# 7

# Regime Switching with Time-Varying Transition Probabilities

FRANCIS X. DIEBOLD, JOON-HAENG LEE, AND GRETCHEN C. WEINBACH*

## 1. Introduction

Models incorporating nonlinearities associated with regime switching have a long tradition in empirical macroeconomics and dynamic econometrics.[1] Key methodological contributions include the early work of Quandt (1958) and Goldfeld and Quandt (1973) and the more recent work of Hamilton (1990). Recent substantive applications include Hamilton (1988) (interest rates), Hamilton (1989) (aggregate output), Cecchetti, Lam and Mark (1990) and Abel (1992) (stock returns), and Engel and Hamilton (1990) (exchange rates), among many others.

Our attention here focuses on Hamilton's Markov switching model, which has become very popular. In Hamilton's model, time-series dynamics are governed by a finite-dimensional parameter vector, which switches (potentially each period) depending upon which of two states is realised, with state transitions governed by a first-order Markov process with constant transition probabilities.

Although the popularity of Hamilton's model is well deserved, it nevertheless incorporates a potentially severely binding constraint, the constancy of state transition probabilities. Economic considerations suggest the desirability of allowing the transition probabilities to vary. As an example, consider the process of exchange rate revaluation. It is plausible that the likelihood of exchange rate revaluation increases under progressively more severe over- or undervaluation on the basis of economic fundamentals, and certainly, one would not want to exclude that possibility from the outset.

* Financial support from the U.K. Economic and Social Research Council under grant R000233447 is gratefully acknowledged by both authors. We are indebted to Neil Ericsson for helpful comments.

[1] For a survey of the nonlinear tradition in empirical macroeconomics, with particular attention paid to regime switching, see Diebold and Rudebusch (1993).

We therefore propose in this paper a class of Markov switching models in which the transition probabilities are endogenous.[2] We discuss the model in Section 2, develop an EM algorithm for parameter estimation in Section 3, and illustrate the methodology with a simulation example in Section 4. We conclude with a discussion of directions for future research in Section 5.

## 2. The Model

Let $\{s_t\}_{t=1}^{T}$ be the sample path of a first-order, two-state Markov process with transition probability matrix illustrated in Figure 7.1. As is apparent in the figure, the two transition probabilities are time-varying, evolving as logistic functions of $x'_{t-1}\beta_i$, $i = 0, 1$, where the $(k \times 1)$ conditioning vector $x_{t-1}$ contains economic variables that affect the state transition probabilities. It will be convenient to stack the two sets of parameters governing the transition probabilities into a $(2k \times 1)$ vector, $\beta = (\beta_0', \beta_1')'$.

It is obvious, but worth noting, that when the last $(k - 1)$ terms of the $(1 \times k)$ transition probability parameter vectors, $\beta_0$ and $\beta_1$, are set to zero, the transition probability functions are time-invariant so that $p_t^{00}$

$$
\begin{array}{lc}
 & \text{Time } t \\
 & \begin{array}{cc} \text{State 0} & \qquad\qquad\qquad\qquad \text{State 1} \end{array} \\
\begin{array}{c} \text{State 0} \\ \\ \\ \text{Time } t-1 \\ \\ \\ \text{State 1} \end{array} &
\left[\begin{array}{c:c}
p_t^{00} & p_t^{01} = (1 - p_t^{00}) \\
P(s_t = 0 \mid s_{t-1} = 0, x_{t-1}; \beta_0) & P(s_t = 1 \mid s_{t-1} = 0, x_{t-1}; \beta_0) \\
\dfrac{\exp(x'_{t-1}\beta_0)}{1 + \exp(x'_{t-1}\beta_0)} & 1 - \dfrac{\exp(x'_{t-1}\beta_0)}{1 + \exp(x'_{t-1}\beta_0)} \\
\hdashline
p_t^{10} = (1 - p_t^{11}) & p_t^{11} \\
P(s_t = 0 \mid s_{t-1} = 1, x_{t-1}; \beta_1) & P(s_t = 1 \mid s_{t-1} = 1, x_{t-1}; \beta_1) \\
1 - \dfrac{\exp(x'_{t-1}\beta_1)}{1 + \exp(x'_{t-1}\beta_1)} & \dfrac{\exp(x'_{t-1}\beta_1)}{1 + \exp(x'_{t-1}\beta_1)}
\end{array}\right]
\end{array}
$$

*Note:* $x_{t-1} = (1, x_{1,t-1}, \ldots, x_{(k-1),t-1})'$ and $\beta_i = (\beta_{i0}, \beta_{i1}, \ldots, \beta_{i(k-1)})'$, $i = 0, 1$.

Fig. 7.1. Transition Probability Matrix

[2] The first work in this area is Lee (1991), from which this paper draws. Related subsequent literature includes Fillardo (1991), who considers Markov-switching business-cycle models with transition probabilities that change with movements in an index of leading indicators, as well as Ghysels (1992) and De Toldi, Gourieroux and Monfort (1992), who consider duration models with hazard rates that vary across seasons.

and $p_t^{11}$ are simply constants; our model collapses to that of Hamilton (1990).

Let $\{y_t\}_{t=1}^T$ be the sample path of a time series that depends on $\{s_t\}_{t=1}^T$ as follows:

$$(y_t \mid s_t = i;\ \alpha_i) \overset{iid}{\sim} N(\mu_i, \sigma_i^2), \tag{1}$$

where $\alpha_i = (\mu_i, \sigma_i^2)'$, $i = 0, 1$.[3] Thus, the density of $y_t$ conditional upon $s_t$ is:

$$f(y_t \mid s_t = i;\ \alpha_i) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left( \frac{-(y_t - \mu_i)^2}{2\sigma_i^2} \right), \tag{2}$$

$i = 0, 1$. It will be convenient to stack the two sets of parameters governing the densities into a $(4 \times 1)$ vector, $\alpha = (\alpha_0', \alpha_1')'$.

As we shall see, a quantity of particular interest in the likelihood function is $P(s_1)$, which denotes $P(S_1 = s_1)$. Regarding $x_t$, there are two cases to consider, stationary and nonstationary. In the stationary case,

$$P(s_1) = P(s_1 \mid \underline{x}_T;\ \theta) = P(s_1;\ \beta).$$

That is, $P(s_1)$ is simply the long-run probability of $S_1 = s_1$, which in turn is determined by $\beta$. In the nonstationary case, the long-run probability does not exist, and so $P(S_1 = s_1)$ must be treated as an additional parameter to be estimated. It turns out, as we show subsequently, that $P(S_1 = 1)$ is all that is needed to construct the first likelihood term. We shall call this quantity '$\rho$' in both the stationary and nonstationary cases, remembering that in the stationary case $\rho$ is *not* an additional parameter to be estimated, but rather is determined by $\beta$, while the nonstationary case $\rho$ *is* an additional parameter to be estimated.[4] In certain situations, computation of $\rho$ in the stationary case may be done via simulation; see Diebold and Schuermann (1992).

[3] Generalisations to allow for more than two states and/or intra-state dynamics are straightforward but tedious, so we shall not consider them here.

[4] Alternatively, if prior information is available, $\rho$ may be set accordingly. The issues are analogous to those that arise with initialisation of the Kalman filter in the nonstationary case.

Let $\theta = (\alpha', \beta', \rho)'$ be the $(2k + 5 \times 1)$ vector of all model parameters. The complete-data likelihood is then[5]

$$\begin{aligned} f(\underline{y}_T, \underline{s}_T \mid \underline{x}_T; \theta) &= f(y_1, s_1 \mid \underline{x}_T; \theta) \prod_{t=2}^{T} f(y_t, s_t \mid \underline{y}_{t-1}, \underline{s}_{t-1}, \underline{x}_T; \theta) \\ &= f(y_1 \mid s_1, \underline{x}_T; \theta) P(s_1) \prod_{t=2}^{T} f(y_t \mid s_t, \underline{y}_{t-1}, \underline{s}_{t-1}, \underline{x}_T; \theta) \\ &\quad \times P(s_t \mid \underline{y}_{t-1}, \underline{s}_{t-1}, \underline{x}_t; \theta) \\ &= f(y_1 \mid s_1; \alpha) P(s_1) \prod_{t=2}^{T} f(y_t \mid s_t; \alpha) P(s_t \mid s_{t-1}, x_{t-1}; \beta), \end{aligned}$$

here $f$ denotes any density and underlining denotes past history of the variable from $t = 1$ to the variable subscript.

It will prove convenient to write the complete-data likelihood in terms of indicator functions,

$$\begin{aligned} f(\underline{y}_T, \underline{s}_T \mid \underline{x}_T; \theta) = {} & [\mathrm{I}(s_1 = 1) f(y_1 \mid s_1 = 1; \alpha_1)\rho \\ & + \mathrm{I}(s_1 = 0) f(y_1 \mid s_1 = 0; \alpha_0)(1 - \rho)] \\ & \times \prod_{t=2}^{T} \{\mathrm{I}(s_t = 1, s_{t-1} = 1) f(y_t \mid s_t = 1; \alpha_1) p_t^{11} \\ & + \mathrm{I}(s_t = 0, s_{t-1} = 1) f(y_t \mid s_t = 0; \alpha_0)(1 - p_t^{11}) \\ & + \mathrm{I}(s_t = 1, s_{t-1} = 0) f(y_t \mid s_t = 1; \alpha_1)(1 - p_t^{00}) \\ & + \mathrm{I}(s_t = 0, s_{t-1} = 0) f(y_t \mid s_t = 0; \alpha_0) p_t^{00}\}. \end{aligned}$$

Conversion to log form yields

$$\begin{aligned} \log f(\underline{y}_T, \underline{s}_T \mid \underline{x}_T; \theta) &= \mathrm{I}(s_1 = 1)[\log f(y_1 \mid s_1 = 1; \alpha_1) + \log \rho] \\ &\quad + \mathrm{I}(s_1 = 0)[\log f(y_1 \mid s_1 = 0; \alpha_0) + \log(1 - \rho)] \\ &\quad + \sum_{t=2}^{T} \{\mathrm{I}(s_t = 1) \log f(y_t \mid s_t = 1; \alpha_1) \end{aligned}$$

[5]'Complete-data' refers to the (hypothetical) assumption that both $\{y_t\}$ and $\{s_t\}$ are observed.

$$
\begin{aligned}
&+\mathrm{I}(s_t = 0)\log f(y_t \mid s_t = 0; \alpha_0) \\
&+\mathrm{I}(s_t = 1, s_{t-1} = 1)\log(p_t^{11}) \\
&+\mathrm{I}(s_t = 0, s_{t-1} = 1)\log(1 - p_t^{11}) \\
&+\mathrm{I}(s_t = 1, s_{t-1} = 0)\log(1 - p_t^{00}) \\
&+\mathrm{I}(s_t = 0, s_{t-1} = 0)\log(p_t^{00})\big\}.
\end{aligned}
$$

The complete-data log likelihood cannot be constructed in practice, because the complete data are not observed. Conceptually, the fact that the states are unobserved is inconsequential, because the incomplete-data log likelihood may be obtained by summing over all possible state sequences,

$$
\log f\big(\underline{y}_T \mid \underline{x}_T; \theta\big) = \log\left(\sum_{s_1=0}^{1}\sum_{s_2=0}^{1}\cdots\sum_{s_T=0}^{1} f\big(\underline{y}_T, \underline{s}_T \mid \underline{x}_T; \theta\big)\right),
$$

and then maximised with respect to $\theta$. In practice, however, construction and numerical maximisation of the incomplete-data log likelihood in this way is computationally intractable, as $\{s_t\}_{t=1}^{T}$ may be realised in $2^T$ ways. Therefore, following Hamilton's (1990) suggestion for the case of constant transition probabilities, we propose an EM algorithm for maximisation of the incomplete-data likelihood.

## 3. Model Estimation: The EM Algorithm

The EM algorithm is a stable and robust procedure for maximising the incomplete-data log likelihood via iterative maximisation of the expected complete-data log likelihood, conditional upon the observable data.[6] The procedure, shown schematically in Figure 7.2, amounts to the following.[7]

[6] Insightful discussions of the EM algorithm may be found in Dempster, Laird and Rubin (1977), Watson and Engle (1983), and Ruud (1991).

[7] Parameter superscripts count iterations.

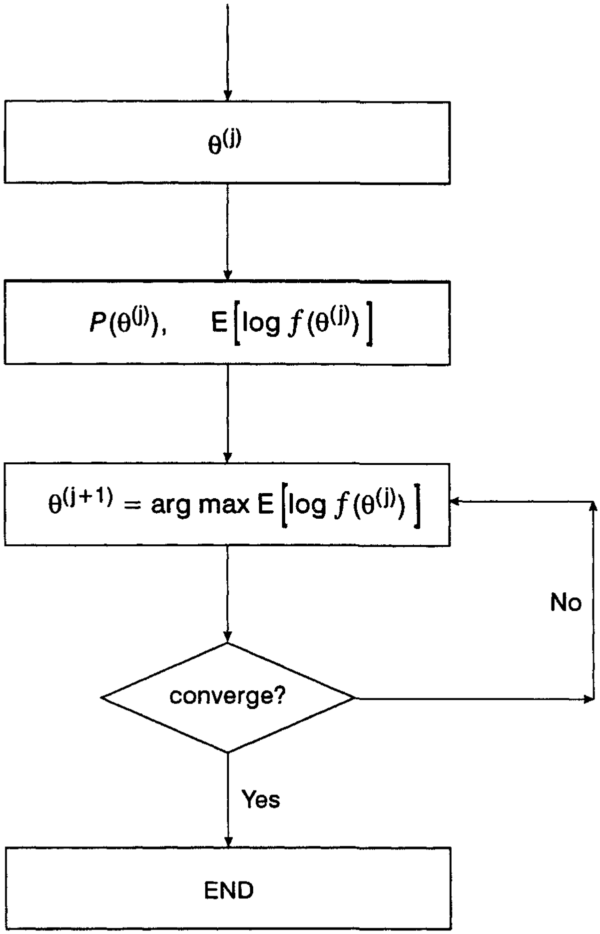


Fig. 7.2. The EM algorithm (notation discussed in the text).

(1) Pick $\theta^{(0)}$,
(2) Get:

$$P\left(s_t = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(0)}\right) \quad \forall t,$$

$$P\left(s_t = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(0)}\right) \quad \forall t,$$

$$P\left(s_t = 1, s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(0)}\right) \quad \forall t,$$

$$P\left(s_t = 0, s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(0)}\right) \quad \forall t,$$

$$P\left(s_t = 1, s_{t-1} = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(0)}\right) \quad \forall t,$$

$$P\left(s_t = 0, s_{t-1} = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(0)}\right) \quad \forall t;$$

construct $E \log f(\underline{y}_T, \underline{s}_T \mid \underline{x}_T; \theta^{(0)})$ by replacing I's with $P$'s.

**(3)** Set $\theta^{(1)} = \arg\max_{\theta} E[\log f(\underline{y}_T, \underline{s}_T \mid \underline{x}_T; \theta^{(0)})]$.

**(4)** Iterate to convergence.

Step (1) simply assigns an initial guess to the parameter vector, $\theta^{(0)}$, in order to start the EM algorithm. Step (2) is the 'E' (expectation) part of the algorithm, which produces smoothed state probabilities conditional upon $\theta^{(0)}$, while step (3) is the 'M' (maximisation) part, which produces an updated parameter estimate, $\theta^{(1)}$, conditional upon the smoothed state probabilities obtained in step (2). The convergence criterion adopted in (4) may be based upon various standard criteria, such as the change in the log likelihood from one iteration to the next, the value of the gradient vector, or $\|\theta^{(j)} - \theta^{(j-1)}\|$, for various norms $\|\bullet\|$.

### *3.1. The Expectation Step*

We wish to take expectations of the complete-data log likelihood, conditional upon the observed data. As in Hamilton (1990), this amounts to substitution of smoothed state probabilities (to be derived below) for indicator functions in the complete-data log likelihood,

$$\begin{aligned}
&\mathrm{E}\left[\log f\left(\underline{y}_T, \underline{s}_T \mid \underline{x}_T; \theta^{(j-1)}\right)\right] \\
&\quad = \rho^{(j-1)}\left[\log f(y_1 \mid s_1 = 1; \alpha_1^{(j-1)}) + \log \rho^{(j-1)}\right] \\
&\quad + (1 - \rho^{(j-1)})\left[\log f(y_1 \mid s_1 = 0; \alpha_0^{(j-1)}) + \log(1 - \rho^{(j-1)})\right] \\
&\quad + \sum_{t=2}^{T} \left\{P\left(s_t = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)\log f(y_t \mid s_t = 1; \alpha_1^{(j-1)})\right. \\
&\quad + P\left(s_t = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)\log f(y_t \mid s_t = 0; \theta_0^{(j-1)}) \\
&\quad + P\left(s_t = 1, s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)\log(p_t^{11}) \\
&\quad + P\left(s_t = 0, s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)\log(1 - p_t^{11}) \\
&\quad + P\left(s_t = 1, s_{t-1} = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)\log(1 - p_t^{00}) \\
&\quad \left.+ P\left(s_t = 0, s_{t-1} = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)\log(p_t^{00})\right\},
\end{aligned} \tag{3}$$

where the smoothed state probabilities are obtained from the optimal nonlinear smoother, conditional upon the current 'best guess' of $\theta$, $\theta^{(j-1)}$.

Given $\theta^{(j-1)}$, $\underline{y}_T$, and $\underline{x}_T$, the algorithm for calculating the smoothed state probabilities for iteration $j$ is as follows:

**1.** Calculate the sequence of conditional densities of $y_t$ given by (2.2) (a $(T \times 2)$ matrix), and transition probabilities given by Figure 7.1 (a $(T-1 \times 4)$ matrix).

**2.** Calculate filtered joint state probabilities (a $(T-1 \times 4)$ matrix) by iterating on steps 2a–2d below for $t = 2, \ldots, T$:

**2a.** Calculate the joint conditional distribution of $(y_t, s_t, s_{t-1})$ given $\underline{y}_{t-1}$ and $\underline{x}_{t-1}$ (four numbers): For $t = 2$, the joint conditional distribution is given by

$$f(y_2, s_2, s_1 \mid y_1, x_1; \theta^{(j-1)})$$
$$= f(y_2 \mid s_2; \alpha^{(j-1)}) P(s_2 \mid s_1, x_1; \beta^{(j-1)}) P(s_1).$$

For subsequent time $t$, the joint conditional distribution is

$$f\left(y_t, s_t, s_{t-1} \mid \underline{y}_{t-1}, \underline{x}_{t-1}; \theta^{(j-1)}\right)$$
$$= \sum_{s_{t-2}=0}^{1} f(y_t \mid s_t; \alpha^{(j-1)}) P(s_t \mid s_{t-1}, x_{t-1}; \beta^{(j-1)})$$
$$\times P\left(s_{t-1}, s_{t-2} \mid \underline{y}_{t-1}, \underline{x}_{t-1}; \theta^{(j-1)}\right)$$

where the conditional density $f(y_t \mid s_t; \alpha^{(j-1)})$ and transition probabilities $P(s_t \mid s_{t-1}, x_{t-1}; \beta^{(j-1)})$ are given by step 1, and $P(s_{t-1}, s_{t-2} \mid \underline{y}_{t-1}, \underline{x}_{t-1}; \theta^{(j-1)})$ is the filtered probability resulting from execution of step 2 for the previous $t$ value.

**2b.** Calculate the conditional likelihood of $y_t$ (one number):

$$f\left(y_t \mid \underline{y}_{t-1}, \underline{x}_{t-1}; \theta^{(j-1)}\right)$$
$$= \sum_{s_t=0}^{1} \sum_{s_{t-1}=0}^{1} f\left(y_t, s_t, s_{t-1} \mid \underline{y}_{t-1}, \underline{x}_{t-1}; \theta^{(j-1)}\right).$$

**2c.** Calculate the time-$t$ filtered state probabilities (four numbers):

$$P\left(s_t, s_{t-1} \mid \underline{y}_t, \underline{x}_t; \theta^{(j-1)}\right) = \frac{f\left(y_t, s_t, s_{t-1} \mid \underline{y}_t, \underline{x}_t; \theta^{(j-1)}\right)}{f\left(y_t \mid \underline{y}_t, \underline{x}_t; \theta^{(j-1)}\right)},$$

where the numerator is the joint conditional distributional of $(y_t, s_t, s_{t-1})$ from step 2a and the denominator is the conditional likelihood of $y_t$ from step 2b above.

**2d.** These four filtered probabilities are used as input for step 2a to calculate the filtered probabilities for the next time period, and steps 2a–2d are repeated $(T-2)$ times.

**3.** Calculate the smoothed joint state probabilities as follows (a $(T-1 \times 6)$ matrix):

**3a.** For $t=2$ and a given valuation of $(s_t, s_{t-1})$, sequentially calculate the joint probability of $(s_\tau, s_{\tau-1}, s_t, s_{t-1})$ given $\underline{y}_\tau$ and $\underline{x}_\tau$, for $\tau = t+2, t+3, \ldots, T$.[8]

$$P\left(s_\tau, s_{\tau-1}, s_t, s_{t-1} \mid \underline{y}_\tau, \underline{x}_\tau; \theta^{(j-1)}\right)$$

$$= \frac{\sum_{s_{\tau-2}=0}^{1} f(y_\tau \mid s_\tau; \alpha^{(j-1)}) P(s_\tau \mid s_{\tau-1}, x_{\tau-1}; \beta^{(j-1)}) \times P\left(s_{\tau-1}, s_{\tau-2}, s_t, s_{t-1} \mid \underline{y}_{\tau-1}, \underline{x}_{\tau-1}; \theta^{(j-1)}\right)}{f\left(y_\tau \mid \underline{y}_{\tau-1}, \underline{x}_{\tau-1}; \theta^{(j-1)}\right)}$$

where the first two terms in the numerator are given by step 1, the third by the previous step 3a computation, and the denominator by step 2b. When $\tau = t+2$, the third term in the numerator is initialised with the following expression:

$$P\left(s_{t+1}, s_t, s_{t-1} \mid \underline{y}_{t+1}, \underline{x}_{t+1}; \theta^{(j-1)}\right)$$

$$= \frac{f(y_{t+1} \mid s_{t+1}; \alpha^{(j-1)}) P(s_{t+1} \mid s_t, x_t; \beta^{(j-1)}) \times P\left(s_t s_{t-1} \mid \underline{y}_t, \underline{x}_t; \theta^{(j-1)}\right)}{f\left(y_{t+1} \mid \underline{y}_t, \underline{x}_t; \theta^{(j-1)}\right)}.$$

For each $\tau$ value we produce a $(4 \times 1)$ vector of probabilities corresponding to the four possible valuations of $(s_\tau, s_{\tau-1})$. Thus, upon reaching $\tau = T$, we have computed and saved a $(T-3) \times 4$ matrix, the last row of which is used in step 3b below.

[8] There are of course four possible $(s_t, s_{t-1})$ sequences: (0, 0), (0, 1), (1, 0) and (1, 1).

**3b.** Upon reaching $\tau = T$, the smoothed joint state probability for time $t$ and the chosen valuation of $(s_t, s_{t-1})$ is calculated as

$$P\left(s_t, s_{t-1} \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)$$
$$= \sum_{s_T=0}^{1} \sum_{s_{T-1}=0}^{1} P\left(s_T, s_{T-1}, s_t, s_{t-1} \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right).$$

**3c.** Steps 3a and 3b are repeated for all possible time $t$ valuations $(s_t, s_{t-1})$, until a smoothed probability has been calculated for each of the four possible valuations. At this point we have a $(1 \times 4)$ vector of smoothed joint state probabilities for $(s_t, s_{t-1})$.

**3d.** Steps 3a–3c are repeated for $t = 3, 4, \ldots, T$, yielding a total of $(T - 1 \times 4)$ smoothed joint state probabilities.

**4.** Smoothed marginal state probabilities are found by summing over the smoothed joint state probabilities. For example,

$$P\left(s_t = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right) = P\left(s_t = 1, s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)$$
$$+ P\left(s_t = 1, s_{t-1} = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right).$$

These $(T - 1 \times 6)$ smoothed state probabilities are used as input for the maximisation step, which we now describe.

### *3.2. The Maximisation Step*

Given the smoothed state probabilities, the expected complete-data log likelihood, given by (3), is maximised directly with respect to the model parameters. The resulting $2k + 5$ first-order conditions are linear both in $\rho$ and the conditional density parameter vector $\alpha$, and nonlinear in the transition probability parameter vector $\beta$. Moreover, these two sets of parameters appear in distinctly different terms in the likelihood function. Due to this separability, five of the first-order conditions are linear in the parameters:

$$\sum_{t=1}^{T} P\left(s_t = i \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)\left(y_t - \mu_i^{(j)}\right) = 0$$

$$\sum_{t=1}^{T} P\left(s_t = i \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)\left(\frac{\left(y_t - \mu_i^{(j)}\right)^2}{\left(\sigma_i^2\right)^{(j)}} - 1\right) = 0$$

$$P\left(s_1 = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)\left(\frac{1}{\rho}\right) - 1 = 0,$$

and yield immediate closed-form expressions for the maximum likelihood estimators.

$$\mu_i^{(j)} = \frac{\sum_{t=1}^{T} y_t P\left(s_t = i \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)}{\sum_{t=1}^{T} P\left(s_t = i \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)}$$

$$(\sigma_i^2)^{(j)} = \frac{\sum_{t=1}^{T} (y_t - \mu_i^{(j)})^2 P\left(s_t = i \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)}{\sum_{t=1}^{T} P\left(s_t = i \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)}$$

$$\rho^{(j)} = P\left(s_1 = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right),$$

$i = 0, 1.$

However, given our use of logit transition probability functions, the remaining $2k$ first-order conditions are *nonlinear* in $\beta$, and are given by[9]

$$\sum_{t=2}^{T} x_{t-1}\Big\{P\left(s_t = 0, s_{t-1} = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)$$

$$-p_t^{00} P\left(s_{t-1} = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)\Big\} = 0$$

$$\sum_{t=2}^{T} x_{t-1}\Big\{P\left(s_t = 1, s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)$$

$$-p_t^{11} P\left(s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\right)\Big\} = 0.$$

Closed-form solutions are found by linearly approximating $p_t^{00}$ and $p_t^{11}$ using a first-order Taylor series expansion around $\beta_0^{(j-1)}$ and $\beta_1^{(j-1)}$,

[9] A variety of alternative functional forms, in addition to the logit, are examined in Lee (1991).

respectively. These linear approximations are given by

$$p_t^{00}(\beta_0^{(j-1)}) \approx p_t^{00}(\beta_0^{(j-1)}) + \left.\frac{\partial p_t^{00}(\beta_0)}{\partial \beta_0}\right|_{\beta_0=\beta_0^{(j-1)}} (\beta_0 - \beta_0^{(j-1)})$$

$$p_t^{11}(\beta_1^{(j-1)}) \approx p_t^{11}(\beta_1^{(j-1)}) + \left.\frac{\partial p_t^{11}(\beta_1)}{\partial \beta_1}\right|_{\beta_1=\beta_1^{(j-1)}} (\beta_1 - \beta_1^{(j-1)}).$$

For simplicity, we adopt the following notation:

$$p_{it}^{00}(\beta_0^{(j-1)}) = \left.\frac{\partial p_t^{00}(\beta_0)}{\partial \beta_{i0}}\right|_{\beta_0=\beta_0^{(j-1)}}, \; i = 0, \ldots, k-1$$

$$p_{it}^{11}(\beta_1^{(j-1)}) = \left.\frac{\partial p_t^{11}(\beta_1)}{\partial \beta_{i1}}\right|_{\beta_1=\beta_1^{(j-1)}}, \; i = 0, \ldots, k-1,$$

so that the vectors of partials are $(1 \times k)$ row vectors given by

$$\left.\frac{\partial p_t^{00}(\beta_0)}{\partial \beta_0}\right|_{\beta_0=\beta_0^{(j-1)}} = \left[p_{0t}^{00}(\beta_0^{(j-1)}), p_{1t}^{00}(\beta_0^{(j-1)}), \ldots, p_{(k-1)t}^{00}(\beta_0^{(j-1)})\right]$$

$$\left.\frac{\partial p_t^{11}(\beta_1)}{\partial \beta_1}\right|_{\beta_1=\beta_1^{(j-1)}} = \left[p_{0t}^{11}(\beta_1^{(j-1)}), p_{1t}^{11}(\beta_1^{(j-1)}), \ldots, p_{(k-1)t}^{11}(\beta_1^{(j-1)})\right],$$

and the individual partials are given by

$$p_{it}^{00}(\beta_0^{(j-1)}) = x_{i,t-1}\left[p_t^{00}(\beta_0^{(j-1)}) - p_t^{00}(\beta_0^{(j-1)})^2\right], \; i = , \ldots, k-1$$

$$p_{it}^{11}(\beta_1^{(j-1)}) = x_{i,t-1}\left[p_t^{11}(\beta_1^{(j-1)}) - p_t^{11}(\beta_1^{(j-1)})^2\right], \; i = 0, \ldots, k-1,$$

where $x_{i,t-1}$ is the $i^{\text{th}}$ element of $x_{t-1}$.

Substituting these linear approximations for the transition probabilities into the $2k$ nonlinear first-order conditions results in $2k$ *linear* first-order conditions given by

$$\sum_{t=2}^{T} x_{t-1}\Bigg\{P\Big(s_t = 0, s_{t-1} = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\Big)$$
$$-P\Big(s_{t-1} = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\Big)$$
$$\times\left[p_t^{00}(\beta_0^{(j-1)}) + \frac{\partial p_t^{00}(\beta_0)}{\partial \beta_0}(\beta_0 - \beta_0^{(j-1)})\right]\Bigg\} = 0$$

$$\sum_{t=2}^{T} x_{t-1}\Bigg\{P\Big(s_t = 1, s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\Big)$$
$$-P\Big(s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\Big)$$
$$\times\left[p_t^{11}(\beta_1^{(j-1)}) + \frac{\partial p_t^{11}(\beta_1)}{\partial \beta_1}(\beta_1 - \beta_1^{(j-1)})\right]\Bigg\} = 0,$$

where all derivatives are understood to be evaluated at $\beta_0^{(j-1)}$ or $\beta_1^{(j-1)}$, as relevant. Solving these, we obtain a closed-form solution for $\beta_0^{(j)}$,

$$\beta_0^{(j)} = \left(\sum_{t=2}^{T} x_{t-1} P\Big(s_{t-1} = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\Big)\frac{\partial p_t^{00}(\beta_0)}{\partial \beta_0}\right)^{-1}$$
$$\times\left(\sum_{t=2}^{T} x_{t-1}\left\{\begin{array}{l} P\Big(s_t = 0, s_{t-1} = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\Big) \\ \quad -P\Big(s_{t-1} = 0 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\Big) \\ \quad \times\left[p_t^{00}(\beta_0^{(j-1)}) - \dfrac{\partial p_t^{00}(\beta_0)}{\partial \beta_0}\beta_0^{(j-1)}\right]\end{array}\right\}\right).$$

Similarly the closed-form solution for $\beta_1^{(j)}$ is

$$\beta_1^{(j)} = \left(\sum_{t=2}^{T} x_{t-1} P\Big(s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\Big)\frac{\partial p_t^{11}(\beta_1)}{\partial \beta_1}\right)^{-1}$$
$$\times\left(\sum_{t=2}^{T} x_{t-1}\left\{\begin{array}{l} P\Big(s_t = 1, s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\Big) \\ \quad -P\Big(s_{t-1} = 1 \mid \underline{y}_T, \underline{x}_T; \theta^{(j-1)}\Big) \\ \quad \times\left[p_t^{11}(\beta_1^{(j-1)}) - \dfrac{\partial p_t^{11}(\beta_1)}{\partial \beta_1}\beta_1^{(j-1)}\right]\end{array}\right\}\right).$$

The cases of $k = 2$ and $k = 3$ are of particular interest in applied work. For this reason, we catalog explicit expressions for $\beta_0^{(j)}$ and $\beta_1^{(j)}$ in those cases in the appendix.

## 4. Simulation Results

In order to demonstrate the methodology, we present the results of a simulation exercise. Sample size is 100. We set $k = 2$ so that the time-varying transition probabilities are driven by one $x$ series. The transition probability parameters, $\alpha$, are chosen and the $x$ series constructed so that the (true) probabilities of staying in state, $p_t^{00}$ and $p_t^{11}$, each alternate between 0.40 and 0.90 over successive sets of twenty sample observations, beginning with $p_t^{00} = 0.40$ and $p_t^{11} = 0.90$. The chosen parameter values are shown in Table 7.1, the simulated $y$ and $s$ sequences are shown in Figure 7.3, the $x$ sequence is shown in Figure 7.4, and the resultant probabilities of staying in state are shown in Figure 7.7 (labeled 'actual').

Parameter estimation using the EM algorithm begins at the true parameter values. Convergence of the EM algorithm is checked as follows. Upon the calculation of each new parameter vector, say the $j^{\text{th}}$, a comparison is make with the previous vector, the $(j - 1)^{\text{st}}$. If the

**TABLE 7.1**
Estimation Results

| *Parameter* | $\theta_0$ | $\hat{\theta}$ | $\hat{\theta}_H$ |
|---|---|---|---|
| $\mu_0$ | −1.00 | −1.62 | −0.38 |
| $\sigma_0^2$ | 4.00 | 2.83 | 3.73 |
| $\mu_1$ | 1.00 | 1.27 | 2.25 |
| $\sigma_1^2$ | 4.00 | 2.97 | 2.24 |
| $\beta_{00}$ | 0.79 | −1.53 | 1.72 |
| $\beta_{01}$ | −2.00 | −2.89 | NA |
| $\beta_{10}$ | 1.00 | 0.94 | 0.49 |
| $\beta_{11}$ | 2.00 | 2.07 | NA |
| $\rho$ | 1.00 | 0.97 | 0.29 |
| MSE | | 0.13 | 0.27 |
| Iterations | | 462 | 208 |
| lnL | | −238.17 | −247.53 |

*Notes:* $\theta_0$ is the true parameter vector; $\hat{\theta}$ is the estimated parameter vector, and $\hat{\theta}_H$ is the estimated parameter vector for the Hamilton model, obtained by constraining $\beta_{01}$ and $\beta_{11}$ to be zero.

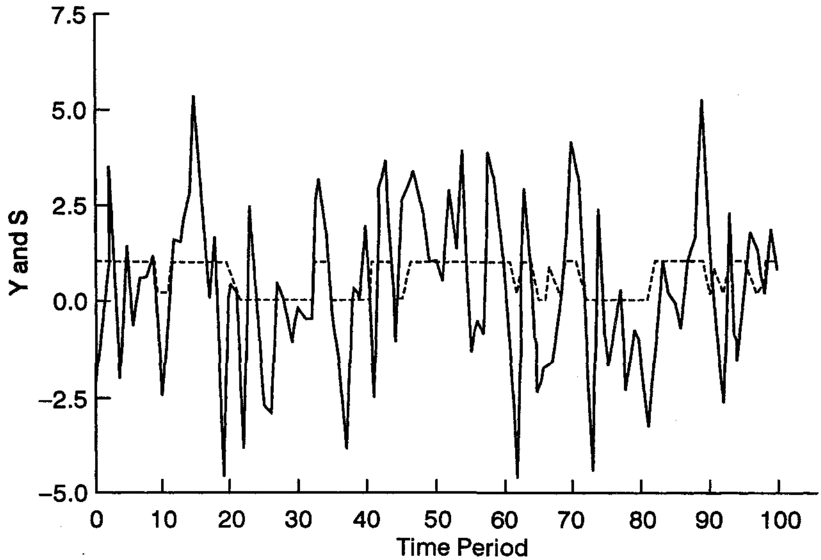


Fig. 7.3. Y and S sequences.

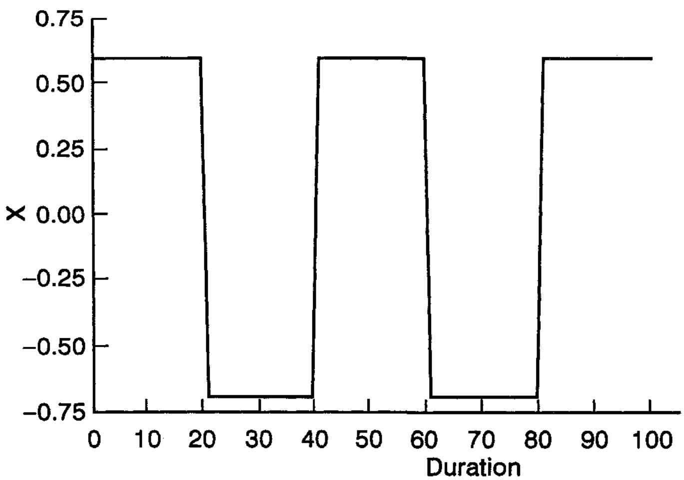


Fig. 7.4. X sequence.

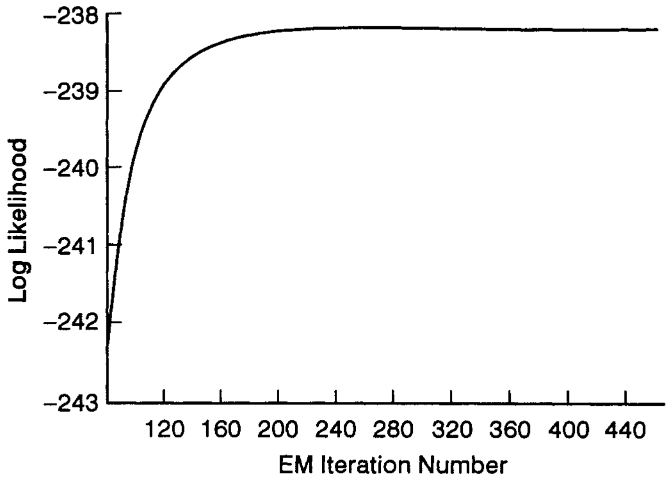


Fig. 7.5. Sequence of log-likelihood values.

absolute value of the maximal difference in like elements of $\theta^{(j)}$ and $\theta^{(j-1)}$ is less than $10e^{-8}$, iteration is terminated. Convergence was obtained in 462 iterations, and the location of the likelihood maximum obtained was robust to a variety of alternative start up parameters. As shown in Figure 7.5, the EM algorithm gets close to the likelihood maximum very quickly, but then takes more iterations to reach convergence.[10]

The resultant maximum likelihood parameter estimates are given in Table 7.1, labelled $\hat{\theta}$. Given the small sample size, the likelihood maximum is fairly close to the true parameter vector, the main exception being the estimate of $\beta_{00}$, which diverges from the true value by a rather large amount.

In Figure 7.6 we graph the time series of true states and smoothed state probabilities produced by the EM algorithm. The smoothed state probabilities, which are the EM algorithm's best guess at the state each period based on our time-varying transition probability model (Figure 7.6(a)), track the true states quite well. The mean-squared state extraction error using our model is 0.13.[11]

Next, we use the EM algorithm to fit a Hamilton model, which does not allow for time-varying transition probabilities, to the same dataset. The Hamilton model parameter estimates are given in Table 7.1,

[10] This behaviour is noted frequently in the literature.

[11] The mean squared state extraction error using the *true* parameter values, which may be viewed as a lower bound, is 0.11.

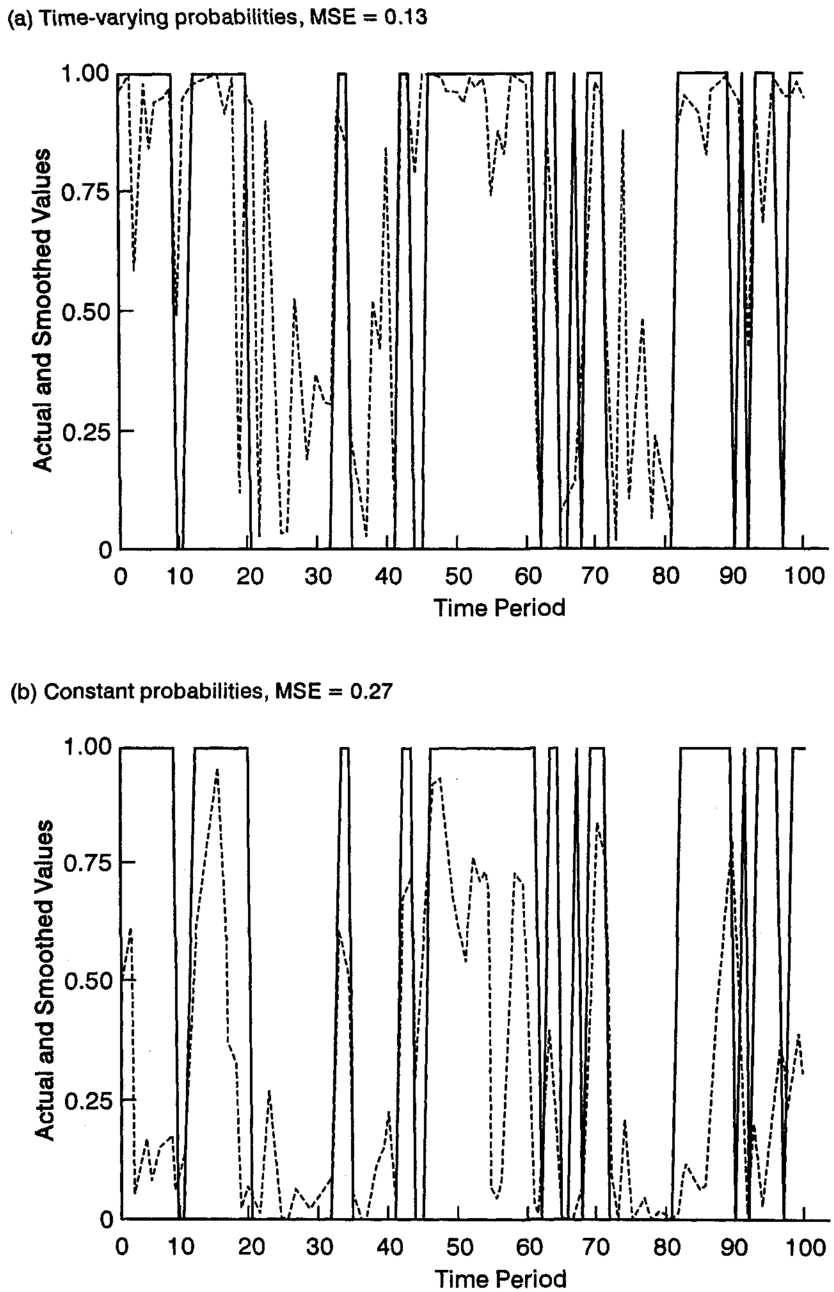


Fig. 7.6. Actual and smoothed state sequences.

labelled $\hat{\theta}_H$. The likelihood ratio test statistic clearly rejects the null of constant transition probabilities. The time series of true states and smoothed state probabilities that result from the fitted Hamilton model appear in Figure 7.6(b). The mean-squared state extraction error is 0.27, more than twice as large as that resulting from our time-varying transition probability model.

The fitted values of the transition probabilities for our model and the Hamilton model appear in Figures 7.7(a) and (b), along with the true probabilities. Our estimates do a reasonable job of tracking the time-varying probabilities, whereas, needless to say, the Hamilton estimates do not. As intuition suggests, the Hamilton estimates lie between the actual 'high' and 'low' values. Their restriction to constancy is responsible for the higher mean-squared state extraction error associated with the Hamilton model.

## 5. Concluding Remarks

This paper has been largely methodological, and numerous additional methodological issues are currently under investigation, including formal asymptotic distribution theory, elimination of the linear approximation employed in solving the first-order conditions, model specification tests, and analytic determination of ergodic probabilities. We shall not dwell on those issues here; instead, we shall briefly discuss two potentially fruitful areas of application.

The first concerns exchange rate dynamics. Engel and Hamilton (1990) have suggested that exchange rates may follow a switching process. We agree. But certainly, it is highly restrictive to require constancy of the transition probabilities. Rather, they should be allowed to vary with fundamentals, such as relative money supplies, relative real outputs, interest rate differentials, and so forth. Moreover, Mark (1992) produces useful indexes of fundamentals, which may be exploited to maintain parsimony. We shall provide a detailed report on this approach in a future paper.

The second concerns aggregate output dynamics. Diebold, Rudebusch and Sichel (1993) have found strong duration dependence in postwar U.S. contractions. That is, the longer a contraction persists, the more likely it is to end. That suggests allowing the transition probabilities in a Markov switching model of aggregate output dynamics to depend on length-to-date of the current regime, which can readily be achieved by expanding the state space of the process.[12]

[12] We thank Atsushi Kajii and Jim Hamilton for pointing this out.

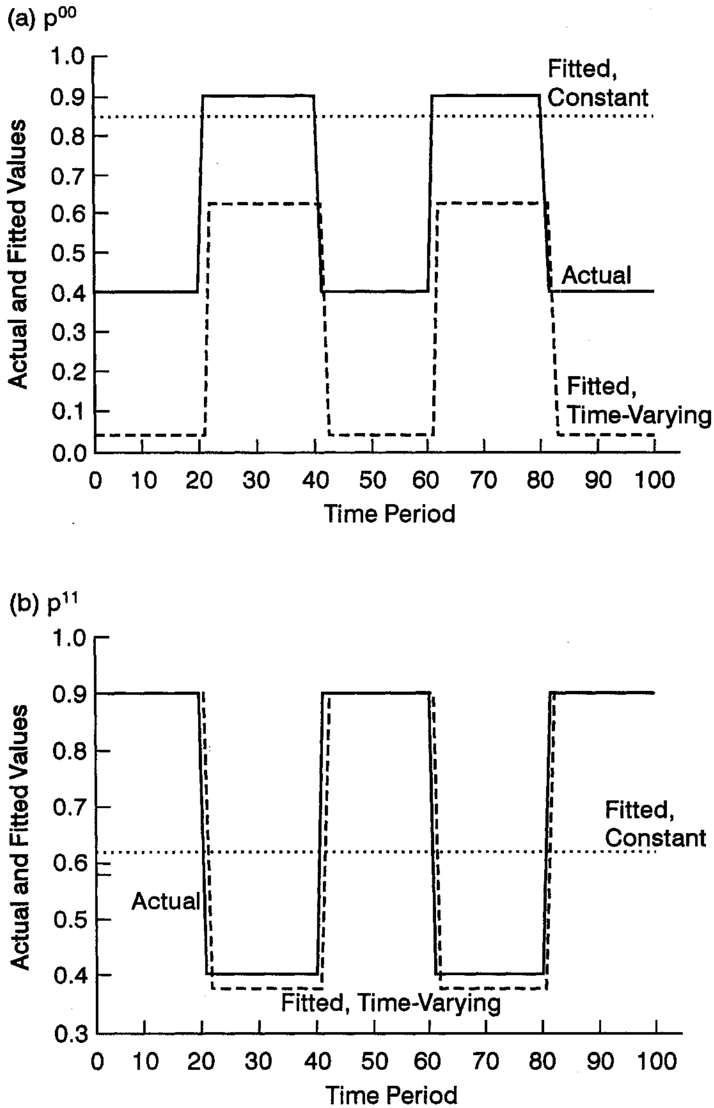


Fig. 7.7. Actual and fitted values.

## Appendix

The general form of the maximum expected complete-data likelihood estimators for the $2k$ transition probability function parameters, $\beta_0^{(j)}$ and $\beta_1^{(j)}$, is given in Section 3.2. Here we include the explicit expressions for the cases of $k = 2$ and $k = 3$, which are of particular interest in applied work. Due to space limitations, it is understood that in the expressions that follow all smoothed probabilities are conditional on $\underline{y}_T$ and $\underline{x}_T$ given $\theta^{(j-1)}$, and that transition probabilities $p_t^{00}$, $p_t^{11}$ and their derivatives are evaluated at $\beta_0^{(j-1)}$ and $\beta_1^{(j-1)}$, respectively.

$\underline{k = 2}$

$$\beta_0^{(j)} = \begin{pmatrix} \beta_{00}^{(j)} \\ \beta_{01}^{(j)} \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{t=2}^{T} x_{0,t-1} P(s_{t-1} = 0) p_{0t}^{00} & \Sigma_{t=2}^{T} x_{0,t-1} P(s_{t-1} = 0) p_{1t}^{00} \\ \Sigma_{t=2}^{T} x_{1,t-1} P(s_{t-1} = 0) p_{0t}^{00} & \Sigma_{t=2}^{T} x_{1,t-1} P(s_{t-1} = 0) p_{1t}^{00} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \Sigma_{t=2}^{T} x_{0,t-1} \left\{ P(s_t = 0, s_{t-1} = 0) - P(s_{t-1} = 0) \left[ p_t^{00} - \dfrac{\partial p_t^{00}}{\partial \beta_0} \beta_0^{(j-1)} \right] \right\} \\ \Sigma_{t=2}^{T} x_{1,t-1} \left\{ P(s_t = 0, s_{t-1} = 0) - P(s_{t-1} = 0) \left[ p_t^{00} - \dfrac{\partial p_t^{00}}{\partial \beta_0} \beta_0^{(j-1)} \right] \right\} \end{pmatrix}$$

$$\beta_1^{(j)} = \begin{pmatrix} \beta_{10}^{(j)} \\ \beta_{11}^{(j)} \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{t=2}^{T} x_{0,t-1} P(s_{t-1} = 1) p_{0t}^{11} & \Sigma_{t=2}^{T} x_{0,t-1} P(s_{t-1} = 1) p_{1t}^{11} \\ \Sigma_{t=2}^{T} x_{1,t-1} P(s_{t-1} = 1) p_{0t}^{11} & \Sigma_{t=2}^{T} x_{1,t-1} P(s_{t-1} = 1) p_{1t}^{11} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \Sigma_{t=2}^{T} x_{0,t-1} \left\{ P(s_t = 1, s_{t-1} = 1) - P(s_{t-1} = 1) \left[ p_t^{11} - \dfrac{\partial p_t^{11}}{\partial \beta_1} \beta_1^{(j-1)} \right] \right\} \\ \Sigma_{t=2}^{T} x_{1,t-1} \left\{ P(s_t = 1, s_{t-1} = 1) - P(s_{t-1} = 1) \left[ p_t^{11} - \dfrac{\partial p_t^{11}}{\partial \beta_1} \beta_1^{(j-1)} \right] \right\} \end{pmatrix}$$

$\underline{k = 3}$

$$\beta_0^{(j)} = \begin{pmatrix} \beta_{00}^{(j)} \\ \beta_{01}^{(j)} \\ \beta_{02}^{(j)} \end{pmatrix}$$

$$= \left( \begin{matrix} \Sigma_{t=2}^{T} x_{0,t-1} P(s_{t-1} = 0) p_{0t}^{00} & \Sigma_{t=2}^{T} x_{0,t-1} P(s_{t-1} = 0) p_{1t}^{00} \\ \Sigma_{t=2}^{T} x_{1,t-1} P(s_{t-1} = 0) p_{0t}^{00} & \Sigma_{t=2}^{T} x_{1,t-1} P(s_{t-1} = 0) p_{1t}^{00} \\ \Sigma_{t=2}^{T} x_{2,t-1} P(s_{t-1} = 0) p_{0t}^{00} & \Sigma_{t=2}^{T} x_{2,t-1} P(s_{t-1} = 0) p_{1t}^{00} \end{matrix} \right.$$

$$
\left.\begin{matrix}
\Sigma_{t=2}^{T} x_{0,t-1} P(s_{t-1}=0) p_{2t}^{00} \\
\Sigma_{t=2}^{T} x_{1,t-1} P(s_{t-1}=0) p_{2t}^{00} \\
\Sigma_{t=2}^{T} x_{2,t-1} P(s_{t-1}=0) p_{2t}^{00}
\end{matrix}\right)^{-1}
$$

$$
\times \begin{pmatrix}
\Sigma_{t=2}^{T} x_{0,t-1} \left\{ P(s_t=0, s_{t-1}=0) - P(s_{t-1}=0) \left[ p_t^{00} - \dfrac{\partial p_t^{00}}{\partial \beta_0} \beta_0^{(j-1)} \right] \right\} \\
\Sigma_{t=2}^{T} x_{1,t-1} \left\{ P(s_t=0, s_{t-1}=0) - P(s_{t-1}=0) \left[ p_t^{00} - \dfrac{\partial p_t^{00}}{\partial \beta_0} \beta_0^{(j-1)} \right] \right\} \\
\Sigma_{t=2}^{T} x_{2,t-1} \left\{ P(s_t=0, s_{t-1}=0) - P(s_{t-1}=0) \left[ p_t^{00} - \dfrac{\partial p_t^{00}}{\partial \beta_0} \beta_0^{(j-1)} \right] \right\}
\end{pmatrix}
$$

$$
\beta_1^{(j)} = \begin{pmatrix} \beta_{10}^{(j)} \\ \beta_{11}^{(j)} \\ \beta_{12}^{(j)} \end{pmatrix}
$$

$$
= \left(\begin{matrix}
\Sigma_{t=2}^{T} x_{0,t-1} P(s_{t-1}=1) p_{0t}^{00} & \Sigma_{t=2}^{T} x_{0,t-1} P(s_{t-1}=1) p_{1t}^{00} \\
\Sigma_{t=2}^{T} x_{1,t-1} P(s_{t-1}=1) p_{0t}^{00} & \Sigma_{t=2}^{T} x_{1,t-1} P(s_{t-1}=1) p_{1t}^{00} \\
\Sigma_{t=2}^{T} x_{2,t-1} P(s_{t-1}=1) p_{0t}^{00} & \Sigma_{t=2}^{T} x_{2,t-1} P(s_{t-1}=1) p_{1t}^{00}
\end{matrix}\right.
$$

$$
\left.\begin{matrix}
\Sigma_{t=2}^{T} x_{0,t-1} P(s_{t-1}=1) p_{2t}^{00} \\
\Sigma_{t=2}^{T} x_{1,t-1} P(s_{t-1}=1) p_{2t}^{00} \\
\Sigma_{t=2}^{T} x_{2,t-1} P(s_{t-1}=1) p_{2t}^{00}
\end{matrix}\right)^{-1}
$$

$$
\times \begin{pmatrix}
\Sigma_{t=2}^{T} x_{0,t-1} \left\{ P(s_t=1, s_{t-1}=1) - P(s_{t-1}=1) \left[ p_t^{11} - \dfrac{\partial p_t^{11}}{\partial \beta_1} \beta_1^{(j-1)} \right] \right\} \\
\Sigma_{t=2}^{T} x_{1,t-1} \left\{ P(s_t=1, s_{t-1}=1) - P(s_{t-1}=1) \left[ p_t^{11} - \dfrac{\partial p_t^{11}}{\partial \beta_1} \beta_1^{(j-1)} \right] \right\} \\
\Sigma_{t=2}^{T} x_{2,t-1} \left\{ P(s_t=1, s_{t-1}=1) - P(s_{t-1}=1) \left[ p_t^{11} - \dfrac{\partial p_t^{11}}{\partial \beta_1} \beta_1^{(j-1)} \right] \right\}
\end{pmatrix}
$$

## References


Cecchetti, S. G., P. Lam and N. C. Mark (1990), "Mean Reversion in Equilibrium Asset Prices," *American Economic Review*, 80, 3, pp. 398–418.

Dempster, A. P., N. M. Laird and D. B. Rubin (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, 39, pp. 1–38.

De Toldi, M., C. Gourieroux and A. Monfort (1992), "On Seasonal Effects in Duration Models," Working Paper #9216, INSEE, Paris.

Diebold, F. X. and G. D. Rudebusch (1993), "Measuring Business Cycles: A Modern Perspective," Manuscript, Dept. of Economics, University of Pennsylvania.

Diebold, F. X., G. D. Rudebusch and D. Sichel (1993), "Further Evidence on Business Cycle Duration Dependence," in J. H. Stock and M. W. Watson (eds.), *New Research on Business Cycles, Indicators and Forecasting*, University of Chicago Press for NBER, Chicago, pp. 255–84.

Diebold, F. X. and T. Schuermann (1992), "Exact Maximum Likelihood Estimation of ARCH Models," Manuscript, Dept. of Economics, University of Pennsylvania.

Engel, C. and J. D. Hamilton (1990), "Long Swings in the Dollar: Are They in the Data and do Markets Know it?," *American Economic Review*, 80, 4, pp. 689–713.

Fillardo, A. J. (1991), "Business Cycle Phases and Their Transitions," Manuscript, Dept. of Economics, University of Chicago.

Ghysels, E. (1992), "A Time Series Models of Growth Cycles and Seasonals with Stochastic Regime Switches," Manuscript, CRDE, Dept. of Economics, University of Montreal.

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

Hamilton, J. D. (1988), "Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates," *Journal of Economic Dynamics and Control*, 12, pp. 385–423.

—— (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, pp. 357–84.

—— (1990), "Analysis of Time Series Subject to Changes in Regime," *Journal of Econometrics*, 45, pp. 39–70.

Lee, J.-H. (1991), "Nonstationary Markov Switching Models of Exchange Rates: The Pound-Dollar Exchange Rate," PhD Dissertation, University of Pennsylvania.

Mark, N. C. (1992), "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability and Overshooting," Manuscript, Dept. of Economics, Ohio State University.

Quandt, R. E. (1958), "The Estimation of Parameters of Linear Regression System Obeying Two Separate Reigmes," *Journal of the American Statistical Association*, 55, pp. 873–80.

Ruud, P. A. (1991), "Extension of Estimation Methods Using the EM Algorithm," *Journal of Econometrics*, 49, pp. 305–41.

Watson, M. W. and R. F. Engle (1983), "Alternative Algorithms for the Estimation of Dynamic Factor, Mimic and Varying Coefficient Models," *Journal of Econometrics*, 15, pp. 385–400.