

**LECTURE ON**  
**THE MARKOV SWITCHING MODEL**

**CHUNG-MING KUAN**

Institute of Economics  
Academia Sinica

This version: April 19, 2002

© Chung-Ming Kuan (all rights reserved).

Address for correspondence: Institute of Economics, Academia Sinica, Taipei 115, Taiwan.

E-mail: [ckuan@econ.sinica.edu.tw](mailto:ckuan@econ.sinica.edu.tw); URL: [www.sinica.edu.tw/as/ssrc/ckuan](http://www.sinica.edu.tw/as/ssrc/ckuan).

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Markov Switching Model of Conditional Mean</b>	<b>3</b>
2.1	A Simple Model . . . . .	3
2.2	Some Extensions . . . . .	5
2.3	Markov Trend . . . . .	6
<b>3</b>	<b>Model Estimation</b>	<b>6</b>
3.1	Quasi-Maximum Likelihood Estimation . . . . .	7
3.2	Estimation via Gibbs Sampling . . . . .	9
<b>4</b>	<b>Hypothesis Testing</b>	<b>11</b>
4.1	Testing for Switching Parameters . . . . .	11
4.2	Testing Other Hypotheses . . . . .	13
<b>5</b>	<b>Application: A Study of Taiwan's Business Cycles</b>	<b>14</b>
<b>6</b>	<b>The Markov Switching Model of Conditional Variance</b>	<b>18</b>
<b>7</b>	<b>Application: A Study of Taiwan's Short-Term Interest Rates</b>	<b>21</b>
<b>8</b>	<b>Concluding Remarks</b>	<b>26</b>
	<b>Appendix I: Estimation of the Model (2.5)</b>	<b>27</b>
	<b>Appendix II: Computation of Hansen's Statistic (4.4)</b>	<b>29</b>
	<b>References</b>	<b>30</b>

## 1 Introduction

It is now common to employ various time series models to analyze the dynamic behavior of economic and financial variables. The leading choices are linear models, such as autoregressive (AR) models, moving average (MA) models, and mixed ARMA models. The linear time series models became popular partly because they have been incorporated into many “canned” statistics and econometrics packages. Although these models are quite successful in numerous applications, they are unable to represent many nonlinear dynamic patterns such as asymmetry, amplitude dependence and volatility clustering. For example, GDP growth rates typically fluctuate around a higher level and are more persistent during expansions, but they stay at a relatively lower level and less persistent during contractions. For such data, it would not be reasonable to expect a single, linear model to capture these distinct behaviors.

In the past two decades, we have witnessed a rapid growth of the development of nonlinear time series models; see e.g., Tong (1990) and Granger and Teräsvirta (1993) for more thorough discussions. Nonlinear time series models are, however, not a panacea and have their own limitations. First, implementing nonlinear models is typically cumbersome. For instance, the nonlinear optimization algorithms are easy to get stuck at a local optimum in the parameter space. Second, most nonlinear models are designed to describe certain nonlinear patterns of data and hence may not be so flexible as one would like. The latter problem suggests that the success of a nonlinear model largely depends on the data set to which it applies. An exception is the so-called *artificial neural network model* which, due to its “universal approximation” property, is capable of characterizing any nonlinear pattern in data; see e.g., Kuan and White (1994). Unfortunately, this model suffers from the identification problem and is hence vulnerable.

The *Markov switching model* of Hamilton (1989), also known as the *regime switching model*, is one of the most popular nonlinear time series models in the literature. This model involves multiple structures (equations) that can characterize the time series behaviors in different regimes. By permitting switching between these structures, this model is able to capture more complex dynamic patterns. A novel feature of the Markov switching model is that the switching mechanism is controlled by an unobservable state variable that follows a first-order Markov chain. In particular, the Markovian property regulates that the current value of the state variable depends on its immediate past value. As such, a structure may prevail for a random period of time, and it will be replaced by another structure when a switching takes place. This is in sharp contrast with the random switching model of Quandt (1972) in which the events of switching

---

are independent over time. The Markov switching model also differs from the models of *structural changes*. While the former allows for frequent changes at random time points, the latter admits only occasion and exogenous changes. The Markov switching model is therefore suitable for describing correlated data that exhibit distinct dynamic patterns during different time periods.

The original Markov switching model focuses on the mean behavior of variables. This model and its variants have been widely applied to analyze economic and financial time series; see e.g., Hamilton (1988, 1989), Engel and Hamilton (1990), Lam (1990), Garcia and Perron (1996), Goodwin (1993), Diebold, Lee and Weinbach (1994), Engel (1994), Filardo (1994), Ghysels (1994), Sola and Driffill (1994), Kim and Yoo (1995), Schaller and van Norden (1997), and Kim and Nelson (1998), among many others. Recently, this model has also been a popular choice in the study of Taiwan's business cycles; see Huang, Kuan and Lin (1998), Huang (1999), Chen and Lin (2000a, b), Hsu and Kuan (2001) and Rau, Lin and Li (2001). Given that the Markov switching model of conditional mean is highly successful, it is natural to consider incorporating this switching mechanism into conditional variance models. A leading class of conditional variance models is the GARCH (generalized autoregressive conditional heteroskedasticity) model introduced by Engle (1982) and Bollerslev (1986). Cai (1994), Hamilton and Susmel (1994) and Gray (1996) study various ARCH and GARCH models with Markov switching. So, Lam and Li (1998) also introduce Markov switching to the stochastic volatility model of Melino and Turnbull (1990), Harvey, Ruiz, and Shephard (1994), and Jacquier, Polson and Rossi (1994). Other financial applications of switching conditional variance models include, among others, Hamilton and Lin (1996), Dueker (1997), and Ramchand and Susmel (1998). Chen and Lin (1999) and Lin, Hung and Kuan (2002) also apply these models to analyze Taiwan's financial time series.

This lecture note is organized as follows. In Section 2, we introduce a simple Markov switching model of conditional mean and its generalizations. We then study two estimation methods (quasi-maximum likelihood method and Gibbs sampling) in Section 3 and discuss how to conduct hypothesis testing in Section 4. Section 5 is an empirical study of Taiwan's business cycles based on a bivariate Markov switching model. Section 6 presents the Markov switching model of conditional variance. Section 7 is an empirical analysis of Taiwan's short term interest rates. Section 8 concludes this note. Readers may also consult Hamilton (1994) for a concise treatment of the Markov switching model. For a more complete discussion of this model and its applications we refer to Kim and Nelson (1999); computer programs (written in GAUSS by C.-J. Kim) for implementing

this model are available from the web site:

<http://www.econ.washington.edu/user/cnelson/SSMARKOV.htm>

Some GAUSS programs written by Y.-L. Huang, S.-H. Hsu and the author are also available upon request.

## 2 The Markov Switching Model of Conditional Mean

Numerous empirical evidences suggest that the time series behaviors of economic and financial variables may exhibit different patterns over time. Instead of using one model for the conditional mean of a variable, it is natural to employ several models to represent these patterns. A Markov switching model is constructed by combining two or more dynamic models via a Markovian switching mechanism. Following Hamilton (1989, 1994), we shall focus on the Markov switching AR model. In this section, we first illustrate the features of Markovian switching using a simple model and then discuss more general model specifications.

### 2.1 A Simple Model

Let  $s_t$  denote an unobservable state variable assuming the value one or zero. A simple switching model for the variable  $z_t$  involves two AR specifications:

$$z_t = \begin{cases} \alpha_0 + \beta z_{t-1} + \varepsilon_t, & s_t = 0, \\ \alpha_0 + \alpha_1 + \beta z_{t-1} + \varepsilon_t, & s_t = 1, \end{cases} \quad (2.1)$$

where  $|\beta| < 1$  and  $\varepsilon_t$  are i.i.d. random variables with mean zero and variance  $\sigma_\varepsilon^2$ . This is a stationary AR(1) process with mean  $\alpha_0/(1 - \beta)$  when  $s_t = 0$ , and it switches to another stationary AR(1) process with mean  $(\alpha_0 + \alpha_1)/(1 - \beta)$  when  $s_t$  changes from 0 to 1. Then provided that  $\alpha_1 \neq 0$ , this model admits two dynamic structures at different levels, depending on the value of the state variable  $s_t$ . In this case,  $z_t$  are governed by two distributions with distinct means, and  $s_t$  determines the switching between these two distributions (regimes).

When  $s_t = 0$  for  $t = 1, \dots, \tau_0$  and  $s_t = 1$  for  $t = \tau_0 + 1, \dots, T$ , the model (2.1) is the model with a single structural change in which the model parameter experiences one (and only one) abrupt change after  $t = \tau_0$ . When  $s_t$  are independent Bernoulli random variables, it is the random switching model of Quandt (1972). In the random switching model, the realization of  $s_t$  is independent of the previous and future states so that  $z_t$

may be “jumpy” (switching back and forth between different states). If  $s_t$  is postulated as the indicator variable  $\mathbf{1}_{\{\lambda_t \leq c\}}$  such that  $s_t = 0$  or  $1$  depending on whether the value of  $\lambda_t$  is greater than the cut-off (threshold) value  $c$ , (2.1) becomes a *threshold model*. It is quite common to choose a lagged dependent variable (say,  $z_{t-d}$ ) as the variable  $\lambda_t$ .

While these models are all capable of characterizing the time series behaviors in two regimes, each of them has its own limitations. For the model with a single structural change, it is very restrictive because only one change is admitted. Although extending this model to allow for multiple changes is straightforward, the resulting model estimation and hypothesis testing are typically cumbersome; see e.g., Bai and Perron (1998) and Bai (1999). Moreover, changes in such models are solely determined by time which is exogenous to the model. The random switching model, by contrast, permits multiple changes, yet its state variables are still exogenous to the dynamic structures in the model. This model also suffers from the drawback that the state variables are independent over time and hence may not be applicable to time series data. On the other hand, switching in the threshold model is dependent and endogenous and results in multiple changes. Choosing a suitable variable  $\lambda_t$  and the threshold value  $c$  for this model is usually a difficult task, however.

One approach to circumventing the aforementioned problems is to consider a different specification for  $s_t$ . In particular, suppose that  $s_t$  follows a first order Markov chain with the following transition matrix:

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} \mathbb{P}(s_t = 0 \mid s_{t-1} = 0) & \mathbb{P}(s_t = 1 \mid s_{t-1} = 0) \\ \mathbb{P}(s_t = 0 \mid s_{t-1} = 1) & \mathbb{P}(s_t = 1 \mid s_{t-1} = 1) \end{bmatrix} \\ &= \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}, \end{aligned} \tag{2.2}$$

where  $p_{ij}$  ( $i, j = 0, 1$ ) denote the transition probabilities of  $s_t = j$  given that  $s_{t-1} = i$ . Clearly, the transition probabilities satisfy  $p_{i0} + p_{i1} = 1$ . The transition matrix governs the random behavior of the state variable, and it contains only two parameters ( $p_{00}$  and  $p_{11}$ ). The model (2.1) with the Markovian state variable is known as a Markov switching model. The Markovian switching mechanism was first considered by Goldfeld and Quandt (1973). Hamilton (1989) presents a thorough analysis of the Markov switching model and its estimation method; see also Hamilton (1994) and Kim and Nelson (1999).

In the Markov switching model, the properties of  $z_t$  are jointly determined by the random characteristics of the driving innovations  $\varepsilon_t$  and the state variable  $s_t$ . In particular, the Markovian state variable yields random and frequent changes of model structures,

and its transition probabilities determine the persistence of each regime. While the threshold model also possesses similar features, the Markov switching model is relatively easy to implement because it does not require choosing *a priori* the threshold variable  $\lambda_t$ . Instead, the regime classification in this model is probabilistic and determined by data. A difficulty with the Markov switching model is that it may not be easy to interpret because the state variables are unobservable.

## 2.2 Some Extensions

The model (2.1) is readily extended to allow for more general dynamic structures. Consider first a straightforward extension of the model (2.1):

$$z_t = \alpha_0 + \alpha_1 s_t + \beta_1 z_{t-1} + \cdots + \beta_k z_{t-k} + \varepsilon_t, \quad (2.3)$$

where  $s_t = 0, 1$  are the Markovian state variables with the transition matrix (2.2), and  $\varepsilon_t$  are i.i.d. random variables with mean zero and variance  $\sigma_\varepsilon^2$ . This is a model with a general AR( $k$ ) dynamic structure and switching intercepts.

For the  $d$ -dimensional time series  $\{z_t\}$ , we write (2.3) as

$$z_t = \alpha_0 + \alpha_1 s_t + \mathbf{B}_1 z_{t-1} + \cdots + \mathbf{B}_k z_{t-k} + \varepsilon_t, \quad (2.4)$$

where  $s_t = 0, 1$  are still the Markovian state variables with the transition matrix (2.2),  $\mathbf{B}_i$  ( $i = 1, \dots, k$ ) are  $d \times d$  matrices of parameters, and  $\varepsilon_t$  are i.i.d. random vectors with mean zero and the variance-covariance matrix  $\Sigma_\varepsilon$ . Clearly, (2.4) is a VAR (vector autoregressive) model with switching intercepts. This generalization is easy, but it may not always be realistic to require  $d$  variables to switch at the same time.

What we have discussed thus far are the 2-state Markov switching model because the state variable is binary. Further generalizations of these models are possible. For example, we may allow the state variable to assume  $m$  values, where  $m > 2$ , and obtain the  $m$ -state Markov switching model. Such models are essentially the same as the models given above, except that the transition matrix  $\mathbf{P}$  must be expanded accordingly. We may also set  $z_t$  to depend on both current and past state variables. Specifically, let  $\tilde{z}_t = z_t - \alpha_0 - \alpha_1 s_t$  and postulate the following model:

$$\tilde{z}_t = \beta_1 \tilde{z}_{t-1} + \cdots + \beta_k \tilde{z}_{t-k} + \varepsilon_t. \quad (2.5)$$

Then,  $\tilde{z}_t$  (and hence  $z_t$ ) depends not only on  $s_t$  but also on  $s_{t-1}, \dots, s_{t-k}$ . As there are  $2^{k+1}$  possible values of the collection  $(s_t, s_{t-1}, \dots, s_{t-k})$ , the model (2.5) can be viewed as (2.3) with  $2^{k+1}$  states. Another generalization is to allow for time-varying transition

probabilities. For example, the transition probabilities may be postulated as functions of some exogenous (or predetermined) variables so that they may vary with time. Clearly, a time-varying Markov switching model is even more flexible but involves more parameters.

### 2.3 Markov Trend

The Markov switching model and its variants discussed in the preceding sections are only suitable for stationary data. Let  $y_t$  be the observed time series which contains a unit root. The Markov switching model should be applied to the differenced series  $z_t = \Delta y_t = y_t - y_{t-1}$ . When  $y_t$  are quarterly data containing a seasonal unit root, we apply the Markov switching model to seasonally differenced series  $z_t = \Delta_4 y_t = y_t - y_{t-4}$ .

When a unit root is present in  $y_t$ , the switching intercept in  $z_t$  results in a deterministic trend with breaks in  $y_t$ . Given  $z_t$  in (2.3),  $y_t$  can be expressed as

$$y_t = \left( \alpha_0 t + \alpha_1 \sum_{i=1}^t s_i \right) + \beta_1 y_{t-1} + \cdots + \beta_k y_{t-k} + \sum_{i=1}^t \varepsilon_t,$$

where the two terms in the parenthesis is a trend function with changes, the second term is a dynamic component, and the last term  $\sum_{i=1}^t \varepsilon_t$  is the stochastic trend. It is clear that the trend function depends on  $s_t$ ; the resulting trend is therefore known as a *Markov trend*. The “basic” slope of this trend function is  $\alpha_0$ . When there is one  $s_i = 1$ , the trend function moves upward (downward) by  $\alpha_1$ ; when  $s_i$  takes the value 1 consecutively, these state variables yield a slope change in the trend function. This function would resume the original slope when  $s_i$  switches to the value 0. In Figure 1, we illustrate two Markov trend lines, where the black boxes signify the periods at which  $s_i = 1$ . The left figure shows the trend with  $\alpha_0 > 0$  and  $\alpha_1 > 0$ ; the right figure shows the one with  $\alpha_0 > 0$  and  $\alpha_1 < 0$ . It can be seen that both lines are kinked.

## 3 Model Estimation

There are various ways to estimate the Markov switching model; see Hamilton (1989, 1990, 1994), Kim (1994), and Kim and Nelson (1999). In this section we focus on the model (2.3) and discuss its quasi-maximum likelihood estimation and estimation via *Gibbs sampling*; estimating (2.4) is completely analogous with minor modifications. We also discuss the estimation of the more general model (2.5) in Appendix I.



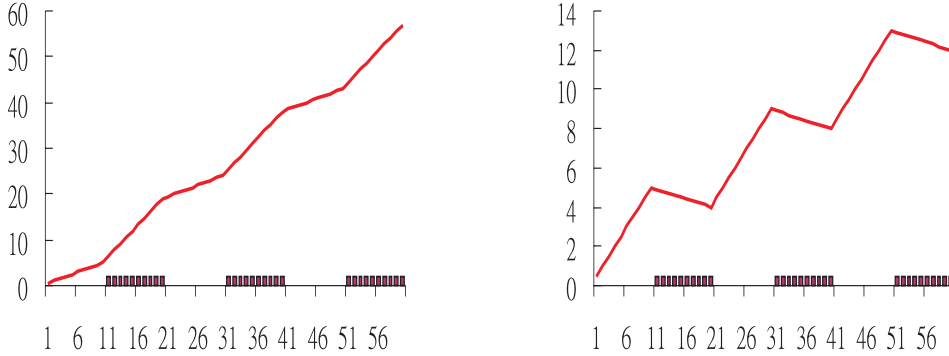


Figure 1: The Markov trend function with  $\alpha_1 > 0$  (left) and  $\alpha_1 < 0$  (right).

### 3.1 Quasi-Maximum Likelihood Estimation

Given the model (2.3), the vector of parameters is

$$\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1, \dots, \beta_k, \sigma_\varepsilon^2, p_{00}, p_{11})'.$$

Let  $\mathcal{Z}^t = \{z_t, z_{t-1}, \dots, z_1\}$  denote the collection of all the observed variables up to time  $t$ , which represents the information set we have at time  $t$ . Then,  $\mathcal{Z}^T$  is the information set based on the full sample. To assess the likelihood of the state variable  $s_t$ , it is important to evaluate its optimal forecasts (conditional expectations) of  $s_t = i$ ,  $i = 0, 1$ , based on different information sets. These forecasts include the *prediction probabilities*  $\mathbb{P}(s_t = i \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})$  which are based on the information prior to time  $t$ , the *filtering probabilities*  $\mathbb{P}(s_t = i \mid \mathcal{Z}^t; \boldsymbol{\theta})$  which are based on the past and current information, and the *smoothing probabilities*  $\mathbb{P}(s_t = i \mid \mathcal{Z}^T; \boldsymbol{\theta})$  which are based on the full-sample information. By deriving the algorithms of these probabilities, we also obtain the quasi-log-likelihood function as a byproduct, from which the quasi-maximum likelihood estimates (QMLE) can be computed.

Under the normality assumption, the density of  $z_t$  conditional on  $\mathcal{Z}^{t-1}$  and  $s_t = i$  ( $i = 0, 1$ ) is

$$\begin{aligned} f(z_t \mid s_t = i, \mathcal{Z}^{t-1}; \boldsymbol{\theta}) \\ = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{ \frac{-(z_t - \alpha_0 - \alpha_1 i - \beta_1 z_{t-1} - \dots - \beta_k z_{t-k})^2}{2\sigma_\varepsilon^2} \right\}. \end{aligned} \tag{3.1}$$

Given the prediction probability  $\mathbb{P}(s_t = i \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})$ , the density of  $z_t$  conditional on

$\mathcal{Z}^{t-1}$  alone can be obtained from (3.1) as

$$\begin{aligned} f(z_t | \mathcal{Z}^{t-1}; \boldsymbol{\theta}) &= \mathbb{P}(s_t = 0 | \mathcal{Z}^{t-1}; \boldsymbol{\theta}) f(z_t | s_t = 0, \mathcal{Z}^{t-1}; \boldsymbol{\theta}) \\ &\quad + \mathbb{P}(s_t = 1 | \mathcal{Z}^{t-1}; \boldsymbol{\theta}) f(z_t | s_t = 1, \mathcal{Z}^{t-1}; \boldsymbol{\theta}). \end{aligned} \quad (3.2)$$

For  $i = 0, 1$ , the filtering probabilities of  $s_t$  are

$$\mathbb{P}(s_t = i | \mathcal{Z}^t; \boldsymbol{\theta}) = \frac{\mathbb{P}(s_t = i | \mathcal{Z}^{t-1}; \boldsymbol{\theta}) f(z_t | s_t = i, \mathcal{Z}^{t-1}; \boldsymbol{\theta})}{f(z_t | \mathcal{Z}^{t-1}; \boldsymbol{\theta})}, \quad (3.3)$$

by the Bayes theorem, and the relationship between the filtering and prediction probabilities is

$$\mathbb{P}(s_{t+1} = i | \mathcal{Z}^t; \boldsymbol{\theta}) = p_{0i} \mathbb{P}(s_t = 0 | \mathcal{Z}^t; \boldsymbol{\theta}) + p_{1i} \mathbb{P}(s_t = 1 | \mathcal{Z}^t; \boldsymbol{\theta}), \quad (3.4)$$

where  $p_{0i} = \mathbb{P}(s_{t+1} = i | s_t = 0)$  and  $p_{1i} = \mathbb{P}(s_{t+1} = i | s_t = 1)$  are transition probabilities. Observe that the equations (3.1)–(3.4) form a recursive system for  $t = k, \dots, T$ .

With the initial values  $\mathbb{P}(s_k = i | \mathcal{Z}^{k-1}; \boldsymbol{\theta})$ ,<sup>1</sup> we can iterate the equations (3.1)–(3.4) to obtain the filtering probabilities  $\mathbb{P}(s_t = i | \mathcal{Z}^t; \boldsymbol{\theta})$  as well as the conditional densities  $f(z_t | \mathcal{Z}^{t-1}; \boldsymbol{\theta})$  for  $t = k, \dots, T$ . The quasi-log-likelihood function is therefore

$$\mathcal{L}_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T \ln f(z_t | \mathcal{Z}^{t-1}; \boldsymbol{\theta}),$$

which is a complex nonlinear function of  $\boldsymbol{\theta}$ . The QMLE  $\hat{\boldsymbol{\theta}}_T$  now can be computed using a numerical-search algorithm. For example, the GAUSS program adopts the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm. The estimated filtering and prediction probabilities are then easily calculated by plugging  $\hat{\boldsymbol{\theta}}_T$  into the formulae for these probabilities.

To compute the smoothing probabilities  $\mathbb{P}(s_t = i | \mathcal{Z}^T; \boldsymbol{\theta})$  we follow the approach of Kim (1994). By noting that

$$\begin{aligned} \mathbb{P}(s_t = i | s_{t+1} = j, \mathcal{Z}^T; \boldsymbol{\theta}) \\ &= \mathbb{P}(s_t = i | s_{t+1} = j, \mathcal{Z}^t; \boldsymbol{\theta}) \\ &= \frac{p_{ij} \mathbb{P}(s_t = i | \mathcal{Z}^t; \boldsymbol{\theta})}{\mathbb{P}(s_{t+1} = j | \mathcal{Z}^t; \boldsymbol{\theta})}, \end{aligned}$$

<sup>1</sup>Hamilton (1994, p. 684) suggests setting the initial value  $\mathbb{P}(s_k = i | \mathcal{Z}^{k-1}; \boldsymbol{\theta})$  to its limiting unconditional counterpart: the third column of the matrix  $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ , where

$$\mathbf{A} = \begin{bmatrix} \mathbf{I} - \mathbf{P} \\ \mathbf{1}' \end{bmatrix},$$

with  $\mathbf{I}$  the identity matrix and  $\mathbf{1}$  the two-dimensional vector of ones.

for  $i, j = 0, 1$ , the smoothing probabilities can be expressed as

$$\begin{aligned}
& \mathbb{P}(s_t = i \mid \mathcal{Z}^T; \boldsymbol{\theta}) \\
&= \mathbb{P}(s_{t+1} = 0 \mid \mathcal{Z}^T; \boldsymbol{\theta}) \mathbb{P}(s_t = i \mid s_{t+1} = 0, \mathcal{Z}^T; \boldsymbol{\theta}) \\
&\quad + \mathbb{P}(s_{t+1} = 1 \mid \mathcal{Z}^T; \boldsymbol{\theta}) \mathbb{P}(s_t = i \mid s_{t+1} = 1, \mathcal{Z}^T; \boldsymbol{\theta}) \\
&= \mathbb{P}(s_t = i \mid \mathcal{Z}^t; \boldsymbol{\theta}) \\
&\quad \times \left( \frac{p_{i0} \mathbb{P}(s_{t+1} = 0 \mid \mathcal{Z}^T; \boldsymbol{\theta})}{\mathbb{P}(s_{t+1} = 0 \mid \mathcal{Z}^t; \boldsymbol{\theta})} + \frac{p_{i1} \mathbb{P}(s_{t+1} = 1 \mid \mathcal{Z}^T; \boldsymbol{\theta})}{\mathbb{P}(s_{t+1} = 1 \mid \mathcal{Z}^t; \boldsymbol{\theta})} \right).
\end{aligned} \tag{3.5}$$

Using the filtering probability  $\mathbb{P}(s_T = i \mid \mathcal{Z}^T; \boldsymbol{\theta})$  as the initial value, we can iterate the equations (3.3), (3.4) and (3.5) backward to get the smoothing probabilities for  $t = T - 1, \dots, k + 1$ . These probabilities are also functions of  $\boldsymbol{\theta}$ ; plugging the QMLE  $\hat{\boldsymbol{\theta}}_T$  into these formulae yields the estimated smoothing probabilities.

### 3.2 Estimation via Gibbs Sampling

An alternative approach to estimating the Markov switching model is the method of Gibbs sampling; see e.g., Albert and Chib (1993) and McCulloch and Tsay (1994). Gibbs sampling is a *Markov Chain Monte Carlo* simulation method (also known as MCMC method) introduced by Geman and Geman (1984) for image processing problems, and it is closely related to the idea of data augmentation of Tanner and Wong (1987).

Similar to the Bayesian analysis, this method treats parameters as random variables. Suppose that the parameter vector  $\boldsymbol{\theta}$  can be classified into  $k$  groups:

$$\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2, \dots, \boldsymbol{\theta}'_k)'$$

Given the observed data  $\mathcal{Z}^T$ , let

$$\pi(\boldsymbol{\theta}_i \mid \mathcal{Z}^T, \{\boldsymbol{\theta}_j, j \neq i\}), \quad i = 1, \dots, k,$$

denote the *full conditional distribution* of  $\boldsymbol{\theta}_i$ , which is also the *conditional posterior distribution* in the Bayesian analysis. By specifying the prior distributions of parameters and likelihood functions, the conditional posterior distributions can be derived.

The Gibbs sampler starts from  $k$  conditional posterior distributions and randomly generated initial values:

$$\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_1^{(0)'}, \boldsymbol{\theta}_2^{(0)'}, \dots, \boldsymbol{\theta}_k^{(0)'})'$$

The  $i$ th realization of  $\boldsymbol{\theta}$  is then obtained via the following procedure.

1. Randomly draw a realization of  $\boldsymbol{\theta}_1$  from the full conditional distribution

$$\pi(\boldsymbol{\theta}_1 \mid \mathcal{Z}^T, \boldsymbol{\theta}_2^{(i-1)}, \dots, \boldsymbol{\theta}_k^{(i-1)}),$$

and denote this realization as  $\boldsymbol{\theta}_1^{(i)}$ .

2. Randomly draw a realization of  $\boldsymbol{\theta}_2$  from the full conditional distribution

$$\pi(\boldsymbol{\theta}_2 \mid \mathcal{Z}^T, \boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_3^{(i-1)}, \dots, \boldsymbol{\theta}_k^{(i-1)}),$$

and denote this realization as  $\boldsymbol{\theta}_2^{(i)}$ .

3. Proceeds similarly to draw  $\boldsymbol{\theta}_3^{(i)}, \dots, \boldsymbol{\theta}_k^{(i)}$ .

The  $i$ th realization of  $\boldsymbol{\theta}$  is then

$$\boldsymbol{\theta}^{(i)} = (\boldsymbol{\theta}_1^{(i)'}, \boldsymbol{\theta}_2^{(i)'}, \dots, \boldsymbol{\theta}_k^{(i)'})'.$$

Repeating the procedure above  $N$  times we obtain a Gibbs sequence  $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(N)}\}$ .

We can then compute  $N$  full conditional distributions of  $\boldsymbol{\theta}_i$  based on the Gibbs sequence.

For example, the full conditional distributions of  $\boldsymbol{\theta}_1$  are

$$\pi(\boldsymbol{\theta}_1 \mid \mathcal{Z}^T, \boldsymbol{\theta}_2^{(i)}, \boldsymbol{\theta}_3^{(i)}, \dots, \boldsymbol{\theta}_k^{(i)}), \quad i = 1, \dots, N.$$

To get rid of the effect of initial values, it is typical to drop the beginning  $N_1$  estimates in the Gibbs sequence and keep the remaining  $N_2$  estimates, where  $N_1 + N_2 = N$ .

Geman and Geman (1984) show that the Gibbs sequence converges in distribution exponentially fast to the true distribution of  $\boldsymbol{\theta}$ , i.e.,

$$\boldsymbol{\theta}^{(N)} \xrightarrow{D} \pi(\boldsymbol{\theta} \mid \mathcal{Z}^T),$$

as  $N$  tends to infinity, and that each subvector  $\boldsymbol{\theta}_i^{(N)}$  also converges in distribution exponentially fast to the true marginal distribution of  $\boldsymbol{\theta}_i$ . Moreover, for any measurable function  $g$ ,

$$\frac{1}{N} \sum_{i=1}^N g(\boldsymbol{\theta}^{(i)}) \xrightarrow{\text{a.s.}} \mathbb{E}[g(\boldsymbol{\theta})],$$

where  $\xrightarrow{\text{a.s.}}$  denotes almost sure convergence. See also Gelfand and Smith (1990), Casella and George (1992), and Chib and Greenberg (1996) for more detailed discussion of the properties of Gibbs sampling.

In the current context, in addition to the parameter vector  $\boldsymbol{\theta}$ , the unobserved state variables  $s_t$ ,  $t = 1, \dots, T$ , are also treated as parameters. The augmented parameter vector now can be classified into four groups: state variables  $s_t$ , transition probabilities  $p_{00}$  and  $p_{11}$ , the intercept and slope parameters  $\alpha_0, \alpha_1, \beta_1, \dots, \beta_k$ , and the variance  $\sigma_\varepsilon^2$ . Random drawings from the conditional posterior distributions yield the Gibbs sequence. The sample average of the Gibbs sequence is the desired estimate of unknown parameters.

## 4 Hypothesis Testing

To justify whether the Markov switching model is appropriate, it is natural to consider the following hypotheses: (1) the switching parameters (intercepts) are in fact the same; (2) the state variables are independent. Rejecting the first hypothesis suggests that switching does occur. Failure to reject the second hypothesis is an evidence against the Markovian structure, yet rejecting this hypothesis provides only a partial support for Markov switching. In addition, one may also want to test the significance of model parameters as well as some linear (nonlinear) hypotheses of these parameters. In this section, we will discuss how to conduct proper tests.

### 4.1 Testing for Switching Parameters

The first hypothesis is that  $\alpha_1 = 0$ . When the null hypothesis is true, one equation suffices to characterize  $z_t$  so that (2.3) reduces to an AR( $k$ ) model. Then, the quasi-log-likelihood value can not be affected by the values of  $p_{00}$  and  $p_{11}$ . That is, the parameters  $p_{00}$  and  $p_{11}$  are *not* identified under the null hypothesis; these parameters are usually referred to as *nuisance parameters*. It is well known that when there are unidentified parameters under the null hypothesis, the quasi-log-likelihood function is flat with respect to these nuisance parameters so that there is no unique maximum. Consequently, the standard likelihood-based tests are no longer valid, as discussed in Davies (1977, 1987) and Hansen (1996b). This is a very serious problem in hypothesis testing.

We now introduce the “conservative” testing procedure proposed by Hansen (1992, 1996a). It is worth noting that the asymptotic theory of the test of Garcia (1998) may not be valid. Partition the parameter vector  $\boldsymbol{\theta}$  as

$$\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\theta}'_1)' = (\alpha_1, p_{00}, p_{11}, \boldsymbol{\theta}'_1)',$$

where  $\boldsymbol{p}' = (p_{00}, p_{11})$  is the vector of the nuisance parameters not identified under the null. Fixing  $\boldsymbol{\gamma} = (\alpha_1, p_{00}, p_{11})'$ , the concentrated QMLE of  $\boldsymbol{\theta}_1$  is

$$\hat{\boldsymbol{\theta}}_1(\boldsymbol{\gamma}) = \operatorname{argmax} L_T(\boldsymbol{\gamma}, \boldsymbol{\theta}_1).$$

which converges in probability to, say,  $\boldsymbol{\theta}_1(\gamma)$ . The concentrated quasi-log-likelihood functions evaluated at  $\hat{\boldsymbol{\theta}}_1(\gamma)$  and  $\boldsymbol{\theta}_1(\gamma)$  are

$$\begin{aligned}\hat{L}_T(\gamma) &= L_T(\gamma, \hat{\boldsymbol{\theta}}_1(\gamma)), \\ L_T(\gamma) &= L_T(\gamma, \boldsymbol{\theta}_1(\gamma)).\end{aligned}$$

The resulting likelihood ratio statistics are then

$$\begin{aligned}\widehat{\mathcal{LR}}_T(\gamma) &= \hat{L}_T(\gamma) - \hat{L}_T(0, p_{00}, p_{11}), \\ \mathcal{LR}_T(\gamma) &= L_T(\gamma) - L_T(0, p_{00}, p_{11}).\end{aligned}$$

where  $\hat{L}_T(0, \mathbf{p})$  and  $L_T(0, \mathbf{p})$  are the concentrated quasi-log-likelihood functions under the null hypothesis. As  $\gamma$  contains the nuisance parameters, it is natural to consider the likelihood ratios for all possible values of  $\gamma$ . This motivates a supremum statistic:  $\sup_{\gamma} \sqrt{T} \widehat{\mathcal{LR}}_T(\gamma)$ ; see also Andrews (1993) for an analogous test in the context of testing structural changes at unknown times.

In view of equations (A.1), (A.2) and (8) of Hansen (1992),<sup>2</sup>

$$\sqrt{T}(\hat{L}_T(\gamma) - L_T(\gamma)) = o_{\mathbb{P}}(1).$$

It follows that under the null hypothesis,

$$\begin{aligned}\sqrt{T} \widehat{\mathcal{LR}}_T(\gamma) &= \sqrt{T} [\widehat{\mathcal{LR}}_T(\gamma) - \mathcal{LR}_T(\gamma)] + \sqrt{T} [\mathcal{LR}_T(\gamma) - M_T(\gamma)] + \sqrt{T} M_T(\gamma) \\ &= \sqrt{T} [\mathcal{LR}_T(\gamma) - M_T(\gamma)] + \sqrt{T} M_T(\gamma) + o_{\mathbb{P}}(1),\end{aligned}$$

where  $M_T(\gamma) = \mathbb{E}[\mathcal{LR}_T(\gamma)]$  is non-positive when the null hypothesis is true. It follows that

$$\sqrt{T} \widehat{\mathcal{LR}}_T(\gamma) \leq \sqrt{T} Q_T(\gamma) + o_{\mathbb{P}}(1), \quad (4.1)$$

where  $Q_T(\gamma) = \mathcal{LR}_T(\gamma) - M_T(\gamma)$ . Under suitable conditions, an empirical-process central limit theorem (CLT) holds in the sense that

$$\sqrt{T} Q_T(\gamma) \Rightarrow Q(\gamma), \quad (4.2)$$

where  $\Rightarrow$  stands for weak convergence (of the associated probability measures), and  $Q$  is a Gaussian process with mean zero and the covariance function  $K(\gamma_1, \gamma_2)$ . Note that an empirical-process CLT is analogous to a functional CLT; see Andrews (1991) for more

<sup>2</sup>It should be noted that in our notations,  $\hat{L}_T$  and  $L_T$  are *averages* of individual log-likelihoods, whereas they are sums of individual log-likelihoods in Hansen (1992).

details. Equations (4.1) and (4.2) together suggest that when  $T$  is sufficiently large,  $Q(\gamma)$  is approximately an upper bound of  $\sqrt{T} \widehat{\mathcal{LR}}_T(\gamma)$  for any  $\gamma$ . We thus have

$$\mathbb{P} \left\{ \sup_{\gamma} \sqrt{T} \widehat{\mathcal{LR}}_T(\gamma) > c \right\} \leq \mathbb{P} \left\{ \sup_{\gamma} Q(\gamma) > c \right\}, \quad (4.3)$$

under the null hypothesis. The result (4.3) shows that for a given significance level, the critical value of  $\sup_{\gamma} \sqrt{T} \widehat{\mathcal{LR}}_T(\gamma)$  is smaller than that of  $\sup_{\gamma} Q(\gamma)$ .

Based on the derivation above, Hansen (1992) proposed using a standardized supremum statistic

$$\sup_{\gamma} \widehat{\mathcal{LR}}_T^*(\gamma) = \sup_{\gamma} \sqrt{T} \widehat{\mathcal{LR}}_T(\gamma) / \hat{V}_T(\gamma)^{1/2}, \quad (4.4)$$

where  $\hat{V}_T(\gamma)$  is a variance estimate; the exact form of  $\hat{V}_T(\gamma)$  is given by Hansen (1992). Let  $V(\gamma)$  be the probability limit of  $\hat{V}_T(\gamma)$ . In the light of (4.1) and (4.2), the Hansen statistic is such that

$$\sup_{\gamma} \widehat{\mathcal{LR}}_T^*(\gamma) \leq \sup_{\gamma} \sqrt{T} Q_T(\gamma) / \hat{V}_T(\gamma)^{1/2} + o_{\mathbb{P}}(1) \Rightarrow \sup_{\gamma} Q^*(\gamma),$$

where  $Q^*(\gamma) = Q(\gamma) / V(\gamma)^{1/2}$  is also a Gaussian process with mean zero and the covariance function  $K^*(\gamma_1, \gamma_2) = K(\gamma_1, \gamma_2) / [V(\gamma_1)^{1/2} V(\gamma_2)^{1/2}]$ . Similar to (4.3), we have

$$\mathbb{P} \left\{ \sup_{\gamma} \widehat{\mathcal{LR}}_T^*(\gamma) > c \right\} \leq \mathbb{P} \left\{ \sup_{\gamma} Q^*(\gamma) > c \right\}.$$

Thus, for a given significance level, the critical value of  $\sup_{\gamma} \widehat{\mathcal{LR}}_T^*(\gamma)$  is also smaller than that of  $\sup_{\gamma} Q^*(\gamma)$ . This suggests that, even when the distribution of the standardized supremum statistic is unknown, we may appeal to  $\sup_{\gamma} Q^*(\gamma)$  and obtain “conservative” critical values for the standardized supremum statistic (the critical values render the true significance level less than the nominal significance level). Such critical values are larger than needed and hence ought to have negative effects on test power. Hansen (1992) also suggested a simulation approach to generate the distribution of  $\sup_{\gamma} Q^*(\gamma)$ . Implementing Hansen’s test is computationally quite intensive; we will discuss related computing issues in Appendix II.

## 4.2 Testing Other Hypotheses

Consider now the test of the independence of state variables. Note that if  $p_{00} = p_{10}$  and  $p_{01} = p_{11}$ , then regardless of the previous state, the variable has the same probability of being in the state 0 (or 1). That is, previous state has no effect on current state so that

---

these state variables are independent. Since  $p_{00} + p_{01} = 1$  and  $p_{10} + p_{11} = 1$ , the null hypothesis of independent state variables can be expressed compactly as

$$H_0: p_{00} + p_{11} = 1.$$

Once we reject the first hypothesis discussed in the preceding subsection, there is no longer the “nuisance parameter” problem. Consequently, the hypothesis of independent state variables can be tested using standard likelihood-based tests, such as the Wald test. Other hypotheses on model parameters can also be tested by standard likelihood-based tests; see also Hamilton (1996). We omit the details of those tests.

## 5 Application: A Study of Taiwan’s Business Cycles

In discussing business cycles, Lucas (1977) emphasizes on the comovement of important macroeconomic variables such as production, consumption, investment and employment. Diebold and Rudebusch (1996) further suggest that a model for business cycles should take into account two features: the comovement of economic variables and persistence of economic states. Clearly, a univariate Markov switching model is able to characterize the latter feature but not the former. It is therefore more appropriate to consider a multivariate model.

There have been numerous applications of the Markov switching model to aggregate output series and business cycles; see e.g., Hamilton (1989), Lam (1990), Goodwin (1993), Diebold, Lee and Weinbach (1994), Durland and McCurdy (1994), Filardo (1994), Ghysels (1994), Kim and Yoo (1995), Filardo and Gordon (1998), and Kim and Nelson (1998). Among the studies of Taiwan’s business cycles, Huang, Kuan and Lin (1998), Huang (1999), and Chen and Lin (2000a) applied univariate Markov switching models to real GNP or GDP data. Blanchard and Quah (1989) pointed out, however, that analyzing GDP alone is not enough to characterize the effects of both supply and demand shocks. Our empirical analysis below is based on Hsu and Kuan (2001) which applied a bivariate Markov switching model to real GDP and employment growth rates. Chen and Lin (2000b) also adopted a bivariate model to analyze real GDP and CP (consumption expenditure). We consider employment rather than CP because CP itself is already a major component of the GDP series.

The quarterly data of real GDP and employment numbers are taken from the AREMOS database of the Ministry of Education. There were total 151 observations for the real GDP (the first quarter of 1962 through the third quarter of 1999) and 87 observations



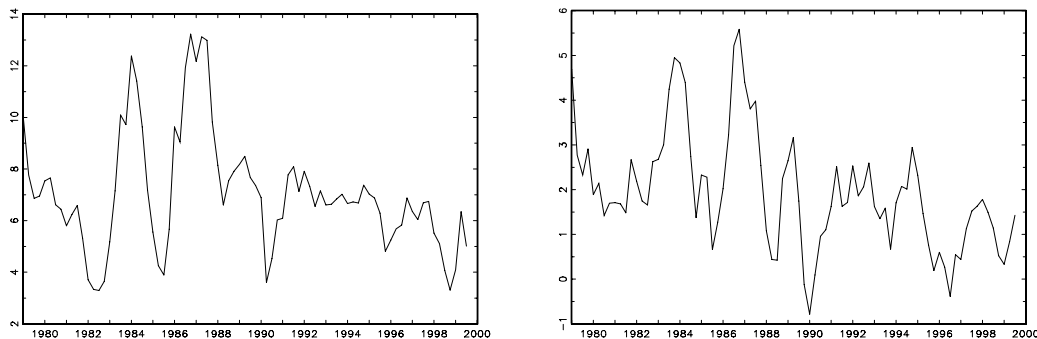


Figure 2: The growth rates of GDP (left) and employment (right): 1979 Q1–1999 Q3.

for employment numbers (the first quarter of 1978 through the third quarter of 1999). In what follows, we will use Q1 to denote the first quarter, Q2 for the second, and so on. Let  $\zeta_t$  denote the vector of GDP and employment. Taking seasonal differences of the log of  $\zeta_t$  yields the annual growth rates of  $\zeta_t$ :

$$z_t = \log(\zeta_t) - \log(\zeta_{t-4}).$$

The GDP and employment growth rates from 1979 Q1 through 1999 Q3 are plotted in Figure 2. It can be seen that these two series do not exhibit any trending behavior.

The estimation results of this section are all computed via Gibbs sampling; see Hsu and Kuan (2001) for the prior distributions and conditional posterior distributions. We first apply the bivariate Markov switching model to  $z_t$  using the observations of the whole sample (1978 Q1 through 1999 Q3). The estimation result shows that  $s_t = 1$  is the state of rapid growth. Figures 3 is the plot of the smoothing probabilities  $\text{IP}(s_t = 1 \mid \mathcal{Z}^T)$  which indicate that these probabilities are almost zero in the past 10 years, where the vertical solid (dashed) lines signify the peaks (troughs) identified by the CEPD (Council of Economic Planning and Development) of the Executive Yuan. That is, it is highly unlikely that the economy is in the state of rapid growth during this period (or it is highly likely that the economy is in the state of low growth). Thus, the Markov switching model based on the information of the full sample fails to identify any cycle in 1990s. By contrast, existing studies show that the Markov switching model is quite successful in identifying Taiwan's business cycles before 1990 and that their results are close to the cycles identified by the CEPD.

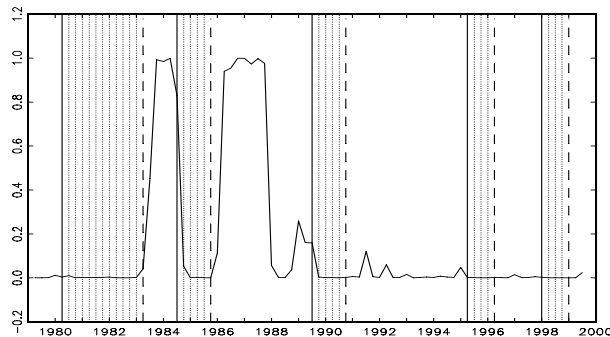


Figure 3: The smoothing probabilities of  $s_t = 1$ : bivariate model, 1979 Q1–1999 Q3.

Examining the data more closely we observe that Taiwan's economy grew rapidly before 1990 but much slower afterwards. For example, the average GDP growth rates in 1960s, 1970s and 1980s were, respectively, 9.82%, 10.27% and 8.16%, whereas the average growth rate in 1990s was only 6.19%. This explains why the Markov switching model classifies all the growth rates in 1990s into the same state when the full sample is considered. Nevertheless, from Figure 2 we can see that Taiwan's economy still experienced some ups and downs during this period. The question is: How can we identify the business cycles in 1990s?

To properly identify the cycles in 1990s, it seems natural to consider only a subsample of more recent observations. To be sure, we test whether there was a structural change (at an unknown time) in these two series using the maximal-Wald test of Andrews (1993) and estimate the change point using the least-squares method. For the GDP and employment growth rates, the maximal-Wald statistics are, respectively, 12.036 and 40.360, which exceed the 5% critical value 9.31. We therefore reject the null hypothesis of no mean change. The least-squares change-point estimates further indicate that the change point for the GDP growth rates was 1989 Q4 and that for the employment growth rates was 1987 Q4. Therefore, we shall concentrate on the after-change sample of  $z_t$  from 1989 Q4 through 1999 Q3.<sup>3</sup> Note that the way we determine the change point is different from that of Rau, Lin and Li (2001).

<sup>3</sup>Choosing this sub-sample is quite reasonable. The average growth rates of GDP and employment are 7.81% resp. 2.56% before 1990 and drop to 6.19% resp. 1.28% after 1990. This amounts to a 21% decrease in the GDP growth rates and a 50% decrease in the employment growth rates.

The estimation results summarized in Table 1 are obtained by applying the bivariate Markov switching model to the after-change sample. In this table, the columns under “prior dist.” are the parameter values of the prior distributions, and the columns under “posterior dist.” give the parameter estimates and their standard errors obtained from Gibbs sampling. From this result we can calculate the estimated average growth rates for GDP: 7.35% for state 1 and 3.26% for state 0. We will therefore term the states 1 and 0 as the rapid- and low-growth states, respectively. These estimates are significantly lower than those reported in other studies. For example, using the real GDP data from 1961 Q1 through 1996 Q4, the 2-state model of Huang (1999) results in the estimated average growth rates as 11.3% and 7.3%.<sup>4</sup> For employment, the estimated average growth rates in the rapid- and low-growth periods are 1.46% and 1.15%, respectively.

From Table 1 we also observe that the transition probabilities are  $p_{00} = 0.5619$  and  $p_{11} = 0.6918$ . These probabilities are also much smaller than those of other studies and suggest that both states are less persistent than before. For example, for the low- and rapid-growth states, the transition probabilities in Huang (1999) are, respectively, 0.927 and 0.804, whereas those in Huang, Kuan and Lin (1998) are, respectively, 0.927 and 0.956. The expected duration is approximately  $1/(1 - p_{11}) \approx 3.2$  quarters for the rapid-growth period and is  $1/(1 - p_{00}) \approx 2.3$  quarters for the low-growth period.<sup>5</sup> These durations are much shorter than those obtained by Huang, Kuan and Lin (1998), which are approximately 22.7 quarters for the period of rapid growth and 13.7 quarters for the period of low growth. Note, however, that the estimated durations of Huang (1999) are approximately 5 quarters for the rapid-growth period and 13.7 quarters for the low-growth period. These estimates contradict the usual wisdom that expansions in Taiwan typically last longer than recessions. To summarize, our estimation results indicate that the expected growth rates of Taiwan’s GDP are much lower and that the phases of business cycles have shorter durations in 1990s.

The smoothing (posterior) probabilities of  $s_t = 1$  are summarized in Table 2 and plotted in Figure 4. We use the smoothing probabilities to determine the peaks and troughs of business cycle and take 0.5 as the cut-off value for  $s_t = 0$  or 1. That is, the periods with the smoothing probabilities of  $s_t = 1$  greater (less) than 0.5 are more

<sup>4</sup>Huang, Kuan and Lin (1998) use the real GNP data from 1962 Q1 through 1995 Q3 and obtain the estimated average growth rates as 10.12% and 5.74%.

<sup>5</sup>The expected duration of the state 0 is

$$\sum_{k=1}^{\infty} k p_{00}^{k-1} (1 - p_{00}) = 1/(1 - p_{00}),$$

and the expected duration of the state 1 is  $1/(1 - p_{11})$ ; see Hamilton (1989, p. 374).

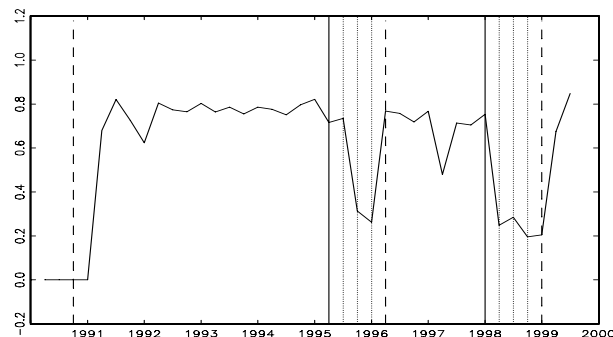


Figure 4: The smoothing probabilities of  $s_t = 1$ : bivariate model, 1990 Q1–1999 Q3.

likely to be in the state of rapid (low) growth. We also adopt the simple rule that the last period with the smoothing probability greater (less) than 0.5 is taken as the peak (trough). According to this rule, there were two complete cycles in 1990s: one with the peak at 1995 Q2 and trough at 1995 Q4, and another one with the peak at 1997 Q4 and trough at 1998 Q4. The former is close to the 8th cycle announced by the CEPD (peak at 1995 Q1 and trough at 1996 Q1) but with a shorter recession period, whereas the latter agrees with the 9th cycle identified by the CEPD. The 9th cycle shows that Taiwan's economy reached the peak while the Asian currency crisis started spreading and was at the trough when this crisis finally came to an end.

Furthermore, we adopt the prior distributions in Table 1 and separately apply the univariate Markov switching model to after-change GDP and employment growth rates, as did in Kim and Nelson (1998). The estimated smoothing probabilities for these two series are plotted in Figure 5. It is interesting to see that the univariate model for the after-change sample still cannot identify any cycle during 1990s. This suggests that the bivariate model does capture important data characteristics that cannot be revealed by a univariate model.

## 6 The Markov Switching Model of Conditional Variance

In addition to the Markov switching model of conditional mean, it is also important to incorporate a Markov switching mechanism into conditional variance models. In this section, we focus on the GARCH model with Markov switching.

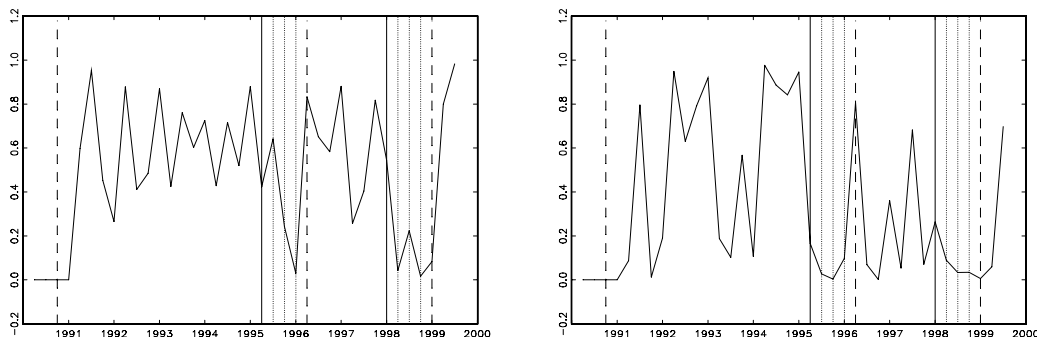


Figure 5: The smoothing probabilities of  $s_t = 1$ : univariate model for GDP (left) and employment (right), 1990 Q1–1999 Q3.

Write a simple GARCH( $p, q$ ) model as  $z_t = \sqrt{h_t} \varepsilon_t$ , where

$$h_t = c + \sum_{i=1}^q a_i z_{t-i}^2 + \sum_{i=1}^p b_i h_{t-i}, \quad (6.1)$$

which is the conditional variance of  $z_t$  given all the information up to time  $t - 1$  and  $\varepsilon_t$  are i.i.d. random variables with mean zero and variance 1. When  $h_t$  does not depend on its lagged values, the model above reduces to an ARCH( $q$ ) model. When  $p = q = 1$ , we have the GARCH(1,1) model:

$$h_t = c + a_1 z_{t-1}^2 + b_1 h_{t-1}.$$

In many empirical studies, it is found that a GARCH(1,1) model usually suffices to describe the volatility patterns in many time series. Interestingly, the sum of the estimated  $a_1$  and  $b_1$  coefficients is typically close to one. From 6.1), we can characterize  $z_t^2$  using an ARMA(1,1) model:

$$z_t^2 = h_t \varepsilon_t^2 = c + (a_1 + b_1) z_{t-1}^2 - b_1 (z_{t-1}^2 - h_{t-1}) + (z_t^2 - h_t), \quad (6.2)$$

where  $z_t^2 - h_t$  is the innovation with mean zero. Thus, when  $a_1 + b_1$  is indeed one,  $z_t^2$  has a unit root so that the resulting  $h_t$  are highly persistent. In this case,  $\{h_t\}$  is said to be an integrated GARCH (IGARCH) process. Lamoureux and Lastrapes (1990) point out that the detected IGARCH pattern does not have theoretical motivation and may well be a consequence of ignoring parameter changes in the GARCH model.

Let  $\Phi_{t-1}$  denote the information set up to time  $t-1$  and  $h_{i,t} = \text{var}(z_t | s_t = i, \Phi_{t-1})$ . Cai (1994) considers an ARCH( $q$ ) model with switching intercepts:  $z_t = \sqrt{h_{i,t}} \varepsilon_t$ , and

$$h_{i,t} = \alpha_0 + \alpha_1 i + \sum_{j=1}^q a_j z_{t-j}^2, \quad i = 0, 1. \quad (6.3)$$

Hamilton and Susmel (1994) proposed the SWARCH( $q$ ) model:  $z_t = \sqrt{h_{i,t}} \varepsilon_t$ , and

$$h_{i,t} = \lambda_i \eta_t = \lambda_i \left( c + \sum_{j=1}^q a_j \zeta_{t-j}^2 \right), \quad i = 0, 1. \quad (6.4)$$

That is, the conditional variances in these two regimes are proportional to each other. Clearly, the conditional variances of (6.3) have left shifts, but those of (6.4) have different scales. Both models are, of course, very special forms of switching conditional variances.

Extending the models (6.3) and (6.4) to allow for lagged conditional variances is not straightforward, however. To see this, observe that when the conditional variance  $h_{i,t}$  depends on  $h_{i,t-1}$ , it is determined not only by  $s_t$  but also by  $s_{t-1}$  due to the presence of  $h_{i,t-1}$ . The dependence of  $h_{i,t-1}$  on  $h_{i,t-2}$  then implies that  $h_{i,t}$  must also be affected by the value of  $s_{t-2}$ , and so on. Consequently, the conditional variance at time  $t$  is in effect determined by the realization of  $(s_t, s_{t-1}, \dots, s_1)$  which has  $2^t$  possible values. This “path dependence” property would result in a very complex model and render model estimation intractable. Gray (1996) circumvents this problem by postulating that  $h_{i,t}$  depends on  $h_t = \mathbb{E}(z_t^2 | \Phi_{t-1})$ , the sum of  $h_{i,t}$  weighted by the prediction probability  $\mathbb{P}(s_t = i | \Phi_{t-1})$ . That is,  $z_t = \sqrt{h_{i,t}} \varepsilon_t$ , and

$$h_{i,t} = c_i + \sum_{j=1}^q a_{i,j} z_{t-j}^2 + \sum_{j=1}^p b_{i,j} h_{t-j}, \quad i = 0, 1, \quad (6.5)$$

$$h_t = h_{0,t} \mathbb{P}(s_t = 0 | \Phi_{t-1}) + h_{1,t} \mathbb{P}(s_t = 1 | \Phi_{t-1}).$$

A salient feature of (6.5) is that  $h_{i,t}$  are no longer path dependent because both  $h_{0,t-j}$  and  $h_{1,t-j}$  have been used to form  $h_{t-j}$ . Thus, this model can be computed without considering all possible values of  $(s_t, \dots, s_1)$ .

Gray’s model is readily generalized to allow both conditional mean and conditional variance to switch. Let  $\mu_{i,t}$  denote the conditional mean  $\mathbb{E}(z_t | s_t = i, \Phi_{t-1})$  and write

$$z_t = \mu_{i,t} + v_{i,t}, \quad v_{i,t} = \sqrt{h_{i,t}} \varepsilon_t, \quad (6.6)$$

$$h_{i,t} = c_i + \sum_{j=1}^q a_{i,j} v_{t-j}^2 + \sum_{j=1}^p b_{i,j} h_{t-j}.$$

In this case, we must compute two weighted sums:

$$h_t = \mathbb{E}(z_t^2 | \Phi_{t-1}) - \mathbb{E}(z_t | \Phi_{t-1})^2,$$

$$v_t = z_t - \mathbb{E}(z_t | \Phi_{t-1}),$$

where  $\mathbb{E}(z_t | \Phi_{t-1})$  and  $\mathbb{E}(z_t^2 | \Phi_{t-1})$  are calculated as

$$\mathbb{E}(z_t | \Phi_{t-1}) = \mu_{0,t} \mathbb{P}(s_t = 0 | \Phi_{t-1}) + \mu_{1,t} \mathbb{P}(s_t = 1 | \Phi_{t-1}),$$

$$\mathbb{E}(z_t^2 | \Phi_{t-1}) = (\mu_{0,t}^2 + h_{0,t}) \mathbb{P}(s_t = 0 | \Phi_{t-1}) + (\mu_{1,t}^2 + h_{1,t}) \mathbb{P}(s_t = 1 | \Phi_{t-1}).$$

Under this specification, neither  $h_t$  nor  $v_t$  is path dependent.

When the state variable assumes  $k$  values ( $k > 2$ ), let  $M_t$  denote the vector whose  $i$ th element is  $\mu_{i,t}$ ,  $H_t$  the vector whose  $i$ th element is  $h_{i,t}$ , and  $\Xi_{t|t-1}$  the vector whose  $i$ th element is the prediction probability  $\mathbb{P}(s_t = i | \Phi_{t-1})$ . Similar to the 2-state model, the conditional means and conditional variances in different states can be combined as

$$h_t = (M_t \odot M_t + H_t)' \Xi_{t|t-1} - (M_t' \Xi_{t|t-1})^2,$$

$$v_t = z_t - M_t' \Xi_{t|t-1},$$

where  $\odot$  denotes the element-by-element product.

Comparing to the models of Cai (1994) and Hamilton and Susmel (1994), the switching GARCH model of Gray (1996) allows all the GARCH parameters to switch and does not impose any constraint on these parameters. Thus, Gray's model offers much more flexibility than Cai's model and the SWARCH model. Gray's model can be estimated using the method discussed in Section 3; see Gray (1996) and Lin, Hung, and Kuan (2002) for details. Note that in practice, one may, instead of assuming conditional normality, postulate  $\varepsilon_t$  as i.i.d. random variables with  $t(n)$  distribution, where  $n$  is the degrees of freedom. Such a specification is capable of describing more erratic conditional variances.

## 7 Application: A Study of Taiwan's Short-Term Interest Rates

In this section we investigate the behaviors of the short-term interest rates in Taiwan. It is well known that Taiwan's short-term interest rates were carefully monitored and controlled by her Central Bank. In general, the Central Bank allows the interest rates to freely fluctuate within a given range. When the interest rates rise sharply in response to a major political or economic shock, the Central Bank usually intervenes the market

to ensure the stability of the interest rates. As such, it is reasonable to believe that Taiwan's interest rates may behave differently during different periods. This observation motivates us to apply the model of Gray (1996).

Let  $r_t$  denote the interest rate. A leading empirical model of  $\Delta r_t$  is

$$\Delta r_t = \alpha_0 + \beta_0 r_{t-1} + v_t, \quad (7.1)$$

where  $v_t$  is typically modeled as a GARCH(1,1) process:  $v_t = \sqrt{h_t} \varepsilon_t$  with

$$h_t = c_0 + a_0 v_{t-1}^2 + b_0 h_{t-1}; \quad (7.2)$$

see e.g., Chan et al. (1992). Letting  $\mu$  denote the long-run level of  $r_t$ ,  $\alpha_0 = \rho\mu$  and  $\beta_0 = -\rho$ , (7.1) becomes

$$\Delta r_t = \rho(\mu - r_{t-1}) + v_t.$$

As long as  $\rho > 0$  (i.e.,  $\beta_0 < 0$ ),  $\Delta r_t$  is positive (negative) when  $r_{t-1}$  is below (above) the long-run level. In this case,  $r_t$  will adjust toward the long-run level and hence exhibit mean reversion. Estimating (7.1) allows us to examine the property of mean reversion by checking the sign of the estimate of  $\beta_0$ . The ratio of the estimates of  $\alpha_0$  and  $\beta_0$  is then an estimate of the long-run level  $\mu$ . The postulated GARCH model (7.2), as usual, is used to characterize the volatility of  $\Delta r_t$ .

To allow for regime switching, we following Gray (1996) to specify:

$$\Delta r_t = \alpha_i + \beta_i r_{t-1} + v_{i,t}, \quad i = 0, 1, \quad (7.3)$$

and  $v_{i,t} = \sqrt{h_{i,t}} \varepsilon_t$  with

$$h_{i,t} = c_i + a_i v_{i,t-1}^2 + b_i h_{i,t-1}, \quad i = 0, 1. \quad (7.4)$$

Gray (1996) also include the additional term  $\omega_i r_{t-1}^{\tau_i}$  in (7.4) so as to capture the *level effect*. We do not pursue this possibility here, however.

In our study, we focus on the market rates of the 30-day Commercial Paper in the monetary market. We choose this data series because Commercial Papers are actively traded and hence their rates are a better index of short-term interest rates. The weekly average interest rates are computed from the daily data of the TEJ (Taiwan Economic Journal) database. There are 258 observations, from Jan. 4, 1994 through Dec. 7, 1998. During this period, Taiwan experienced numerous major shocks; for a comprehensive list of these events see Lin, Hung, and Kuan (2002). Our study thus allows us to evaluate how



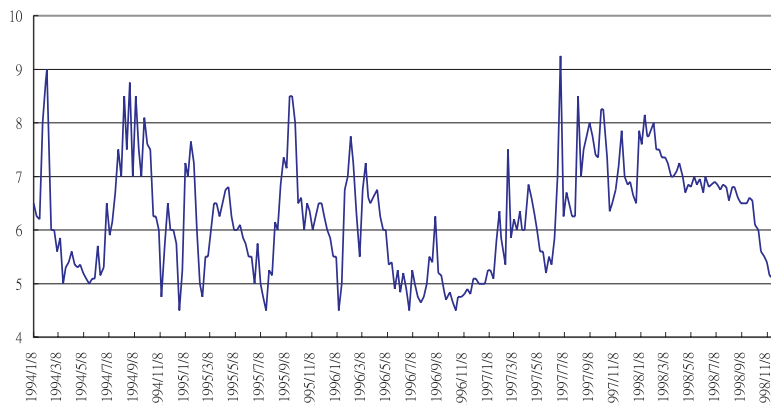


Figure 6: The weekly interest rates  $r_t$ : Jan. 1994–Dec. 1998.

the market responds to different shocks. We plot  $r_t$  and  $\Delta r_t$  in Figure 6 and Figure 7, respectively. Some summary statistics of  $\Delta r_t$  are: the sample average  $-0.0055\%$ , the standard deviation  $0.441$ , the skewness coefficient  $-0.8951$ , and the kurtosis coefficient  $4.2229$ . We also find that the sample correlation coefficient of  $\Delta r_t$  and  $r_{t-1}$  is  $-0.3087$ , suggesting that  $r_t$  may be mean reverting.

We consider three cases: (i) model (7.1) with the standard GARCH(1,1) errors (7.2); (ii) model (7.1) with the switching GARCH(1,1) errors (7.4), and (iii) model of switching mean (7.3) with the switching GARCH(1,1) errors (7.4). All models are estimated under the assumption that  $\varepsilon_t$  are i.i.d.  $\mathcal{N}(0, 1)$  random variables. As there are many very small and insignificant estimates for case (ii), we report only the result based on a special form of (7.4):

$$h_{0,t} = c_0 + a_0 v_{t-1}^2 + b_0 h_{t-1},$$

$$h_{1,t} = b_1 h_{t-1}.$$

Similarly, we report only the result of case (iii) based on:

$$h_{0,t} = c_0 + a_0 v_{t-1}^2,$$

$$h_{1,t} = c_1 + b_1 h_{t-1}.$$

The estimation results are summarized in Table 3 in which the numbers are taken from Lin, Hung, and Kuan (2002). The smoothing probabilities of  $s_t = 0$  and the estimated conditional variances of case (iii) are plotted in Figure 8 and Figure 9, respectively.

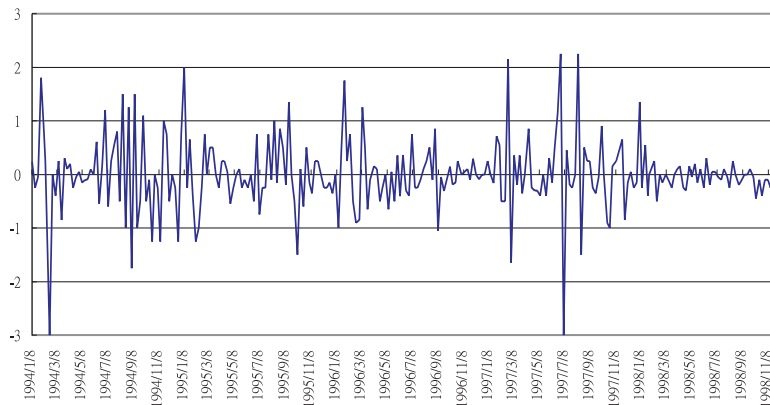


Figure 7: The first differences of weekly interest rates  $\Delta r_t$ : Jan. 1994–Dec. 1998.

It can be seen from Table 3 that both AIC and SIC improve when the model allows for regime switching. For case (iii), the Hansen (1992, 1996a) test statistic is 3.257, which rejects the null hypothesis of no switching at 1% level. For cases (ii) and (iii), both transition probabilities are highly significant. The Wald tests of state independence are 59.53 for case (ii) and 140.39 for case (iii), which are significant at any level. These results support using the Markov switching model. We also observe the following.

1. All models yield negative estimates of  $\beta$  and hence exhibit mean reversion. For case (iii), the magnitude of the estimate of  $\beta_0$  is much greater than that of  $\beta_1$ , indicating a much quicker adjustment speed in state 0. When there is no switching in mean, as in cases (i) and (ii), the magnitude of the estimated  $\beta$  is small and close to that of  $\beta_1$  in case (iii).
2. For case (iii), the estimated long-run levels in state 0 and state 1 are, respectively, 6.6% and 5%, whereas the estimated long-run levels are 6.2% in case (i) and 5.53% in case (ii).
3. We find the IGARCH-type volatility persistence for case (i) but not for cases (ii) and (iii) when the conditional variances are allowed to switch. This is similar to the finding of Gray (1996). Also, the GARCH parameters in different regimes do not appear to be proportional, contradicting the assumption of Hamilton and Susmel (1994).
4. The estimation result of case (iii) suggests that  $h_{0,t}$  is approximately a constant (i.e.,

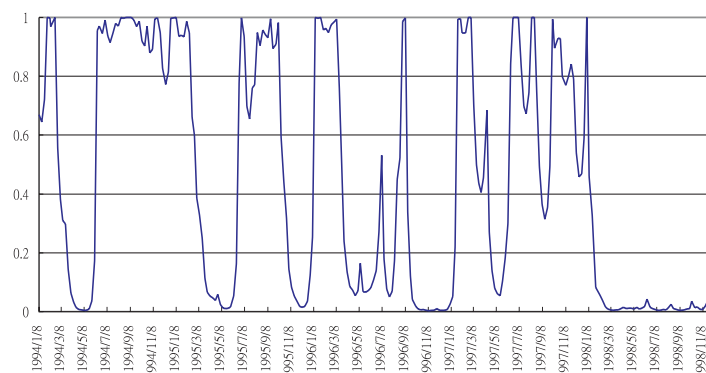


Figure 8: The estimated smoothing probabilities of  $s_t = 0$ .

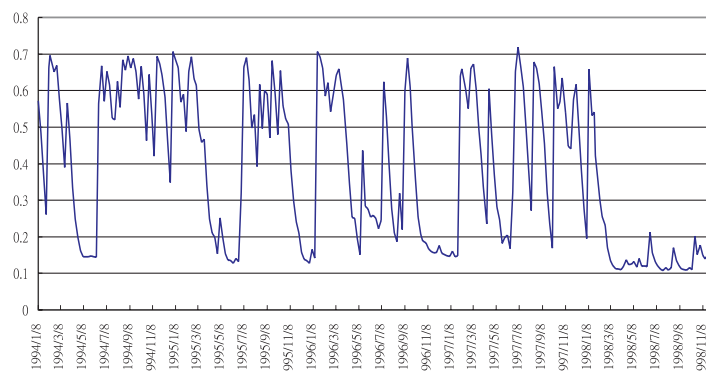


Figure 9: The estimated conditional variances  $h_t$ .

conditional homoskedasticity) and  $h_{1,t}$  are mainly determined by  $h_{t-1}$  (i.e., GARCH effect only). This volatility pattern is quite different from that of Gray (1996).

Based on the result of case (iii), the state 0 may be interpreted as the state of high long-run level with a quick adjustment speed and high volatility level without persistence. In contrast, the state state 1 is the state of low long-run level with a very slow adjustment speed and low volatility level with a quickly diminishing GARCH effect. One possible explanation of this result is that, when the short-term interest rates are in the high-level regime, the Central Bank's intervention successfully suppressed their volatility so that no ARCH or GARCH effect exists. Without intervention in the low-level regime, the interest rates exhibit a GARCH effect, yet volatility clustering may last for only a short

time period.

## 8 Concluding Remarks

This lecture note presents the Markov switching models of the conditional mean and conditional variance behaviors of time series. Although these models are well known in the literature, research on this topic is still promising. In addition to empirical applications of these models, there is still room for theoretical development. From Section 4.1 and Appendix II we can see that the Hansen test is not completely satisfactory because it is conservative and computationally very demanding by construction. A proper and simpler test for the Markov switching model is highly desirable. If a new test can be derived, it may also be applicable to other models that involve unidentified nuisance parameters. Such a test would surely be a major contribution to the econometrics literature. Furthermore, what is important in the Markov switching model is its Markovian switching mechanism. This switching mechanism may also be imposed on other models to yield new models and different applications. Research along this line is definitely worth exploring.

## Appendix I: Estimation of the Model (2.5)

In this Appendix we consider quasi-maximum likelihood estimation of the model (2.5).

Let  $\tilde{z}_t = z_t - \alpha_0 - \alpha_1 s_t$ , model (2.5) is

$$\tilde{z}_t = \beta_1 \tilde{z}_{t-1} + \cdots + \beta_k \tilde{z}_{t-k} + \varepsilon_t.$$

We first define a new state variable  $s_t^* = 1, 2, \dots, 2^{k+1}$  such that each of these values represents a particular realization of  $(s_t, s_{t-1}, \dots, s_{t-k})$ . For example, when  $k = 2$ ,

$$\begin{aligned} s_t^* = 1 & \text{ if } s_t = s_{t-1} = s_{t-2} = 0, \\ s_t^* = 2 & \text{ if } s_t = 0, s_{t-1} = 0, \text{ and } s_{t-2} = 1, \\ s_t^* = 3 & \text{ if } s_t = 0, s_{t-1} = 1, \text{ and } s_{t-2} = 0, \\ & \vdots \\ s_t^* = 8 & \text{ if } s_t = s_{t-1} = s_{t-2} = 1. \end{aligned}$$

It is easy to see that  $s_t^*$  is also a first-order Markov chain. We can also arrange the values of  $s_t^*$  so that the transition matrix is

$$P^* = \begin{bmatrix} P_{00} & \mathbf{0} \\ \mathbf{0} & P_{10} \\ P_{01} & \mathbf{0} \\ \mathbf{0} & P_{11} \end{bmatrix},$$

with  $P_{ji}$  ( $j, i = 0, 1$ ) is a  $2^{k-1} \times 2^k$  block diagonal matrix:

$$P_{ji} = \begin{bmatrix} p_{ji} & p_{ji} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & p_{ji} & p_{ji} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & p_{ji} & p_{ji} \end{bmatrix},$$

a  $2^{k-1} \times 2^k$  block-diagonal matrix.

Let  $\boldsymbol{\xi}_{t,i} = (\tilde{z}_t, \tilde{z}_{t-1}, \dots, \tilde{z}_{t-k})'$ , where the values of  $\tilde{z}_t, \dots, \tilde{z}_{t-k}$  depend on the realization of  $(s_t, s_{t-1}, \dots, s_{t-k})$ , and this realization is such that  $s_t^* = i, i = 1, 2, \dots, 2^{k+1}$ . For  $\mathbf{b} = (1, -\beta_1, \dots, -\beta_k)'$ , we have

$$\mathbf{b}'\boldsymbol{\xi}_{t,i} = \tilde{z}_t - \beta_1 \tilde{z}_{t-1} - \cdots - \beta_k \tilde{z}_{t-k}.$$

When  $k = 2$  and  $i = 3$ , for example, the realization of  $(s_t, s_{t-1}, s_{t-2})$  is  $(0, 1, 0)$  so that

$$\mathbf{b}'\boldsymbol{\xi}_{t,i} = (z_t - \alpha_0) - \beta_1(z_{t-1} - \alpha_0 - \alpha_1) - \beta_2(z_{t-2} - \alpha_0).$$

Under the normality assumption, the density of  $z_t$  conditional on  $s_t^* = i$  and  $\mathcal{Z}^{t-1}$  is

$$f(z_t | s_t^* = i, \mathcal{Z}^{t-1}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp \left\{ \frac{-(\mathbf{b}'\boldsymbol{\xi}_{t,i})^2}{2\sigma_\varepsilon^2} \right\}, \quad (8.5)$$

where  $i = 1, 2, \dots, 2^{k+1}$  and  $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1, \dots, \beta_k, \sigma_\varepsilon, p_{00}, p_{11})'$ .

The derivation below is similar to that in Section 3.1. Given  $\mathbb{P}(s_t^* = i | \mathcal{Z}^{t-1}; \boldsymbol{\theta})$ , the density of  $z_t$  conditional on  $\mathcal{Z}^{t-1}$  can be obtained via (8.5) as

$$f(z_t | \mathcal{Z}^{t-1}; \boldsymbol{\theta}) = \sum_{i=1}^{2^{q+1}} \mathbb{P}(s_t^* = i | \mathcal{Z}^{t-1}; \boldsymbol{\theta}) f(z_t | s_t^* = i, \mathcal{Z}^{t-1}; \boldsymbol{\theta}). \quad (8.6)$$

To compute the filtering probabilities  $\mathbb{P}(s_t^* = i | \mathcal{Z}^t; \boldsymbol{\theta})$ , note that

$$\mathbb{P}(s_t^* = i | \mathcal{Z}^t; \boldsymbol{\theta}) = \frac{\mathbb{P}(s_t^* = i | \mathcal{Z}^{t-1}; \boldsymbol{\theta}) f(z_t | s_t^* = i, \mathcal{Z}^{t-1}; \boldsymbol{\theta})}{f(z_t | \mathcal{Z}^{t-1}; \boldsymbol{\theta})}, \quad (8.7)$$

and that the  $(j, i)$ th element of  $\mathbf{P}^*$  is

$$p_{ji}^* = \mathbb{P}(s_t^* = i | s_{t-1}^* = j) = \mathbb{P}(s_t^* = i | s_{t-1}^* = j, \mathcal{Z}^t),$$

by the Markov property. These in turn yield

$$\mathbb{P}(s_{t+1}^* = i | \mathcal{Z}^t; \boldsymbol{\theta}) = \sum_{j=1}^{2^{k+1}} p_{ji}^* \mathbb{P}(s_t^* = j | \mathcal{Z}^t; \boldsymbol{\theta}). \quad (8.8)$$

Thus, with the initial values  $\mathbb{P}(s_k^* = i | \mathcal{Z}^{k-1}; \boldsymbol{\theta})$ , we can iterate the equations (8.5)–(8.8) to obtain  $\mathbb{P}(s_t^* = i | \mathcal{Z}^t; \boldsymbol{\theta})$  for  $t = k + 1, \dots, T$ . The quasi-log-likelihood function can also be constructed using (8.6), from which the QMLE can be computed.

Then for each  $t$ , the desired filtering probability of  $s_t$  is

$$\mathbb{P}(s_t = 1 | \mathcal{Z}^t; \boldsymbol{\theta}) = \sum \mathbb{P}(s_t^* = i | \mathcal{Z}^t; \boldsymbol{\theta}),$$

where the summation is taken over all  $i$  that associated with  $s_t = 1$ , and

$$\mathbb{P}(s_t = 0 | \mathcal{Z}^t; \boldsymbol{\theta}) = 1 - \mathbb{P}(s_t = 1 | \mathcal{Z}^t; \boldsymbol{\theta}).$$

For the initial values  $\mathbb{P}(s_k^* = i | \mathcal{Z}^{k-1}; \boldsymbol{\theta})$ , we can set it to its limiting unconditional counterpart: the  $(2^{k+1} + 1)$ th column of the matrix  $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ , where

$$\mathbf{A} = \begin{bmatrix} \mathbf{I} - \mathbf{P}^* \\ \mathbf{1}' \end{bmatrix},$$

with  $\mathbf{I}$  the identity matrix and  $\mathbf{1}$  the  $2^{k+1}$ -dimensional vector of ones.

## Appendix II: Computation of Hansen's Statistic (4.4)

In this Appendix we discuss some computational issues in implementing Hansen's statistic (4.4). To compute the concentrated QMLE, the concentrated quasi-log-likelihood function must be maximized for each value of  $\gamma$ . Note that  $p_{00}$  and  $p_{11}$  in  $\gamma$  could take any value in  $[0, 1]$  and that  $\alpha_1$  could be any value on the real line. A practical way is to consider only finitely many values of  $\gamma$ . This amounts to setting up a set of grid points in the parameter space and computing the concentrated QMLE only with respect to these points. For example, Hansen (1992) restricts  $\alpha_1$  to be in the range  $[0, 2]$  and set 20 grid points:  $0.1, 0.2, \dots, 2$ , and his Grid 3 for  $p_{00}$  (and  $p_{11}$ ) contains 8 grid points:  $0.12, 0.23, \dots, 0.89$ . This results in a total of 1280 ( $= 8 \times 8 \times 20$ ) grid points for  $\gamma$  so that there will be 1280 optimizations. This is computationally demanding; finer grid points of course require even more intensive computation.

Hansen (1992, 1996a) suggest a simulation method to generate the distribution of  $\sup_{\gamma} Q^*(\gamma)$ , where  $Q^*$  is a Gaussian process with mean zero and the covariance function  $K^*(\gamma_1, \gamma_2)$ . Let  $\hat{K}(\gamma_1, \gamma_2)$  denote a consistent estimate of  $K^*(\gamma_1, \gamma_2)$ ; see Hansen (1996a) for the exact expression of  $\hat{K}$ . One can then repeatedly generate Gaussian processes whose covariance functions are all  $\hat{K}(\gamma_1, \gamma_2)$ . As a Gaussian process is completely determined by its covariance function, the supremum of each generated process has approximately the same distribution as  $\sup_{\gamma} Q^*(\gamma)$ . These supremum values together form a simulated distribution of  $\sup_{\gamma} Q^*(\gamma)$ , from which the critical values and  $p$ -values of Hansen's statistic can be calculated. Following this idea, Hansen (1996a) proposes to generate a sample of i.i.d.  $N(0, 1)$  random variables  $\{u_1, \dots, u_{T+M}\}$  and compute

$$\frac{\sum_{j=0}^M \sum_{t=1}^T q_t(\gamma, \hat{\theta}(\gamma)) u_{t+k}}{\sqrt{1 + M \hat{V}_T(\gamma)}},$$

where  $\gamma$  takes the grid points discussed in the preceding paragraph, and  $q_t$  are the summands of  $Q_T(\gamma)$ . Then conditional on the data used in model estimation, the process so generated has mean zero and exact covariance function  $\hat{K}^*(\gamma_1, \gamma_2)$ . This is precisely the process we need.

---

## References

- Albert, J. and S. Chib (1993). Bayesian inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts, *Journal of Business & Economic Statistics*, **11**, 1–15.
- Andrews, D.W.K. (1991). An empirical process central limit theorem for dependent non-identically distributed random variables, *Journal of Multivariate Analysis*, **38**, 187–203.
- Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change points, *Econometrica*, **61**, 821–856.
- Bai, J. (1999). Likelihood ratio tests for multiple structural changes, *Journal of Econometrics*, **91**, 299–323.
- Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes, *Econometrica*, **66**, 47–78.
- Blanchard, O.J. and D. Quah (1989). The dynamic effects of aggregate demand and supply disturbance, *American Economic Review*, **79**, 655–673.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.
- Cai, J. (1994). A Markov model of switching-regime ARCH, *Journal of Business & Economic Statistics*, **12**, 309–316.
- Casella, G. and E. George (1992). Explaining the Gibbs sampler, *American Statistician*, **46**, 167–174.
- Chan, K.C., G.A. Karolyi, F.A. Longstaff, and A.B. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance*, **47**, 1209–1227.
- Chen, S.-W. and J.-L. Lin (1999). Switching ARCH models of stock market volatility in Taiwan, Working Paper.
- Chen, S.-W. and J.-L. Lin (2000a). Modeling business cycles in Taiwan with time-varying Markov-switching models, *Academia Economic Papers*, **28**, 17–42.
- Chen, S.-W. and J.-L. Lin (2000b). Identifying turning points and business cycles in Taiwan: A multivariate dynamic Markov-switching factor model approach, *Academia*



- 
- Economic Papers*, **28**, 289–320.
- Chib, S. and E. Greenberg (1996). Markov chain Monte Carlo simulation methods in econometrics, *Econometric Theory*, **12**, 409–431.
- Davies, R.B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, **64**, 247–254.
- Davies, R.B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, **74**, 33–43.
- Diebold, F.X., J.-H. Lee and G.C. Weinbach (1994). Regime switching with time-varying transition probabilities, in C. Hargreaves (ed.) *Nonstationary Time Series Analysis and Cointegration*, pp. 283–302, Oxford: Oxford University Press.
- Diebold, F.X. and G.D. Rudebusch (1996). Measuring business cycles: A modern perspective, *Review of Economics and Statistics*, **78**, 67–77.
- Dueker, M.J. (1997). Markov switching in GARCH processes and mean-reverting stock market volatility, *Journal of Business & Economic Statistics*, **15**, 26–34.
- Durland, J.M. and T.H. McCurdy (1994). Duration-dependent transitions in a Markov model of U.S. GNP growth, *Journal of Business & Economic Statistics*, **12**, 279–288.
- Engel, C. (1994). Can the Markov switching model forecast exchange rates?, *Journal of International Economics*, **36**, 151–165.
- 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**, 689–713.
- Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1007.
- Filardo, A.J. (1994). Business-cycle phases and their transitional dynamics, *Journal of Business & Economic Statistics*, **12**, 299–308.
- Filardo, A.J. and S.F. Gordon (1998). Business cycle durations, *Journal of Econometrics*, **85**, 99–123.
- Garcia, R. (1998). Asymptotic null distribution of the likelihood ratio test in Markov switching model, *International Economic Review*, **39**, 763–788.
- Garcia, R. and P. Perron (1996). An analysis of the real interest rate under regime shifts, *Review of Economics and Statistics*, **78**, 111–125.

- 
- Gelfand, A.E. and A.F.M. Smith (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**, 398–409.
- Ghysels, E. (1994). On the periodic structure of the business cycle, *Journal of Business & Economic Statistics*, **12**, 289–298.
- Goldfeld, S.M. and R.E. Quandt (1973). A Markov model for switching regressions, *Journal of Econometrics*, **1**, 3–16.
- Goodwin, T.H. (1993). Business-cycle analysis with a Markov switching model, *Journal of Business & Economic Statistics*, **11**, 331–339.
- Granger, C.W.J. and T. Teräsvirta (1993). *Modelling Nonlinear Economic Relationships*, New York, NY: Oxford University Press.
- Gray, S.F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics*, **42**, 27–62.
- Hamilton, J.D. (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**, 385–423.
- Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, **57**, 357–384.
- Hamilton, J.D. (1990). Analysis of time series subject to changes in regime, *Journal of Econometrics*, **45**, 39–70.
- Hamilton, J.D. (1994). *Time Series Analysis*, Princeton, NJ: Princeton University Press.
- Hamilton, J.D. (1996). Specification testing in Markov-switching time series models, *Journal of Econometrics*, **70**, 127–157.
- Hamilton, J.D. and G. Lin (1996). Stock market volatility and the business cycle, *Journal of Applied Econometrics*, **11**, 573–593.
- Hamilton, J.D. and R. Susmel (1994). Autoregressive conditional heteroscedasticity and changes in regime, *Journal of Econometrics*, **64**, 307–333.
- Hansen, B.E. (1992). The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP, *Journal of Applied Econometrics*, **7**, S61–S82.
- Hansen, B.E. (1996a). Erratum: The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP, *Journal of Applied Econometrics*, **11**,

195–198.

- Hansen, B.E. (1996b). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, **64**, 413–430.
- Harvey, A.C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models, *Review of Economic Studies*, **61**, 247–264.
- Hsu, S.-H. and C.-M. Kuan (2001). Identifying Taiwan's business cycles in 1990s: An application of the bivariate Markov switching model and Gibbs sampling (in Chinese), *Journal of Social Sciences and Philosophy*, **13**, 515–540.
- Huang, C.H. (1999). Phases and characteristics of Taiwan business cycles: A Markov switching analysis, *Taiwan Economic Review*, **27**, 185–214.
- Huang, Y.-L., C.-M. Kuan, and K.S. Lin (1998). Identifying the turning points of business cycles and forecasting real GNP growth rates in Taiwan (in Chinese), *Taiwan Economic Review*, **26**, 431–457.
- Jacquier, E., N.G. Polson, and P. Rossi (1994). Bayesian analysis of stochastic volatility models (with discussion), *Journal of Business & Economic Statistics*, **12**, 371–417.
- Kim, C.J. (1994). Dynamic linear models with Markov-switching, *Journal of Econometrics*, **60**, 1–22.
- Kim, C.J. and C.R. Nelson (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching, *Review of Economics and Statistics*, **80**, 188–201.
- Kim, C.J. and C.R. Nelson (1999). *State Space Models with Regime Switching, Classical and Gibbs Sampling Approaches with Applications*, Cambridge, MA: MIT Press.
- Kim, M.-J. and J.-S. Yoo (1995). New index of coincident indicators: A multivariate Markov switching factor model approach, *Journal of Monetary Economics*, **36**, 607–630.
- Kuan, C.-M. and H. White (1994). Artificial neural networks: An econometric perspective (with reply), *Econometric Reviews*, **13**, 1–91 and 139–143.
- Lam, P.S. (1990). The Hamilton model with a general autoregressive component *Journal of Monetary Economics*, **26**, 409–432.
- Lamoureux, C.G. and W.D. Lastrapes (1990). Persistence in variance, structural change and the GARCH model, *Journal of Business & Economic Statistics*, **8**, 225–234.

- 
- Lin, C.-C., M.-W. Hung, and C.-M. Kuan (2002). The dynamic behavior of short term interest rates in Taiwan: An application of the regime switching model (in Chinese), *Academia Economic Papers*, **30**, 29–55, 2002.
- Lucas, R.E. (1977). Understanding business cycles, in K. Brunner and A. Metzler (eds.), *Stabilization of the Domestic and International Economy*, 7–29, Carnegie–Rochester series on Public Policy 5.
- McCulloch, R.E. and R.S. Tsay (1994). Statistical analysis of economic time series via Markov switching models, *Journal of Time Series Analysis*, **15**, 523–539.
- Melino, A. and S.M. Turnbull (1990). Pricing foreign currency options with stochastic volatility, *Journal of Econometrics*, **45**, 239–265.
- Quandt, R.E. (1972). A new approach to estimating switching regressions, *Journal of the American Statistical Association*, **67**, 306–310.
- Ramchand, L. and R. Susmel (1998). Volatility and cross correlation across major stock markets, *Journal of Empirical Finance*, **5**, 397–416.
- Rau, H.-H., H.-W. Lin, and M.-Y. Li (2001). Examining Taiwan’s business cycle via two-period MS Models (in Chinese), *Academia Economic Papers*, forthcoming.
- Schaller, H. and S. van Norden (1997). Regime switching in stock market returns, *Applied Financial Economics*, **7**, 177–191.
- So, M.K.P., K. Lam and W.K. Li (1998). A stochastic volatility model with Markov switching, *Journal of Business & Economic Statistics*, **16**, 244–253.
- Sola, M. and J. Driffill (1994). Testing the term structure of interest rates using a stationary vector autoregression with regime switching, *Journal of Economic Dynamics and Control*, **18**, 601–628.
- Tanner, M. and W.H. Wong (1987). The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, **82**, 528–550.
- Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*, New York, NY: Oxford University Press.



Table 1: The estimation results of the bivariate Markov switching model on GDP and employment growth rates.

parameter	Prior dist.		Posterior dist.	
	average	s.d.	average	s.d.
$\alpha_{01}$	2	1	1.1073	0.9113
$\alpha_{02}$	2	1	0.8738	0.8332
$\alpha_{11}$	2	1	2.3538	1.1602
$\alpha_{12}$	2	1	1.4057	1.0251
$b_{11}$	0	1	0.8753	0.2600
$b_{12}$	0	1	-0.2166	0.2590
$b_{13}$	0	1	0.0172	0.3204
$b_{14}$	0	1	0.6703	0.3119
$b_{21}$	0	1	-0.1205	0.3136
$b_{22}$	0	1	0.2590	0.3122
$b_{23}$	0	1	-0.2654	0.3743
$b_{24}$	0	1	0.0434	0.3655
$b_{31}$	0	1	0.1176	0.3004
$b_{32}$	0	1	0.0029	0.2891
$b_{33}$	0	1	0.3936	0.3964
$b_{34}$	0	1	0.2330	0.3422
$b_{41}$	0	1	-0.1672	0.2075
$b_{42}$	0	1	-0.1363	0.2073
$b_{43}$	0	1	-0.2740	0.3192
$b_{44}$	0	1	-0.4507	0.2823
$\sigma_{11}$	1	.	0.9781	4.1622
$\sigma_{12}$	0	.	-0.1305	8.1463
$\sigma_{22}$	1	.	2.1819	35.936
$p_{00}$	0.5	0.0012	0.5619	0.1725
$p_{11}$	0.5	0.0012	0.6918	0.1488

Note:  $\alpha_{i1}$  and  $\alpha_{i2}$  are, respectively, the intercepts for the GDP and employment growth rates when  $s_t = i$ ,  $i = 0, 1$ ; for  $j = 1, \dots, 4$ ,

$$\mathbf{B}_j = \begin{bmatrix} b_{j1} & b_{j3} \\ b_{j2} & b_{j4} \end{bmatrix}; \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}.$$

Table 2: The estimated smoothing probabilities of the bivariate Markov switching model on GDP and employment growth rates.

quarter	prob.	quarter	prob.
1990 Q1	N/A	1995 Q1	0.7158
Q2	N/A	Q2	0.7349
Q3	N/A	Q3	0.3133
Q4	N/A	Q4	0.2622
1991 Q1	0.6792	1996 Q1	0.7680
Q2	0.8206	Q2	0.7579
Q3	0.7257	Q3	0.7185
Q4	0.6235	Q4	0.7674
1992 Q1	0.8042	1997 Q1	0.4798
Q2	0.7730	Q2	0.7135
Q3	0.7647	Q3	0.7045
Q4	0.8029	Q4	0.7534
1993 Q1	0.7638	1998 Q1	0.2484
Q2	0.7856	Q2	0.2851
Q3	0.7549	Q3	0.1952
Q4	0.7864	Q4	0.2041
1994 Q1	0.7766	1999 Q1	0.6745
Q2	0.7510	Q2	0.8483
Q3	0.7957	Q3	N/A
Q4	0.8218	Q4	N/A

Table 3: The estimation results of various models on the interest rates.

Model	No switching in mean and variance		No switching in mean Switching in variance		Switching in mean and variance	
	estimate	<i>t</i> ratio	estimate	<i>t</i> ratio	estimate	<i>t</i> ratio
$\alpha_0$	0.5737	3.42	0.3989	3.09	2.3356	4.05
$\alpha_1$	—	—	—	—	0.2566	1.81
$\beta_0$	-0.0923	-3.46	-0.0721	-3.49	-0.3505	-4.13
$\beta_1$	—	—	—	—	-0.0507	-2.25
$c_0$	0.0155	1.70	0.5450	2.11	0.6699	3.57
$c_1$	—	—	—	—	0.0056	0.41
$a_0$	0.3540	4.60	0.1143	1.17	0.0820	0.54
$a_1$	—	—	—	—	—	—
$b_0$	0.6756	13.57	0.3821	0.80	—	—
$b_1$	—	—	0.2363	2.73	0.2175	2.99
$p_{00}$	—	—	0.8660	13.26	0.8891	19.38
$p_{11}$	—	—	0.8827	20.73	0.9052	27.50
AIC	442.45		394.73		382.25	
SIC	460.22		422.15		417.78	