999999	0.999992
67.4	0.999423	0.998807	0.999825	0.999330
68.1	0.999974	0.999953	0.999984	0.999974
68.2	0.999995	0.999989	0.999999	0.999994
68.3	0.999687	0.999345	0.999931	0.999612
68.4	0.996268	0.993957	0.995836	0.996602
69.1	0.999994	0.999981	0.999999	0.999989
69.2	0.997060	0.992085	0.998863	0.995511
69.3	0.999721	0.999393	0.999976	0.999497
69.4	0.983517	0.967607	0.998839	0.972155
70.1	0.963577	0.937956	0.997368	0.960104
70.2	0.990049	0.980572	0.999166	0.988743
70.3	0.999868	0.999734	0.999803	0.998283
70.4	0.276312	0.260709	0.029695	0.386064
71.1	0.999950	0.999952	0.999997	0.999905
71.2	0.772706	0.774267	0.985142	0.849904
71.3	0.967311	0.977317	0.998665	0.984220
71.4	0.965649	0.971081	0.989702	0.983573
72.1	0.999979	0.999974	0.999997	0.999985
72.2	0.999920	0.999899	0.999984	0.999942
72.3	0.997836	0.998467	0.998593	0.999140
72.4	0.999971	0.999977	0.999968	0.999987
73.1	0.999995	0.999994	0.999997	0.999995
73.2	0.978165	0.976752	0.992196	0.985541
73.3	0.985618	0.987532	0.992710	0.992921
73.4	0.999786	0.999743	0.999837	0.999538
74.1	0.908802	0.876432	0.988689	0.871150
74.2	0.990882	0.987743	0.993198	0.892386
74.3	0.298118	0.275369	0.018681	0.044806
74.4	0.160343	0.192083	0.013512	0.024134

Table 2 (continued)

Quarter	$\Pr[S_t = 1 \mid \psi_t]$ Lam (1990)	$\Pr[S_t = 1 \mid \psi_t]$ State–space	$\Pr[S_t = 1 \mid \psi_T]$ Lam (1990)	$\Pr[S_t = 1 \mid \psi_T]$ State–space
75.1	0.003546	0.002771	0.001756	0.004916
75.2	0.989702	0.994498	0.997941	0.996904
75.3	0.999823	0.999881	0.999977	0.999933
75.4	0.999572	0.999656	0.999899	0.999807
76.1	0.999955	0.999968	0.999997	0.999980
76.2	0.988662	0.990046	0.999119	0.994253
76.3	0.995953	0.997146	0.999536	0.998394
76.4	0.999713	0.999734	0.999946	0.999851
77.1	0.999894	0.999885	0.999960	0.999935
77.2	0.999933	0.999933	0.999970	0.999962
77.3	0.999984	0.999983	0.999992	0.999980
77.4	0.883378	0.890713	0.922268	0.935172
78.1	0.998324	0.998336	0.993722	0.999066
78.2	1.000000	1.000000	1.000000	1.000000
78.3	0.996157	0.993774	0.995772	0.996495
78.4	0.999898	0.999883	0.999948	0.999928
79.1	0.993224	0.989548	0.995534	0.993816
79.2	0.995722	0.994242	0.995418	0.996757
79.3	0.999905	0.999831	0.999875	0.999875
79.4	0.978756	0.969439	0.984264	0.981413
80.1	0.999672	0.999466	0.997989	0.994610
80.2	0.003766	0.003325	0.007848	0.005559
80.3	0.829849	0.892513	0.878688	0.936291
80.4	0.997858	0.998353	0.994643	0.999074
81.1	0.999959	0.999960	0.999572	0.999850
81.2	0.646298	0.711570	0.260692	0.665050
81.3	0.939218	0.969000	0.322511	0.758463
81.4	0.120700	0.134809	0.008262	0.019557
82.1	0.013820	0.017459	0.017573	0.022276
82.2	0.839975	0.861506	0.931052	0.708689
82.3	0.517556	0.504434	0.907270	0.612544
82.4	0.880608	0.902779	0.985670	0.941776
83.1	0.993801	0.995123	0.998395	0.997259
83.2	0.999980	0.999983	0.999996	0.99999
83.3	0.999290	0.999097	0.999554	0.999493
83.4	0.999897	0.999905	0.999870	0.999947
84.1	0.999991	0.999989	0.999997	0.999994
84.2	0.999164	0.998940	0.999762	0.999392
84.3	0.997934	0.997701	0.999227	0.998657
84.4	0.998358	0.997585	0.998358	0.997585

compares the probabilities that the economy is in the high growth state from both Lam's estimation and the state–space estimation, based on current and full information. The relative magnitudes of these probabilities are almost the same. Inferences on the periods of low growth are almost the same as in Lam (1990).[7]

In fig. 1, the estimated stochastic trend based on smoothing algorithms (2.24) through (2.28) is plotted against actual real GNP. Again, this is almost indistinguishable from Lam's estimated trend.

It has to be admitted that Lam's estimation procedure may be more efficient than the state–space estimation procedure, as the latter is based on an approximation while the former is not. Actually, the standard errors of the estimates are somewhat larger and the log-likelihood value is somewhat smaller for the state–space estimation results. In employing the algorithm introduced in this paper, however, there is a significant advantage in terms of the computation costs, while the loss in efficiency is only marginal. The following explains the differences in the computation time in the two algorithms.

Consider the approximation to the likelihood of the observation conditional on its past from the state–space model in this paper. It is given by (2.16′) for the model that involves no lagged dependent variables. Assuming a two-state, first-order Markov process, it can be represented by

$$f(y_t|\psi_{t-1}) = \sum_{s_t=0}^{1} \sum_{s_{t-1}=0}^{1} f(y_t, S_t = s_t, S_{t-1} = s_{t-1}|\psi_{t-1}). \tag{3.9}$$

It is easy to see that, for each t ($t = 1, \ldots, T$), the number of cases to be evaluated is 2^2, regardless of the order of the autoregressive process in (3.3). The likelihood of the observation conditional on its past from Lam's original algorithm without approximation [Lam (1990, p. 415)] is given by

$$\begin{aligned} f(y_t|\psi_{t-1}) &= \sum_{s_t=0}^{1} \cdots \sum_{s_{t-r+1}=0}^{1} \sum_{\omega=0}^{t} f\Bigg(y_t, S_t = s_t, S_{t-1} = s_{t-1}, \ldots, S_{t-r+1} = s_{t-r+1}, \\ &\qquad \sum_{h=1}^{t} S_h = \omega|\psi_{t-1} \Bigg). \end{aligned} \tag{3.10}$$

[7] In table 1, notice that, while the filtered probabilities (1st and 2nd columns) are nearly always very close, there are some differences in the smoothing columns (3rd and 4th columns). These differences are distinguished in the period 1981:2–3. Smoothed probabilities of boom during that period are less than 0.5 in Lam's case, but they are 0.665 and 0.758 in my case. However, if we consider the actual average output growth rates during that period (0.05%) as compared to those during the recession and boom periods detected by Lam (– 1.264% and 0.916%, respectively), the inferences based on my state–space model are not unreasonable. The period 1981:2–3 was a period of minor recession. Due to the approximations employed, the state–state model results in somewhat inconclusive inferences during that period.

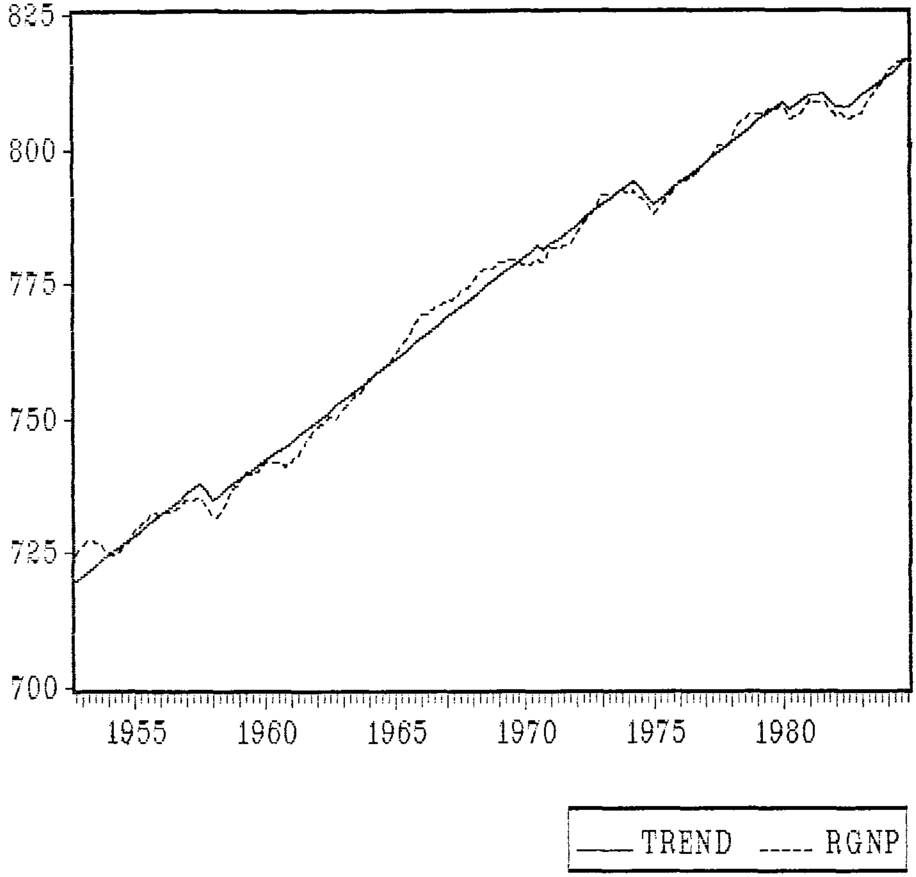

Fig. 1. Quarterly real GNP vs. state–space estimation of its trend component.

Table 3

Comparison of computation time: Lam's (1990) algorithm and our state–space algorithm.[a]

	Lam (1990)	State–space
One pass through basic filter	15 seconds	3 seconds
Full-sample smoothing	57 hours and 11 minutes	9 seconds

[a](1) Programming language used is GAUSS. (2) An 80386 IBM compatible PC (25 Mhz) with a math co-processor is used.

Thus, the number of cases to be evaluated is $2^r \times (t-r)$ for each t ($t = r+1, \dots, T$), where r is the order of the autoregressive process in (3.3). As $r = 2$ and $T = 131$ in our example, the total number of cases to be evaluated for all t ($t = 3, \dots, T$) is 33,540 for Lam's algorithm and 516 for the state–space algorithm with an approximation introduced in this paper. Table 3 compares actual computation time between the two algorithms. For the basic filter, the state–space algorithm is only 5 times as efficient as Lam's algorithm in terms of computation time. This is because the state–space algorithm involves heavier computation per case to be evaluated than Lam's original algorithm. For the full-sample smoothing, however, the state–space algorithm is vastly more efficient than Lam's algorithm. While full-sample smoothing took only 9 seconds for the state–space algorithm, it took more than 57 hours for Lam's original algorithm without approximation.

In addition to the considerable advantage in the computation time, while Lam's original approach without an approximation may hardly be tractable when we assume a general M-state Markov process, our present approach based on an approximation can easily handle this case.

4. Summary and discussion

In this paper, Hamilton's (1988, 1989) Markov-switching model has been extended to the state–space models. This paper also complements Shumway and Stoffer's (1991) dynamic linear model with switching, by introducing dependence in the switching process and by allowing switching in both transition and measurement equations. The dynamic linear model with Markov-switching considered in this paper is a general one that includes ARIMA models and classical regression models as special cases. It also includes, as a special case, the Hamilton model with a general autoregressive component proposed by Lam (1990).

When we introduce Markov-switching to the measurement and transition equations of a state–space model, the estimation of the model is virtually intractable. This is because each iteration of the Kalman filtering in eqs. (2.6) through (2.12) produces an M-fold increase in the number of cases to consider, where M is the total number of states or regimes at each point in time. Thus, to make the estimation of the model tractable, the approximations similar to those introduced by Harrison and Stevens (1976) were employed.

The basic filtering, smoothing, and maximum likelihood estimation procedures presented in this paper operate quite well. To prove the effectiveness of these algorithms based on the approximation, they were applied to the state–space representation of Lam's (1990) generalized Hamilton model as an example. The estimation results and inferences on the unobserved states from the two different algorithms were close, with the algorithm here enjoying a considerable advantage in terms of the computation time. For more general state–space models, approximations to the Kalman filtering are unavoidable if the estimation of the models are to be tractable.

References

Andrews, D.W.K., 1990, Tests for parameter instability and structural change with unknown change point, Cowles Foundation discussion paper no. 943 (Yale University, New Haven, CT).

Bar-Shalom, Y., 1978, Tracking methods in a multi-target environment, IEEE Transactions on Automatic Control AC-23, 618–626.

Brown, R.L., J. Durbin, and J.M. Evans, 1975, Techniques for testing the constancy of regression relationships over time, Journal of the Royal Statistical Society B 37, 149–192.

Cecchetti, S.G., P.-S. Lam, and N.C. Mark, 1990, Mean reversion in equilibrium asset prices, American Economic Review 80, 398–418.

Chow, G., 1960, Tests of the equality between two sets of coefficients in two linear regressions, Econometrica 28, 561–605.

Chu, C.J., 1989, New tests for parameter constancy in stationary and nonstationary regression models (Department of Economics, University of California, San Diego, CA).

Cosslett, S.R. and L.-F. Lee, 1985, Serial correlation in latent discrete variable models, Journal of Econometrics 27, 79–97.

Engel, Charles and James Hamilton, 1990, Long swings in the dollar: Are they in the data and do markets know it?, American Economic Review 80, 689–713.

Engle, Robert F. and Mark W. Watson, 1985, The Kalman filter: Applications to forecasting and rational-expectations models, in: Truman Bewley, ed., Advances in econometrics, Vol. I, Fifth World Congress of the Econometric Society.

Farley, J.U. and M.J. Hinich, 1970, A test for a shifting slope coefficient in a linear model, Journal of the American Statistical Association 65, 1320–1329.

Garcia, R. and P. Perron, 1990, An analysis of the real interest rate under regime shifts, Mimeo. (Princeton University, Princeton, NY).

Goldfeld, S.M. and R.E. Quandt, 1973, A Markov model for switching regression, Journal of Econometrics 1, 3–16.

Gorden, K. and A.F.M. Smith, 1988, Modeling and monitoring discontinuous changes in time series, in: J.C. Spall, ed., Bayesian analysis of time series and dynamic linear models (Marcel Dekker, New York, NY) 359–392.

Hamilton, James, 1988, Rational expectations econometric analysis of changes in regimes: An investigation of the term structure of interest rates, Journal of Economic Dynamics and Control 12, 385–432.

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

Hamilton, James, 1991, States–space models, in: R. Engle and D. McFadden, eds., Handbook of econometrics, Vol. 4 (North-Holland, Amsterdam) forthcoming.

Harrison, P.J. and C.F. Stevens, 1976, Bayesian forecasting, Journal of the Royal Statistical Society B 38, 205–247.

Harvey, A.C., 1985, Applications of the Kalman filter in econometrics, in: Truman Bewley, ed., Advances in econometrics, Vol. I, Fifth World Congress of the Econometric Society.

Harvey, A.C., 1991, Forecasting, structural time series models and the Kalman filter (Cambridge University Press, Cambridge).

Highfield, R.A., 1990, Bayesian approaches to turning point prediction, in: 1990 Proceedings of the Business and Economics Section (American Statistical Association, Washington, DC) 89–98.

Kim, H.J. and D. Siegmund, 1989, The likelihood ratio test for a change-point in simple linear regression, Biometrika 76, 409–423.

Kim, I.-M. and G.S. Maddala, 1991, Multiple structural breaks and unit roots in the nominal and real exchange rates (Department of Economics, University of Florida, Gainesville, FL).

Kitagawa, Genshiro, 1987, Non-Gaussian state–space modeling of nonstationary time series, Journal of the American Statistical Association 82, no. 400, Theory and Method, 1032–1041.

Lam, Pok-sang, 1990, The Hamilton model with a general autoregressive component: Estimation and comparison with other models of economic time series, Journal of Monetary Economics 26, 409–432.

Neftci, S.N., 1984, Are economic time series asymmetric over the business cycle?, Journal of Political Economy 92, 306–328.

Ploberger, W., W. Kramer, and K. Kontrus, 1989, A new test for structural stability in the linear regression model, Journal of Econometrics 40, 307–318.

Quandt, R.E., 1958, The estimation of the parameters of a linear regression system obeying two separate regimes, Journal of the American Statistical Association 53, 873–880.

Quandt, R.E., 1960, Tests of the hypothesis that a linear regression system obeys two separate regimes, Journal of the American Statistical Association 55, 324–330.

Quandt, R.E., 1972, A new approach to estimating switching regressions, Journal of the American Statistical Association 67, 306–310.

Sclove, S.L., 1983, Time-series segmentation: A model and a method, Information Sciences 29, 7–25.

Shumway, R.H. and D.S. Stoffer, 1991, Dynamic linear models with switching, Journal of the American Statistical Association, forthcoming.

Smith, A.F.M. and U.E. Makov, 1980, Bayesian detection and estimation of jumps in linear systems, in: O.L.R. Jacobs, M.H.A. Davis, M.A.H. Dempster, C.J. Harris, and P.C. Parks, eds., Analysis and optimization of stochastic systems (Academic Press, New York, NY) 333–345.

Watson, Mark W. and Robert F. Engle, 1983, Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models, Journal of Econometrics 23, 385–400.

Wecker, W.E., 1979, Predicting the turning points of a time series, Journal of Business 52, 35–50.