ISSN 1440-771X
ISBN 0 7326 1079 6

# A Structural Time Series Model with Markov Switching

**Roland G. Shami and Catherine S. Forbes**

**Working Paper 10/2000**

**December 2000**

**DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS**

MONASH UNIVERSITY

**AUSTRALIA**

# A Structural Time Series Model

# with Markov Switching


## Roland G. Shami and Catherine S. Forbes



**Abstract:** We propose an innovations form of the structural model underlying exponential smoothing that is further augmented by a latent Markov switching process. A particular case of the new model is the local level model with a switching drift, where the switching component describes the change between high and low growth rate periods. This new model is used to analyse the US business cycle using US Quarterly real GNP data. Model parameters are estimated using a Gibbs sampling algorithm and subsequently used for forecasting purposes. In addition, the stability of the new model is tested against Hamilton's model over a range of observation periods.



**Key words:** Structural model - Markov switching regime - Gibbs sampling - Business cycle.

**JEL Classification:** C11, C22, C53

# *1. Introduction*

Hamilton (1989) proposed a model for the analysis of an observed time series subject to irregular transitions between two different regimes, and applied his model to characterise the US Business cycle using US GNP data. His model uses a simple autoregressive structure to characterise the evolution of the observed series whose conditional mean is determined by a latent, binary Markov switching variable that takes a value of unity during expansionary periods and a value of zero during recessions. Hamilton also provided an algorithm for estimating the probability of a recession at each time period based on a maximum likelihood approach. Since that time, several other authors have investigated modifications to the model specification (Lam, 1990; Hansen, 1992; Kim, 1994), computation of the recession probabilities (Albert and Chib, 1993) and the application of the models to various other data sources (Cecchetti et al, 1990; Hamilton and Lin, 1996). In this paper we propose a new model based on the structural time series model underlying exponential smoothing (Snyder, 1985; Forbes, Snyder and Shami, 2000) that is augmented by a latent binary switching variable. We call this model the switching structural model (SSM).

The characterisation of an economic time series using linear structural models (LSM) is based on a traditional decomposition of the observed series into level, growth, seasonal and irregular components (Harvey, 1984). These unobserved components are assumed to evolve dynamically according to a linear relationship, traditionally made stochastic by the inclusion of an additive error term that is uncorrelated with the observation error. Statistical analysis of an observed time series using this traditional state space model form requires the use of a Kalman filter.

Another equally general state space framework involves only a single source of error (Snyder, 1985; Ord, Koehler and Snyder, 1997). Called the innovations form by Aoki (1987), the calculation of the likelihood function for this model requires simple exponential smoothing methods rather than the more cumbersome Kalman filter. It also has a more direct equivalence relationship to the popular ARIMA models than does the traditional LSM (Shami and Snyder, 1998).

Linear state space models in both the traditional form (Harvey, 1985; Watson, 1986; and Clark, 1987) and the innovations form (Aoki, 1988 and 1993) have been used to characterise economic time series. Notably, Harvey (1985) uses the LSM on US GNP data to analyse the business cycle. However, as Hamilton (1989) suggests, activities during the business cycle's expansionary phase may be different from those that take place during the recessionary phase. As such, nonlinear models are used in preference to linear models to characterise this distinction.

In addition to proposing the new SSM model, we provide a complete approach to computing a Bayesian analysis of the proposed model. The Gibbs sampling (Gelfand and Smith, 1990) based algorithm builds on the work of Forbes, Snyder and Shami (2000), who demonstrate the use of Monte Carlo composition to compute Bayesian posterior parameter and forecasting distributions for the linear structural model based on the innovations form. Others, notably Albert and Chib (1993), Kim and Nelson (1999) and Luginbuhl and De Vos (1999) have used Bayesian methods on various traditional switching models.

The plan of the paper is as follows. The new model is presented in Section 2, along with a brief description of the conditional distribution of the data. An outline of the Bayesian approach and our prior distributional assumptions are given in Section 3. Also, an algorithm for producing a sample from the joint posterior distribution of the parameters in the model, along with a method for calculating the marginal posterior switching probabilities is presented. The new model and technique are applied to US quarterly real GNP data from 1954 to 1984 in Section 4, and the stability of the model is discussed in Section 5. The paper concludes with a brief discussion and directions for further research.

## 2. A Switching Structural Model

Structural time series models can provide useful descriptions of business and economic series by decomposing the observed series at different time points into level, growth, seasonal and irregular components. To handle the statistical aspects of the structural time series models, we put them in state space forms. The innovations form state space model, called the single source of error (SSOE) model in Forbes, Snyder and Shami (2000), is given by the two equations

$$y_t = x'b_{t-1} + e_t \qquad \text{measurement equation} \qquad (1)$$

$$b_t = Tb_{t-1} + \alpha e_t \qquad \text{transition equation} \qquad (2)$$

where $x$ is a fixed $k$-vector and $T$ is a $k$x$k$ transition matrix typically taken to contain known constants. The $e_t$'s are assumed to be independent normally distributed disturbances with common mean zero and variance $\sigma^2$, and $\alpha$ is a $k$-vector of 'smoothing' parameters. Here, the value of the observed series at time $t$, $y_t$, is described as arising from a known linear combination of unobservable components, $b_{t-1}$ and an independent disturbance term, $e_t$. This SSOE model is linear, and provides a statistical framework encompassing general exponential smoothing.

One special case of this general SSOE model is the structural model containing a local level and constant growth term, $g_0$,

$$y_t = l_{t-1} + g_0 + e_t \qquad (3)$$

$$l_t = l_{t-1} + g_0 + \alpha_l e_t \qquad (4)$$

which can be put into the general SSOE state space form by taking $b_t' = (l_t, g_0)$, $b_0' = (l_0, g_0)$, $x' = (1,1)$, $T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, and $\alpha' = (\alpha_l, 0)$. The model with a constant growth term, however, does not seem to adequately describe business cycle fluctuations, and so it seems useful to adjust the model by introducing switching behaviour into the growth coefficient to distinguish between periods of contraction and expansion. A new nonlinear structural model, which we call the Switching Structural Model (SSM), is defined now by the equations

$$y_t = l_{t-1} + g_{t-1} + e_t \tag{5}$$

$$l_t = l_{t-1} + g_{t-1} + \alpha e_t \tag{6}$$

$$g_t = \mu_1 s_t + \mu_0 , \tag{7}$$

where $y_t$ is the observed value at time $t$, $l_t$ represents the unobserved level at time $t$, $g_t$ is the unobserved growth at time $t$, the $e_t$'s are independent and normally distributed disturbances with mean 0 and variance $\sigma^2$, and $\alpha$ is the level smoothing parameter. Now the growth coefficient at time $t$, $g_t$, depends on an unobserved random variable that determines the 'state' of the system (or economy) at time $t$, $s_t$. This variable $s_t$ takes a value of 1 or 0 according to a Markov chain of order 1. Importantly, it is also assumed that when given the value of the true state during the last time period, $s_{t-1}$, the state variable at time $t$, $s_t$, is independent of the observations $Y_t = \left(y_1, y_2, \ldots, y_t\right)'$. Hence we have

$$P(s_t = 1 | s_{t-1} = 1) = p ,$$

$$P(s_t = 0 | s_{t-1} = 0) = q ,$$

and

$$P(s_t | s_{t-1}, \cdots, s_1, Y_t) = P(s_t | s_{t-1}) . \tag{8}$$

The transition probabilities can be represented by a matrix of the form

$$M = \begin{bmatrix} p & 1-q \\ 1-p & q \end{bmatrix}, \tag{9}$$

where $m_{ij}$ gives the probability $P(S_t = 2-i | S_{t-1} = 2-j)$. Also, we take as the initial state probabilities

$$P\left(s_0 = 1\right) = \frac{1-q}{2-p-q} \quad \text{and} \quad P\left(s_0 = 0\right) = \frac{1-p}{2-p-q}, \tag{10}$$

which are the ergodic probabilities of the Markov chain component of the model. Note that due to the nonconstant growth term, the model is no longer linear and therefore cannot be put into the SSOE form given by (1) and (2).

This idea of using a Markov chain with two states to represent the expansion and contraction phases of the business cycle was introduced by Hamilton (1989). The parameters $\mu_1$ and $\mu_0$ together define the two levels of growth. During an expansion, $s_t$ = 1 and the growth rate is given by $g_t = \mu_1 + \mu_0$, whereas during a recession, $s_t$ = 0 and the growth rate is given by $g_t = \mu_0$. As we require 'expansion' to have a higher growth rate than 'recession', we impose the constraint $\mu_1 > 0$.

The difference between Hamilton's model and the SSM is that in the former model, the observation is regressed on a limited number of past observations and the latter exploits the property of exponential smoothing by regressing on the whole set of observations weighted by a decaying parameter. Also the number of parameters to estimate decreases from nine in Hamilton (1989) to seven in the SSM.

The likelihood function for this SSM model can be constructed from consideration of the joint probability distribution of the observed data and unobserved state variables. For convenience, we collect the model parameters into two groups, $\lambda_1 = \left(l_0, \sigma^2, \alpha\right)'$ and $\lambda_2 = \left(\mu_0, \mu_1, p, q\right)'$. Partitioning the parameter set into these two blocks is convenient because $\lambda_1$ is associated with the linear structural model corresponding to exponential smoothing with a constant growth term, whereas $\lambda_2$ is associated with the Markov switching component.

The joint probability of the observed data, $Y_n = \left(y_1, y_2, \ldots, y_n\right)'$, and the unobserved state vector, $S_{n-1} = \left(s_0, s_1, \ldots, s_{n-1}\right)'$, given $\lambda' = \left(\lambda_1', \lambda_2'\right)$ has the form

$$f\left(Y_n, S_{n-1}|\lambda\right) = f\left(y_1|s_0, \lambda\right) f\left(s_0|\lambda\right) \prod_{t=2}^{n} f\left(y_t|Y_{t-1}, S_{t-1}, \lambda\right) f\left(s_{t-1}|S_{t-2}, \lambda\right). \quad (11)$$

The likelihood function of the parameters, $L(\lambda|Y_n) = p(Y_n|\lambda)$, is calculated for any particular value of $\lambda$ by averaging (11) over all possible $2^n$ values of $S_{n-1}$.

We will show how to compute the Bayesian posterior probability distribution in the next section utilising the special structure of model. In particular, for fixed values of $\lambda_2$ and $S_{t-1}$, the model retains the essential features of an SSOE model, and hence $f(y_t|Y_{t-1}, S_{t-1}, \lambda)$ can be computed using a minor modification of the procedure given in Forbes, Snyder and Shami (2000). Their procedure is based on a transformation from (5) and (6) into a reduced form regression relationship, which we now detail.

From the SSM measurement equation in (5) and conditional on $\lambda$, $S_{t-1}$ and $Y_{t-1}$, the distribution of $y_t$ is normal with mean $l_{t-1} + g_{t-1}$ and variance $\sigma^2$. Let $\delta = 1 - \alpha$. By substituting the value of the noise term, $e_t = (y_t - l_{t-1} - g_{t-1})$, from the measurement equation (5) into the level transition equation (6) yields

$$l_t = \delta l_{t-1} + \delta g_{t-1} + \alpha y_t . \tag{12}$$

Backsolving to time $t$ = 1, we obtain

$$l_t = \delta^t l_0 + \sum_{j=1}^{t} \delta^j g_{t-j} + \sum_{j=0}^{t-1} \alpha \delta^j y_{t-j} , \tag{13}$$

where $g_0 = \mu_1 s_0 + \mu_0$, and hence

$$y_t = \delta^{t-1} l_0 + \sum_{j=1}^{t} \delta^{j-1} g_{t-j} + \sum_{j=1}^{t-1} \alpha \delta^{j-1} y_{t-j} + e_t . \tag{14}$$

Note (14) can be conveniently rearranged as

$$\tilde{y}_t = \tilde{x}_t l_0 + e_t , \tag{15}$$

where

$$\tilde{y}_t = y_t - \sum_{j=1}^{t} \delta^{j-1} g_{t-j} - \sum_{j=1}^{t-1} \alpha \delta^{j-1} y_{t-j} \quad \text{and} \quad \tilde{x}_t = \delta^{t-1} . \tag{16}$$

Computing these transformed values is fast due to the recurrence relationships $\tilde{y}_{t+1} = y_{t+1} - y_t + \delta\tilde{y}_t$ - $\mu_1 s_t$ - $\mu_0$ and $\tilde{x}_{t+1} = \delta\tilde{x}_t$. Therefore, $f\left(y_t | Y_{t-1}, S_{t-1}, \lambda\right)$ can be computed in (11) using

$$f\left(y_t | Y_{t-1}, S_{t-1}, \lambda\right) \propto \sigma^{-1} \exp\left\{-\frac{1}{2\sigma^2}\left(\tilde{y}_t - \tilde{x}_t l_0\right)^2\right\}. \tag{17}$$

# 3. A Bayesian Analysis

To complete a Bayesian analysis, a joint prior distribution for the unknown parameters must be specified. As parameters will be sampled in blocks, we specify the general form of

$$P(\lambda_1, \lambda_2) \propto P(\lambda_1) P(\lambda_2). \tag{18}$$

We follow Forbes, Snyder and Shami (2000) by imposing

$$P(\lambda_1) = P\left(l_0, \alpha, \sigma^2\right) \propto \sigma^{-2} P(\alpha), \tag{19}$$

for $-\infty < l_0 < \infty$, $\sigma^2 > 0$ and $0 < \alpha < 2$. The limits of $\alpha$ are derived by writing the SSM as an ARIMA model. By taking the first difference of the level in (6) and substituting the result into the first difference of (5), we obtain an ARIMA(0,1,1) with drift. The moving average coefficient is equal to $\alpha - 1$. The invertibility condition of the ARIMA process leads us to impose that the absolute value of $\alpha - 1$ should be less than one which translates into the constraint $0 < \alpha < 2$.

As the algorithm we detail is not sensitive to the choice of $P(\alpha)$, we leave the notation general at this stage. In our example, we choose a uniform distribution so that $P(\alpha) = 0.5$ for all $0 < \alpha < 2$. The marginal prior for $\lambda_2$ is chosen to simplify the Gibbs sampling algorithm we propose, with

$$P(\lambda_2) = g\left(p; u_{11}, u_{10}\right) g\left(q; u_{00}, u_{01}\right) I_{\mu_1 > 0}, \tag{20}$$

where

$$g(x; u, v) = \frac{\Gamma(u+v)}{\Gamma(u)\Gamma(v)} x^{u-1} (1-x)^{v-1} I_{0<x<1} \text{ with } u > 0, v > 0. \tag{21}$$

That is, *p* and *q* are assumed to have independent Beta marginal prior distributions and the marginal prior for the growth coefficients is (improper and) uniform over the region where $\mu_0 < \mu_0 + \mu_1$. This restriction is imposed so that observations at times corresponding to an expansion have a higher growth rate than those corresponding to a recession. In our example, we choose both of Beta prior distributions as uniform distributions, corresponding to $u_{11} = u_{10} = u_{00} = u_{01} = 1$.

Taking the likelihood function, as discussed in Section 2, and the above joint prior distribution, we can construct the posterior distribution for the unknown parameters contained in $\lambda$ using to Bayes' theorem

$$P(\lambda | Y_n) = \frac{L(\lambda | Y_n) P(\lambda)}{P(Y_n)}. \quad (22)$$

However, direct Bayesian inference about $\lambda$ in the SSM is not an attractive option, since it entails the computation of the complicated likelihood. As there is no available analytical expression for the posterior distribution, we resort to using a Gibbs sampling simulation method to sample from this joint posterior. Indeed, we actually obtain a sample from the joint posterior distribution of $(\lambda, S_{n-1})$, and marginalise appropriately to obtain useful summaries of the distribution of these variables conditionally on the observed data.

Gibbs sampling (Gelfand and Smith, 1990; Geweke, 1999) traditionally refers to the process of repeated sequential sampling of each unknown, typically univariate, parameter in a model from its complete conditional posterior distribution; i.e. the distribution of that parameter given previously sampled values of all other unknown parameters and the observed data. Under some regularity conditions (Tierney, 1994) this sample of parameter values, although not independent, does converge to its full joint posterior distribution. It has been shown (Liu, Wong and Kong, 1994) that Gibbs sampling schemes that collect together individual parameters into groups, or blocks, are more efficient as they generally reduce the correlation between successive sampled parameter values.

The sampling scheme we suggest here utilises fully the special structure of the model and prior distribution by sampling from the following distributions

$$P(\lambda_1 | Y_n, S_{n-1}, \lambda_2), \tag{23}$$

$$P(S_{n-1} | Y_n, \lambda_1, \lambda_2), \tag{24}$$

$$P(\lambda_2 | Y_n, S_{n-1}, \lambda_1). \tag{25}$$

We now consider each of these in detail.

**Complete Conditional Distribution of $\lambda_1$**

As discussed in Section 2, when all of the switching states are known the model reduces to a minor variant of the linear model considered by Forbes, Snyder and Shami (2000) and Bayesian inference is easy to execute. In particular, it can be shown that

$$P(\lambda_1 | Y_n, S_{n-1}, \lambda_2) \propto P(\alpha | Y_n, S_{n-1}, \lambda_2) P(\sigma^2 | Y_n, S_{n-1}, \lambda_2, \alpha) \\ .P(l_0 | Y_n, S_{n-1}, \lambda_2, \alpha, \sigma^2). \tag{26}$$

Thus, a sampled value from the complete conditional distribution of $\lambda_1$ may be sampled using simple composition. First sample a value of $0 < \alpha < 2$ with probabilities given by

$$P(\alpha | Y_n, S_{n-1}, \lambda_2) \propto \left( \sum_{t=0}^{n-1} \delta^{2t} \right)^{-1/2} SSE^{-(n-1)/2}, \tag{27}$$

with

$$SSE = \sum_{t=1}^{n} \left( \tilde{y}_t - \delta^{t-1} \hat{l}_0 \right)^2 \text{ and } \hat{l}_0 = \left( \sum_{t=0}^{n-1} \delta^{2t} \right)^{-1} \sum_{t=1}^{n} \delta^{t-1} \tilde{y}_t . \tag{28}$$

No analytical normalising constant is available. However, a sampled value can be obtained by evaluating (27) numerically on a grid of J+1 points $\alpha_j$ over $0 < \alpha_j < 2$, normalising, and using an inverse cumulative distribution function approach to sample $\alpha$.

Having obtained this value of $\alpha$, the conditional distribution for $\sigma^2$ can be shown to be an inverted gamma distribution with shape parameter $a = (N-1)/2$ and scale parameter $c = 2/SSE$, so that

$$P\left(\sigma^2 | \alpha, Y_n, S_{n-1}, \mu_1, \mu_0\right) = \Gamma\left(a\right)^{-1} c^{-a} \sigma^{2(a+1)} \exp\left\{-1/c\sigma^2\right\}. \quad (29)$$

To sample a value for $\sigma^2$, we sample a random variable, $x$, from a gamma distribution with shape parameter $a$ and unit scale parameter, and then compute $\sigma^2 = \left(cx\right)^{-1}$.

Finally, given values of $\alpha$ and $\sigma^2$, $l_0$ can be shown to have a univariate normal distribution with mean $\hat{l}_0$ and variance $\hat{V}_{l_0}$, where

$$\hat{V}_{l_0} = \sigma^2 \left[\sum_{t=0}^{n-1} \delta^{2t}\right]^{-1} = \sigma^2 \frac{1-\delta^{2n}}{1-\delta^2}. \quad (30)$$

**Complete Conditional Distribution of $S_{n-1}$**

We generate the entire vector of switching states as a block from its complete conditional distribution

$$P(S_{n-1} | Y_n, \lambda) = P(s_{n-1} | Y_n, \lambda) \prod_{t=0}^{n-2} P(s_t | Y_n, \lambda, s_{t+1}). \quad (31)$$

Looking at any term from the product of the right hand side of the equation, $P(s_t | Y_n, \lambda, s_{t+1})$; conditional on $Y_{t+1}$, $\left(Y_{t+2}, \cdots, Y_n\right)$ contains no information about $s_t$ beyond that contained in $Y_{t+1}$. Hence,

$$P(s_t | Y_n, \lambda, s_{t+1}) = P(s_t | Y_{t+1}, \lambda, s_{t+1}),$$

and the complete conditional distribution for $S_{n-1}$ can be written

$$P(S_{n-1} | Y_n, \lambda) = P(s_{n-1} | Y_n, \lambda) \prod_{t=0}^{n-2} P(s_t | Y_{t+1}, \lambda, s_{t+1}). \quad (32)$$

Although the required computations are different, we follow the basic strategy of Carter and Kohn (1994) by first filtering forward recursively to obtain the probabilities $P\left(s_t | Y_{t+1}, \lambda\right)$ until the final state

$$P(s_{n-1}|Y_n, \lambda), \tag{33}$$

and subsequently sampling in reverse order using the smoothed probabilities

$$P(s_{n-t}|Y_{t+1}, \lambda, s_{n-t+1}). \tag{34}$$

Before demonstrating how to compute (33) and (34), we recall for notational convenience, that the level at time $t$ can be described by a linear combination of past data, $Y_t$, the initial level, $l_0$, and the past growth values, $G_{t-1} = (g_0, g_1, \ldots, g_{t-1})'$ according to

$$l_t = \delta^t l_0 + \sum_{j=1}^{t} \delta^j g_{t-j} + \sum_{j=0}^{t-1} \delta^j \alpha y_{t-j}. \tag{35}$$

Hence, conditional on all of the values $(Y_t, S_{t-1}, \lambda)$, the level vector, $L_{t-1} = (l_1, l_2, \ldots, l_{t-1})'$, can be computed exactly.

**Forward Filtering Probabilities**

Let $Y_t$ be the observed data up to time $t$, $L_t$ be the level vector up to time $t$ and $S_t$ be the switching vector up to time $t$ including the initial value $s_0$. Following the spirit of Hamilton's approach (1989), filtering probabilities are derived here by an iterative procedure. The input value of the filter is the conditional probability of the state at time $t$, $P(s_t|Y_{t+1}, L_{t+1}, \lambda)$, and the output is the conditional probability of the state at time $t$+1, $P(s_{t+1}|Y_{t+2}, L_{t+2}, \lambda)$. Each of the input and the output values is a vector consisting of two elements, one for each regime. These two elements are probabilities and always sum to unity.

To set up the iteration, the procedure needs an initial value $P(s_0|Y_0, L_0, \lambda)$, which in the absence of observed data is equal to the unconditional probability $P(s_0|\lambda)$. These probabilities are computed from (10). The filtering algorithm proceeds as follows:

**Step 1:** Assume $P(s_t|Y_{t+1},L_{t+1},\lambda)$ is known. By adding another state $s_{t+1}$, calculate the joint probability of $(s_{t+1},s_t)$ conditional on currently available data $Y_{t+1}$, currently available level $L_{t+1}$, and parameter vector $\lambda$, using

$$P(s_{t+1},s_t|Y_{t+1},L_{t+1},\lambda)=P(s_{t+1}|s_t,Y_{t+1},L_{t+1},\lambda)P(s_t|Y_{t+1},L_{t+1},\lambda)\,. \tag{36}$$

However, $P(s_{t+1}|s_t,Y_{t+1},L_{t+1},\lambda)=P(s_{t+1}|s_t,\lambda)$ by assumption (2), and hence

$$P(s_{t+1},s_t|Y_{t+1},L_{t+1},\lambda)=P(s_{t+1}|s_t,\lambda)P(s_t|Y_{t+1},L_{t+1},\lambda)\,. \tag{37}$$

**Step 2:** Sum over the possible values of the state $s_t$,

$$P(s_{t+1}|Y_{t+1},L_{t+1},\lambda)=\sum_{s_t=0}^{1}P(s_{t+1},s_t|Y_{t+1},L_{t+1},\lambda)\,. \tag{38}$$

**Step 3:** Calculate the conditional distribution of $s_{t+1}$ given $Y_{t+2}$, $L_{t+2}$ and $\lambda$ using Bayes' theorem

$$P(s_{t+1}|Y_{t+2},L_{t+2},\lambda)=\frac{f(l_{t+2},y_{t+2}|s_{t+1},Y_{t+1},L_{t+1},\lambda)P(s_{t+1}|Y_{t+1},L_{t+1},\lambda)}{f(l_{t+2},y_{t+2}|Y_{t+1},L_{t+1},\lambda)}\,,$$

or

$$\begin{aligned}&P(s_{t+1}|Y_{t+2},L_{t+2},\lambda)\propto\\&\qquad f(l_{t+2}|y_{t+2},s_{t+1},l_{t+1},\lambda)f(y_{t+2}|s_{t+1},l_{t+1},\lambda)P(s_{t+1}|Y_{t+1},L_{t+1},\lambda),\end{aligned} \tag{39}$$

where the first term of the right hand side is degenerate, allowing the value to be derived from (12),

$$l_{t+2}=\alpha y_{t+2}+\delta(l_{t+1}+\mu_1 s_{t+1}+\mu_0)\,.$$

The second term is calculated from the measurement equation (5), $y_{t+2}=l_{t+1}+\mu_1 s_{t+1}+\mu_0+e_{t+2}$,

$$f(y_{t+2}|s_{t+1},l_{t+1},\lambda)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y_{t+2}-l_{t+1}-\mu_1 s_{t+1}-\mu_0)^2}{2\sigma^2}}\,, \tag{40}$$

and the third term from (38).

**Step 4:** Normalise the right hand side of (39), which is symbolised by $P^*(s_{t+1}|Y_{t+2},L_{t+2})$, to obtain the output

$$P(s_{t+1}|Y_{t+2},L_{t+2},\lambda)=\frac{P^*(s_{t+1}|Y_{t+2},L_{t+2})}{\sum_{s_{t+1}=0}^{1}P^*(s_{t+1}|Y_{t+2},L_{t+2})}\,. \tag{41}$$

Note here that the first iteration begins at step 3 and subsequent iterations at the first step.

**Backward Sampling Probabilities**

Once all the filtered probabilities $P(s_t|Y_{t+1}, L_{t+1}, \lambda)$ for all $t = 0, \cdots, n-1$ are known, a backward filtering and sampling procedure is needed to find the probabilities of $s_t$ conditional on $Y_{t+1}$, $L_{t+1}$ and $s_{t+1}$. An application of Bayes theorem gives

$$P(s_t|Y_{t+1}, L_{t+1}, s_{t+1}, \lambda) = \frac{P(s_{t+1}|s_t, Y_{t+1}, L_{t+1}, \lambda) P(s_t|Y_{t+1}, L_{t+1}, \lambda)}{P(s_{t+1}|Y_{t+1}, L_{t+1}, \lambda)}. \qquad (42)$$

The two probabilities in the numerator of (42) are available because $P(s_{t+1}|s_t, Y_{t+1}, L_{t+1}, \lambda) = P(s_{t+1}|s_t, \lambda)$ from (8), and $P(s_t|Y_{t+1}, L_{t+1}, \lambda)$ is obtained from the forward filter. Hence, (42) simplifies to

$$P(s_t|Y_{t+1}, L_{t+1}, s_{t+1}, \lambda) \propto P(s_{t+1}|s_t) P(s_t|Y_{t+1}, L_{t+1}, \lambda), \qquad (43)$$

the right hand side of which, symbolised by $P^*(s_t|Y_{t+1}, L_{t+1}, s_{t+1})$, can be normalised to obtain

$$P(s_t|Y_{t+1}, L_{t+1}, s_{t+1}, \lambda) = \frac{P^*(s_t|Y_{t+1}, L_{t+1}, s_{t+1})}{\sum_{s_t=0}^{1} P^*(s_t|Y_{t+1}, L_{t+1}, s_{t+1})}. \qquad (44)$$

Thus, after filtering forward to obtain $P(s_t|Y_{t+1}, L_{t+1}, \lambda)$ for $t = 0, 1, \ldots, n-1$, a sampled value of $s_{n-1}$ can obtained, and subsequently sampled values of $s_t$ can be obtained using the backward sampling probabilities in (44), for $t = n-2, \ldots, 0$.

**Complete Conditional Distribution of $\lambda_2$**

It can be shown that the complete conditional distribution of $\lambda_2$ can be decomposed into the product

$$P(\lambda_2|Y_n, S_{n-1}, \lambda_1) = P(\mu|Y_n, S_{n-1}, \lambda_1) P(p|Y_n, S_{n-1}, \lambda_1) P(q|Y_n, S_{n-1}, \lambda_1), \quad (45)$$

primarily due to the availability of $S_{n-1}$. Each of $\mu = (\mu_1, \mu_0)'$, $p$ and $q$ can then be sampled independently from the relevant distribution given $Y_n$, $S_{n-1}$ and $\lambda_1$.

Consider first the distribution for $\mu$. Substituting $g_{t\text{-}1}$ into the measurement equation in (5), we have

$$y_t - l_{t-1} = \mu_1 s_{t-1} + \mu_0 + e_t\,. \tag{46}$$

Given values of $\lambda_1, S_{n-1}$, this is a regression of the form

$$z_t = w_t{}' \mu + e_t\,, \tag{47}$$

where $z_t = y_t - l_{t-1}$, $w_t = (s_{t-1},1)'$. Under the assumption that $\mu_1 > 0$, standard Bayesian algebraic manipulations indicate that $\mu$ has a truncated normal distribution having density

$$p(\mu | Y_n, S_{n-1}, \lambda_1, p, q) \propto \exp\left\{-\tfrac{1}{2}\left(\mu - \mu_\mu\right)' \Sigma_\mu{}^{-1}\left(\mu - \mu_\mu\right)\right\} I_{\mu_1 > 0}\,, \tag{48}$$

where $\mu_\mu = (W'W)^{-1} W'Z$ and $\Sigma_\mu = (W'W)^{-1}\sigma^2$, with

$$W' = \left(w_1, w_2, \cdots, w_n\right) \text{ and } X' = \left(x_1, x_2, \cdots, x_n\right).$$

We next consider the transition probabilities $p$ and $q$. Conditional on $Y_n, S_{n-1}$, and $\lambda_1$, $p$ and $q$ each have a conjugate prior distribution as given by (21), and hence have independent beta posterior distributions

$$P(p, q | Y_n, S_{n-1}, \lambda_1) = g\left(p; v_p, w_p\right) g\left(q; v_q, w_q\right) \tag{49}$$

where $v_p = n_{11} + u_{11}, w_p = n_{10} + u_{10}$, $v_q = n_{00} + u_{00}, w_q = n_{01} + u_{01}$ and $n_{ij}$ is the number of transitions from $s_{t-1} = i$ to $s_t = j$.

**Computing Posterior Marginal Switching Probabilities**

Once the Gibbs sampler algorithm has converged, a sample of size $r$ from the posterior distribution is available and estimates of numerous features of the posterior are available. Forbes, Snyder and Shami (2000) detail how to obtain forecast distributions for the SSOE model, and those calculations can be directly extended for the SSM model.

Of particular interest is the posterior marginal switching probabilities $P(s_t|Y_N)$, which can be computed using Rao-Blackwellised estimators as follows

$$P(s_t = 1|Y_n) = \frac{1}{r}\sum_{k=1}^{r} P(s_t = 1|Y_n, \lambda^{(k)}), \tag{50}$$

where $P(s_t = 1|Y_n, \lambda^{(k)})$ is the $(t+1)^{th}$ element of the vector $P(S_{n-1} = 1|Y_n, \lambda^{(k)})$ and $\lambda^{(k)}$ is the sampled parameter value for $k = 1, 2, ..., r$.

**Computing Regime Duration**

The expected duration of an expansion is estimated using the sample of size $r$ of generated probabilities $p^{(k)}$. Let $D_1$ be the duration of an expansion beginning at time $t$. Its expected value is given by

$$E(D_1|s_t = 1) = \frac{1}{r}\sum_{k=1}^{r} 1/(1 - p^{(k)}), \tag{51}$$

where $p^{(k)}$ is the $k^{th}$ iteration of $p$ from the Gibbs sampler. Similarly, the expected value of $D_0$, the duration of a recession, is

$$E(D_0|s_t = 0) = \frac{1}{r}\sum_{k=1}^{r} 1/(1 - q^{(k)}). \tag{52}$$

# *4. Application to US GNP*

In this section the quarterly US real GNP from the first quarter of 1951 to the fourth quarter of 1984 will be modelled using an SSM and our algorithm detailed in the last section. The observations $y_t$ are the natural logarithm of real GNP multiplied by 100. The first 2000 iterations from the Gibbs sampling were discarded, although it appeared that convergence was obtained after only a few hundred iterations. An additional 5000 iterations were saved and used to draw inferences on the parameters. The estimated posterior means are shown in Table 1 below along with the standard error of the corresponding marginal posterior distributions.

| | Mean | Standard Error |
|---|---|---|
| $\alpha$ | 1.154 | 0.103 |
| $\sigma^2$ | 0.716 | 0.13 |
| $l_0$ | 714.853 | 0.953 |
| $\mu_1$ | 1.646 | 0.418 |
| $\mu_0$ | -0.566 | 0.466 |
| $p$ | 0.874 | 0.111 |
| $q$ | 0.594 | 0.158 |

Table 1 - Parameter estimates and standard errors - SSM

The smoothing parameter value of 1.15 is close to one. The equivalent ARIMA(0,1,1) of the SSM collapses to a random walk with switching drift. This is in line with previous studies suggesting existence of unit root in GNP time series (Nelson and Plosser; 1982, Stock and Watson; 1986, Perron and Phillips; 1987). The growth during the expansion period is around 1.08% ($\mu_1 - \mu_0$) and during the recessionary period is around -0.56% ($\mu_0$). The expected duration of expansions and recessions, which can be derived from (51) and (52) are equal to 11.72 and 2.97 quarters respectively.

The Rao-Blackwellised marginal posterior density estimates for the parameters are given in Figures 1 to 7. The marginal posterior distribution of $\alpha$ is symmetrically distributed around its estimated mean (Figure 1) as is the marginal posterior distribution for the initial level (Figure 3). Figure 2 shows a skewness in the distribution of $\sigma^2$ as do Figures 4 and 5 which display the marginal posterior distributions of the switching components $\mu_1$ and $\mu_0$. Moreover, the effect of truncating the Gaussian distribution appears to the left in Figure 4 and to the right in Figure 5. The marginal posterior distributions of *p* and *q* are given in Figures 6 and 7 and they also show skewness in their distributions.

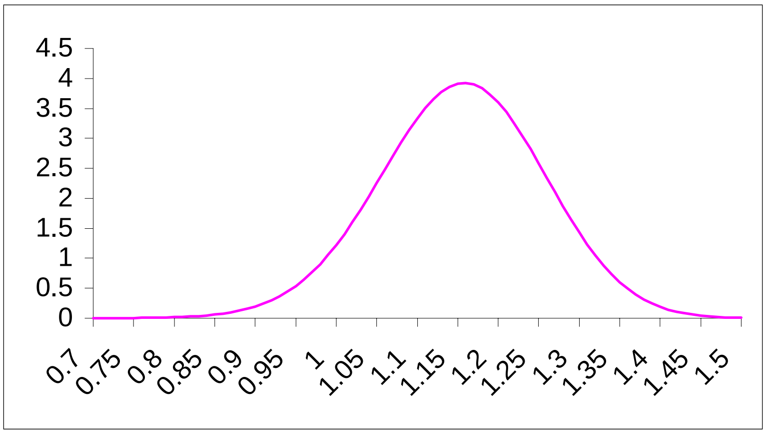


Figure 1 - Marginal Posterior Distribution of $\alpha$

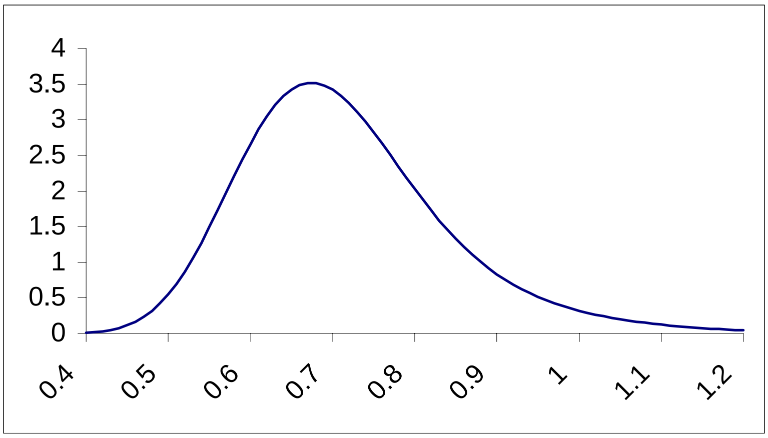


Figure 2 - Marginal Posterior Distribution of $\sigma^2$

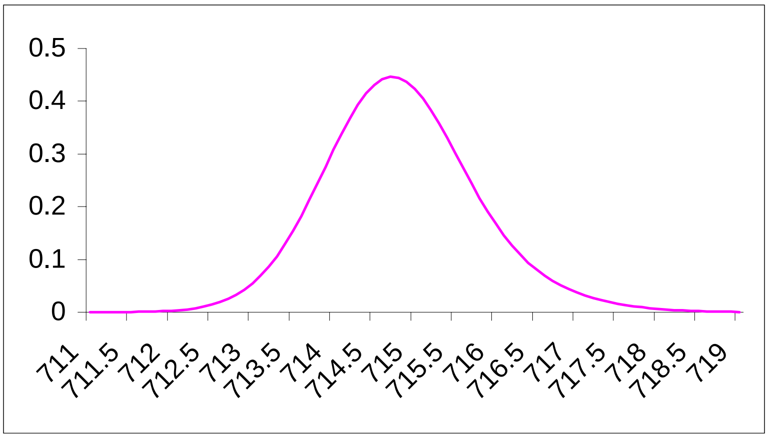


Figure 3 - Marginal Posterior Distribution of $l_0$

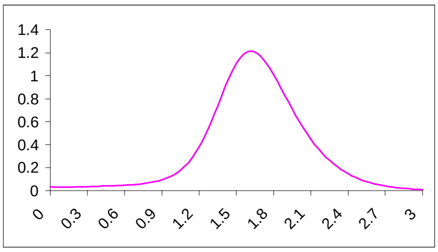


Figure 4 - Marginal Posterior Distribution of $\mu_1$

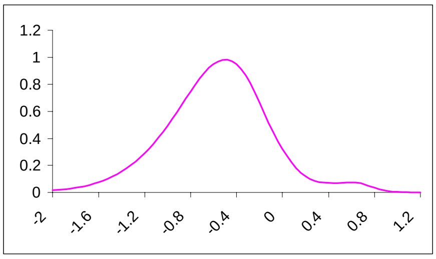


Figure 5 - Marginal Posterior Distribution of $\mu_0$

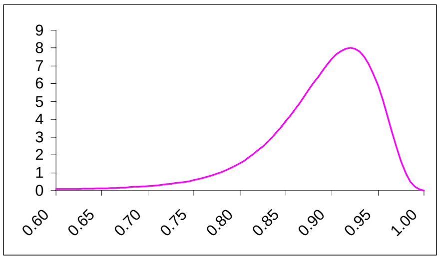


Figure 6 - Marginal Posterior Distribution of $p$

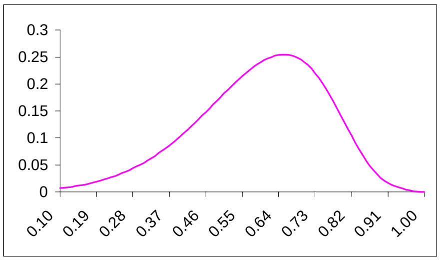


Figure 7 - Marginal Posterior Distribution of $q$

Figure 8 shows the observations $y_t$ (dark colour) and the estimated state vectors, which represent the levels $l_{t-1}$ (light colour). Notice that the level $l_t$ closely follows the observed data series, which is a main feature of a structural model.

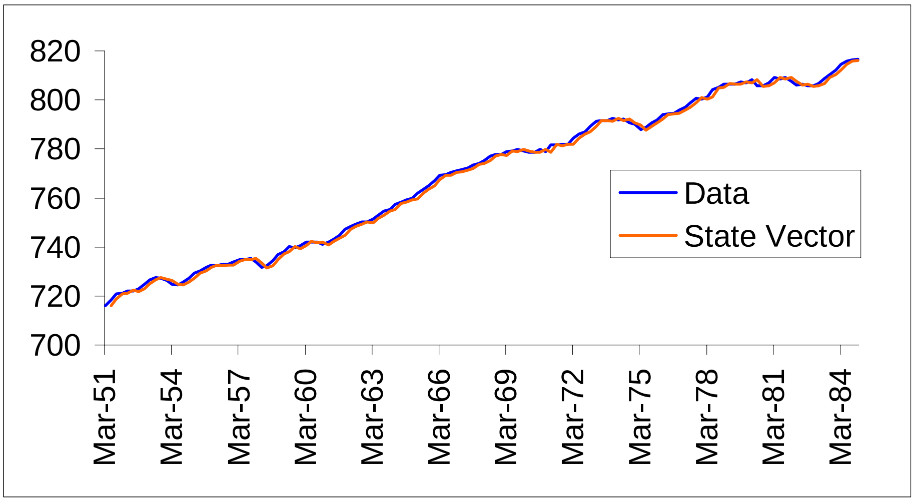


Figure 8 - Observations and levels.

Once the model is estimated, the parameters are used to forecast for 2 years ahead or 8 quarters up to the fourth quarter 1986. Rao-Blackwellised estimates are used to compute the mean and variance of the predictive values. The forecast distributions for all horizons are illustrated in Figure 9. These show how both the mean and the variance increase when the horizon increases. These distributions, which are a mixture of Gaussian distributions, appear symmetric.

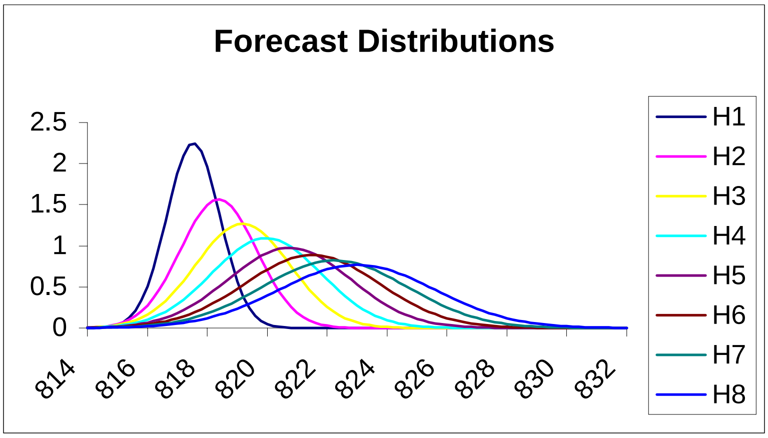


Figure 9 - Marginal posterior distribution of the forecasts up to 8 horizons.

Table 2 shows these predictions (Mean) along with the standard error of the corresponding posterior distributions (Std), 90% HPD intervals (L90% and U90%) and 95% HPD intervals (L95% and U95%) where HPD means highest posterior density. They are also illustrated in Figure 10 with the 90% and 95% HPDs.

| | Data | Mean | Std | L 95% | L 90% | U 90% | U 95% |
|---|---|---|---|---|---|---|---|
| H1 | 817.39 | 817.38 | 0.89 | 815.59 | 815.86 | 818.70 | 818.96 |
| H2 | 817.96 | 818.15 | 1.36 | 815.44 | 815.85 | 820.23 | 820.65 |
| H3 | 818.97 | 818.91 | 1.71 | 815.51 | 816.04 | 821.55 | 822.08 |
| H4 | 819.49 | 819.65 | 2.01 | 815.69 | 816.31 | 822.76 | 823.38 |
| H5 | 820.41 | 820.39 | 2.27 | 815.94 | 816.64 | 823.91 | 824.60 |
| H6 | 820.56 | 821.12 | 2.51 | 816.24 | 817.00 | 825.00 | 825.77 |
| H7 | 821.24 | 821.85 | 2.73 | 816.57 | 817.40 | 826.07 | 826.90 |
| H8 | 821.56 | 822.58 | 2.93 | 816.92 | 817.81 | 827.11 | 828.00 |

Table 2 - Observations and estimates - SSM

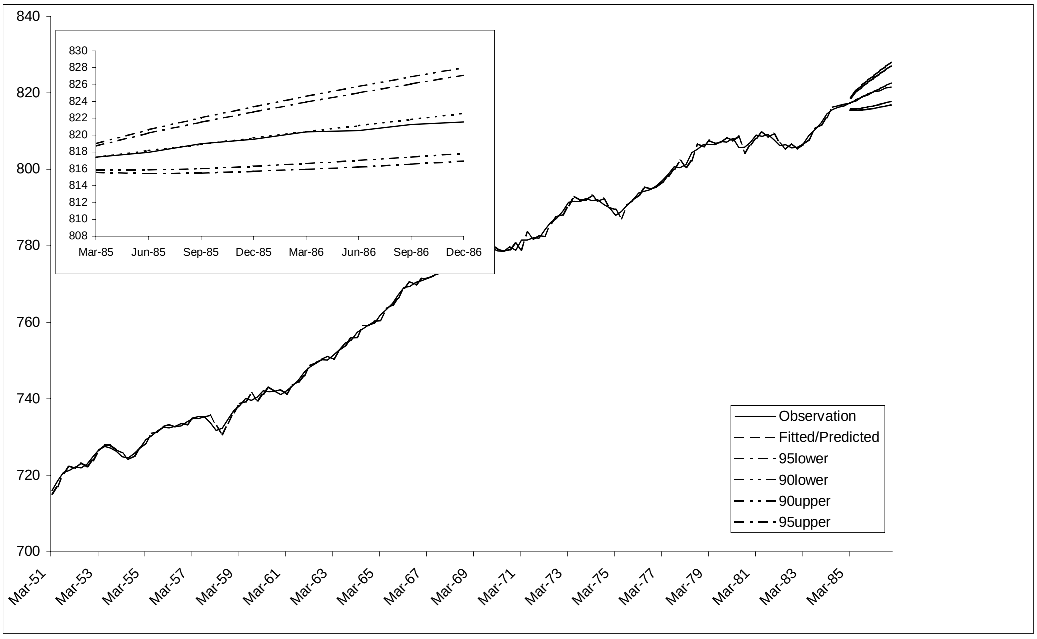


Figure 10 - Observations, estimates and 90% and 95% interval estimates

The forecasted values are well inside the HPD intervals which give an indication about the usefulness of this model in forecasting. Furthermore the one step forecasts are given in Table 3 and Figure 11.

| | Data | Estimate |
|---|---|---|
| Q1-1985 | 817.39 | 817.38 |
| Q2-1985 | 817.96 | 818.17 |
| Q3-1985 | 818.97 | 818.67 |
| Q4-1985 | 819.49 | 819.73 |
| Q1-1986 | 820.41 | 820.15 |
| Q2-1986 | 820.56 | 821.14 |
| Q3-1986 | 821.24 | 821.16 |
| Q4-1986 | 821.56 | 821.94 |

Table 3: Observations and one step ahead estimates

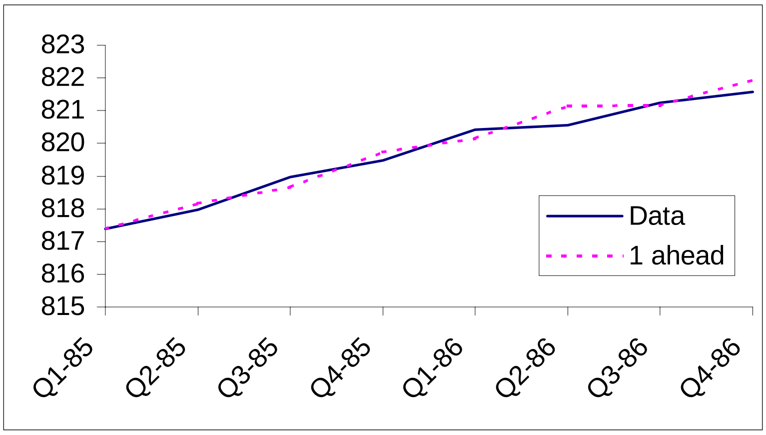


Figure 11: Observations and one step ahead estimates

The SSM model is not only used to estimate the population parameters and to forecast the values in the coming horizons, but it encompasses the estimation of the switching regime. In the case of GNP, the two regimes represent the expansion ($s_t$=1) and recession ($s_t$=0) of the business cycle. Figure 12 shows the marginal filtered probability $P(s_t=0|Y_N)$. The straight lines represent the trough and the peak times according to NBER.

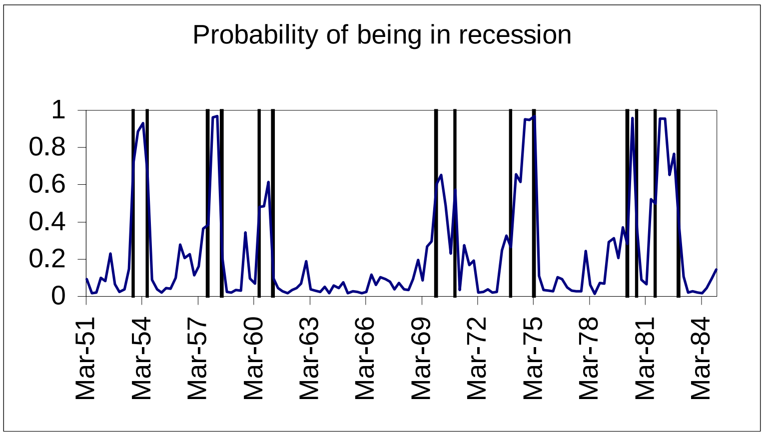


Figure 12 - $P(s_t=0|Y_{t-1})$.

## *5. Stability of the SSM Model*

One concern raised in the switching regime literature is model stability. That is, what are the consequences of applying a model to a shortened or lengthened series? In particular, if the estimates of the switching probabilities $P(s_t=0|Y_N)$ change greatly for a different sample size, then the model (or possibly the estimation method) is thought to be unstable. Hamilton's model has been criticised by many researchers for its instability, for example, Boldin (1996) observed a breakdown of the Hamilton model for data, which includes the end of World War II and the Korean War.

Also, Kim and Nelson (1999) found that Hamilton's model fails to provide reasonable inferences on the probabilities of a recession or a boom when the data set is extended until 1992 by using GDP. To correct for this, they added a dummy variable from 1983 to account for a structural break in the growth rate. This idea is in line with the structural models including the SSM, which allows for "structural breaks" at each $t$ through a change in the level $l_t$. In another application, by estimating Hamilton generalised model, Lam (1990) in using maximum likelihood approach and Kim (1994) in using state space forms and Kalman filter approach failed to capture all seven recession periods mentioned by the NBER. The model captures only five recession periods and the low growth phases were shorter than those of NBER recessions.

As a benchmark for the switching regime literature, Hamilton's model (1989) is applied to four different sets of GNP and estimated by a maximum likelihood estimation procedure. The first is the original set from 1951 to 1984 and the three others are extended sets from the original set backwards and forwards in time. Consequently, the second set runs from 1951 to 1986, the third set runs from 1947 to 1984 and the fourth set runs from 1947 to 1986. In addition, The same data sets have been analysed as SSM models and estimated by the Bayesian procedure proposed in this paper.

The results are shown in Figure 13 where the panels at left show Hamilton's model and the panels at right show SSM model. Note that there appears to be little difference between the two columns in the first row, as they both capture the behaviour of the business cycle almost identically to the dates set by the NBER. Nevertheless, differences appear in the subsequent rows, as the recession estimates given by the Hamilton model does not persist in other data sets. One of the reasons for such instability may be owed to the computational difficulty of maximising numerically an often ill-behaved likelihood surface with respect to a large number of unknown parameters (Hamilton, 1990). In contrast to this instability in Hamilton model, the SSM appears to behave well and captures the switching process in each of the GNP data set considered here.

Another way to compare different models is to evaluate the probability estimates. This can be done by many procedures. Here the two well known measures that are described in Diebold and Rudebush (1989) are used. The first is the quadratic probability score (QPS) defined by Brier (1950) and given by the following

$$QPS = \frac{1}{N}\sum_{t=1}^{N} 2(p_{e,t} - p_{o,t})^2 , \tag{53}$$

where $p_{e,t}$ is the estimated value of the probability at time $t$ and $p_{o,t}$ is the observed value. Like the usual mean squared error measure, the QPS provides a similar measure: a lower QPS implies that the prediction is more accurate. The other common measure is the log probability score (LPS), which is defined by

$$LPS = -\frac{1}{N}\sum_{t=1}^{N} [p_{o,t} \ln p_{e,t} + (1 - p_{o,t}) \ln(1 - p_{e,t})]. \tag{54}$$

Like QPS, a lower LPS implies that the prediction is more accurate. However, LPS penalises large mistakes more heavily than QPS, and while QPS is bounded by 0 and 2 ($0 < QPS < 2$), LPS has no upper bound ($0 < LPS < \infty$).

To compare Hamilton's model and the SSM, the QPS and LPS measures are used. The observed data used are the dates set by the NBER as contraction and expansion periods. Table 4 shows the results obtained for the different sets of data. Hamilton's model performs slightly better than the SSM on the original set (1951-1984) in both the QPR and LPS measures but is outclassed on the other three sets which is consistent with the results in Figure 13.

| | | Data Set | | | |
|---|---|---|---|---|---|
| | | 51-84 | 51-86 | 47-84 | 47-86 |
| QPS | Hamilton | 0.103 | 0.167 | 0.293 | 0.272 |
| | SSM | 0.119 | 0.109 | 0.229 | 0.152 |
| LPS | Hamilton | 0.177 | 0.274 | 0.558 | 0.516 |
| | SSM | 0.213 | 0.187 | 0.390 | 0.289 |

Table 4: Probability Scores

# *6. Conclusion*

In this paper, a switching structural model is proposed and the Bayesian analysis developed in Forbes, Snyder and Shami (2000) of linear state space model is extended to incorporate the switching part of the growth component. The algorithm, based on Gibbs sampler, uses a mixture of filtering and smoothing and Monte Carlo composition. Subsequently, the marginal distributions of the parameters and the forecast distributions are obtained.

The advantages of this approach include production of exact, small sample prediction distributions that are computed very quickly via Monte Carlo composition. The SSM was applied on quarterly US GNP data by defining the expansion and recession phases of the business cycle as the two switching states and its stability was also checked using other data time periods.

The promising results of the SSM support different variations to it. One extension is to look at the case where the switching parameters $\mu_1$ and $\mu_0$ follow a random walk similar to Luginbuhl and De Vos (1999) using an SSOE framework. Another extension to the SSM would be to explore the case where the transition probabilities are not constant, but are modelled instead as probabilities linked to different input data such as a leading indicator.

Hamilton SSM

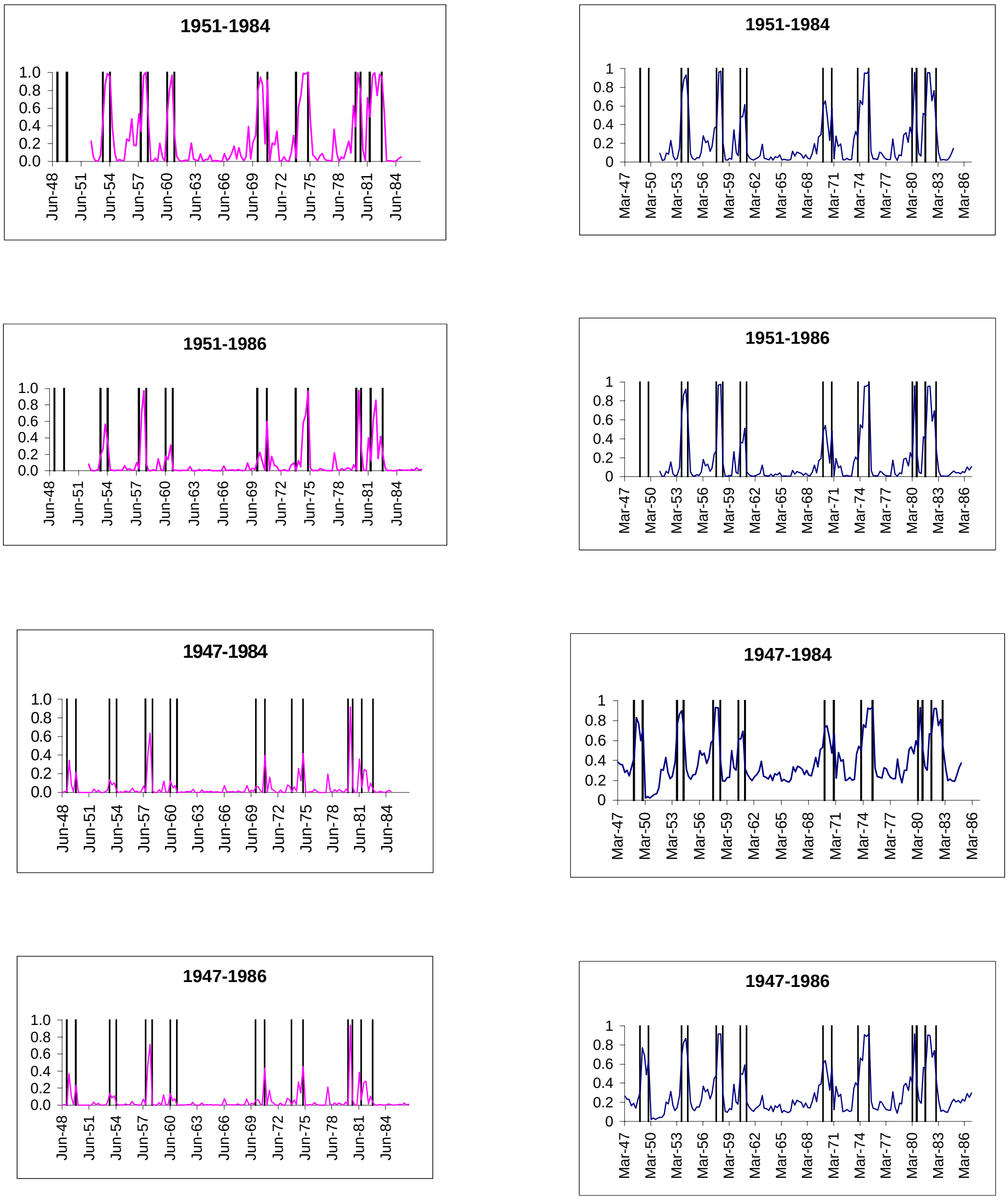


Figure 13 - $P(s_t=0|Y_{t-1})$.

# 7. References

Albert J. H. and Chib S. (1993): "Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts", *Journal of Business and Economic Statistics*, V11, pp 1-15.

Aoki M. (1987): "*State Space Modelling of Time Series*", Springer-Verlag, Berlin.

Aoki M. (1988): "On alternative state space representations of time series models", *Journal of Economic Dynamics and Control*, V12, pp 595-607.

Aoki M. (1993): "Interactions of real GNP business cycles in a three country time series model", *Journal of Forecasting*, V12, pp 331-344.

Boldin M. D. (1996): "A check on the robustness of Hamilton's Markov switching model approach to the economic analysis of the business cycle", *Studies in Nonlinear Dynamics and Econometrics*, V1, pp 35-46.

Brier G. W. (1950): "Verification of Forecasts Expressed in Terms of Probability", *Monthly Weather Review*, V75, pp 1-3.

Carter C. K. and Kohn R. (1994) "On Gibbs Sampling for State Space Models", *Biometrika*, V61, N3, pp 541-553.

Cecchetti S. G., Lam P. S. and Mark N. (1990): "Mean reversion in equilibrium asset prices", *The American Economic Review*, V80, pp 398-418.

Clark P. K. (1987): "The cyclical component of U.S. economic activity", *Quarterly Journal of Economics*, V102, pp 797-814.

Diebold F. X. and Rudebush G. D. (1989): "Scoring the Leading Indicators", *Journal of Business*, V62, pp 369-391.

Forbes C. S., Snyder R. D and Shami R. G. (2000): *Bayesian Exponential Smoothing*, Working paper 7/00, Department of Econometrics and Business Statistics, Monash University.

Gelfand A. E. and Smith A. F. M. (1990): "Sampling Based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association*, V85, pp 398-409.

Geweke J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication", *Economic Reviews*, V18, pp 1-126.

Hamilton J. D. (1989): “A new approach to the economic analysis of nonstationary time series and the business cycle”, *Econometrica*, V57, N2, pp 357-384.

Hamilton J. D. (1990): "Analysis of Time Series Subject to Changes in Regime", *Journal of Econometrics*, V45, pp 39-70.

Hamilton J. D. and Lin G. (1996): "Stock market volatility and the business cycle", *Journal of Applied Econometrics*, V11, pp 573-593.

Hansen B. E. (1992): “The Likelihood Ratio Test Under Nonstandard Conditions: Testing The Markov Switching Model of GNP”, *Journal of Applied Econometrics*, V7, pp S61-S82.

Harvey A. C. (1984): "A Unified View of Statistical Forecasting Procedures", *Journal of Forecasting* V3, pp 245-75.

Harvey A. C. (1985): “Trends and cycles in macroeconomic time series”, *Journal of Business and Economic Statistics*, V3, N3, pp 216-227.

Kim C.-J. (1994): "Dynamic Linear Models with Markov Switching " *Journal of Econometrics*, V60, pp 1-22.

Kim C.-J. and Nelson R. N. (1999) State Space Models with Regime Switching. The MIT Press, Cambridge

Lam P-S. (1990): "The Hamilton Model with a General Autoregressive Component" *Journal of Monetary Economics*, V26, pp 409-432.

Liu J., Wong W. H. and Kong A. (1994): "Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes", *Biometrika*, V81, pp 27-40.

Luginbuhl R. and De Vos A. (1999):“Bayesian Analysis of an Unobserved Component Time Series Model of GDP with Markov Switching and Time Varying Growths”, *Journal of Business and Economic Statistics*, V17, N4, pp 456-465.

Nelson C. R. and Plosser C. I. (1982): "Trends and random walks in macroeconomic time series: Some evidence and implications", *Journal of Monetary Economics*, V10, pp 139-162.

Ord J. K., Koehler A. B. and Snyder R. D. (1997): "Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models", *Journal of the American Statistical Association*, V92, pp 1621-29.

Perron P. and Phillips P. C. B. (1987): "Does GNP have a unit root? A re-evaluation", *Economic Letters*, V23, pp 139-145.

Shami R. G. and Snyder R. D. (1998): *Exponential Smoothing Methods of Forecasting and General ARMA Time Series Representations*, Working paper 3/98, Department of Econometrics and Business Statistics, Monash University.

Snyder R. D. (1985): "Recursive Estimation of Dynamic Linear Models", *The Journal of the Royal Statistical Society - Series B*, V47, pp 272-76.

Stock J. H. and Watson M. W. (1986): "Does GNP have a unit root? A re-evaluation", *Economic Letters*, V22, pp 147-151.

Tierney L. (1994): "Markov Chains for Exploring Posterior Distributions", *Annals of Statistics*, V22, N4, pp 1701-1728.

Watson M. W. (1986): "Univariate detrending methods with stochastic trends", *Journal of Monetary Economics*, V18, pp 49-75.