

博士論文

袴田守一

Extended Markov Switching Models and Their Applications

Morikazu Hakamata

June 24, 2002

Contents

1	Introduction and Overview	13
1.1	Motivation and Background	13
1.2	Review of Literature	18
1.3	Contributions of this Thesis	20
2	Markov Switching Model	23
2.1	Markov Chain	23
2.2	Ordinary Markov Switching Model	28
2.2.1	Markov Switching Stochastic Trend Model	28
2.2.2	Multivariate Markov Switching Statistic Trend Model	30
2.2.3	Transition Probability	33
2.2.4	Exogenous Time-varying Transition Probabilities	39
2.2.5	Switching ARCH Model	41
2.3	State Space Representation	42
2.3.1	Linear Gaussian State Space Modeling	42
2.3.2	Nonlinear Non-Gaussian State Space Modeling	43
2.3.3	General State Space Modeling	44
3	State Estimation and Model Identification	45
3.1	Non-Gaussian Filter and Smoother	45
3.2	Gaussian Sum Filter	46
3.3	Gaussian Sum Smoother	49
3.3.1	Kitagawa's two-filter Formula and Gaussian Sum Smoother	49
3.3.2	Kim's Smoother	50
3.4	Monte Carlo Filter and Smoother	51

3.5	Parameter Estimation	53
3.6	Model Selection by AIC	54
3.6.1	Akaike Information Criterion	54
3.6.2	AIC for Model Selection of the Markov Switching Model	54
3.7	Numerical Examples	55
4	Empirical Analysis	61
4.1	Trend Identification and Trading Strategy	61
4.1.1	Introduction	61
4.1.2	Markov Switching Slope Change and ARCH Model	62
4.1.3	Trading Strategy	66
4.1.4	Application and Comparison of Trading Performance	67
4.1.5	Conclusions	69
4.2	Time-series and Cross-sectional Volatility Analysis	77
4.2.1	Introduction	77
4.2.2	Time-series and Cross-sectional Properties of the Japanese Stock Market	78
4.2.3	Markov Switching Models for Volatility	79
4.2.4	Application and Model Comparison	84
4.2.5	Conclusions	86
4.3	Analysis of Japanese Business Cycles	93
4.3.1	Introduction	93
4.3.2	Multivariate Stochastic Trend Model with Simultaneous Markov Switching Drift	94
4.3.3	Variable Selection	97
4.3.4	Application and Model Comparison	98
4.3.5	Conclusions	101
4.4	Japan Premium and Japanese Banks' Stock Volatility	106
4.4.1	Introduction	106
4.4.2	The Bivariate Markov Switching Model	107
4.4.3	Application and Model Comparison	109
4.4.4	Conclusions	114
4.5	Transmission of Volatility	122
4.5.1	Introduction	122
4.5.2	The Bivariate Markov Switching ARCH Model	123

4.5.3	Application and Model Comparison	127
4.5.4	Conclusions	131
4.6	Self-organizing Markov Switching State Space Model	142
4.6.1	Introduction	142
4.6.2	The Self-organizing Markov Switching State Space Model with Time-varying Transition Probability	143
4.6.3	The Bivariate Self-organizing Markov Switching State Space Model	145
4.6.4	Application I: Foreign Exchange Volatility and Intervention	146
4.6.5	Application II: International Business Cycle Transmission	148
4.6.6	Conclusions	149
5	Summary of Extended Markov Switching Models	157
5.1	Modified Markov Switching ARCH Model	158
5.2	Modified Multivariate Markov Switching ARCH Model	160
5.3	Markov Switching Slope Change and ARCH Model	162
5.4	Markov Switching Models with Exogenous Leverage Effects	164
5.4.1	Markov switching heteroskedasticity model with an exogenous leverage effect . . .	164
5.4.2	Markov switching ARCH model with an exogenous leverage effect	166
5.5	Self-organizing Markov Switching State-Space Modeling	167
5.6	Semi-Markov Switching Model	169
5.6.1	Univariate Semi-Markov Switching Stochastic Trend Model	169
5.6.2	Multivariate Semi-Markov Switching Slope Change Model	171
6	Conclusions	173

List of Figures

1.1	Example 1 — Tokyo Stock Price Index (TOPIX)	14
1.2	Example 2 — Logarithmic return of Japan/U.S. exchange rate	14
1.3	Example 3 — Japanese industrial production and business cycles	15
1.4	Example 3 — U.S. and Germany industrial production	15
2.1	Duration distribution of Markov chain.	27
2.2	Duration distribution of a semi-Markov chain.	27
2.3	Example 1 — Joint distributions.	31
2.4	Example 2 — Joint distributions.	32
2.5	Illustration of general transition probabilities.	36
2.6	Illustration of independent (b) and simultaneous (c) transition probabilities.	38
3.1	One cycle of the Monte Carlo filter ($m = 9$).	52
4.1	Weekly TOPIX: 1986/1 – 2001/12.	73
4.2	Logarithmic return of weekly TOPIX: 1986/1 – 2001/12.	73
4.3	Filtered slope Δt_n : 1986/1 – 2001/12.	74
4.4	Filtered regime probability for slope change and high volatility: 1986/1 – 2001/12.	74
4.5	Filtered slope Δt_n : 1992/1 – 2001/12.	75
4.6	Average trading returns of m period moving average trading strategies: 1992/1 – 2001/12.	75
4.7	Cumulative sums of profits for MS-SC/ARCH trading strategy: 1992/1 – 2001/12.	76
4.8	TOPIX500 index: 1999/1 – 2001/12.	89
4.9	Logarithmic return of daily TOPIX500 index: 1999/1 – 2001/12.	89
4.10	Cross-sectional standard deviation of TOPIX500.	90
4.11	Cross-sectional skewness of TOPIX500.	90
4.12	Cross-sectional kurtosis of TOPIX500.	91

4.13 Smoothed regime probability of $S_n = 1$ (high volatility regime) of MSH model.	91
4.14 Smoothed regime probability of $S_n = 1$ (high volatility regime) of MSH- $L_{S,D}$ model. . . .	92
4.15 Coincident indicators.	104
4.16 Filtered regime probability of contractions in the in-sample period.	105
4.17 Filtered regime probability of contractions in the out-of-sample period.	105
4.18 LIBOR spread.	116
4.19 TSEBK return.	116
4.20 Sample autocorrelation function of the LIBOR spread.	117
4.21 Sample autocorrelation function of the TSEBK return.	117
4.22 Smoothed regime probability of regime 1	118
4.23 Smoothed regime probability of regime 2	118
4.24 Smoothed regime probability of regime 3	118
4.25 Smoothed regime probability of regime 4	118
4.26 Two joint density functions for the LIBOR spread.	119
4.27 Two joint density functions for the TSEBK return.	120
4.28 Relationship among regimes of the LIBOR spread and those of the TSEBK return.	121
4.29 Daily Japan/U.S. exchange rate.	135
4.30 Daily TOPIX.	135
4.31 Logarithmic return of the daily Japanese/U.S. exchange rate.	136
4.32 Logarithmic return of the daily TOPIX.	136
4.33 Two joint densities and the unconditional density for the JPY.	137
4.34 Two joint densities and the unconditional density for TOPIX.	138
4.35 Smoothed probability of regime $R_n = 1$	139
4.36 Smoothed probability of regime $R_n = 2$	139
4.37 Smoothed probability of regime $R_n = 3$	139
4.38 Smoothed probability of regime $R_n = 4$	139
4.39 Transmission of volatility process between the JPY and the TOPIX.	140
4.40 Japan Premium.	141
4.41 Japan/U.S. exchange rate	150
4.42 Logarithmic return of the Japan/U.S. exchange rate	150
4.43 Interventions by the Bank of Japan	151
4.44 Filtered probability of high volatility regime	152

LIST OF FIGURES

9

4.45 Time-varying parameter 1	152
4.46 Time-varying parameter 2	153
4.47 Industrial production indices for the U.S. and Germany.	154
4.48 Transition probabilities of the approximate maximum likelihood estimate	155
4.49 Smoothed joint probabilities of the SOMS model	156

List of Tables

3.1	Simulation Results 1 — Statistics —	57
3.2	Simulation Results 1 — Estimated Parameters of Markov switching Gaussian	58
3.3	Simulation Results 2 — Statistics —	59
3.4	Simulation Results 2 — Estimated Parameters of Markov switching Gaussian	60
4.1	Estimated results of MS-SC/ARCH model.	71
4.2	Basic statistical summary of the trading returns.	71
4.3	Performance evaluation of the MS-SC/ARCH trading strategy.	72
4.4	Basic statistical summary of the TOPIX500 index returns.	87
4.5	Estimated parameters and statistics of models with no exogenous leverage effect.	87
4.6	Estimated parameters and statistics of MSH-L models.	88
4.7	Log likelihood, number of parameters and AIC of no-switching models.	88
4.8	Coincident indicators and corresponding QPSs.	102
4.9	Estimated parameters and statistics for the multivariate MSST model.	103
4.10	Regime Definitions	108
4.11	Model comparison.	115
4.12	Estimated parameters and corresponding standard errors	115
4.13	Regime definitions of the bivariate MS-ARCH model.	125
4.14	Estimated parameters and corresponding statistics.	132
4.15	Estimated transition probability.	133
4.16	Estimated parameters and corresponding statistics.	134
4.17	Log likelihood, number of parameters and AIC of no-switching models.	134
5.1	Corresponding Empirical Analysis Topics.	157

Chapter 1

Introduction and Overview

1.1 Motivation and Background

Issues of financial and economic time series

This thesis deals with the Markov switching models and their empirical analysis in economic and financial time series. Economic and financial time series display typical nonlinear characteristics. One of their important features is the existence of regimes within which the observations, and the corresponding processes show different dynamic behaviors. In the statistical analysis of economic and financial time series, in some cases, to capture the reality of the time series, it is more appropriate to divide and classify the whole period into parts as “*regime*”, and to apply different statistical models or processes to each regime. In addition, approximation of time series by discrete regimes of stochastic process may be useful for several real decision making problems, including economic and monetary policy, financial investment and risk management. On the other hand, many observed economic indicators, such as stock prices and foreign exchange rates, are considered as proxy variables of some economic activities, economic conditions, investor’s behaviours and several other factors. Therefore, it is difficult if not impossible to know perfectly in advance the determinative regime in economic and financial time series. It may be more practical to statistically identify the probability of the regime.

We can see some examples of such a regime dependent dynamic behavior. The following examples illustrate that regime switching is relevant for real economic and financial time series.

Figure 1.1: Example 1 — Tokyo Stock Price Index (TOPIX)

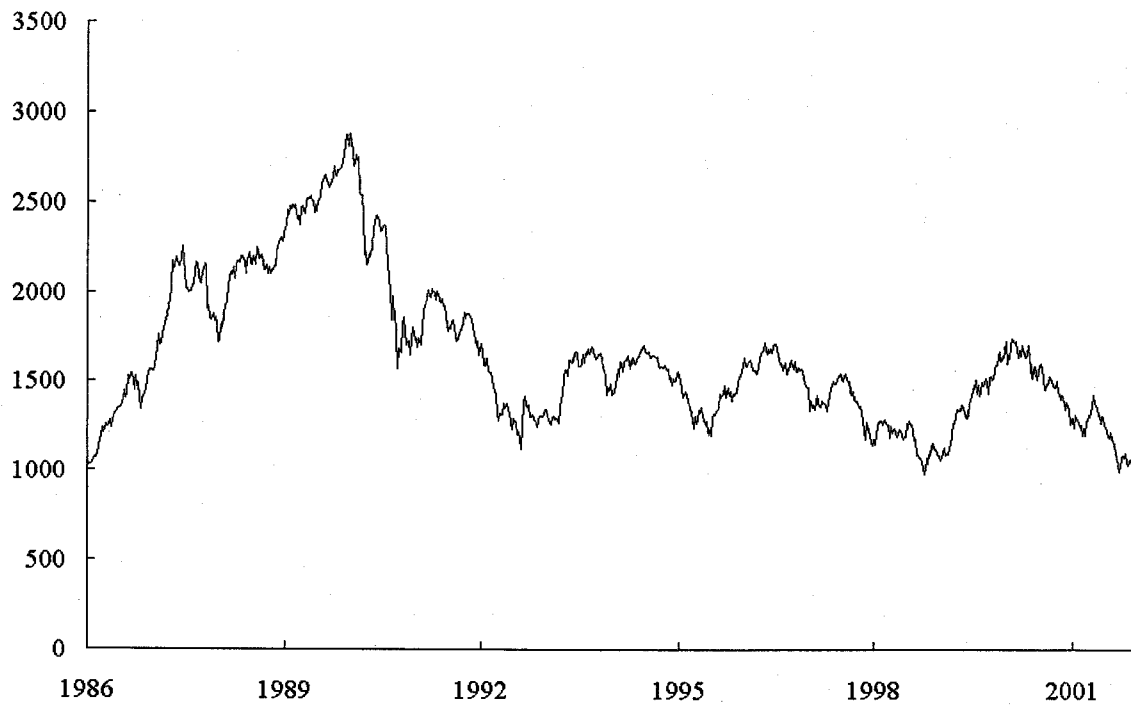


Figure 1.2: Example 2 — Logarithmic return of Japan/U.S. exchange rate

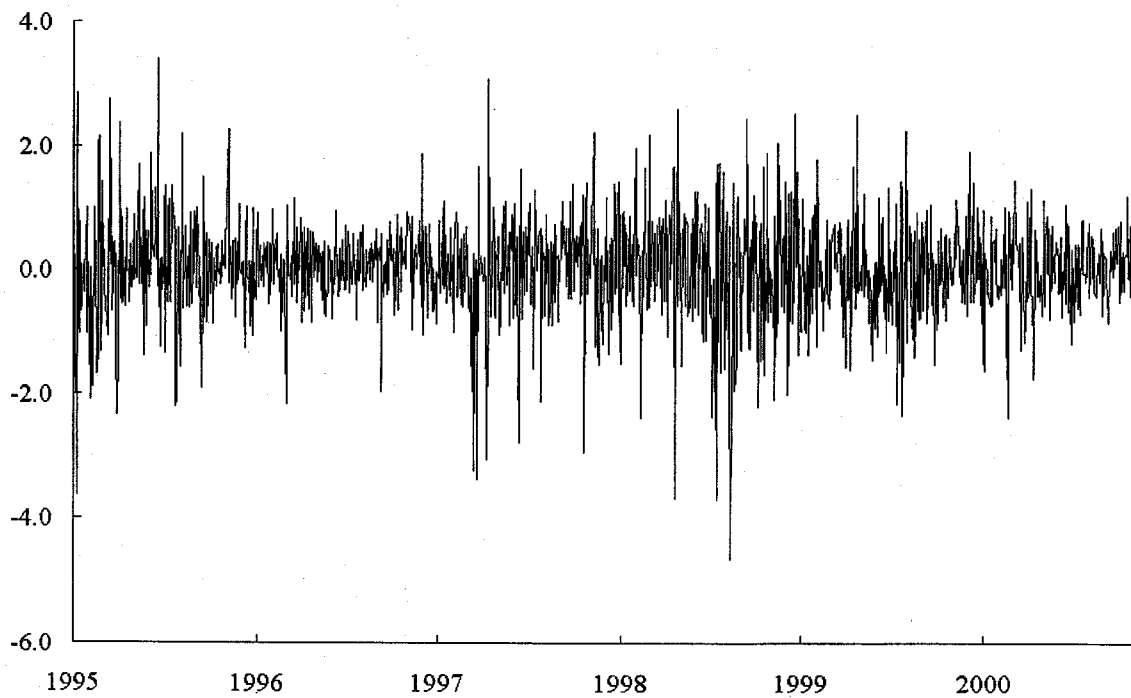


Figure 1.3: Example 3 — Japanese industrial production and business cycles

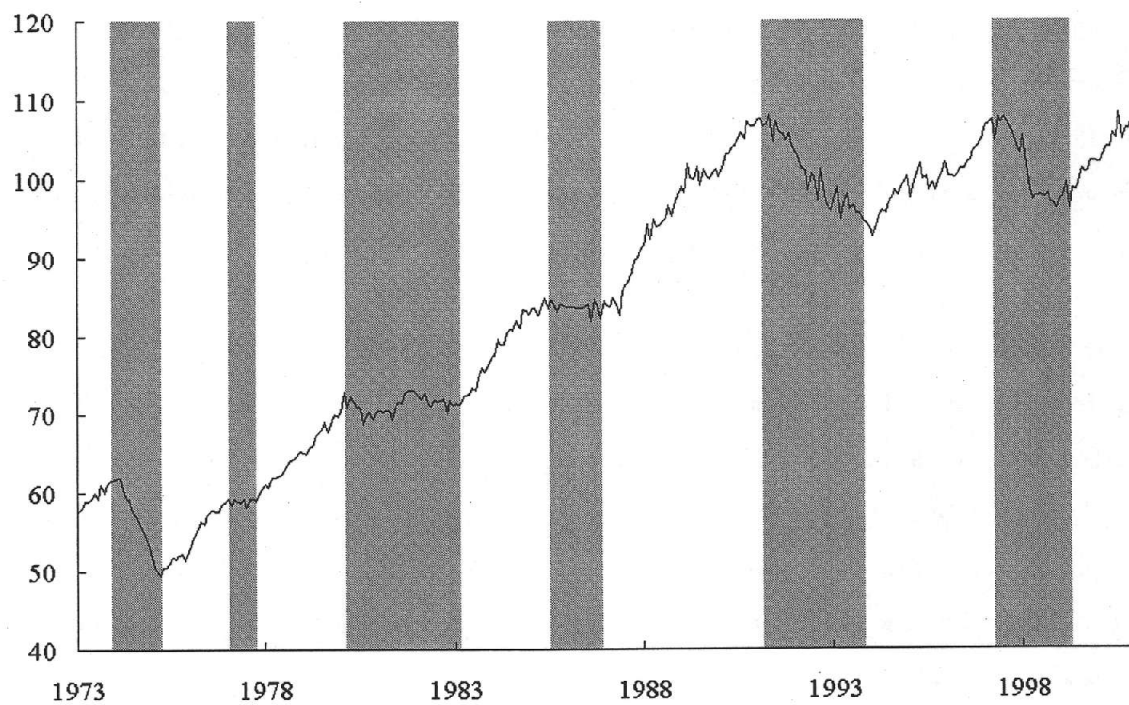
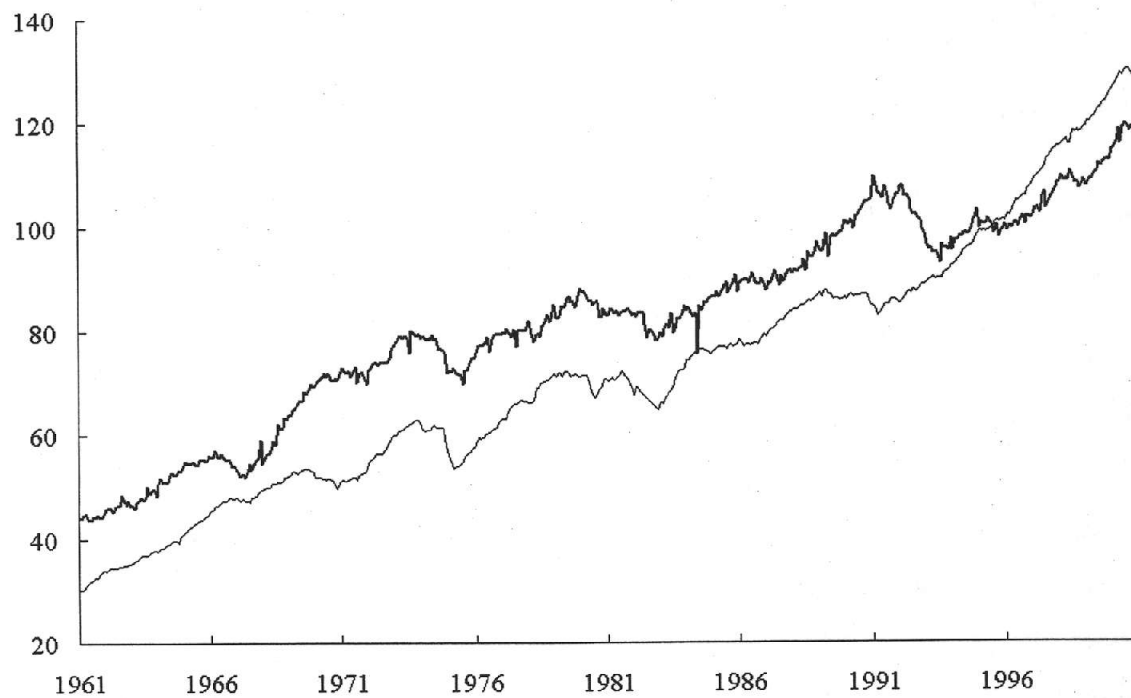


Figure 1.4: Example 4 — U.S. and Germany industrial production



Example 1. Trend identification

Figure 1.1 plots the weekly Tokyo stock price index (TOPIX) from 1986 to 2001. We may find some piecewise linear movements of TOPIX. In the traditional technical analysis of financial markets, the chartist identifies “trend” by empirical methods. However, these approaches are not necessarily scientifically reproducible. In the framework of time series analysis, some models to estimate the trend of time series are proposed. West and Harrison (1997) and Kitagawa and Gersch (1996) summarize in detail with polynomial trend models and state space trend models.

Example 2. Volatility of asset price returns

Figure 1.2 plots the logarithmic return of Japanese/U.S. exchange rates. Obviously, this level of volatility differs in the periods of 1998 and others. From this figure, we may find the presence of some heteroscedasticity and its persistence in the time series. Estimating and forecasting volatility of asset price returns are important matters in risk management, derivative pricing and hedging, portfolio selection and many other financial activities. There is now an enormous body of research on volatility models: for example, ARCH, GARCH, stochastic volatility models and their derivatives.

Example 3. Business cycles analysis

Figure 1.2 plots Japanese industrial products, the shadowed bars showing recession periods defined by the Japanese Cabinet Office. Industrial production is an important factor in identifying business cycles with expansions and recessions. This time series seems to have upward and downward trend regimes. One regime corresponds to the expansion and the other to the contraction of production. It is important to estimate and forecast business cycle turning points because such turning points allow us to make economic and business policy decisions more flexibly and rapidly. Komaki (2001) and Honda and Matsuoka (2001) review some methods of estimating and forecasting the business cycle turning points: for example, the diffusion index (DI), Neftci model, Probit model, Stock-Watson model and so on.

Example 4. Transmission of time series properties

Figure 1.3 plots U.S. and German industrial production. Both time series display common upward and downward trend regimes, but also a period when they move in different directions. It is possible to consider that there is some unobserved relationships between these regimes in both time series. At the present time, the globalization of the economy seems to emphasize the linkage of business cycles and financial markets between countries. Vector autoregressive (VAR) and multivariate ARCH-type models are methods to analyze relations of cause and effect in a multivariate linear time series.

Markov switching approach

The available regime-switching models differ in the way the regime evolves over time. Two main classes of models can be distinguished. The models in the first class assume that the regimes can be characterized by an observable variable. Consequently, the regimes that have occurred in the past and present are known with certainty. The models in the second class assume that the regime cannot actually be observed but is determined by an underlying unobservable stochastic process. This implies that one can never be certain that a particular regime has occurred at a particular point in time, but can only assign probabilities to the occurrence of the different regimes.

In this thesis, we focus on situations in which the regime shifts are *unknown* and the regime processes are *stochastic* and *recursive*. In recent years several time series models have been proposed that formalize the idea of the existence of different regimes generated by the stochastic process. One stochastic model that deals with the nonlinear time series including the stochastic regimes is the *Markov switching model*. The purpose of this Markov switching approach to modeling economic and financial time series seems to be to identify different regimes and their changing points, and to allow for the possibility that the dynamic behavior of economic variables depends on the regime that occurs at any unknown point in time.

The primary purpose of this thesis is to modify and extend the ordinary Markov switching models in order to capture the regime-switching and non-linear characteristics of real economic and financial time series. The secondary purpose is to examine the empirical analysis by using the proposed Markov switching models.

We consistently use the Markov switching models in the empirical analysis of economic and financial time series (in the Chapter 4) through this thesis. It seems reasonable to suppose that the approximation of time series by discrete regimes and corresponding stochastic process is one of useful approaches to capture economic and financial dynamic behaviors. Because economic and financial time series are not necessarily generated under the common condition and by the common mechanism over time, differently from physical observations or simulation data. The industrial index of products provides an example. Industrial companies as the economic agent seem to take different action for their management strategy (*e.g.*, business investment and marketing strategy), depending on regimes of economic expansion and contraction. In addition, the government adopts different economic policy, which has different effects on the industrial index of products. Financial asset price is another illustration of the same point. Market participants, trading volume and investors' risk tolerance seem to differ according to volatility levels (*e.g.*, low and high volatility levels) and price directions (*e.g.*, upward and downward trends). In financial markets, the monetary policy and market inventory by authorities cannot be said to be

symmetric obviously. In the empirical analysis of the Chapter 4, we compare some Markov switching models with their alternative continuous-type stochastic models and ordinary analytical methods.

1.2 Review of Literature

Markov switching model

Goldfred and Quandt (1973), Lindgren (1978), Tyssedal and Tjostheim (1998) and Komaki (1993) proposed the switching models with some regime processes depending on the finite Markov chain. Hamilton (1989) generalized the existing Markov switching autoregressive models, and applied it to the U.S. business cycle analysis using GNP data. Many modifications and extensions of Hamilton's Markov switching model have been proposed in the business cycle analysis: Birchenhall *et al.* (1999), Filardo (1994, 1998), Kaufmann (2000), Kim (1994), Lahiri (1994), Lam (1990) and Luginbuhl (1999). Applications for foreign exchange rates were by Dacco *et al.* (1999), Engel and Hamilton (1990), Engel and Kim (1999), Kaminisky (1993) and Fruhwirth (2001); those for interest rates, monetary policies and inflation were Bidarkota (2001), Garcia and Perron (1996), Hall *et al.* (1999) and Kim (1993). While there are several studies on stock markets using Markov switching models with regime shifts in mean, variance and autoregressive terms, some alternative models were proposed for stock market volatility, as shown below. Franses and Dijk (2000), Hamilton (1994), and Kim and Nelson (1999) summarized the Markov switching models and their empirical applications in economic and financial time series.

To capture the property of volatility in the economic and financial time series, ARCH models with Markov switching structures were proposed by Hamilton and Susmel (1994) and Cai (1994). Fong (1997) applied Hamilton and Susmel's (1994) SWARCH model to the Japanese stock market. Hamilton and Lin (1996) and Susmel (200) extended the univariate SWARCH model to the bivariate versions to capture international volatility transmission in stock markets. Other Markov switching models for heteroscedasticity were given by Dueker (1997) and So *et al.* (1998).

Shumway and Stoffer (1991) and Kim (1993, 1994) represented the dynamic switching and Markov switching models in the state space form, respectively, and their models are considered as extensions of the Multi process model by Harrison and Stevens (1976). Filardo (1994) and Filardo and Gordon (1998) allowed an update of the transition probability with not only past regimes but also exogenous variables. The semi-Markov chain is another regime transition mechanism, and Durland and McCurdy (1994), Maheu and McCurdy (2000) and Xianping and Smyth (2000) used the duration-dependent transitions and semi-Markov chain.

Phillips (1991) extended the univariate Markov switching model to the bivariate version with several types of transition mechanism, and applied it to the international business cycle transmission. Kim and Yoo (1995), Kontolemis (2001) and Krolzig (1997) studied business cycle analysis by using the multivariate Markov switching model with simultaneous transition probability. In the analysis of volatility, there are also some bivariate Markov switching models (Hamilton and Lin 1996, Susmel 2000).

The self-organizing state space model was first introduced by Kitagawa (1996). The implementation of the self-organizing state space model is achieved by the Monte Carlo filter algorithm (Kitagawa 1996). Kitagawa and Hakamata (2001) and Higuchi (2001) incorporated the Markov and semi-Markov switching models into the self-organizing model, while Hakamata and Kitagawa (2002) proposed its modifications and extensions.

Other approaches for latent regime and change point detection

There are many studies for latent regime and change point analysis in statistics and engineering (Basseville and Nikiforov 1993, Lai 1995), while the CUSUM algorithm is a well-known classical method (Page 1954). Some models allowing the capture of a piecewise linear property of time series are proposed; *e.g.*, piecewise regression (Hawkins 1976, Gustafsson 1996), segmented regression (Lerman 1980) and multi-phase regression (Hinkley 1971). In a Bayesian framework, Broemeling and Tsurumi (1987), Smith and Cook (1980) and Stephens (1994) deal with such a problem.

Stochastic models with Poisson white noise (Kontorovich and Lyandres 1996, Snyder and Miller 1991) attempt to approximate the low-frequency probability of regime changes or shifts, which resemble the sound of a gunshot or a sudden noise. A variety of applications to analyze volatility of some financial asset price returns have recently been proposed. Ball and Torous (1985), Jorion (1988) and Ball and Roma (1993) examine empirical analysis by using models with independent jump and diffusion components. Bates (1996, 2000), Iino and Ozaki (1999) and Ozaki and Iino (2001) have pointed out problems of ordinary modeling and have proposed new models. It can be said that this Poisson jump-type approach is more effective to modeling economic and financial time series including very large and rare outliers like financial market crashes and economic events: *e.g.*, the September 11 synchronized terrorist attack, Enron bankruptcy and European monetary union. On the other hand, the Markov switching model wields its power in the case of where regime shifts occur relatively slowly or gradually, or where it would be difficult to identify changing points due to several unstable noises. We can see easily such a phenomenon in real economy and financial markets: *e.g.*, business cycle and trend change of financial asset prices.

The duration-dependent Markov switching model (Durland and McCurdy 1994, Maheu and McCurdy

2000, Lam 1997) or semi-Markov switching model (Kitagawa and Hakamata 2001) is one of the statistical models for regime changes. In this model, the probability of regime shift depends on both the previous regime and its duration time. Thus, the Markov switching model is regarded as a special case of the semi-Markov switching mode. We consider semi-Markov switching model as one of the subjects of our future works. Univariate and multivariate semi-Markov switching models are described briefly in the Chapter 5.

1.3 Contributions of this Thesis

This thesis, which covers 1) empirical analysis in economics and financial markets by using newly proposed Markov switching models, and 2) descriptions for extensions of the ordinary Markov switching models concerning model specification and time-varying parameterization, proposes two contributions.

The first contribution is to undertake empirical economic and financial time series analysis. To capture the characteristics of complex time series processes in economic and financial markets, We apply the ordinary and newly modified and extended Markov switching models to the empirical economic and financial time series data in Chapter 4. A notable feature of this application is the novelty in scope, object and method of economic and financial analysis. The empirical analysis covers the following subjects and issues:

- (1) Trend identification and trading strategy.
- (2) Time-series and cross-sectional volatility analysis.
- (3) Japanese business cycle analysis.
- (4) Japan premium and Japanese banks' stock volatility.
- (5) Transmission of volatility.
- (6) Foreign exchange volatility and intervention.
- (7) International business cycle transmission.

The second contribution is the suggested extension of some of the Markov switching models used in the empirical analysis in Chapter 5. Those that include constant parameters are generalizations of Markov switching ARCH model, Markov switching slope change and ARCH model, and semi-Markov switching models. In addition, We introduce the self-organizing Markov switching model (Kitagawa and

Hakamata, 2001). This model is obtained by incorporating the conventional Markov switching model with constant parameters into a self-organizing state space model (Kitagawa 1998).

Chapter 2

Markov Switching Model

2.1 Markov Chain

The concept of the Markov chain plays a central role in the Markov switching model. In this section, we briefly review the theory of Markov chains. Kijima (1997) studied the general theory of the Markov processes, and Hamilton (1994) discussed the Markov chain for the Markov switching model.

Discrete-time Markov Chain

The Markov property asserts that the distribution of the sequence of stochastic processes $\{s_n\}$ depends only on the state $s_{n-1} = i_{n-1}$ at time $n - 1$, not on the whole history. Formally, for each n and every i_0, \dots, i_{n-1} and $j \in S$, the process $\{s_n\}$ with state space S is called *Markov process*, if the following equality holds

$$\Pr(s_n = j | s_0 = i_0, \dots, s_{n-1} = i_{n-1}) = \Pr(s_n = j | s_{n-1} = i_{n-1}). \quad (2.1)$$

Given the history $\{s_0 = i_0, \dots, s_{n-1} = i_{n-1}\}$, the Markov property in equation (2.1) suggests that the current state $s_n = j$ is enough to determine all distributions of the future. We refer to a Markov process as a *Markov chain* for a discrete-time case ($n = 1, 2, \dots, N$), when the state space is finite $S \in \{1, 2, \dots, m\}$. The conditional probability in the right-hand side of the Markov property in equation (2.1) is called the *one-step transition probability* from state i to state j at time n .

If the one-step transition probability in equation (2.1) is independent of time n , the transition probability is *stationary*, and the Markov chain $\{s_n\}$ is said to be *time-homogeneous*. Thus, we can define the

transition probability in equation (2.1) as follows:

$$p_{ij} \equiv \Pr(s_n = j | s_{n-1} = i), \quad n = 0, 1, \dots, N. \quad (2.2)$$

It is often convenient to express the transition probabilities p_{ij} in an $(m \times m)$ matrix P known as the *transition matrix*:

$$P = \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{m1} \\ p_{12} & p_{22} & \cdots & p_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1m} & p_{2m} & \cdots & p_{mm} \end{pmatrix}, \quad (2.3)$$

where

$$0 \leq p_{ij} \leq 1, \quad \sum_{j=1}^m p_{ij} = 1, \quad i, j \in \{1, 2, \dots, m\}. \quad (2.4)$$

In this expression, the (i, j) element of P is the transition probability p_{ij} ; i.e., the probability that state i will be followed by state j . More generally, the l -step transition probability at time n is defined by

$$p_{ij}^{(l)} = \Pr(s_{n+l} = j | s_n = i) = \sum_{k=1}^m p_{jk}^{(l-1)} p_{ki}, \quad (2.5)$$

and the corresponding l -step transition matrix at time n becomes

$$P^{(l)} = P^l, \quad (2.6)$$

which is independent of time n .

When the transition probability in equation (2.1) is not stationary (nonstationary), the definition of equation (2.2) is replaced by

$$p_{n,ij} \equiv \Pr(s_n = j | s_{n-1} = i). \quad (2.7)$$

and the corresponding l -step transition matrix at time n ,

$$p_n^{(l)} = P_n P_{n+1} \cdots P_{n+l-1}, \quad (2.8)$$

can be represented by the product of l time-varying transition matrices P_n, \dots, P_{n+l-1} , which is a generalization of the result of the stationary transition probability such as equation (2.6).

Reducibility and Irreducibility

As an example, we consider a two-state Markov chain with the transition matrix

$$P = \begin{pmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{pmatrix}. \quad (2.9)$$

Suppose that $p_{11} = 1$, so that the matrix P is upper triangular. Then, once the process enters state 1, there is no possibility of ever returning to state 2. In such a case we would say that state 1 is an *absorbing state* and that the Markov chain is *reducible*.

In general, an m -state Markov chain is said to be reducible if there exists a way to label the states (that is, a way to choose which state to call state 1, which to call state 2, and so on) such that the transition matrix can be written in the form

$$P = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}, \quad (2.10)$$

where B denotes a $(k \times k)$ matrix for some $1 \leq k < m$. If P is upper block-triangular matrix, then so is P^l for any l . Hence, once such a process enters a state j such that $j \leq k$, there is no possibility of ever returning to one of the states $k+1, k+2, \dots, m$.

On the other hand, a Markov chain that is not reducible is said to be *irreducible*. For example, a two-state chain is irreducible, if $p_{11} < 1$ and $p_{22} < 1$.

Periodicity and Aperiodicity

A Markov chain is said to be *periodic* with d , if $d \leq 2$ is the greatest common divisor of all integers $l > 1$ for which $p_{jj}^{(l)} > 0$. If there is no such $d \leq 2$, then a Markov chain is *aperiodic*. Such chains have the property that the states can be classified into d distinct classes.

Ergodic Markov Chain and Stationary Distribution

A Markov chain is called *ergodic* if it is irreducible and aperiodic. Consider an m -state ergodic Markov chain with transition matrix P . Suppose that one of the eigenvalues of P is unity and that all other eigenvalues of P are inside the unit circle. The $(m \times 1)$ vector of *stationary distribution* or *ergodic probability* for an ergodic Markov chain is denoted by η . This vector η is defined as the eigenvector of P associated with the unit eigenvalue; that is, the vector of stationary distribution η satisfies

$$P\eta = \eta. \quad (2.11)$$

The eigenvector π is normalized so that its elements sum to unity ($\mathbf{1}^T \pi = 1$, where $\mathbf{1}$ denotes an $(m \times 1)$ vector of 1s). It can be shown that if P is the transition matrix for an ergodic Markov chain,

$$\lim_{m \rightarrow \infty} P^m = \eta \mathbf{1}^T. \quad (2.12)$$

Calculating Stationary Distribution

For a general ergodic m -state process, the vector of unconditional probabilities represents a vector η with the properties that $P\eta = \eta$ and $\mathbf{1}^T \eta = 1$. We thus seek a vector η satisfying

$$A\eta = e_{m+1}. \quad (2.13)$$

where e_{m+1} denotes the $(m+1)$ th column of I_{m+1} and the $((m+1) \times m)$ matrix A is defined by

$$A = \begin{pmatrix} I_m - P \\ \mathbf{1}^T \end{pmatrix}. \quad (2.14)$$

Such a solution can be found by premultiplying equation (2.13) by $(A^T A)^{-1} A^T$:

$$\pi = (A^T A)^{-1} A^T e_{m+1}. \quad (2.15)$$

In other words, η is the $(m+1)$ th column of the matrix $(A^T A)^{-1} A^T$.

Expected Duration of a State

The expected duration of each state can easily be obtained from the diagonal elements of the transition probability; that is, the self-loop transition probability. When d is defined as the number of time-steps spent in state j , the probability distribution $D_j(t_d)$ of the duration time t_d of the system in state j is given by

$$D_j(t_d = d) = p_{jj}^{d-1} (1 - p_{jj}). \quad (2.16)$$

The Markov chain constrains the state-duration distributions to be geometric in form. Figure 2.1 plots their illustrative example.

Then, the expected duration of regime j can be derived as

$$\begin{aligned} E(d) &= \sum_{d=1}^{\infty} d D_j(t_d = d) \\ &= 1 \times (1 - p_{jj}) + 2 \times p_{jj} (1 - p_{jj}) + 3 \times p_{jj}^2 (1 - p_{jj}) + \dots \\ &= \frac{1}{1 - p_{jj}}. \end{aligned}$$

Semi-Markov Chain

This subsection discusses a semi-Markov chain as the modification of the standard Markov chain (Xianping and Padhraic 2000). The process $\{s_n\}$ is called *semi-Markov chain* if it has the following generative description:

Figure 2.1: Duration distribution of Markov chain.

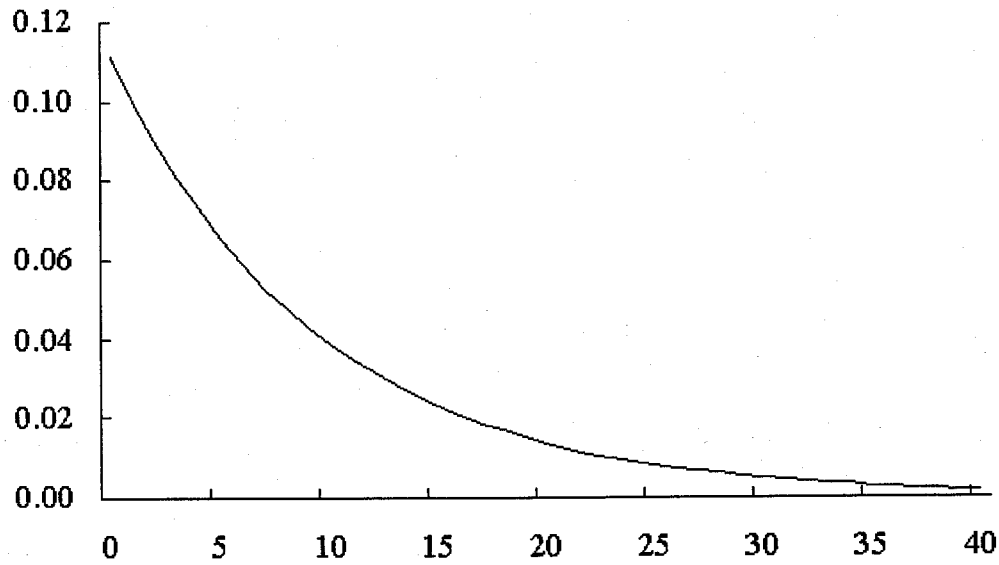
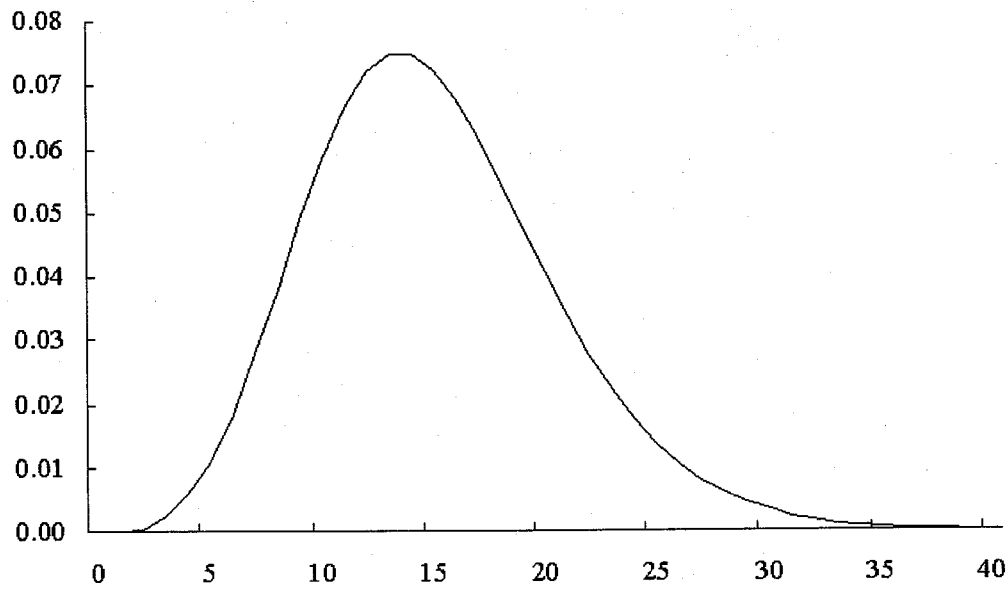


Figure 2.2: Duration distribution of a semi-Markov chain.



- (1) On entering state i , a duration time t_d is drawn from an arbitrary probability distribution $D_j(t_d)$.
Note that t_d is constrained to take only integer values.
- (2) The process remains in state j for time t_d .
- (3) At time t_d the process transitions to another state according to a transition matrix P , and the process repeats.

Here, $D_i(t_d)$, $1 \leq i \leq m$, can be modeled using parametric distributions (such as log-normal, Gamma, negative binomial etc) or non-parametrical by mixtures, kernel densities, etc. If t_d is constrained to take only integer values we get a discrete-time semi-Markov process. The semi-Markov chain can be represented by a non-stationary Markov model where the transition probabilities are a deterministic function of the probability distribution $D_j(t_d)$ of the duration time t_d . Figure 2.2 illustrates an example of the duration probability of a semi-Markov chain with negative binomial distribution.

2.2 Ordinary Markov Switching Model

2.2.1 Markov Switching Stochastic Trend Model

We consider a first-order stochastic trend process in which the means and variances of innovation terms can change as the result of a regime-shift. Let y_n , ($n = 1, 2, \dots, N$) be a first-order difference (or logarithmic difference) of a univariate observed variable. The univariate Markov switching stochastic trend model is given by

$$y_n \sim N(\mu_{s_n}, \sigma_{s_n}^2), \quad (2.17)$$

where y_n follows a normal white noise whose distribution changes according to the regime. The regime itself will be described as the outcome of a latent discrete Markov chain $s_n \in \{1, 2, \dots, m\}$. y_n depends on s_n over time as follows:

$$\begin{aligned} \text{if } s_n = 1, & \quad y_n \sim N(\mu_1, \sigma_1^2) \\ \text{if } s_n = 2, & \quad y_n \sim N(\mu_2, \sigma_2^2) \\ & \vdots \\ \text{if } s_n = m, & \quad y_n \sim N(\mu_m, \sigma_m^2) \end{aligned}$$

First, the joint conditional distribution of y_n , s_n and s_{n-1} on Ψ_{n-1} is obtained by

$$f(y_n, s_n = j, s_{n-1} = i | \Psi_{n-1}) = f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1}) \Pr(s_n = j, s_{n-1} = i | \Psi_{n-1}), \quad (2.18)$$

where Ψ_{n-1} denotes the information available up to time $n-1$. Here, the density of y_n conditional on $s_n, s_{n-1} = i$ and Ψ_{n-1} taking on the value j is

$$f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1}) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left\{ -\frac{(y_n - \mu_j)^2}{2\sigma_j^2} \right\}, \quad (2.19)$$

for all $j = 1, 2, \dots, m$. According to the property of the first-order Markov chain, the transition probability of s_n in the Markov switching model is given by

$$\Pr(s_n = j | s_{n-1} = i) = p_{ij}, \quad (2.20)$$

with

$$\sum_{i=1}^m \Pr(s_n = j | s_{n-1} = i) = 1 \quad j = 1, 2, \dots, m. \quad (2.21)$$

Thus, the conditional probability of s_n and s_{n-1} in equation (2.18) can be updated via the transition probability in equation (2.20) as follows

$$\Pr(s_n = j, s_{n-1} = i | \Psi_{n-1}) = \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}), \quad (2.22)$$

and the predictive probability of s_n is given by

$$\begin{aligned} \Pr(s_n = j | \Psi_{n-1}) &= \sum_{i=1}^m \Pr(s_n = j, s_{n-1} = i | \Psi_{n-1}) \\ &= \sum_{i=1}^m \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}). \end{aligned} \quad (2.23)$$

Summarizing equations (2.18) to (2.23), the joint conditional distribution of y_n, s_n is given by

$$\begin{aligned} f(y_n, s_n = j, s_{n-1} = i | \Psi_{n-1}) \\ = \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}) \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left\{ -\frac{(y_n - \mu_j)^2}{2\sigma_j^2} \right\}, \end{aligned} \quad (2.24)$$

and that of s_{n-1} on Ψ_{n-1} , and that of y_n on Ψ_{n-1} is given by

$$\begin{aligned} f(y_n | \Psi_{n-1}) &= \sum_{j=1}^m \sum_{i=1}^m f(y_n, s_n = j, s_{n-1} = i | \Psi_{n-1}) \\ &= \sum_{j=1}^m \sum_{i=1}^m \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}) \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left\{ -\frac{(y_n - \mu_j)^2}{2\sigma_j^2} \right\} \\ &= \sum_{j=1}^m \Pr(s_n = j | \Psi_{n-1}) \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left\{ -\frac{(y_n - \mu_j)^2}{2\sigma_j^2} \right\}. \end{aligned} \quad (2.25)$$

The log likelihood for the observed data of the Markov switching model can be calculated from equation (5.17) as

$$\begin{aligned}
L(\theta) &= \sum_{n=1}^N \log f(y_n | \Psi_{n-1}) \\
&= \sum_{n=1}^N \log \left[\sum_{j=1}^m \sum_{i=1}^m f(y_n, s_n = j, s_{n-1} = i | \Psi_{n-1}) \right] \\
&= \sum_{n=1}^N \log \left[\sum_{j=1}^m \sum_{i=1}^m \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}) \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left\{ -\frac{(y_n - \mu_j)^2}{2\sigma_j^2} \right\} \right] \\
&= \sum_{n=1}^N \log \left[\sum_{j=1}^m \Pr(s_n = j | \Psi_{n-1}) \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left\{ -\frac{(y_n - \mu_j)^2}{2\sigma_j^2} \right\} \right] \tag{2.26}
\end{aligned}$$

where θ denotes the unknown parameters. The maximum likelihood estimation of θ is obtained by maximizing equation (2.26).

By the non-Gaussian filter (Kitagawa 1987), the filtering probability of s_n

$$\begin{aligned}
\Pr(s_n = j | \Psi_n) &= \frac{f(y_n, s_n = j | \Psi_{n-1})}{f(y_n | \Psi_{n-1})} \\
&= \frac{f(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1})}{f(y_n | \Psi_{n-1})} \\
&= \frac{\sum_{i=1}^m \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}) f(y_n | s_n = j, \Psi_{n-1})}{f(y_n | \Psi_{n-1})}, \tag{2.27}
\end{aligned}$$

and the predictive probability

$$\begin{aligned}
\Pr(s_{n+1} = k | \Psi_n) &= \sum_{j=1}^m \Pr(s_{n+1} = k, s_n = j | \Psi_n) \\
&= \sum_{j=1}^m \Pr(s_{n+1} = k | s_n = j) \Pr(s_n = j | \Psi_n), \tag{2.28}
\end{aligned}$$

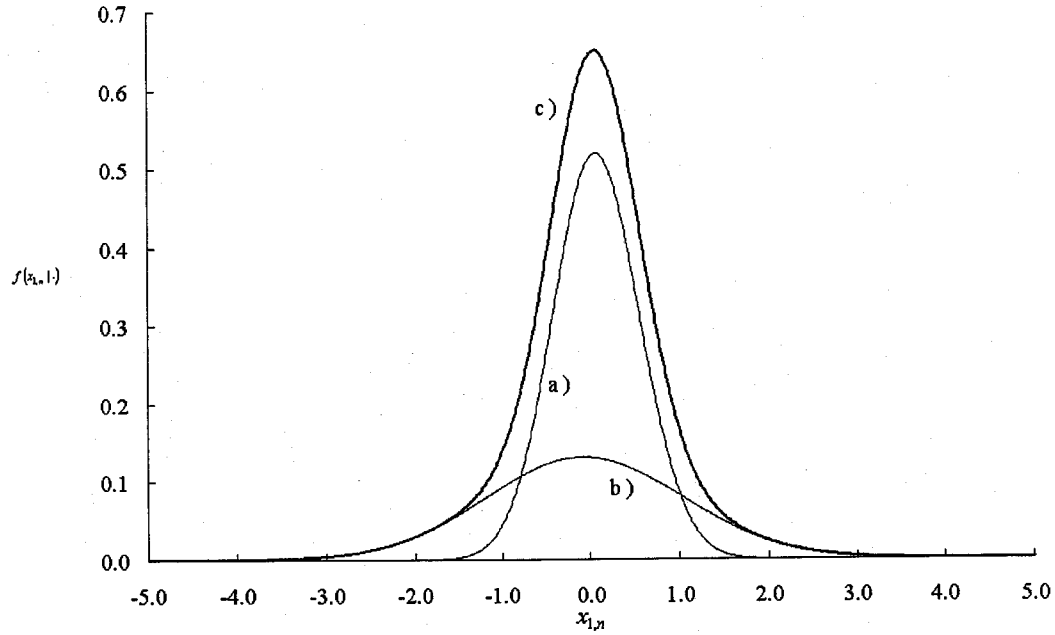
are recursively obtained.

Two illustrative examples for the two-regime Markov switching model ($m = 2$) can be found in Figures 2.3 (example 1) and 2.4 (example 2). The joint distribution $f(y_n, s_n = i), i \in \{1, 2\}$, is obtained as η_i times a $N(\mu_i, \sigma_i^2)$ density. The unconditional distribution $f(y_n)$ for the observed variable is the sum of two magnitudes, $f(y_n, s_n = 1)$ and $f(y_n, s_n = 2)$. Figure 2.3 shows the case of Gaussian mixtures with the Markov switching variance, allowing skew or kurtosis of variable y_n . Figure 2.4 shows the case of Gaussian mixtures with the Markov switching mean.

2.2.2 Multivariate Markov Switching Statistic Trend Model

We extend the Markov switching process for the univariate time series mentioned in the previous section to the multivariate time series. Let $y_n \equiv (y_{1,n}, y_{2,n}, \dots, y_{k,n})^T, n = 1, 2, \dots, N$, be the k -dimensional first-order difference (or logarithmic difference) of time series at time n . The k -dimensional first-order

Figure 2.3: Example 1 — Joint distributions a: $f(y_n, s_n = 1)$ and b: $f(y_n, s_n = 1)$, and unconditional distribution c: $f(y_n)$.



stochastic trend model (random walk model) with m -state first-order Markov switching drift and variance is given by

$$y_n \sim N(\mu, \Sigma) \quad n = 1, 2, \dots, N, \quad (2.29)$$

where y_n is assumed to be a k -dimensional Gaussian noise sequence with $(k \times 1)$ mean vector and $(k \times k)$ variance-covariance matrix as follows

$$\mu = \begin{pmatrix} \mu_{1,s_{1,n}} \\ \mu_{2,s_{2,n}} \\ \vdots \\ \mu_{k,s_{k,n}} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{1,s_{1,n}}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{2,s_{2,n}}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{k,s_{k,n}}^2 \end{pmatrix}, \quad (2.30)$$

Here, the elements of μ and the diagonal elements of Σ depend upon a k -dimensional m -state and first-order Markov chain vector

$$s_n = \begin{pmatrix} s_{1,n} \\ s_{2,n} \\ \vdots \\ s_{k,n} \end{pmatrix}, \quad n = 1, 2, \dots, N, \quad (2.31)$$

where $s_{l,n} \in \{1, 2, \dots, k\}$.

$$\begin{array}{ll} s_n^* = 1, & \text{if } s_{1,n} = 1, \dots, s_{k,n} = 1, \\ s_n^* = 2, & \text{if } s_{1,n} = 2, \dots, s_{k,n} = 1, \\ \vdots & \vdots \\ s_n^* = m^k - 1, & \text{if } s_{1,n} = m, \dots, s_{k,n} = m - 1, \\ s_n^* = m^k, & \text{if } s_{1,n} = m, \dots, s_{k,n} = m. \end{array}$$
$$\begin{array}{ll} \text{if } s_n^* = 1, & \mu = (\mu_{1,1}, \dots, \mu_{k,1})^T \text{ and } \sigma = (\sigma_{1,1}, \dots, \sigma_{k,1})^T \\ \text{if } s_n^* = 2, & \mu = (\mu_{1,1}, \dots, \mu_{k,1})^T \text{ and } \sigma = (\sigma_{1,1}, \dots, \sigma_{k,1})^T \\ \vdots & \vdots \\ \text{if } s_n^* = m^k - 1, & \mu = (\mu_{1,m^k}, \dots, \mu_{k,m^k-1})^T \text{ and } \sigma = (\sigma_{1,m^k}, \dots, \sigma_{k,m^k-1})^T \\ \text{if } s_n^* = m^k, & \mu = (\mu_{1,m^k}, \dots, \mu_{k,m^k})^T \text{ and } \sigma = (\sigma_{1,m^k}, \dots, \sigma_{k,m^k})^T \end{array}$$

we can replace the density of y_n conditional on s_n , s_{n-1} and Ψ_{n-1} in equation (2.19) for the univariate version by

$$f(y_n | s_n^* = j^*, s_{n-1} = i^*, \Psi_{n-1}) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \bar{y}_n^T \Sigma^{-1} \bar{y}_n \right\}, \quad (2.32)$$

for all $j^* = 1, 2, \dots, m^k$. Note that the $(k \times 1)$ vector $\bar{y}_n \equiv (y_n - \mu)$. The joint conditional distribution of y_n , s_n^* and s_{n-1}^* on Ψ_{n-1} is given by

$$\begin{aligned} f(y_n, s_n^* = j^*, s_{n-1}^* = i^* | \Psi_{n-1}) &= \Pr(s_n^* = j^* | s_{n-1}^* = i^*) \Pr(s_{n-1}^* = i^* | \Psi_{n-1}) \\ &\times (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \bar{y}_n^T \Sigma^{-1} \bar{y}_n \right\}. \end{aligned} \quad (2.33)$$

and the conditional density of y_n can be obtained via summation of equation (5.21) over all possible states:

$$\begin{aligned} f(y_n | \Psi_{n-1}) &= \sum_{j^*=1}^{m^k} \sum_{i^*=1}^{m^k} f(y_n, s_n^* = j^*, s_{n-1}^* = i^*) \\ &= \sum_{j^*=1}^{m^k} \sum_{i^*=1}^{m^k} \Pr(s_n^* = j^* | s_{n-1}^* = i^*) \Pr(s_{n-1}^* = i^* | \Psi_{n-1}) \\ &\quad \times (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \bar{y}_n^T \Sigma^{-1} \bar{y}_n \right\} \\ &= \sum_{j^*=1}^m \Pr(s_n^* = j | \Psi_{n-1}) (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \bar{y}_n^T \Sigma^{-1} \bar{y}_n \right\}. \end{aligned} \quad (2.34)$$

The log likelihood of the multivariate Markov switching model is given by

$$\begin{aligned} L(\theta) &= \sum_{n=1}^N \log f(y_n | \Psi_{n-1}) \\ &= \sum_{n=1}^N \log \left[\sum_{j^*=1}^{m^k} \sum_{i^*=1}^{m^k} f(y_n, s_n^* = j^*, s_{n-1}^* = i^* | \Psi_{n-1}) \right] \\ &= \sum_{n=1}^N \log \left[\sum_{j^*=1}^{m^k} \sum_{i^*=1}^{m^k} \Pr(s_n^* = j^* | s_{n-1}^* = i^*) \Pr(s_{n-1}^* = i^* | \Psi_{n-1}) \right. \\ &\quad \left. \times (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \bar{y}_n^T \Sigma^{-1} \bar{y}_n \right\} \right] \\ &= \sum_{n=1}^N \log \left[\sum_{j^*=1}^{m^k} \Pr(s_n^* = j^* | \Psi_{n-1}) (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \bar{y}_n^T \Sigma^{-1} \bar{y}_n \right\} \right]. \end{aligned} \quad (2.35)$$

where θ denotes the unknown parameters. The maximum likelihood estimation of θ is obtained by maximizing equation (2.35).

2.2.3 Transition Probability

In the multivariate Markov switching model, the type of regime transitions can be classified into three categories depending on their transition probabilities — general (unrestricted), independent, and simul-

taneous transitions. The last two are restricted versions of the first. Each type of transition probability can characterize the relationship among observed time series and their corresponding process.

General Transition Probability

In an unrestricted multivariate Markov switching model, the property of $(k \times 1)$ Markov chain vector $S_n \equiv (s_{1,n}, s_{2,n}, \dots, s_{k,n})^T$ is given by

$$\Pr(S_n = J | S_{n-1} = I^1, S_{n-2} = I^2, \dots) = \Pr(S_n = J | S_{n-1} = I^1), \quad (2.36)$$

where the $(k \times 1)$ state are given by $J \equiv (j_1, j_2, \dots, j_k)^T, I \equiv (i_1^1, i_2^1, \dots, i_k^1)^T, I' \equiv (i_1^2, i_2^2, \dots, i_k^2)^T, \dots, \in \{1, 2, \dots, m\}$. $s_{l,n}, l = 1, 2, \dots, k$, is dependent on not only the previous state $s_{l,n-1}$ for itself, but all the previous state $S_{n-1} \equiv (s_{1,n-1}, s_{2,n-1}, \dots, s_{k,n-1})^T$ for $y_n \equiv (y_{1,n}, y_{2,n}, \dots, y_{k,n})^T$. Since the $(k \times 1)$ transition probability in equation (2.36) is independent of time n , the transition probability of the Markov chain vector S_n is to be stationary, and is defined by

$$p_{i_1 \dots i_k, j_1 \dots j_k} = \Pr(s_{1,n} = j_1, \dots, s_{k,n} = j_k | s_{1,n-1} = i_1, \dots, s_{k,n-1} = i_k), \quad (2.37)$$

with

$$\sum_{j_1=1}^m \dots \sum_{j_k=1}^m p_{i_1 \dots i_k, j_1 \dots j_k} = 1, \quad (2.38)$$

for all $i_1, i_2, \dots, i_k = 1, 2, \dots, m$. Therefore, we may consider the k -dimensional and m -state Markov chain $s_{l,n}, l = 1, 2, \dots, k$ as the univariate and m^k -state Markov chain s_n^* as follows

$$\begin{aligned} s_n^* &= 1, & \text{if } s_{1,n} &= 1, s_{2,n} = 1, \dots, s_{k,n} = 1, \\ s_n^* &= 2, & \text{if } s_{1,n} &= 2, s_{2,n} = 1, \dots, s_{k,n} = 1, \\ &\vdots & &\vdots \\ s_n^* &= m^k - 1, & \text{if } s_{1,n} &= m, s_{2,n} = m, \dots, s_{k,n} = m - 1, \\ s_n^* &= m^k, & \text{if } s_{1,n} &= m, s_{2,n} = m, \dots, s_{k,n} = m, \end{aligned}$$

Since this newly indicated Markov chain s_n^* follows the first-order and m^k -state Markov process, the transition between regimes can be defined with transition probability

$$p_{i^* j^*}^* = \Pr(s_n^* = j^* | s_{n-1}^* = i^*), \quad i^*, j^* = 1, 2, \dots, m^k. \quad (2.39)$$

This transition probability $p_{j^*i^*}^*$ is represented by the following $(m^k \times m^k)$ transition matrix form

$$P = \begin{pmatrix} p_{11}^* & p_{21}^* & \cdots & p_{m^k 1}^* \\ p_{12}^* & p_{22}^* & \cdots & p_{m^k 2}^* \\ \vdots & \vdots & \ddots & \vdots \\ p_{1m^k}^* & p_{2m^k}^* & \cdots & p_{m^k m^k}^* \end{pmatrix}, \quad (2.40)$$

where the (i, j) element shows the transition probability from state j to i . The diagonal elements indicate the self-loop transition probabilities. We can easily see the relationships and transmission among the states with the transition diagram below.

As an example, see the bivariate and two-state Markov chain $s_{l,n} \in \{1, 2\}, l = 1, 2$ with the general transition probability. This can be re-defined as the univariate four-state Markov chain $s_n^* \in \{1, 2, 3, 4\}$ as follows

$$\begin{aligned} s_n^* &= 1, & \text{if } s_{1,n} = 1 \text{ and } s_{2,n} = 1, \\ s_n^* &= 2, & \text{if } s_{1,n} = 2 \text{ and } s_{2,n} = 2, \\ s_n^* &= 3, & \text{if } s_{1,n} = 1 \text{ and } s_{2,n} = 2, \\ s_n^* &= 4, & \text{if } s_{1,n} = 2 \text{ and } s_{2,n} = 1. \end{aligned}$$

From equations (2.39) and (2.40), the transition matrix P with the transition probability $p_{i^*j^*}^*$ is given by

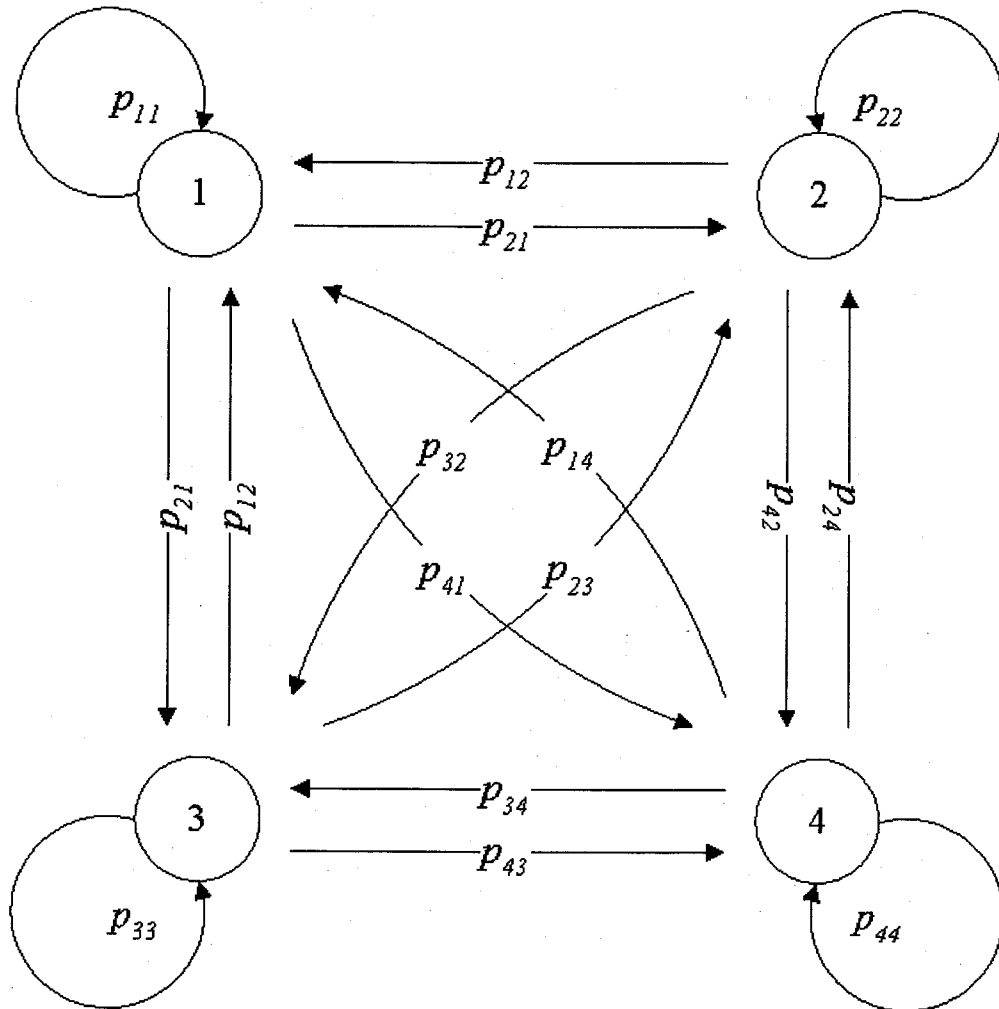
$$P = \begin{pmatrix} p_{11}^* & p_{21}^* & p_{31}^* & p_{41}^* \\ p_{12}^* & p_{22}^* & p_{32}^* & p_{42}^* \\ p_{13}^* & p_{23}^* & p_{33}^* & p_{43}^* \\ p_{14}^* & p_{24}^* & p_{34}^* & p_{44}^* \end{pmatrix}. \quad (2.41)$$

Figure 2.5 easily and visually illustrates the relationships and transition mechanism among regimes. In this Figure 2.5, circles denote four regimes for $s_n^* \in \{1, 2, 3, 4\}$, and arrows denote the transitions between two regimes, including the self-loop transitions.

Independent Transition Probability

Next we consider the multivariate Markov switching model with independent transition probability. This model is a restricted version of the original multivariate Markov switching model described in the previous subsection. The state of the observation vector $y_n \equiv (y_{1,n}, y_{2,n}, \dots, y_{k,n})^T$ depends on the Markov chain vector $s_n \equiv (s_{1,n}, s_{2,n}, \dots, s_{k,n})^T$ as well as the original unrestricted model. The difference between these two multivariate Markov switching models is in the structure of the transition probability. In the Markov

Figure 2.5: Illustration of general transition probabilities (bivariate two-state case)



a: General transition probability

switching model with independent transition probability, the transition probability can be written by multiplying together those for the independent Markov chains governing $s_{1,n}, s_{2,n}, \dots, s_{k,n}$. Therefore, the transition probability in equation (2.37) is replaced by

$$p_{i_1 \dots i_k, j_1 \dots j_k} = \prod_{l=1}^k p_{ij}^l, \quad (2.42)$$

where

$$p_{ij}^l = \Pr(s_{l,n} = j_l | s_{l,n-1} = i_l), \quad l = 1, 2, \dots, k. \quad (2.43)$$

This transition probability can be written as the $(m^k \times m^k)$ transition probability matrix (a Kronecker product of independent transition probability matrices for $l = 1, 2, \dots, k$)

$$P = P^1 \otimes P^2 \otimes \dots \otimes P^{k-1} \otimes P^k \quad (2.44)$$

$$= \begin{pmatrix} p_{11}^1 & p_{21}^1 & \dots & p_{m1}^1 \\ p_{12}^1 & p_{22}^1 & \dots & p_{m2}^1 \\ \vdots & \vdots & \ddots & \vdots \\ p_{1m}^1 & p_{2m}^1 & \dots & p_{mm}^1 \end{pmatrix} \otimes \dots \otimes \begin{pmatrix} p_{11}^k & p_{21}^k & \dots & p_{m1}^k \\ p_{12}^k & p_{22}^k & \dots & p_{m2}^k \\ \vdots & \vdots & \ddots & \vdots \\ p_{1m}^k & p_{2m}^k & \dots & p_{mm}^k \end{pmatrix}, \quad (2.45)$$

$$= \begin{pmatrix} \prod_{l=1}^k p_{11}^l & \prod_{l=1}^{k-1} p_{11}^l p_{21}^k & \dots & \prod_{l=1}^k p_{1m}^l \\ \prod_{l=1}^{k-1} p_{11}^l p_{12}^k & \prod_{l=1}^{k-1} p_{11}^l p_{22}^k & \dots & \prod_{l=1}^{k-1} p_{1m}^l p_{m2}^k \\ \vdots & \vdots & \ddots & \vdots \\ \prod_{l=1}^k p_{1m}^l & \prod_{l=1}^{k-1} p_{1m}^l p_{2m}^k & \dots & \prod_{l=1}^k p_{mm}^l \end{pmatrix}. \quad (2.46)$$

The top (b) of Figure 2.6 illustrates relationship and transition for a bivariate and two-state Markov chain $s_{l,n} \in \{1, 2\}, l = 1, 2$ with independent transition probability. Note that the definitions of Markov chains are the same as those in the general one.

Simultaneous Transition Probability

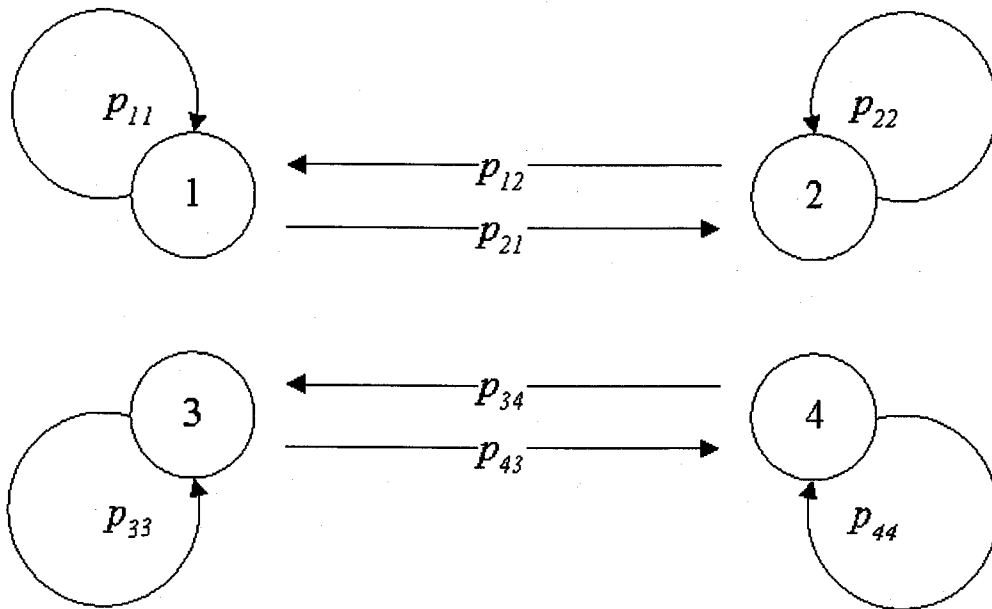
Finally, we consider another restricted multivariate Markov switching model, which includes a simultaneous transition probability. This restriction of simultaneity of state shifts is defined by

$$s_{1,n} = s_{2,n} = \dots = s_{k,n}, \quad (2.47)$$

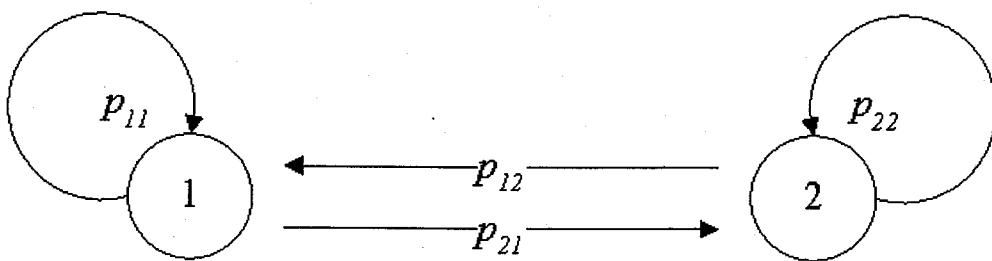
and the transition probability is denoted as follows

$$\begin{aligned} p_{ij} &= \Pr(s_n = j | s_{n-1} = i) \\ &= \Pr(s_{1,n} = s_{2,n} = \dots = s_{k,n} = j | s_{1,n-1} = s_{2,n-1} = \dots = s_{k,n-1} = i), \end{aligned} \quad (2.48)$$

Figure 2.6: Illustration of independent (b) and simultaneous (c) transition probabilities (bivariate two-state case)



b: Independent transition probability



c: Simultaneous transition probability

where $s_n = s_{l,n}$ for all $l = 1, 2, \dots, k$, and

$$\sum_{j=1}^m p_{ij} = 1, \quad (2.49)$$

for all $i = 1, 2, \dots, m$. This simultaneous transition probability can be written as the $(m \times m)$ transition probability matrix

$$P = \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{m1} \\ p_{12} & p_{22} & \cdots & p_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1m} & p_{2m} & \cdots & p_{mm} \end{pmatrix}. \quad (2.50)$$

In this multivariate Markov switching model with simultaneous transition probability, all the observed time series shift among the same state over time. The bottom (c) in Figure 2.6 illustrates relationship and transition for a bivariate and two-state Markov chain $s_{l,n} \in \{1, 2\}, l = 1, 2$ with simultaneous transition probability. Note that the definitions of Markov chains are the same as those in the general one. In simultaneous transition probability, we consider only two regimes, according to equationa (2.47) and (2.48).

2.2.4 Exogenous Time-varying Transition Probabilities

In both the univariate and multivariate Markov switching state space models described in the previous subsections, we assume that the transition probabilities are *time-homogeneous* over time. We extend the time-homogeneity transition probability of Markov switching to exogenous time-varying transition probability of Markov switching. That is, the transition probability changes depending on exogenous variables. The time-homogeneous transition probability from $s_{n-1} = i$ to $s_n = j$ is defined by

$$p_{ij} = \Pr(s_n = j | s_{n-1} = i), \quad i, j = 1, 2, \dots, m, \quad (2.51)$$

and the time-homogeneous transition probability matrix is given by

$$P = \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{m1} \\ p_{12} & p_{22} & \cdots & p_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1m} & p_{2m} & \cdots & p_{mm} \end{pmatrix}. \quad (2.52)$$

To extend the transition probability from time-homogeneity to exogenous time-varying, we can replace equation (2.51) by

$$p_{ij}(z_n) = \Pr(s_n = j | s_{n-1} = i, z_n), \quad i, j = 1, 2, \dots, m, \quad (2.53)$$

where in the case of the univariate Markov switching model $z_n \equiv \{z_{1,n}, z_{2,n}, \dots, z_{i,n}\}$ is the l -dimensional exogenous time series governing the transition probability. The transition probability matrix in equation (2.52) is replaced by

$$P(z_n) = \begin{pmatrix} p_{11}(z_n) & p_{21}(z_n) & \cdots & p_{m1}(z_n) \\ p_{12}(z_n) & p_{22}(z_n) & \cdots & p_{m2}(z_n) \\ \vdots & \vdots & \ddots & \vdots \\ p_{1m}(z_n) & p_{2m}(z_n) & \cdots & p_{mm}(z_n) \end{pmatrix}. \quad (2.54)$$

The exogenous time-varying transition probabilities may have the following logistic form

$$p_{ij}(z_n) = \Pr(s_n = j | s_{n-1} = i, z_n) = \frac{\exp(\alpha_{ij} + \sum_{k=1}^l \beta_{k,ij} z_{k,n})}{1 + \sum_{i=1}^{m-1} \exp(\alpha_{ij} + \sum_{k=1}^l \beta_{k,ij} z_{k,n})}, \quad (2.55)$$

for all $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, m-1$, and

$$p_{im}(z_n) = \Pr(s_n = m | s_{n-1} = i, z_n) = 1 - \sum_{j=1}^{m-1} p_{ij}(z_n), \quad (2.56)$$

for $i = 1, 2, \dots, m$. Note that α_{ij} and $\beta_{k,ij}$ are unknown variables which determine the effects of exogenous variables on the transition probability.

The conditional joint density-distribution of y_n , s_n and s_{n-1}

$$\begin{aligned} f(y_n, s_n = j, s_{n-1} = i | \Psi_{n-1}, z_n) &= \Pr(s_n = j | s_{n-1} = i, z_n) \Pr(s_{n-1} = i | \Psi_{n-1}, z_{n-1}) \\ &\quad \times f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1}), \end{aligned} \quad (2.57)$$

summarizes the information in the data and explicitly links the transition probabilities to the estimation method. The conditional density is

$$\begin{aligned} f(y_n | \Psi_{n-1}, z_n) &= \sum_{j=1}^m \sum_{i=1}^m f(y_n, s_n = j, s_{n-1} = i | \Psi_{n-1}, z_n) \\ &= \sum_{j=1}^m \sum_{i=1}^m f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1}) \Pr(s_n = j | s_{n-1} = i, z_n) \\ &\quad \times \Pr(s_{n-1} = i | \Psi_{n-1}, z_{n-1}) \end{aligned} \quad (2.58)$$

and the log-likelihood function is

$$\begin{aligned} L(\theta) &= \sum_{n=1}^N \log f(y_n | \Psi_{n-1}, z_n) \\ &= \sum_{n=1}^N \log \left[\sum_{j=1}^m \sum_{i=1}^m f(y_n, s_n = j, s_{n-1} = i | \Psi_{n-1}, z_n) \right] \\ &= \sum_{n=1}^N \log \left[\sum_{j=1}^m \sum_{i=1}^m f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1}) \Pr(s_n = j | s_{n-1} = i, z_n) \right. \\ &\quad \left. \times \Pr(s_{n-1} = i | \Psi_{n-1}, z_{n-1}) \right]. \end{aligned} \quad (2.59)$$

The inferred probability of the state at time n can be calculated by integrating out the effects of the past states in the joint density-distribution as follows

$$\begin{aligned} \Pr(s_n = j | \Psi_n, z_n) &= \sum_{i=1}^m \Pr(s_n = j, s_{n-1} = i | \Psi_n, z_n) \\ &= \sum_{i=1}^m \frac{f(y_n, s_n = j, s_{n-1} = i | \Psi_{n-1}, z_n)}{f(y_n | \Psi_{n-1}, z_n)}. \end{aligned} \quad (2.60)$$

Time-varying transition probabilities also imply time-varying expected duration of a regime. A closely related issue in the business cycle literature is duration dependence, or whether the probability of a transition between regimes depends on the length of time the economy has been in a recession or boom. It is a special form of time-varying transition probabilities in a regime-switching model of the business cycle, in which z_{n-1} is replaced by length to date of the current regime (boom or recession). Durland and McCurdy (1994), for example, examine the nature of business cycle duration dependence within Hamilton's (1989) univariate Markov switching model of the business cycle. Kim and Nelson (1998) provide a Bayesian analysis of business cycle duration dependence based on a dynamic factor model with Markov switching.

2.2.5 Switching ARCH Model

By incorporating the Markov switching structure into the autoregressive conditionally heteroskedastic (ARCH) model, Hamilton and Susmel (1994) propose a new ARCH model, the switching ARCH or SWARCH model. This model is superior to some simple ARCH and GARCH models and their modifications because it captures more realistically the economic and financial time series properties (mainly volatility), including dramatic structural change.

Let y_n be the first-order difference (or logarithmic difference) of observed time series. We consider the univariate two-state SWARCH process for y_n as follows

$$y_n = \mu + \varepsilon_n, \quad \varepsilon_n | \Psi_{n-1} \sim N(0, h_n), \quad (2.61)$$

$$\frac{h_n}{g_{s_n}} = \alpha_0 + \sum_{i=1}^q \alpha_i \frac{\varepsilon_{n-i}^2}{g_{s_{n-i}}}, \quad (2.62)$$

where g is a scale parameter that captures the change in regime, and determines depending on regime

$$\begin{aligned} \text{if } s_n = 1, & \quad g_1 = 1, \\ \text{if } s_n = 2, & \quad g_2 \text{ takes a constant variable.} \end{aligned}$$

Here, the unobserved variable s_n takes a value of 1 or 2, and its evolution depends on the first-order and

two-state Markov switching process as follows

$$\Pr(s_n = 1 | s_{n-1} = 1) = p_{11},$$

$$\Pr(s_n = 2 | s_{n-1} = 1) = p_{12},$$

$$\Pr(s_n = 1 | s_{n-1} = 2) = p_{21},$$

$$\Pr(s_n = 2 | s_{n-1} = 2) = p_{22},$$

where p_{ij} is the transition probability that the process switches at time n from state i to state j . The log likelihood for the observed data of the SWARCH model can be calculated as follows

$$\begin{aligned} L(\theta) &= \sum_{n=1}^N \log f(y_n | \Psi_{n-1}). \\ &= \sum_{n=1}^N \log \left[\sum_{j=1}^m \Pr(y_n, s_n = j, s_{n-1} = i | \Psi_{n-1}) \right] \\ &= \sum_{n=1}^N \log \left[\sum_{j=1}^m \sum_{i=1}^m \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}) \right. \\ &\quad \left. \times \frac{1}{\sqrt{2\pi h_n}} \exp \left\{ -\frac{(y_n - \mu)^2}{2h_n} \right\} \right]. \end{aligned} \quad (2.63)$$

where θ denotes the unknown parameters. The maximum likelihood estimation of θ is obtained by maximizing equation (2.63). Here, since the Markov chain s_n is not observed, we can obtain the $\Pr(s_{n-1} | \Phi_{n-1})$ in equation (2.63) by using the method of equations (2.27) and (2.28) of the univariate Markov switching stochastic trend model.

2.3 State Space Representation of the Markov Switching Model

State space models have a wide range of potential applications in econometrics to deal with dynamic time series models. The univariate and multivariate Markov switching models also can be represented by the state space form. In this section, we briefly review linear Gaussian sum, nonlinear and general state space modeling for the Markov switching model.

Note that the state space representations for the modified and extended Markov switching models described in the previous sections are discussed in detail in Chapter 4.

2.3.1 Linear Gaussian State Space Modeling

Shumway and Stoffer (1991) and Kim (1993) incorporated a Markov switching model into the state space model. In this subsection, we briefly review their Markov switching state space model. Let y_n be the l -dimensional observed time series at time n . The ordinary Gaussian state space representation of a

Markov switching model is given by:

$$x_n = \mu_{s_{1,n}} + F_{s_{2,n}} x_{n-1} + G w_n, \quad (2.64)$$

$$y_n = H x_n + v_n, \quad (2.65)$$

and

$$\begin{pmatrix} w_n \\ \varepsilon_n \end{pmatrix} = N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R_{s_{3,n}} & 0 \\ 0 & Q_{s_{4,n}} \end{bmatrix} \right), \quad (2.66)$$

where x_n and $\mu_{s_{1,n}}$ are the m -dimensional state and drift vectors at time n , respectively. v_n is a k -dimensional system noise or state noise with zero mean and variance-covariance matrix Q_{s_n} . w_n is a l -dimensional observation noise with zero mean and variance-covariance matrix $R_{s_{3,n}}$. $F_{s_{2,n}}$, G and H are $m \times m$, $m \times k$ and $l \times m$ matrices, respectively. Also for convenience, we assume that $E(w_n, v_m) = 0$, for all n and m . Equations (2.64) and (2.65) are the system and observation models, respectively.

Here, $F_{s_{1,n}}$, H , $R_{s_{3,n}}$ and $Q_{s_{4,n}}$ are all dependent on discrete-valued k -state Markov chains $s_{1,n}$, $s_{2,n}$, $s_{3,n}$ and $s_{4,n} \in \{1, 2, 3, \dots, k\}$, respectively. The elements of $F_{s_{2,n}}$, H and values of $R_{s_{3,n}}$ and $Q_{s_{4,n}}$ can be specified as

$$f_{s_{1,n}}^{(a,b)} = f_1^{(a,b)} s_{1,1,n} + \dots + f_k^{(a,b)} s_{1,k,n}, \quad (2.67)$$

$$h_{s_{2,n}}^{(a,b)} = h_1^{(a,b)} s_{2,1,n} + \dots + h_k^{(a,b)} s_{2,k,n}, \quad (2.68)$$

$$r_{s_{3,n}}^{(a,a)} = r_1^{(a,a)} s_{3,1,n} + \dots + r_k^{(a,a)} s_{3,k,n}, \quad (2.69)$$

$$q_{s_{4,n}}^{(a,a)} = q_1^{(a,a)} s_{4,1,n} + \dots + q_k^{(a,a)} s_{4,k,n}. \quad (2.70)$$

where $f_k^{(a,b)}$, $h_k^{(a,b)}$, $r_k^{(a,a)}$ and $q_k^{(a,a)}$ are the elements of matrices $F_{s_{1,n}}$, H , $R_{s_{3,n}}$ and $Q_{s_{4,n}}$, respectively. $s_{k,n}$, $k = 1, 2, 3, 4$ takes the value 1 when $s_{k,n}$ is equal to k and 0 otherwise. $s_{1,n}$, $s_{2,n}$, $s_{3,n}$ and $s_{4,n}$ follow k -state and first-order Markov processes.

2.3.2 Nonlinear Non-Gaussian State Space Modeling

The Markov switching model can be represented by using a nonlinear non-Gaussian state space model (Kitagawa 1987) as follows

$$x_n = f_{s_{1,n}}(x_{n-1}, v_n) \quad (2.71)$$

$$y_n = h_{s_{2,n}}(x_n, w_n), \quad (2.72)$$

where x_n is an unknown state vector, and v_n and w_n are the system noise and the observation noise, respectively. Equations (2.71) and (2.72) are called the system model and the observation model, respectively. The initial state x_0 is assumed to be distributed according to the density $p_0(x)$. $f_{s_{1,n}}(x, v)$ and $h_{s_{2,n}}(x, w)$ are possibly nonlinear functions of the state and the noise inputs with different functions and

moments, depending on the k -state and first-order Markov chains $s_{1,n}$ and $s_{1,n} \in \{1, 2, \dots, k\}$, respectively. Also, densities $q_{s_{3,n}}(v)$ and $r_{s_{4,n}}(w)$ of v_n and w_n are dependent on the k -state and first-order Markov chains $s_{3,n}$ and $s_{4,n} \in \{1, 2, \dots, k\}$, respectively.

2.3.3 General State Space Modeling

As a generalization, we consider a general state space model (Kitagawa 1996) for the Markov switching model as follows

$$x_n \sim q(x_n | x_{n-1}) \quad (2.73)$$

$$y_n \sim r(y_n | x_n), \quad (2.74)$$

where y_n is the observed time series and x_n is the unknown state vector. q and r are conditional distributions of x_n given x_{n-1} and of y_n given x_n , respectively. The initial state vector x_0 is distributed according to the distribution $p(x_0 | Y_0)$. Markov switching models can be expressed in this general state space model for (Kitagawa 1996) as follows

$$x_n \sim q(x_n, s_n | x_{n-1}, s_{n-1}) \quad (2.75)$$

$$y_n \sim r(y_n | x_n, s_n). \quad (2.76)$$

Here, $q(x_n, s_n | x_{n-1}, s_{n-1})$ is the conditional distribution, and $r(y_n | x_n, s_n)$ is the conditional density of the observation given the state.

Chapter 3

State Estimation and Model Identification

3.1 Non-Gaussian Filter and Smoother

For the standard linear-Gaussian state space model, predictive, filter and smoother densities can be expressed by Gaussian densities and their mean vectors and the variance-covariance matrices by using the well-known Kalman filter algorithm (Kalman 1960). However, for the nonlinear and non-Gaussian state space models including Markov switching structure, the predictive, filter and smoother distributions cannot be completely specified by the mean vectors and the variance-covariance matrices, since the conditional distributions become non-Gaussian. Here, the filter and smoother of the nonlinear and non-Gaussian state space model with Markov switching (Kitagawa 1994, Kitagawa and Hakamta 2001) are respectively represented as follows:

[Non-Gaussian One-step-ahead Prediction]

$$\begin{aligned} p(x_n, s_n = j | \Psi_{n-1}) &= \sum_{i=1}^m \int p(x_{n-1}, s_{n-1} = i | \Psi_{n-1}) \Pr(s_n = j | s_{n-1} = i) \\ &= \sum_{i=1}^m \int p(x_n | x_{n-1}, s_n = j, s_{n-1} = i) \\ &\quad \times p(x_{n-1}, s_{n-1} = i | \Psi_{n-1}) \Pr(s_n = j | s_{n-1}) dx_{n-1}. \end{aligned} \tag{3.1}$$

[Non-Gaussian Filter]

$$\begin{aligned}
p(x_n, s_n = j | \Psi_n) &= p(x_n, s_n = j | y_n, \Psi_{n-1}) \\
&= \frac{p(y_n | x_n, \Psi_{n-1}) p(x_n, s_n = j | \Psi_{n-1})}{p(y_n | \Psi_{n-1})}.
\end{aligned} \tag{3.2}$$

where the predictive distribution of y_n is evaluated by

$$p(y_n | \Psi_{n-1}) = \int p(y_n | x_n) p(x_n | \Psi_{n-1}) dx_n. \tag{3.3}$$

[Non-Gaussian Smoother]

Using the results of the non-Gaussian filter, the smoothed density $p(x_n | \Psi_N)$ is obtained by

$$\begin{aligned}
p(x_n, s_n = j | \Psi_N) &= \sum_{j=1}^m \int p(x_n, x_{n+1}, s_n = j, s_{n+1} = k | \Psi_N) dx_{n+1} \\
&= \sum_{j=1}^m \int P(x_{n+1}, s_{n+1} = k | \Psi_N) p(x_n | x_{n+1}, s_n = j, s_{n+1} = k, \Psi_N) \\
&\quad \times \Pr(s_{n+1} = k | s_n = j) dx_{n+1} \\
&= p(x_n, s_n = j | \Psi_n) \sum_{j=1}^m \int \frac{p(x_{n+1}, s_{n+1} = k | \Psi_N) p(x_{n+1} | x_n, s_{n+1} = k, s_n = j)}{p(x_{n+1}, s_{n+1} = k | \Psi_n)} \\
&\quad \times \Pr(s_{n+1} = k | s_n = j) dx_{n+1}.
\end{aligned} \tag{3.4}$$

The recursive filter and smoother for the two dimensional state space model can be implemented by using numerical integration based on step function approximation of the related distribution. The marginal conditional distributions of x_n and s_n are obtained by integrating $p(x_n | \Psi_l) = p(x_n, s_n = j | \Psi_l)$, i.e.,

$$p(x_n | \Psi_l) = \sum_{j=1}^m \int p(x_n, s_n = j | \Psi_l) dx_n, \tag{3.5}$$

$$p(s_n = j | \Psi_l) = \int p(x_n, s_n = j | \Psi_l) dx_n. \tag{3.6}$$

The complete implementation of the nonlinear and non-Gaussian filter needs massive computational power. Kitagawa (1987) and Kramer and Sorenson (1988) proposed approximating the densities numerically by a piecewise linear function. In the multi-dimensional case, however, even if we use such an approximation, the computational burden becomes an impediment to applying empirical data analysis.

3.2 Gaussian Sum Filter

Some approximating methods to mitigate this computational burden of the nonlinear and non-Gaussian case are proposed by Gelb (1974), Jazwinski (1996), and Anderson and Moore (1979). For the Markov switching model, the application of a Gaussian-sum filter (Sorenson and Alspach 1971, Alspach and Sorenson 1972, Harrison and Stevens 1976, Anderson and Moore 1979) is one of the most practical

ways. Harrison and Stevens (1976) show the filter algorithm for a mixture model. Shumway and Stoffer (1992) introduce the state space dynamic switching model. Kim (1994) extend Hamilton's (1989) Markov switching model to the standard state-space form, and propose the recursive filter and the modified smoother. First, we describe this recursive filter algorithm for the state vector x_n and the discrete-valued unobserved variable s_n as follows:

Step 1: Filter for the state vector x_n

[One step ahead prediction]

$$\begin{aligned} x_{n|n-1}^{(ji)} &= \mu_{s_n} + Fx_{n-1|n-1}^{(i)}, \\ v_{n|n-1}^{(ji)} &= Fv_{n-1|n-1}^{(i)}F^T + GQG^T, \end{aligned}$$

where

$$\begin{aligned} x_{n|n-1}^{(ji)} &= E(x_n | \Psi_{n-1}, s_n = j, s_{n-1} = i), \\ v_{n|n-1}^{(ji)} &= E[(x_n - x_{n|n-1})(x_n - x_{n|n-1})^T | \Psi_{n-1}, s_n = j, s_{n-1} = i], \end{aligned}$$

and Ψ_{n-1} is an information available at time $n-1$.

[Filtering]

$$\begin{aligned} K_n^{(ji)} &= v_{n|n-1}^{(ji)}H^T(Hv_{n|n-1}^{(ji)}H^T + R)^{-1}, \\ x_{n|n}^{(ji)} &= x_{n|n-1}^{(ji)} + K_n^{(ji)}(y_n - Hx_{n|n-1}^{(ji)}), \\ v_{n|n}^{(ji)} &= (I - K_n^{(ji)}H)v_{n|n-1}^{(ji)}, \end{aligned}$$

where

$$\begin{aligned} x_{n|n}^{(ji)} &= E[x_n | \Psi_n, s_n = j, s_{n-1} = i], \\ v_{n|n}^{(ji)} &= E[(x_n - x_{n|n-1})(x_n - x_{n|n-1})^T | \Psi_n, s_n = j, s_{n-1} = i]. \end{aligned}$$

Step 2: Filter for the unobserved variable s_n

It is possible to represent the filter for the unobserved variable s_n as the special discrete version of the non-Gaussian filter (Kitagawa 1987, 1994). Hamilton (1989) introduced it into estimations of the Markov switching model, and this is well-known as the Hamilton's filter.

[One step ahead prediction]

$$\begin{aligned} \Pr(s_n = j | \Psi_{n-1}) &= \sum_{j=1}^m \Pr(s_n = j, s_{n-1} = i | \Psi_{n-1}) \\ &= \sum_{j=0}^m \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}), \end{aligned}$$

[Filtering]

$$\begin{aligned}\Pr(s_n = j | \Psi_{n-1}) &= \sum_{i=1}^m \Pr(s_n = j, s_{n-1} = i | \Psi_{n-1}) \\ &= \sum_{i=1}^m \frac{f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1}) \Pr(s_n = j, s_{n-1} = i | \Psi_{n-1})}{f(y_n | \Psi_{n-1})}.\end{aligned}$$

Note that $f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1})$ and $f(y_n | \Psi_{n-1})$ denote the following:

$$\begin{aligned}f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1}) \\ = \frac{1}{(2\pi)^{\frac{k}{2}} (r_n)^{\frac{1}{2}}} \exp \left\{ -\frac{(y_n - Hx_{n|n-1}^{(ji)})^T (y_n - Hx_{n|n-1}^{(ji)})}{2r_n} \right\}\end{aligned}$$

with $r_n = H v_{n|n-1}^{(ji)} H^T + r^{(i)}$, and

$$f(y_n | \Psi_{n-1}) = \sum_{j=1}^m \sum_{i=1}^m f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1}) \Pr(s_n = j, s_{n-1} = i | \Psi_{n-1}).$$

Step 3: Collapsing

To avoid an explosion in the number of Gaussian components, the approximation techniques, proposed by Harrison and Stevens (1976), Hamilton (1989), Kitagawa (1994) and Kim (1994), are used:

$$x_{n|n}^{(j)} = \frac{\sum_{i=1}^m \Pr(s_{n-1} = i, s_n = j | \Psi_n) x_{n|n}^{(ji)}}{\Pr(s_n = j | \Psi_n)}, \quad (3.7)$$

$$v_{n|n}^{(j)} = \frac{\sum_{i=1}^m \Pr(s_{n-1} = i, s_n = j | \Psi_n) \left\{ v_{n|n}^{(ji)} + (x_{n|n}^{(j)} - x_{n|n}^{(ji)})(x_{n|n}^{(j)} - x_{n|n}^{(ji)})^T \right\}}{\Pr(s_n = j | \Psi_n)}. \quad (3.8)$$

To avoid an explosion in the number of Gaussian components, we re-approximate the densities by a reduced number of Gaussian components at each time step, (Harrison and Stevens 1976). The re-approximation is motivated by the observation that a relatively small number of Gaussian densities can reasonably approximate a large class of distributions and by the expectation that the complexity of the density will not increase very significantly with the evolution of the time step.

Harrison and Stevens (1976) and Kitagawa (1994) proposed a precise measurement to evaluation approximation by using the Kullback-Leibler information number. General representation of collapsing method in equations (3.7) and (3.8) to approximate the filtering densities by a fixed number of components at each step of the recursion is given by:

$$f(x_n | \Psi_n) \rightarrow \sum_{l=1}^L \Pr(\cdot | \Psi_n) N(x_{n|n}^{(j)}, v_{n|n}^{(j)}). \quad (3.9)$$

In principle, this collapsing should be realized by finding the minimizer of a criterion for the dissimilarity of the true and approximated densities. They exploit the Kullback-Leibler information number as follows

$$I(f^*) = \int \log \frac{f(x)}{f^*(x)} f(x) dx, \quad (3.10)$$

with $f(x)$ and $f^*(x)$ the true and the approximated densities, respectively. If $I(f^*) = 0$, this approximated densities are completely in consistency with the true ones. Closer to zero $I(f^*)$ is, more precise this approximation is.

Kitagawa (1994) shows illustrative examples of this Gaussian sum approximation for the Gaussian mixture model and its evaluation by the Kullback=Leibler information. In addition, Kim (1994) show that enough efficient approximation can be obtained by using these collapsing methods concerning the Markov switching model and the mixtured model, respectively.

Step 4: Iterate step 1 to 3 for $n = 1, 2, \dots, N$

As a by-product of iterating these filter algorithms from set 1 to set 3, we can easily obtain the approximate log likelihood function for unknown parameter vector $\theta \equiv \{\mu_{k,0}, \mu_{k,1}, \tau_k, \sigma_k, p_{00}, p_{11}; k = 1, 2, \dots, K\}$ by:

$$\begin{aligned} L(\theta) &= \sum_{n=1}^N \log f(y_n | \Psi_{n-1}) \\ &= \sum_{n=1}^N \log \left[\sum_{i=1}^m \sum_{j=1}^m f(y_n, s_n = j, s_{n-1} = i | \Psi_{n-1}) \right] \\ &= \sum_{n=1}^N \log \left[\sum_{i=1}^m \sum_{j=1}^m f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1}) \Pr(s_n = j, s_{n-1} = i | \Psi_{n-1}) \right]. \end{aligned}$$

3.3 Gaussian Sum Smoother

3.3.1 Kitagawa's two-filter Formula and Gaussian Sum Smoother

In the Gaussian sum smoothing algorithm using the two-filter formula proposed by Kitagawa (1994), the smoothed density $p(x_n | \Psi_N)$ can be expressed by

$$p(x_n | \Psi_N) = p(x_n | \Psi_{n-1}) p(\Psi^n | x_n) p(\Psi^n | \Psi_{n-1})^{-1}. \quad (3.11)$$

Here Ψ^n is the information from present and future observations. From (6.21), obtaining the smoothed density requires $p(\Psi^n | x_n)$. This term is evaluated by the indicated backward filtering operation.

[Initialization]

$$p(\Psi^N | x_N) = p(y_N | x_N). \quad (3.12)$$

[Backward Filtering]

$$p(\Psi^{n+1} | x_n) = \int_{-\infty}^{\infty} p(\Psi^{n+1} | x_{n+1}) p(x_{n+1} | x_n) dx_{n+1}, \quad (3.13)$$

$$p(\Psi^n | x_n) = p(\Psi^{n+1} | x_n) p(y_n | x_n). \quad (3.14)$$

If we assume that $p(x_n|\Psi_{n-1})$ and $p(\Psi^n|x_n)$ are expressed by

$$p(x_n|\Psi_{n-1}) = \sum_{j=1}^M \Pr(s_n = j) \varphi_j(x_n|\Psi_{n-1}), \quad (3.15)$$

$$p(\Psi^n|x_n) = \sum_{k=1}^M \Pr(s_{n+1} = k) \varphi_k(\Psi^n|x_n), \quad (3.16)$$

where $\varphi_j(x_n|\Psi_{n-1}) \equiv N(x_{n|n-1}^{(j)}, V_{n|n-1}^j)$ and $\varphi_k(\Psi^n|x_n) \equiv N(z_{n|n}^{(k)}, U_{n|n}^k)$, the Gaussian-sum smoother is obtained by

$$\begin{aligned} p(x_n|\Psi_N) &\propto p(\Psi^n|x_n)p(x_n|\Psi_{n-1}) \\ &= \sum_{k=1}^M \sum_{j=1}^M \Pr(s_n = j, s_{n+1} = k|\Psi_N) \varphi_k(\Psi^n|x_n) \varphi_j(x_n|\Psi_{n-1}) \\ &\equiv \sum_{k=1}^M \sum_{j=1}^M \Pr(s_n = j, s_{n+1} = k|\Psi_N) \varphi_k(x_n|\Psi_N). \end{aligned} \quad (3.17)$$

Here $\varphi_k(x_n|\Psi_N)$ is the Gaussian density with mean $x_{n|N}^{(kj)}$ and covariance $V_{n|N}^{(kj)}$ as follows:

$$J_n^{(kj)} = V_{n|n-1}^{(j)} (V_{n|n-1}^{(j)} + U_{n|n}^{(k)})^{-1}, \quad (3.18)$$

$$x_{n|N}^{(kj)} = x_{n|n-1}^{(j)} + J_n^{(kj)} (z_{n|n}^{(k)} - x_{n|n-1}^{(j)}), \quad (3.19)$$

$$V_{n|N}^{(kj)} = (I - J_n^{(kj)}) V_{n|n-1}^{(j)}. \quad (3.20)$$

3.3.2 Kim's Smoother

Kim (1994) proposed another smoothing algorithm for a Gaussian sum approximation of the Markov switching model. In this subsection, we briefly review Kim's smoother.

Step 1: Smoothing for the state vector x_n :

$$A_n^{(j,k)} = v_{n|n}^{(j,k)} F^T v_{n+1|n}^{(j,k)-1}, \quad (3.21)$$

$$x_{n|N}^{(j,k)} = x_{n|n}^{(j,k)} + A_n^{(j,k)} (x_{n+1|N}^{(j,k)} - x_{n+1|n}^{(j,k)}), \quad (3.22)$$

$$v_{n|N}^{(j,k)} = v_{n|n}^{(j,k)} + A_n^{(j,k)} (v_{n+1|N}^{(j,k)} - v_{n+1|n}^{(j,k)}) A_n^{(j,k)T}, \quad (3.23)$$

and

$$x_{n|N} = \sum_{j=1}^m \sum_{k=1}^m \Pr(s_n = j, s_{n+1} = k|\Psi_N) x_{n|N}^{(j,k)}, \quad (3.24)$$

where

$$x_{n|N}^{(j,k)} = E(x_n|\Psi_N, s_{n+1} = k, s_n = j), \quad (3.25)$$

$$v_{n|N}^{(j,k)} = E[(x_n - x_{n|N})(x_n - x_{n|N})^T|\Psi_N, s_{n+1} = k, s_n = j]. \quad (3.26)$$

Step 2: Smoothing for the unobserved variable s_n :

$$\Pr(s_n = j, s_{n+1} = k | \Psi_N) = \Pr(s_n = j | \Psi_n) \frac{\Pr(s_{n+1} = k) \Pr(s_{n+1} = k | s_n = j)}{\Pr(s_{n+1} = k | \Psi_n)} \quad (3.27)$$

and

$$\Pr(s_n = j | \Psi_N) = \Pr(s_n = j | \Psi_n) \sum_{k=1}^m \frac{\Pr(s_{n+1} = k | \Psi_N) \Pr(s_{n+1} = k | s_n = j)}{\Pr(s_{n+1} = k | \Psi_n)} \quad (3.28)$$

where $j, k = 1, 2, \dots, m$ and $\Psi_N \equiv \{y_1, y_2, \dots, y_N\}$.

3.4 Monte Carlo Filter and Smoother

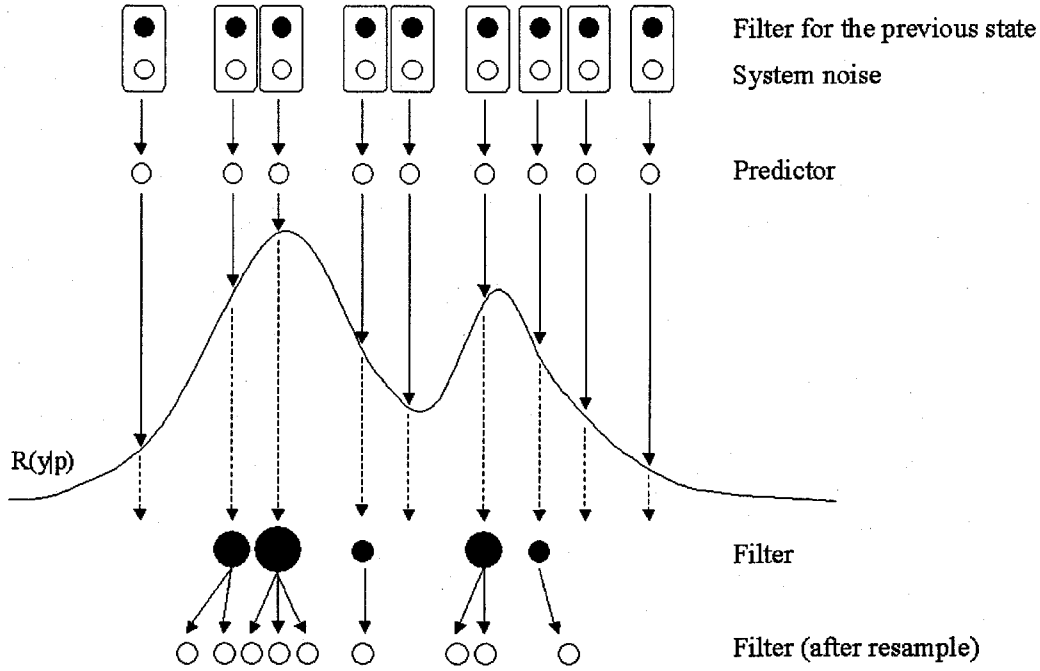
Some filtering methods based on the sequential Monte Carlo approach (for example, Gordon *et al.* 1993, Carlin *et al.*, Carter and Kohn 1994, Kitagawa 1996) have recently been proposed for nonlinear and non-Gaussian models. The filtering and smoothing formula for the Markov switching model can easily be implemented by the sequential Monte Carlo method (Kitagawa 1996). In this method, we approximate each distribution by many “particles”, which can be considered as realizations from that distribution. Specifically, assume that each distribution is expressed by using m “particles” as follows:

$\{v_{i,n}^{(1)}, \dots, v_n^{(m)}\}$	\sim	$p(v_n)$	System noise
$\{p_n^{(1)}, \dots, p_n^{(m)}\}$	\sim	$p(x_n Y_{n-1})$	Predictor
$\{f_n^{(1)}, \dots, f_n^{(m)}\}$	\sim	$p(x_n Y_n)$	Filter
$\{s_{n N}^{(1)}, \dots, s_{n N}^{(m)}\}$	\sim	$p(x_n Y_N)$	Smoother

That is, we approximate the distributions with the empirical distributions determined by the m particles. Then it can be shown that a set of realizations expressing the one step ahead predictor $p(x_n, S_n | Y_{n-1}, z_n)$ and the filter $p(x_n, S_n | Y_n, z_n)$ can be obtained recursively as follows:

[The Monte Carlo Filter]

1. Approximate the initial distribution by $f_0^{(j)} \sim p_0(x)$ for $j = 1, \dots, m$.
2. Repeat the following steps for $n = 1, \dots, N$.
 - (a) Generate system noise $v_{0,n}^{(j)} \sim p(v_0)$ and $v_{1,n}^{(j)} \sim p(v_1)$ for $j = 1, \dots, m$.
 - (b) Generate a latent Markov chain $S_n^{(j)} \in \{0, 1\}$ given S_{n-1} and z_n , for $j = 1, \dots, m$.
 - (c) Approximate the predictive distribution by $p_n^{(j)} = Q(\cdot | f_{n-1}^{(j)})$ and $S_n^{(j)}$, for $j = 1, \dots, m$.
 - (d) Compute the Bayes importance weight by $\alpha_n^{(j)} = R(y_n | p_n^{(j)})$ for $j = 1, \dots, m$.

Figure 3.1: One cycle of the Monte Carlo filter ($m = 9$).

- (e) Approximate the filter distribution, generating $f_n^{(j)}$, $j = 1, \dots, m$ by the re-sampling of $p_n^{(1)}, \dots, p_n^{(m)}$ with weight proportional to $\alpha_n^{(j)}$, $j = 1, \dots, m$.

Figure 3.1 illustrates one cycle of the Monte Carlo filtering with $m = 9$ for simplicity. The particle approximating the predictor is generated from a pair of particles approximating the previous filtered state and the system noise. Then the predictive distribution is approximated by the empirical distribution $m^{-1} \sum_{j=1}^m I(x; p_n^{(j)})$. In the filtering step, the filter distribution is obtained by changing these equal weights to ones proportional to $\alpha_n^{(j)}$. Finally, by re-sampling (sampling with replacement), the filter distribution is re-approximated by using the particles $f_n^{(j)}$ with equal weights, $1/m$.

An algorithm for smoothing (Kitagawa 1996) is obtained by replacing Step 2 (e) of the filtering algorithm by

- (e-L) For fixed L , generate $\{(s_{n-L|n}^{(j)}, \dots, s_{n-1|n}^{(j)}, s_{n|n}^{(j)})^T, j = 1, \dots, m\}$ by the re-sampling of $\{(s_{n-L|n-1}^{(j)}, \dots, s_{n-1|n-1}^{(j)}, p_n^{(j)})^T, j = 1, \dots, m\}$ with $f_n^{(j)} = s_{n|n}^{(j)}$.

This is equivalent to applying the L -lag fixed lag smoother. Increasing the lag, L , will improve the accuracy of the $p(x_n, S_n | Y_{n+L})$ as an approximation to $p(x_n, S_n | Y_N)$, however it is very likely to decrease the accuracy of $\{s_{n|N}^{(1)}, \dots, s_{n|N}^{(m)}\}$ as representatives of $p(x_n, S_n | Y_{n+L})$. Because $p(x_n, S_n | Y_{n+L})$ usually converges rather quickly to $p(x_n, S_n | Y_N)$ as L increases, it is recommended to use a smaller L .

The conditional density can be approximated by

$$p(y_n|Y_{n-1}, z_n; \theta) \approx \frac{1}{m} \sum_{j=1}^m \alpha_n^{(j)}, \quad (3.29)$$

where $\alpha_n^{(j)}$ is the importance weight of the j -th particle obtained in the Monte Carlo filter. In this case, because the likelihood contains the sampling error due to the approximation, it is difficult to obtain precise maximum likelihood estimates of the parameters. This problem will be considered later in the next section.

3.5 Parameter Estimation

The Markov switching models described in the previous sections contain unknown parameters. The vector that consists of such unknown parameters is denoted by θ . The likelihood of the Markov switching models specified by the parameter vector θ is obtained by

$$L(\theta) = f(y_1, y_2, \dots, y_N | \theta) = \prod_{n=1}^N f(y_n | \Psi_{n-1}, \theta), \quad (3.30)$$

where $f(y_n | \Psi_{n-1}, \theta)$ is the conditional density of y_n given Ψ_{n-1} . As a by-product of the filter, we get this conditional density. Then the approximate log likelihood function is given by

$$LL(\theta) = \log [f(y_1, y_2, \dots, y_N)] = \sum_{n=1}^N \log [f(y_n | \Psi_{n-1})]. \quad (3.31)$$

To estimate the parameters of the model, we use a nonlinear optimization procedure (in GAUSS and FORTRAN) to maximize the approximate log likelihood function with respect to the underlying unknown parameters.

Many recent papers propose other estimating techniques of Markov switching model: Expectation maximization (EM) algorithm and Markov Chain Monte Carlo (MCMC). EM algorithm (Dempster *et al.* 1977) has been widely used to estimate parameters for Hidden Markov models (HMMs) in the engineering field. The well-known Baum-Welch algorithm (Baum *et al.* 1970) is the relevant version of the EM algorithm. Hamilton (1989, 1990, 1994) applies it to a Markov switching model in econometric fields. Applications of MCMC methods to inference of Markov switching models are given by Albert and Chib (1993) and Chib (1996). Applications of MCMC to a Markov switching state space model appear in Engel and Kim (1996). Fruhwirth-Schnatter (1998) argued problems of ordinary MCMC methods, and proposed a new MCMC method called permutation sampling.

3.6 Model Selection by AIC

3.6.1 Akaike Information Criterion

The maximum log-likelihood method can be used to estimate the values of parameters. However, it cannot be used to compare different models without some corrections. Using $N^{-1}l(\hat{\theta})$ as the estimation of $E_Y \log f(Y|\hat{\theta})$, the expectation value of bias is

$$C \equiv E \left\{ E_Y \log f(Y|\hat{\theta}) - \frac{1}{N} \sum_{n=1}^N \log f(y_n|\hat{\theta}) \right\} \quad (3.32)$$

$$\approx \frac{K}{N}. \quad (3.33)$$

Here, an approximate correction of the bias is reflected in the definition of the Akaike information criterion (AIC) given below.

$$\begin{aligned} \text{AIC}(m) &= -2(\text{maximized log likelihood of the model}) \\ &\quad + 2(\text{number of estimated parameters in the model}) \\ &= -2 \sum_{n=1}^N \log f_m(y_n|\hat{\theta}_m) + 2|\hat{\theta}_m|, \end{aligned} \quad (3.34)$$

where $|\hat{\theta}_m|$ denotes the dimension of the vector $\hat{\theta}_m$.

3.6.2 AIC for Model Selection of the Markov Switching Model

Identifying an adequate Markov switching model for a given dataset involves selecting many parameters. The most difficult problem may be the specification of the number of regimes. The available methods of model selection and identification for Markov switching models are primarily:

- (1) Information criteria such as AIC and BIC as in Leroux (1992) and Ryden (1995);
- (2) Hypothesis testing based on comparisons of likelihood of the competing models, e.g., Hansen (1992, 1996), Gong and Mariano (1995), and Hamilton (1996).

This section describes the application of the AIC to the univariate and multivariate Markov switching models. Let the univariate Markov switching model be defined by

$$y_n \sim N(\mu_{s_n}, \sigma_{s_n}^2). \quad (3.35)$$

s_n is the k -state Markov chain with state space $S \in \{1, 2, \dots, k\}$. μ_{s_n} and $\sigma_{s_n}^2$ take k different values according to s_n , respectively. The conditional joint probability density function for this model is

$$\begin{aligned} f(y_1, \dots, y_N | s_n = j) &= \eta_j \prod_{n=1}^N f(y_n | \theta) \\ &= \prod_{n=1}^N \eta_j \left(\frac{1}{2\pi\sigma_{s_n}^2} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_{s_n}^2} (y_n - \mu_{s_n})^2 \right\}, \end{aligned} \quad (3.36)$$

where $\theta \equiv \{\mu_1, \mu_2, \dots, \mu_k, \sigma_1^2, \sigma_2^2, \dots, \sigma_k^2, p_{11}, p_{12}, \dots, p_{1k-1}, \dots, p_{kk-1}\}$ is the unknown parameter vector, and η_j is the unconditional probability of the state $s_n = j$. Note that $m = 3k + k^2 - k + 1$ is the number of unknown parameters. The corresponding log-likelihood function is

$$\begin{aligned} L_j(\theta|s_n = j) &= \sum_{n=1}^N \log f(y_n|\theta, s_n = j) \\ &= -\eta_j \left[\frac{N}{2} \log 2\pi\sigma_{s_n}^2 + \frac{1}{2\sigma_{s_n}^2} \sum_{n=1}^N \left\{ -\frac{1}{2\sigma_{s_n}^2} (y_n - \mu_{s_n})^2 \right\} \right]. \end{aligned} \quad (3.37)$$

Given the unknown parameter vector θ , the AIC of a k -state Markov switching model is

$$\text{AIC}(m) = \sum_{j=1}^k \eta_j \left\{ 2 \log L_j(\theta|s_n = j) \right\} + 2(3k + k^2 - k + 1). \quad (3.38)$$

3.7 Numerical Examples

In this section, we provide numerical examples to evaluate how precisely the Markov switching model can capture the unknown regime shifts and corresponding stochastic process. We apply three types of stochastic models with Gaussian, Gaussian mixture and Markov switching Gaussian distributions to some synthesized data set time series with regime shifts for two-regimes ($\in \{0, 1\}$), according to a first-order Markov chain. For the simulation studies, we consider the data generated from the following model:

$$\text{if } s_n = 0, \quad y_n \sim N(0, 1), \quad (3.39)$$

$$\text{if } s_n = 1, \quad y_n \sim N(\mu, \sigma^2), \quad (3.40)$$

and transition probabilities are defined as follows,

$$p_{00} = p_{11} = \Pr(s_n = i | s_{n-1} = i), \quad i = 0, 1, \quad (3.41)$$

where the sample size $N \in \{1, 2, \dots, N\}$ is 10,000 and the initial regime $s_0 = 0$. To evaluate the goodness of fit of each model, we use the log likelihood value (LL), the Akaike information criterion (AIC) and the Quadratic Probability Score (QPS) by Brier (1950). Here QPS is a measure for the accuracy of the probability closeness on average to the realization of the regime by a zero-one dummy variable. Note that QPS is denoted by

$$\text{QPS} = \frac{1}{N} \sum_{n=1}^N 2(P_n - s_n)^2, \quad (3.42)$$

where P_n is the smoothed probability of regime $s_n = 1$ ¹. QPS ranges from 0 to 2. If QPS = 0, it represents perfect accuracy.

¹In general, the predictive probability is often described as P_n

Tables 3.1 to 3.4 summarize statistics for model comparisons and estimated parameters of simulation studies. Tables 3.1 and 3.2 provide data with Markov switching means, and Tables 3.3 and 3.4 provide data with Markov switching variances.

Table 3.1: Simulation Results 1 — Statistics —

Case 1: $\mu = 2.0$ and $\sigma = 1.0$								
Tran Prob	Gaussian		Gaussian mixture		Markov switching Gaussian			
$p_{00} = p_{11}$	LL	AIC	LL	AIC	LL	AIC	QPS	$\Pr(s_n = 1)$
0.95	17576.96	35157.91	17477.16	34964.31	15535.65	31083.29	0.0520	0.4825
0.90	17542.05	35088.10	17457.96	34925.92	16182.54	32377.07	0.0994	0.4858
0.80	17614.12	35232.24	17524.81	35059.63	16903.97	33819.94	0.1658	0.4780
0.70	17839.13	35682.26	17731.36	35472.73	17453.37	34918.74	0.1925	0.4979
0.60	17714.60	35433.20	17583.79	35177.57	17526.92	35065.85	0.2116	0.4983

Case 2: $\mu = 1.0$ and $\sigma = 1.0$								
Tran Prob	Gaussian		Gaussian mixture		Markov switching Gaussian			
$p_{00} = p_{11}$	LL	AIC	LL	AIC	LL	AIC	QPS	$\Pr(s_n = 1)$
0.95	15292.08	30588.17	15288.36	30586.73	14871.67	29755.33	0.1646	0.4878
0.90	15189.72	30383.43	15185.56	30381.11	14968.51	29949.02	0.2688	0.4979
0.80	15277.49	30558.98	15275.62	30561.23	15184.18	30380.37	0.3320	0.4813
0.70	15385.09	30774.18	15383.85	30777.69	15333.56	30679.11	0.3703	0.4944
0.60	15236.25	30476.49	15233.11	30476.21	15230.60	30473.20	0.4818	0.5019

Case 3: $\mu = 0.5$ and $\sigma = 1.0$								
Tran Prob	Gaussian		Gaussian mixture		Markov switching Gaussian			
$p_{00} = p_{11}$	LL	AIC	LL	AIC	LL	AIC	QPS	$\Pr(s_n = 1)$
0.95	14401.19	28806.38	14397.64	28805.29	14341.36	28694.73	0.3406	0.5072
0.90	14545.20	29094.40	14542.66	29095.31	14516.16	29044.32	0.3902	0.5073
0.80	14466.94	28937.89	14466.94	28943.88	14456.97	28925.94	0.4699	0.5061
0.70	14496.33	28996.66	14495.68	29001.36	14488.44	28988.87	0.4778	0.5004
0.60	14542.10	29088.19	14539.56	29089.13	14537.62	29087.23	0.5955	0.5026

Case 4: $\mu = 0.25$ and $\sigma = 1.0$								
Tran Prob	Gaussian		Gaussian mixture		Markov switching Gaussian			
$p_{00} = p_{11}$	LL	AIC	LL	AIC	LL	AIC	QPS	$\Pr(s_n = 1)$
0.95	14361.64	28727.29	14360.91	28731.82	14349.98	28711.95	0.6004	0.4707
0.90	14245.62	28495.25	14244.89	28499.77	14244.48	28500.96	0.5409	0.4959
0.80	14213.29	28430.57	14212.62	28435.25	14208.86	28429.73	0.5099	0.4863
0.70	14311.54	28627.07	14311.53	28633.07	14311.54	28635.07	0.5000	0.4993
0.60	14305.81	28615.61	14305.08	28620.15	14303.19	28618.37	0.6829	0.4958

Table 3.2: Simulation Results 1 — Estimated Parameters of Markov switching Gaussian

Case 1: $\mu = 2.0$ and $\sigma = 1.0$						
Tran Prob	μ_1	μ_2	σ_1	σ_2	p_{00}	p_{11}
0.95	0.0077	1.9902	1.0009	0.9863	0.9486	0.9444
0.90	0.0051	1.9677	0.9911	1.0019	0.8955	0.8904
0.80	0.0219	1.9989	1.0105	0.9979	0.8086	0.7878
0.70	0.0197	2.0651	1.0332	0.9954	0.7170	0.6922
0.60	-0.0837	1.9878	0.9647	0.9863	0.5757	0.6029

Case 2: $\mu = 1.0$ and $\sigma = 1.0$						
Tran Prob	μ_1	μ_2	σ_1	σ_2	p_{00}	p_{11}
0.95	-0.0072	1.0033	0.9942	0.9972	0.9446	0.9442
0.90	0.0424	1.0127	1.0066	0.9786	0.8955	0.8881
0.80	0.0816	0.9725	1.0237	1.0226	0.8587	0.8248
0.70	-0.0259	1.0189	1.0080	0.9893	0.7053	0.7080
0.60	0.3047	1.4608	1.0422	0.9499	0.8752	0.3757

Case 3: $\mu = 0.5$ and $\sigma = 1.0$						
Tran Prob	μ_1	μ_2	σ_1	σ_2	p_{00}	p_{11}
0.95	-0.0678	0.4440	0.9874	0.9925	0.9430	0.9642
0.90	-0.0398	0.5184	0.9891	1.0060	0.8721	0.8837
0.80	-0.0295	0.4449	0.9906	1.0089	0.8094	0.8824
0.70	0.0111	0.5778	1.0068	0.9726	0.7736	0.6936
0.60	0.1650	0.3055	1.0004	1.0448	0.9496	0.9728

Case 4: $\mu = 0.25$ and $\sigma = 1.0$						
Tran Prob	μ_1	μ_2	σ_1	σ_2	p_{00}	p_{11}
0.95	0.0404	0.3743	1.0205	0.9619	0.9857	0.9490
0.90	0.0114	0.1876	0.9441	1.0303	0.8694	0.9340
0.80	-0.1495	0.4983	0.9642	0.9303	0.6673	0.5809
0.70	0.1257	0.1259	1.0123	1.0123	0.9523	0.9523
0.60	0.1073	0.1749	1.0206	1.0153	0.9523	0.9526

Table 3.3: Simulation Results 2 — Statistics —

Case 1: $\mu = 0.0$ and $\sigma = 3.0$								
Tran Prob	Gaussian		Gaussian mixture		Markov switching Gaussian			
$p_{00} = p_{11}$	LL	AIC	LL	AIC	LL	AIC	QPS	$\Pr(s_n = 1)$
0.95	22290.66	44585.32	21685.98	43381.95	20800.43	41612.86	0.1099	0.5205
0.90	22452.73	44909.46	21716.55	43443.11	21136.60	42285.19	0.1819	0.5124
0.80	22010.14	44024.28	21352.73	42715.46	21126.19	42264.39	0.2666	0.4842
0.70	22178.51	44361.02	21491.42	42992.84	21403.47	42818.93	0.3101	0.4963
0.60	22221.85	44447.70	21538.53	43087.06	21516.63	43045.26	0.3387	0.5054

Case 2: $\mu = 0.0$ and $\sigma = 2.0$								
Tran Prob	Gaussian		Gaussian mixture		Markov switching Gaussian			
$p_{00} = p_{11}$	LL	AIC	LL	AIC	LL	AIC	QPS	$\Pr(s_n = 1)$
0.95	18660.88	37325.77	18474.11	36958.23	18164.43	36340.87	0.1840	0.4888
0.90	18747.85	37499.71	18548.95	37107.90	18418.71	36849.41	0.2705	0.4923
0.80	18799.72	37603.44	18604.26	37218.52	18531.75	37075.49	0.3653	0.4883
0.70	18820.10	37644.21	18598.77	37207.54	18576.46	37164.92	0.3971	0.5066
0.60	18792.21	37588.43	18629.27	37268.53	18622.05	37256.11	0.4186	0.5104

Case 3: $\mu = 0.0$ and $\sigma = 1.5$								
Tran Prob	Gaussian		Gaussian mixture		Markov switching Gaussian			
$p_{00} = p_{11}$	LL	AIC	LL	AIC	LL	AIC	QPS	$\Pr(s_n = 1)$
0.95	16628.53	33261.05	16608.43	33226.86	16553.39	33118.78	0.3239	0.5134
0.90	16593.26	33190.51	16565.61	33141.22	16541.21	33094.42	0.4268	0.5219
0.80	16615.60	33235.20	16582.28	33174.55	16571.09	33154.17	0.4772	0.4783
0.70	16524.16	33052.31	16480.29	32970.59	16478.72	32969.44	0.4935	0.4989
0.60	16618.81	33241.62	16596.48	33202.96	16594.98	33201.96	0.5047	0.4852

Case 4: $\mu = 0.0$ and $\sigma = 1.25$								
Tran Prob	Gaussian		Gaussian mixture		Markov switching Gaussian			
$p_{00} = p_{11}$	LL	AIC	LL	AIC	LL	AIC	QPS	$\Pr(s_n = 1)$
0.95	15513.30	31030.60	15509.70	31029.40	15497.90	31007.80	0.4434	0.5220
0.90	15440.95	30885.90	15434.45	30878.90	15427.22	30866.45	0.6656	0.5005
0.80	15284.24	30572.48	15276.39	30562.78	15275.72	30563.43	0.8089	0.4807
0.70	15501.16	31006.31	15493.51	30997.03	15492.27	30996.54	0.6124	0.4919
0.60	15434.41	30872.82	15433.84	30877.67	15429.53	30871.05	0.6700	0.5015

Table 3.4: Simulation Results 2 — Estimated Parameters of Markov switching Gaussian

Case 1: $\mu = 2.0$ and $\sigma = 1.0$						
Tran Prob	μ_1	μ_2	σ_1	σ_2	p_{00}	p_{11}
0.95	0.0124	0.0081	1.0023	2.9552	0.9519	0.9564
0.90	0.0001	-0.0318	0.9911	3.0668	0.9025	0.9038
0.80	0.0085	-0.0308	0.9790	2.9159	0.8027	0.8076
0.70	0.0084	-0.0002	0.9699	2.9540	0.6998	0.7161
0.60	0.0372	-0.0112	0.9863	2.9772	0.6000	0.6134

Case 2: $\mu = 1.0$ and $\sigma = 1.0$						
Tran Prob	μ_1	μ_2	σ_1	σ_2	p_{00}	p_{11}
0.95	-0.0122	-0.0054	0.9952	1.9784	0.9522	0.9517
0.90	-0.0146	-0.0196	1.0013	1.9962	0.8835	0.8827
0.80	0.0053	0.0365	0.9551	1.9798	0.7867	0.8128
0.70	0.0238	-0.0114	0.9596	2.0111	0.6747	0.6916
0.60	0.0050	-0.0205	1.0439	2.0064	0.6361	0.6120

Case 3: $\mu = 0.5$ and $\sigma = 1.0$						
Tran Prob	μ_1	μ_2	σ_1	σ_2	p_{00}	p_{11}
0.95	-0.0143	-0.0200	1.0257	1.4917	0.9485	0.9466
0.90	0.0090	-0.0212	0.9767	1.4166	0.9263	0.9567
0.80	-0.0085	-0.0493	1.0890	1.5871	0.8882	0.7717
0.70	0.0286	-0.0020	1.0403	1.5764	0.7316	0.5339
0.60	0.0926	-0.0404	0.9616	1.4279	0.5536	0.7325

Case 4: $\mu = 0.5$ and $\sigma = 1.0$						
Tran Prob	μ_1	μ_2	σ_1	σ_2	p_{00}	p_{11}
0.95	-0.0022	0.0013	1.0284	1.2895	0.9620	0.9444
0.90	0.0172	0.0089	0.8090	1.1918	0.8341	0.9642
0.80	0.1422	-0.0191	0.6962	1.1581	0.5852	0.9457
0.70	0.1085	-0.0190	0.8826	1.2153	0.6924	0.8931
0.60	0.0722	-0.2795	1.1093	1.1669	0.9671	0.8905

Chapter 4

Empirical Analysis

4.1 Trend Identification and Trading Strategy

- Trend identification and financial trading strategy using a stochastic trend model with a Markov switching slope change and ARCH

4.1.1 Introduction

In real-world financial markets, some practitioners, including dealers and money managers of foreign exchange rates, stocks, commodities, etc., regard “traditional technical analysis” as an important analytical tool. Taylor (1992) reported that more than 90 percent of foreign exchange dealers surveyed in London used some form of technical analysis to formulate their trading decisions. One of the most important principles in technical analysis is that *“asset prices move in trends.”* In other words, the unobserved trends of asset prices tend to remain in motion unless acted upon by another force. The “trend” in the framework of traditional technical analysis is obtained by connecting the lowest or highest values of the asset’s price with a straight line. The “slope” of the trend plays an important role in a practical trading strategy. When it is positive (negative), the practitioner continues to take a long (short) position despite any impulsive fluctuations in the asset’s prices. Traditional technical analysis offers various methods of identifying this unobserved trend, but every method requires not only scientific skill, but also artistic sensitivity based on long experience in financial markets. Therefore, not everyone can identify the useful trend and slope for a practical trading strategy. The traditional technical analysis is briefly reviewed in Murphy (1986).

This section has two purposes. The first is to identify the unobserved trend and slope of real financial

time series using a nonlinear stochastic approach — the stochastic trend model with Markov switching slope changes and ARCH (MS-SC/ARCH). The second purpose is to examine whether the trend and slope obtained by the MS-SC/ARCH model are useful in a practical trading strategy, and to analyze the profit profile of this strategy.

Several studies have analyzed economic and financial time series with linear and nonlinear dynamic systems. Harrison and Stevens (1976) and Harvey and Todd (1983) proposed time series models consisting of trend and slope-changing components described by random walk processes. Kitagawa and Gersch (1996) performed simulations with various types of stochastic trend models, involving Gaussian and non-Gaussian distributions. Hamilton (1989) introduced the Markov switching model into the econometric analysis to capture the time series, including some unknown structural changes. Kim (1993, 1994) extended Hamilton's model to a general state-space model. Cai (1994) and Hamilton and Susmel (1994) examined the heteroskedasticity of financial data using Markov switching models. Finally, Kim and Nelson (1999) reviewed the several types of Markov switching models with economic and financial applications.

This section is organized as follows. Subsection 2 describes the MS-SC/ARCH model. Subsection 3 shows the trading strategy based on the MS-SC/ARCH model. Subsection 4 presents an empirical analysis including estimations and evaluations of a trading strategy based on the MS-SC/ARCH model. Conclusions are presented in Subsection 5.

4.1.2 Markov Switching Slope Change and ARCH Model

Model Specification

In this section, we describe the stochastic trend model with simultaneous Markov switching slope changes and ARCH (MS-SC/ARCH). Let y_n , $n = 1, 2, \dots, N$, be the observed financial asset price — for example, stock prices, exchange rates and commodity prices. The observation model of the MS-SC/ARCH model is given by

$$y_n = t_n + \varepsilon_n, \quad \varepsilon_n | \Psi_{n-1} \sim N(0, h_n), \quad (4.1)$$

$$h_n = \gamma_0(1 - s_n) + \gamma_1 s_n + \alpha \varepsilon_{n-1}^2, \quad (4.2)$$

where Ψ_{n-1} denotes the information up to time $n - 1$, and $\gamma_1 > \gamma_0 > 0$, and $\alpha \geq 0$, and t_n is a trend component. ε_n is a fluctuation around the trend component t_n defined by the ARCH process with a Markov switching structure. s_n is an unobserved Markov chain, and takes the value of zero or one as determined by a first-order two-regime Markov process. When $s_n = 0$, the system is in a low volatility

regime. When $s_n = 1$, it is in a high volatility regime. The system models are given by

$$t_n = t_{n-1} + \Delta t_n + v_n, \quad v_n \sim N(0, \tau^2), \quad (4.3)$$

$$\Delta t_n = \Delta t_{n-1} + w_n, \quad w_n \sim N(0, s_n \sigma^2), \quad (4.4)$$

where $v_1 > v_0 > 0$. Here t_n follows the random walk process with a drift Δt_n and an innovation term v_n . Δt_n is interpreted as the permanent change in the slope, and w_n produces a permanent shift in the level with a subsequent undisturbed growth. When $s_n=0$, Δt_n takes the value of the previous Δt_{n-1} , and t_n follows the random walk with the same slope or drift Δt_n . When $s_n=1$, Δt_n follows a random walk process with a normal distribution w_n with zero mean and variance σ^2 ; that is, $s_n = 0$ represents the regime of slope changes.

We assume that in the MS-SC/ARCH model changes in slope and increases in volatility occur simultaneously depending on s_n . The unobserved Markov chain s_n is updated via the transition probability as follows:

$$p_{ij} = \Pr(s_n = j | s_{n-1} = i), \quad i, j = 0, 1, \quad (4.5)$$

where $\sum_{j=0}^1 p_{ij} = 1$ for $i = 0, 1$. This transition probability p_{ij} is assumed to be homogeneous, and it can be written as the following matrix:

$$P = \begin{pmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{pmatrix} = \begin{pmatrix} p_{00} & 1 - p_{11} \\ 1 - p_{00} & p_{11} \end{pmatrix}. \quad (4.6)$$

The log-likelihood function of the MS-SC/ARCH model is given by:

$$\begin{aligned} L(\theta) &= \sum_{n=1}^N \log f(y_n | \Psi_{n-1}), \\ &= \sum_{n=1}^N \log \left[\sum_{j=1}^m \sum_{i=1}^m f(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1}) \right], \end{aligned} \quad (4.7)$$

where θ denotes the unknown parameters and

$$f(y_n | s_n = j, \Psi_{n-1}) = \frac{1}{\sqrt{2\pi h_{j,n}}} \exp \left\{ -\frac{(y_n - \mu)^2}{2h_{j,n}} \right\}. \quad (4.8)$$

Here $\Pr(s_n = j | \Psi_{n-1})$ in equation (4.7) can be obtained recursively by using a non-Gaussian filter (Kitagawa 1987, Hamilton 1989) as follows:

$$\Pr(s_n = j | \Psi_{n-1}) = \sum_{i=0}^1 \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} | \Psi_{n-1}), \quad (4.9)$$

$$\Pr(s_n = j | \Psi_n) = \frac{f(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1})}{f(y_n | \Psi_{n-1})}. \quad (4.10)$$

Several methods for estimating the unknown parameters have been proposed, for example, the EM (expectation maximization) algorithm (Hamilton 1990, 1994) and the Gibbs sampler (Albert and Chib 1992). For the purpose of this section, the quasi-maximum likelihood method (Kim 1994) is preferred owing to its computational ease.

Estimations of States and Parameters

Let $x_n \equiv (t_n, \Delta t_n, \varepsilon_n)^T$ be a (3×1) state vector. Given the Markov chain S_n , equations (4.1) to (4.4) are represented in the state space form as follows:

$$x_n = Fx_{n-1} + G\omega_n, \quad (4.11)$$

$$y_n = Hx_n. \quad (4.12)$$

Here when Ψ_{n-1} is information available up to time $n-1$, Equation (4.11) representing the system noise is

$$\omega_n | \Psi_{n-1} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau^2 & 0 & 0 \\ 0 & v_1^2 S_n & 0 \\ 0 & 0 & h_n \end{bmatrix} \right). \quad (4.13)$$

F, G and H are respectively defined as follows:

$$F = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}. \quad (4.14)$$

Once the MS-SC/ARCH model is represented in the state space form, we can obtain estimations of the states and parameters by using the Kalman filter algorithm. Kim (1994) proposes the filter for the state space model with the Markov switching structures. Harvey, Ruiz and Sentana (1992) propose the filter for the state-space model with ARCH disturbances. We combine both filters and apply it to the MS-SC/ARCH model. The filter algorithms for the state vector x_n and the unobserved variable s_n are described as follows:

Filter for the state vector x_n

[Prediction]

$$x_{n|n-1}^{(i,j)} = Fx_{n-1|n-1}^{(i)}, \quad (4.15)$$

$$\omega_{n|n-1}^{(i,j)} = F\omega_{n-1|n-1}^{(i)}F^T + GQ^{(j)}G^T, \quad (4.16)$$

where

$$x_{n|n-1}^{(i,j)} = E[x_n | \Psi_{n-1}, S_n = j, S_{n-1} = i], \quad (4.17)$$

$$\omega_{n|n-1}^{(i,j)} = E[(x_n - x_{n|n-1})(x_n - x_{n|n-1})^T | \Psi_{n-1}, S_n = j, S_{n-1} = i], \quad (4.18)$$

$$Q^{(j)} = E[\omega_n, \omega_n^T | \Psi_{n-1}]. \quad (4.19)$$

[Filter]

$$K_n^{(i,j)} = v_{n|n-1}^{(i,j)} H^T (H v_{n|n-1}^{(i,j)} H^T)^{-1}, \quad (4.20)$$

$$x_{n|n}^{(i,j)} = x_{n|n-1}^{(i,j)} + K_n^{(i,j)} (y_n - H x_{n|n-1}^{(i,j)}), \quad (4.21)$$

$$\omega_{n|n}^{(i,j)} = (I - K_n^{(i,j)} H) \omega_{n|n-1}^{(i,j)}, \quad (4.22)$$

where

$$x_{n|n}^{(i,j)} = E[x_n | \Psi_n, S_n = j, S_{n-1} = i], \quad (4.23)$$

$$\omega_{n|n}^{(i,j)} = E[(x_n - x_{n|n})(x_n - x_{n|n})^T | \Psi_n, S_n = j, S_{n-1} = i]. \quad (4.24)$$

However, to process equation (4.20), we need to calculate $h_n = \gamma_{s_n} + \alpha \varepsilon_{n-1}^2$, which is a function of the square of past unobserved shock ε_{n-1}^2 . Thus, the above Kalman filter is not operable. Harvey, Ruiz and Sentana (1992) solve the problem of the unobserved ε_{n-1}^2 by replacing it with its conditional expectation:

$$h_n = \gamma_0(1 - S_n) + \gamma_1 S_n + \alpha E[\varepsilon_{n-1}^{*2} | \Psi_{n-1}], \quad (4.25)$$

Under this specification, the calculation of the terms $E[\varepsilon_{n-1}^2 | \Psi_{n-1}]$ is straightforward. Because we know

$$\varepsilon_{n-1}^* = E[\varepsilon_{n-1}^* | \Psi_{n-1}] + (\varepsilon_{n-1}^* - E[\varepsilon_{n-1}^* | \Psi_{n-1}]), \quad (4.26)$$

it is easy to show that

$$E[\varepsilon_{n-1}^{*2} | \Psi_{n-1}] = E[\varepsilon_{n-1}^* | \Psi_{n-1}]^2 + E[(\varepsilon_{n-1}^* - E[\varepsilon_{n-1}^* | \Psi_{n-1}])^2], \quad (4.27)$$

where $E[\varepsilon_{n-1} | \Psi_{n-1}]$ is obtained from the last two elements of $x_{n-1|n-1}$ and its mean squared errors $E[(\varepsilon_{n-1} - E(\varepsilon_{n-1} | \Psi_{n-1}))^2]$ is obtained from the last two diagonal elements of $w_{n-1|n-1}$. From a by-product of the filtering algorithm for the state space x_n , we can obtain the conditional distribution y_n on s_n, s_{n-1} and Ψ_{n-1} as follows

$$f(y_n | S_n = j, S_{n-1} = i, \Psi_{n-1}) = \frac{1}{\sqrt{2\pi r_n}} \exp \left\{ -\frac{(y_n - H x_{n|n-1}^{(i,j)})^2}{2r_n} \right\}, \quad (4.28)$$

with $r_n = H w_{n|n-1}^{(i,j)} H^T + R^{(j)}$ from the predictive distribution in equation (4.20).

Filter for the unobserved variable s_n **[Prediction]**

$$\Pr(S_n = j | \Psi_{n-1}) = \sum_{i=0}^1 \Pr(S_n = j, S_{n-1} = i | \Psi_{n-1}) \quad (4.29)$$

$$= \sum_{i=0}^1 \Pr(S_n = j, S_{n-1} = i) \Pr(S_{n-1} = i | \Psi_{n-1}). \quad (4.30)$$

[Filter]

$$\Pr(S_n = j | \Psi_n) = \sum_{i=0}^1 \Pr(S_n = j, S_{n-1} = i | \Psi_{n-1}, y_n) \quad (4.31)$$

$$= \sum_{i=0}^1 \frac{f(y_n | S_n = j, S_{n-1} = i, \Psi_{n-1}) \Pr(S_n = j, S_{n-1} = i | \Psi_{n-1})}{f(y_n | \Psi_{n-1})}, \quad (4.32)$$

where

$$f(y_n | \Psi_{n-1}) = \sum_{j=0}^1 f(y_n | S_n = j, \Psi_{n-1}) \Pr(S_n = j | \Psi_{n-1}) \quad (4.33)$$

$$= \sum_{i=0}^1 \sum_{j=0}^1 f(y_n | S_n = j, S_{n-1} = i, \Psi_{n-1}) \Pr(S_n = j, S_{n-1} = i | \Psi_{n-1}). \quad (4.34)$$

To avoid an explosion of the number of Gaussian components, the approximation technique that is proposed by Harrison and Stevens (1976) is used:

$$x_{n|n}^{(j)} = \frac{\sum_{i=0}^1 \Pr(S_{n-1} = i, S_n = j | \Psi_n) x_{n|n}^{(i,j)}}{\Pr(S_n = j | \Psi_n)}, \quad (4.35)$$

$$v_{n|n}^{(j)} = \frac{\sum_{i=0}^1 \Pr(S_{n-1} = i, S_n = j | \Psi_n) [v_{n|n}^{(i,j)} + (x_{n|n}^{(j)} - x_{n|n}^{(i,j)})(x_{n|n}^{(j)} - x_{n|n}^{(i,j)})^T]}{\Pr(S_n = j | \Psi_n)} \quad (4.36)$$

Kim (1994) and Kitagawa (1994) report that such an approximation has only small biased estimations, and is an effective method. In addition, nonlinear filter (Kitagawa 1987) and MCMC (Markov chain Monte Carlo) methods (Albert and Chib 1993) can provide precise estimates.

By running this filter algorithm, the approximate log-likelihood is obtained as its by-product. The approximate log-likelihood function for the unknown parameter vector θ is given by

$$\begin{aligned} L(\theta) &= \sum_{n=1}^N \log f(y_n | \Psi_{n-1}) \\ &= \sum_{n=1}^N \left[\sum_{j=0}^1 \sum_{i=0}^1 \log f(y_n, S_n = j, S_{n-1} = i | \Psi_{n-1}) \right] \\ &= \sum_{n=1}^N \left[\sum_{j=0}^1 \sum_{i=0}^1 \log f(y_n | S_n = j, S_{n-1} = i, \Psi_{n-1}) \Pr(S_n = j, S_{n-1} = i | \Psi_{n-1}) \right] \end{aligned} \quad (4.37)$$

Several methods for estimating the unknown parameters have been proposed, including the EM algorithm (Hamilton 1990, 1994) and the Gibbs sampler (Albert and Chib 1992). For the purpose of this section, the maximum likelihood method is preferred due to its computational ease.

4.1.3 Trading Strategy

We consider a trading strategy based on the MS-SC/ARCH model. The investment position is taken depending on the buy and sell signals generated by the MS-SC/ARCH model. When buy (sell) is signaled at time n by the MS-SC/ARCH model, a long (short) position on a financial asset is taken. The same

investment position remains until the opposite case of sell (buy) is signaled. When buy (sell) is signaled, the investor closes the previous long (short) position, and takes the short (long) position. The trading signal ψ_n at time n is defined depending on a one-step-ahead prediction of the slope component $\Delta t_{n+1|n}$ of the MS-SC/ARCH model as follows:

$$\psi_n = \begin{cases} 1 & \Delta t_{n+1|n} \geq 0 \\ -1 & \Delta t_{n+1|n} < 0. \end{cases} \quad (4.38)$$

where $\Delta t_{n+1|n}$ is drawn from

$$\begin{aligned} x_{n+1|n} &= \sum_{k=0}^1 \sum_{j=0}^1 \Pr(S_{n+1} = k, S_n = j | \Psi_n) x_{n+1|n}^{(j,k)} \\ &= \sum_{k=0}^1 \sum_{j=0}^1 \Pr(S_{n+1} = k | S_n = j) \Pr(S_n = j | \Psi_n) F x_{n|n}^{(j)}. \end{aligned} \quad (4.39)$$

This MS-SC/ARCH trading strategy is based on the general idea of “trend following” in the framework of traditional technical analysis. Therefore, identification of the unobserved slope and removal of some fluctuations around the trend are important for success in generating profits.

The cumulative sum of return to measure the performance for the MS-SC/ARCH trading strategy is given by $\sum_{n=1}^{N-1} \psi_n r_{n+1}$, where $r_{n+1} = \log(y_{n+1}) - \log(y_n)$ is the logarithmic return of the original financial asset price from n to $n+1$.

4.1.4 Application and Comparison of Trading Performance

Data and Estimations

The sample data used in this study are the weekly Tokyo Stock Price Index (TOPIX) over the period from 1986/1 to 2000/12. The number of observations is 835. Figure 4.41 shows plots of the original series of the TOPIX. Figure 4.42 plots the logarithmic return series of the TOPIX. For the parameter estimation below, the logarithmic transformation of the TOPIX multiplied by 100 is used.

Figure 4.43 shows the plot of the filtered trend slope $\Delta_{n|n}$. We find that the trend with a positive or negative slope has some persistence. Table 4.1 shows the estimation results of MS-SC/ARCH model for the TOPIX (1986/1–2000/12). Note that the unknown parameters $\theta \equiv \{\tau, \sigma, p_{00}, p_{11}\}$ are estimated by the maximum likelihood method given a variance of $w_n = 2.00$ in equation (4.4). The trend component t_n follows a random walk process with a time-varying drift Δt_n with normal white noise $N(0, 2.314)$. The fluctuation ε_n around the trend follows the ARCH process with constant terms 0.039 ($s_n = 0$) and 4.167 ($s_n = 1$), and common ARCH coefficient 0.895. From these estimations in the ARCH process, it can be said that $s_n = 0$ and $s_n = 1$ represent the regimes for “no-slope change and low volatility” and for “slope

change and high volatility", respectively. The self-loop transition probabilities p_{00} and p_{11} are 0.993 and 0.796, respectively. $s_n = 0$ tends to be more persistent than $s_n = 1$. This is clear from Figure 4.44, which plots the filtered regime probability of $s_n = 1$. While the amplitude of the slope changes frequently with an increase in volatility in the first half of the sample period, slope changes are relatively stable in the second half.

Next, we test whether regime shifts occur according to a Markov process in the MS-SC/ARCH model. Under the null hypothesis ($\gamma_0 = \gamma_1, \sigma = 0$), p_{00} and p_{11} can be identified and the asymptotic distribution of likelihood ratio statistics does not follow the standard chi-square distribution. To deal with such a problem, Garcia (1998) and Hansen (1992, 1996) propose hypothesis test methods based on a grid search. Since these methods need computational burden, we use the hypothesis test of Engel and Hamilton (1990) as follows

$$H_0 : p_{00} = 1 - p_{11}, \quad \gamma_0 \neq \gamma_1 \quad \sigma_1 \neq 0. \quad (4.40)$$

Null hypothesis H_0 corresponds to the model in which the time series process follows a simple mixed distribution, and regime shifts do not follow the Markov process. Under this null hypothesis H_0 , all the parameters can be identified, and the asymptotic distribution of likelihood ratio statistics does not follow the standard chi-square distribution with one-degree-of-freedom. Table 4.1 includes estimated parameters and corresponding statistics on the null hypothesis H_0 model. The LR (Likelihood Ratio) statistic for the hypothesis H_0 is 3.87, and the null hypothesis H_0 can be rejected at a five percent critical level. In addition, the MS-SC/ARCH model is selected as the best AIC model.

Profitability of the Trading Strategy

To evaluate the profitability of the MS-SC/ARCH trading strategy, we estimate the unknown parameters $\theta \equiv \{\tau^2, \sigma^2, p_{00}, p_{11}\}$ using the first 313 observations from January 1986 to December 1991 as the formation period. Using these parameter estimations, the trading strategy can be constructed based on a one-step-ahead slope forecast $\Delta t_{n+1|n}$ by the MS-SC/ARCH model in the subsequent simulation period from January 1987 to December 2000. According to the trading strategy described in the previous section, we take a long (buy TOPIX) or short (sell TOPIX) investment position.

Table 4.2 summarizes the trading performance of the trading strategies based on the MS-SC/ARCH model and three alternative models as follows

- **Model 1.** Constant volatility,
- **Model 2.** No-switching and simple ARCH volatility,

- **Model 3.** Random walk trend slope.

The cumulative sum of the trading return of the MS-SC/ARCH trading strategy is 95.704 percent (average: 0.183 percent). It is larger than the trading returns for the three alternative models, which were: 57.728 percent (average: 0.111 percent) for Model 1; 70.266 percent (average: 0.135 percent) for Model 2; and 64.069 percent (average: 0.123 percent) for Model 3. In addition, a comparison of the 1 and 5 percentage points of the trading return distributions shows that those of the MS-SC/ARCH model are no smaller than for the other models. This implies that in alternative trading strategies there is a possibility of suffering a loss that is larger than in the MS-SC/ARCH trading strategy.

Figure 4.46 compares the average trading returns under the MS-SC/ARCH trading strategy and some traditional moving average trading strategies with different moving average periods. In this case, when z_n and m define TOPIX at time n and the moving average period respectively, the trading signal at time n of m period moving average trading strategy is given by

$$\psi_n^m = \begin{cases} 1 & z_n \geq \frac{\sum_{i=0}^{m-1} z_{n-i}}{m}, \\ -1 & z_n < \frac{\sum_{i=0}^{m-1} z_{n-i}}{m}. \end{cases} \quad (4.41)$$

$$(4.42)$$

From the comparisons of the average trading returns in Figure 4.46, it is evident that the performance of moving average trading strategies is sensitive to how the moving average period is set. Regardless of out-of-sample simulation, the average trading return of the MS-SC/ARCH trading strategy, which is indicated by a crossbar in Figure 4.46, can be said to obtain a relatively higher performance.

Table 4.3 shows the trading returns and trading frequencies according to year. The MS-SC/ARCH trading strategy obtains stable performance over the sample period, excluding 2000. The cumulative sum of the trading returns of the MS-SC/ARCH trading strategy in Figure 4.7 seems to be on the upside. We can say that this strategy succeeds in remaining stable with a positive performance over time. In regard to the trading frequency, the average position-holding period is 7.46 weeks. Unlike a random walk model, the MS-SC/ARCH model avoids excessive position changes.

4.1.5 Conclusions

In this section, we propose an MS-SC/ARCH model in which the slope of the trend and the ARCH process for the movement around the trend changes simultaneously as described by a first-order and two-regime Markov process. Our MS-SC/ARCH model is useful for estimating the unobserved trend and its slope for the time series, which has some piecewise linearity. In the empirical analysis, we applied the model to the TOPIX. We found sufficient statistical evidence that the trading strategy based on the

estimated trend and its slope makes an excess profit over alternative strategies.

Table 4.1: Estimated results of MS-SC/ARCH model.

	MS-SC/ARCH		$H_0 : p_{00} = 1 - p_{11}$	
	Estimation	(Standard error)	Estimation	(Standard error)
τ	2.314	(0.125)	2.237	(0.101)
γ_0	0.039	(0.055)	0.046	(0.050)
γ_1	4.167	(3.909)	16.975	(14.366)
α	0.895	(0.115)	0.919	(0.076)
p_{00}	0.993	(0.005)	0.987	(0.009)
p_{11}	0.796	(0.125)	0.013	
Log likelihood	-1999.81		-2001.75	
LR statistics			3.87	
p -value			0.049	
AIC	4011.63		4013.50	

Table 4.2: Basic statistical summary of the trading returns.

	MS-SC/ARCH	Model.1	Model.2	Model.3
Cumulative sum	95.704	57.728	70.266	64.069
Average	0.183	0.111	0.135	0.123
Standard deviation	2.681	2.685	2.684	2.684
Skewness	-0.092	-0.321	-0.049	-0.074
Kurtosis	0.924	0.947	0.905	0.908
Downside 1% point	-6.311	-7.289	-6.311	-6.311
Downside 5% point	-4.043	-4.323	-4.254	-4.323

Table 4.3: Performance evaluation of the MS-SC/ARCH trading strategy.

Year	Cumulative sum of trading returns	Trading frequency
1992	28.57	5
1993	20.16	5
1994	-2.04	6
1995	23.13	5
1996	-1.16	5
1997	15.18	10
1998	-3.14	10
1999	16.69	7
2000	-17.49	11
2001	15.79	6
Total	95.70	70

Figure 4.1: Weekly TOPIX: 1986/1 – 2001/12.



Figure 4.2: Logarithmic return of weekly TOPIX: 1986/1 – 2001/12.

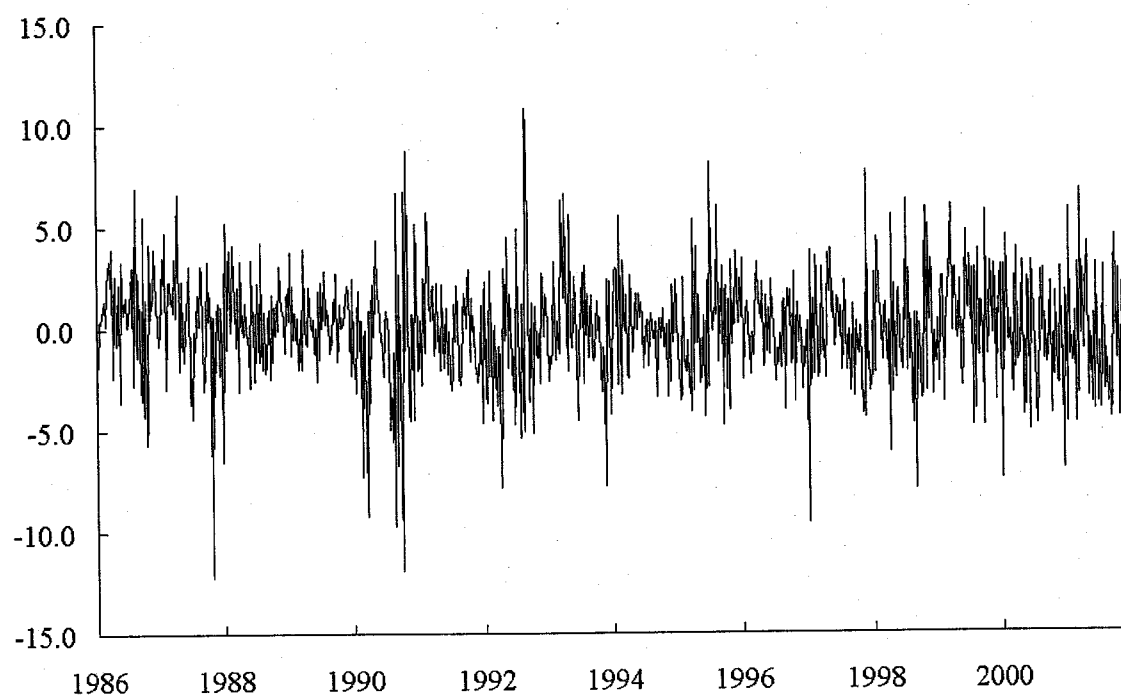


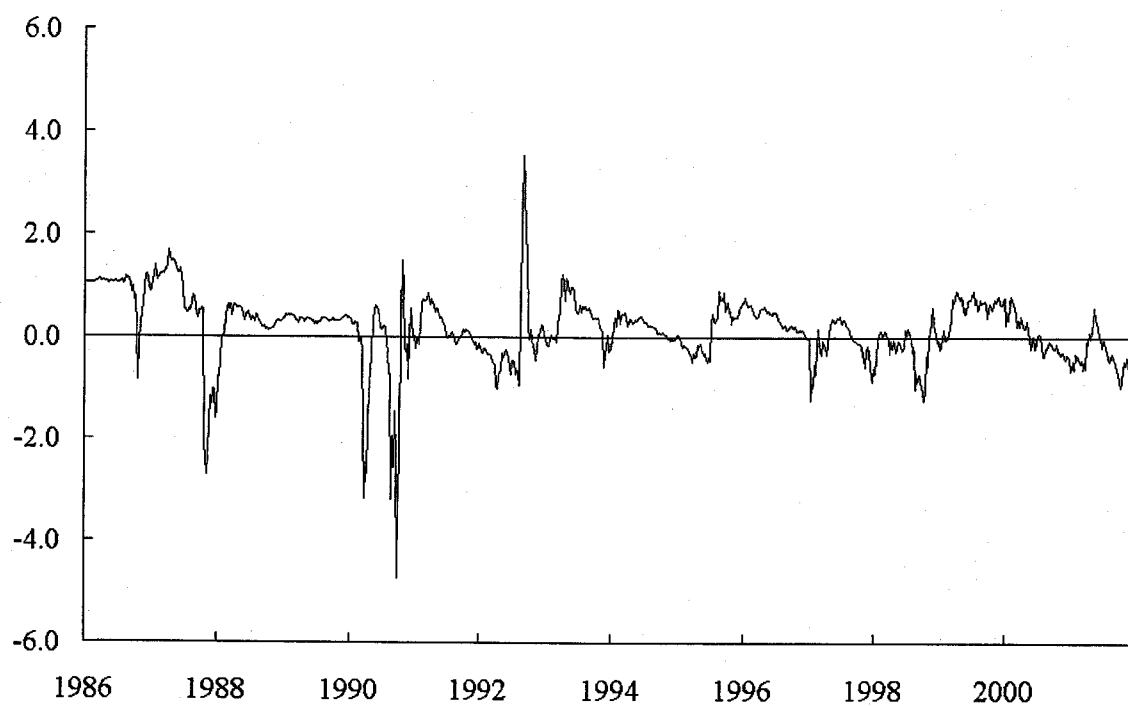
Figure 4.3: Filtered slope Δt_n : 1986/1 – 2001/12.

Figure 4.4: Filtered regime probability for slope change and high volatility: 1986/1 – 2001/12.

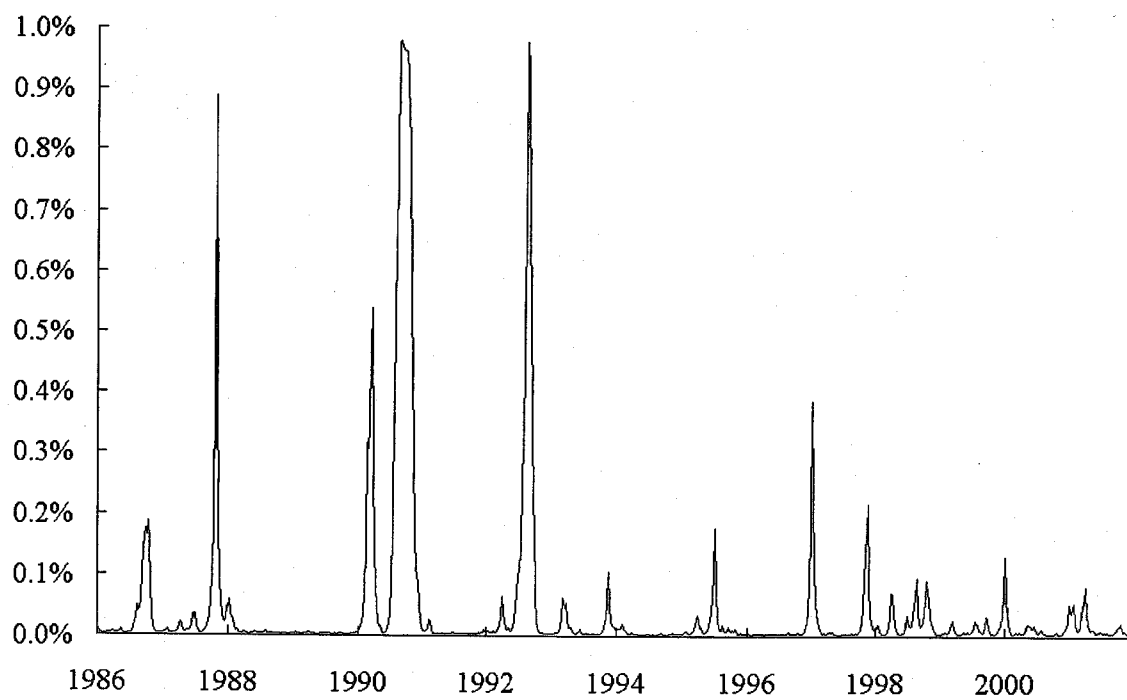


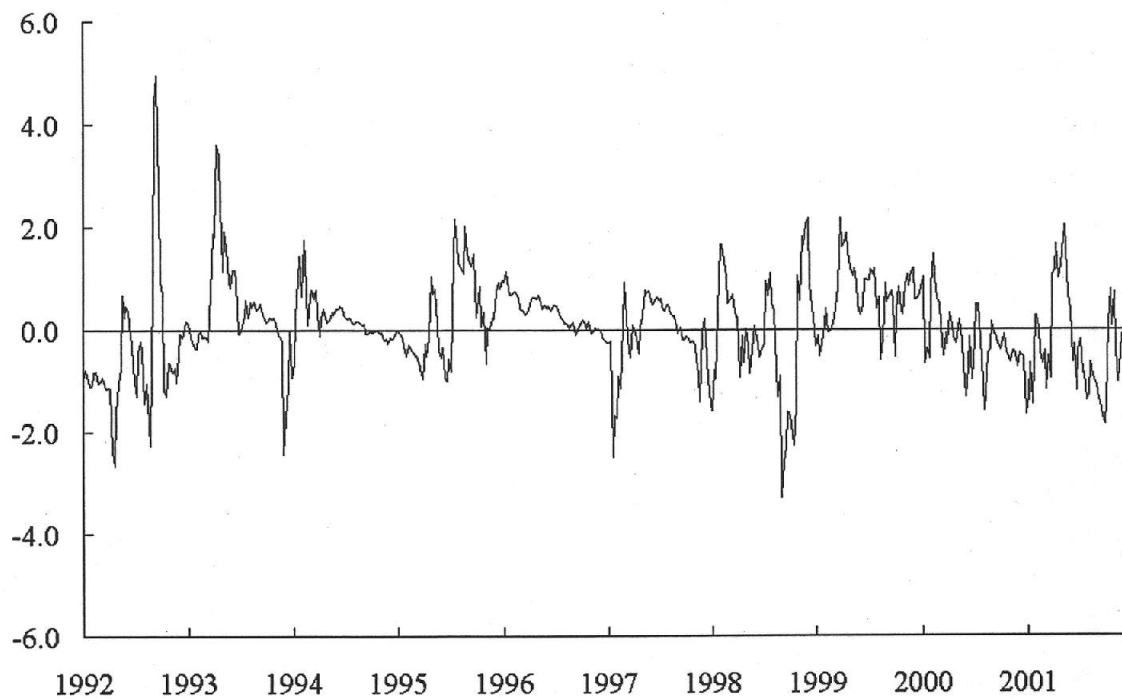
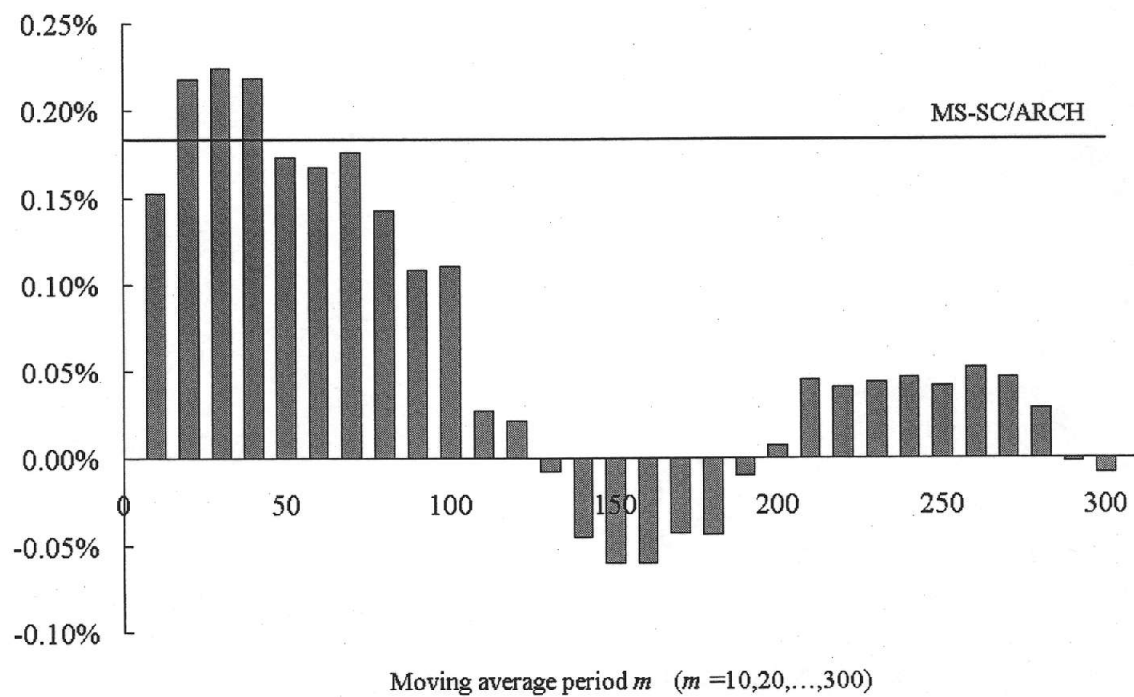
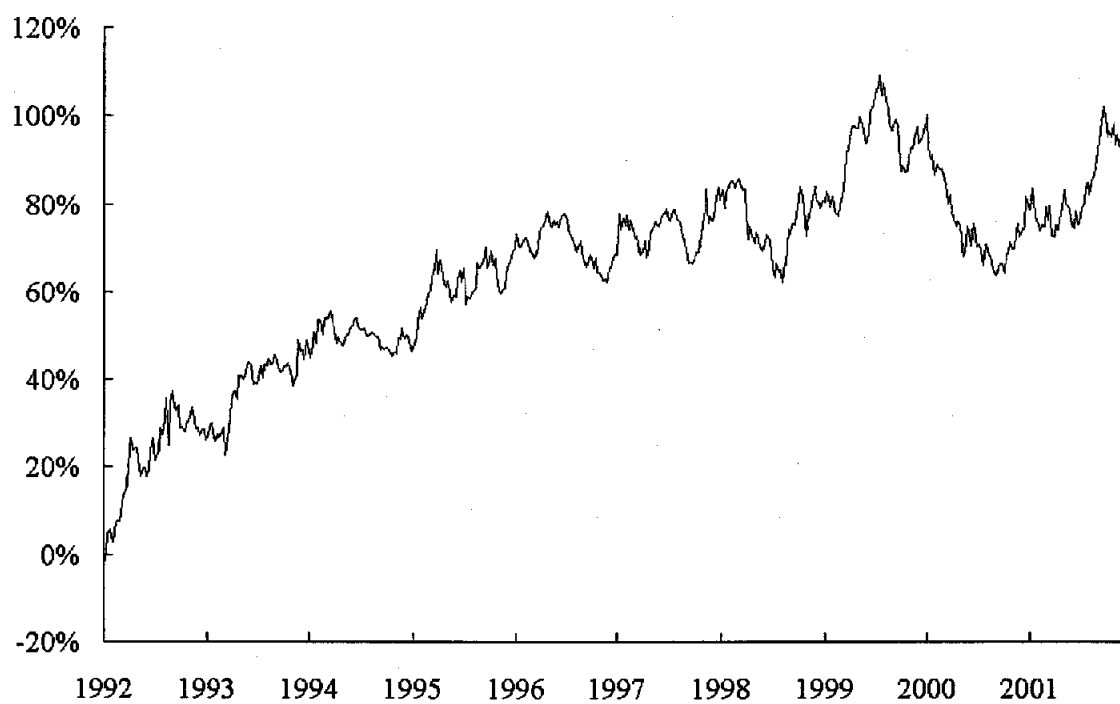
Figure 4.5: Filtered slope Δt_n : 1992/1 – 2001/12.Figure 4.6: Average trading returns of m period moving average trading strategies: 1992/1 – 2001/12.

Figure 4.7: Cumulative sums of profits for MS-SC/ARCH trading strategy: 1992/1 – 2001/12.



4.2 Time-series and Cross-sectional Volatility Analysis

— Time-series Stock Market Volatility and Cross-sectional Distribution of Individual Stock Returns: Empirical Analysis in the Japanese Stock Market

4.2.1 Introduction

The time series volatility of the stock market index return has long been used as a measure of risk in the stock market. This derives from modeling the time series distribution and dynamic characteristics of stock market index returns. On the other hand, the information in cross-sectional distributions of individual stock returns is also important for capturing stock market volatility on a given day; that is, variance, skewness, and kurtosis (MacDonald and Shawky 1995).

Our primary purpose in this section is to propose Markov switching models for evaluating exogenous leverage effects on time series volatility. A secondary purpose is to analyze empirically whether the cross-sectional distribution of individual stock returns affected the dynamic behavior and the process of time-series volatility in the Japanese stock market from 1999 to 2001.

A widely used class of time series models for conditional volatility is the autoregressive conditionally heteroskedastic (ARCH) class of models introduced by Engle (1982), and extended by Bollerslev (1986), Engle *et al.* (1987), Nelson (1991), and Glosten *et al.* (1993), among others. Bollerslev *et al.* (1992, 1994) summarizes this family of models. Another approach to capturing time-series volatility is to use the Markov switching heteroskedasticity model, which was first introduced into the study of economics and finance by Hamilton (1989). Hamilton and Susmel (1994) and Cai (1994) proposed a new ARCH model with Markov switching structures. Engel and Kim (1999) used the Markov switching variance model in state-space form. These Markov switching models more realistically capture the time series properties of dramatic economic events, such as stock market crashes. In these models, volatility tends to depend on past news and the state of the economy. Kim (1999) suggests differences between ARCH-type heteroskedasticity and Markov switching-type heteroskedasticity.

Most of the volatility characteristics outlined above are univariate, and relate the volatility of a time series only to the information contained in the history of that particular series. Of course, it is not credible that financial asset prices evolve independently of the associated market, and so we expect that other variables may contain relevant information about the volatility of a series. Such evidence has been found by Bollerslev and Melvin (1994), Engle and Mezrich (1996), and Engle *et al.* (1990 a, b). Apart from the possibility that other assets affect the volatility of a series, it is also possible that deterministic events

have an impact. For example, scheduled company announcements, macroeconomic announcements, and even deterministic time-of-day effects may influence the volatility process. Andersen and Bollerslev (1998) found that the volatility of the Deutschmark-U.S. Dollar exchange rate increases markedly around the times when U.S. macroeconomic data, such as the Employment Report, the Producer Price Index, and quarterly GDP figures, are released. Glosten *et al.* (1993) found that indicator variables for October and January help to explain some of the dynamics of the conditional volatility of equity returns.

The section is organized as follows. Subsection 2 shows both the time-series and cross-sectional statistical properties of the Japanese stock market using the Tokyo Stock Exchange 500 Index (TOPIX500) index and its component parts. Subsection 3 describes the model specifications used to capture the time-series volatility and to evaluate the influence of the cross-sectional distributions of individual stock returns. Subsection 4 presents the empirical analysis, including estimation and model comparisons. Concluding remarks are given in Subsection 5.

4.2.2 Time-series and Cross-sectional Properties of the Japanese Stock Market

In this section, we use the index return series of the TOPIX500 and the individual stock return series included in the TOPIX500 index. Daily logarithmic returns are calculated for the period from January 1999 to December 2001, and the sample size is 738. The TOPIX500 index comprises the 500 stocks that are the most liquid and have the largest market capitalization of those on the Tokyo stock exchange (first section), and have been listed for a period of at least six months. It covers 90 percent of the market capitalization of the Tokyo stock exchange. Note that, because the composition of the TOPIX500 index has changed during the sample period owing to factors such as absorptions and bankruptcies, for each monthly observation, for simplicity, we use only those stocks that are included for the entire month. Figures 4.8 and 4.9 plot the TOPIX500 index and its logarithmic return $\times 100$. The TOPIX500 index exhibits both upward and downward movements, and there seems to be some heteroskedasticity in the logarithmic return series during the sample period. Table 4.4 presents statistical summary measures of the index returns of the TOPIX500. As is conventional, we calculate cross-sectional moments at the market level using all the components of individual stock returns in the same sample period. r_n^m denotes the m -th stock return at time n . That is, the empirical cross-sectional mean, variance, skewness, and

kurtosis are defined as:

$$\text{Mean:} \quad \mu_{cs,n} = E[r_n^m], \quad (4.43)$$

$$\text{Variance:} \quad \sigma_{cs,n}^2 = E_n[(r_n^m - \mu_{cs,n})^2], \quad (4.44)$$

$$\text{Skewness:} \quad Sk_{cs,n} = \frac{E_n[(r_n^m - \mu_{cs,n})^3]}{(\sigma_{cs,n}^2)^{\frac{3}{2}}}, \quad (4.45)$$

$$\text{Kurtosis:} \quad Kw_{cs,n} = \frac{E_n[(r_n^m - \mu_{cs,n})^4]}{(\sigma_{cs,n}^4)}. \quad (4.46)$$

Figures 4.9, 4.10, 4.11, and 4.12 plot histories of the cross-sectional mean, standard deviation, skewness, and kurtosis for all components of the TOPIX500 index. Note that in calculating these statistics, we exclude outliers (classified as observations that are more than three cross-sectional standard deviations away from the mean). Figure 4.10 suggests that the cross-sectional standard deviation is time varying. Figures 4.11 and 4.12 show that throughout the sample period, the cross-sectional distributions of the components of the TOPIX500 index tend to be skewed to the left (i.e., positively skewed) and fat-tailed (i.e., have negative kurtosis).

4.2.3 Markov Switching Models for Volatility

Models Specification

Markov Switching Heteroskedasticity Model

The first model is the Markov switching heteroskedasticity (MSH) model, which is a special case of the model introduced by Engel and Hamilton (1990) and Kim (1994). Let $y_n, n = 1, 2, \dots, N$, be the observed logarithmic return series. The MSH model is given by:

$$y_n = \mu + \varepsilon_n, \quad \varepsilon_n \sim N(0, \sigma_{s_n}^2), \quad (4.47)$$

where

$$\text{if } s_n = 0, \quad \sigma_{s_n} = \sigma_0, \quad (4.48)$$

$$\text{if } s_n = 1, \quad \sigma_{s_n} = \sigma_1, \quad \sigma_1 > \sigma_0 > 0. \quad (4.49)$$

Here s_n is a latent variable indicating the volatility regime that takes a value of zero or one. When $s_n = 0$, y_n follows a normal distribution with mean μ and variance σ_0^2 . When $s_n = 1$, y_n follows a normal distribution with mean μ and variance σ_1^2 . Since $\sigma_0 < \sigma_1$, $s_n = 0$ indicates a low volatility regime and $s_n = 1$ indicates a high volatility regime. We assume that the latent variable s_n shifts between zero and one according to a first-order Markov process with a constant transition probability as follows:

$$p_{ij} = \Pr(s_n = j | s_{n-1} = i), \quad i, j = 0, 1, \quad (4.50)$$

where $\sum_{j=0}^1 p_{ij} = 1$ for $i = 0, 1$. This transition probability p_{ij} is assumed to be homogeneous over time, and can be written in the form of the following transition probability matrix:

$$P = \begin{pmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{pmatrix} = \begin{pmatrix} p_{00} & 1 - p_{11} \\ 1 - p_{00} & p_{11} \end{pmatrix}. \quad (4.51)$$

Markov Switching ARCH Model

The second model is the Markov switching ARCH (MS-ARCH) model. Similar models have been proposed by Hamilton and Susmel (1994) and Cai (1994). Our MS-ARCH model has Markov switching structures in both the ARCH constant terms and the ARCH coefficients, and is given by:

$$y_n = \mu + \varepsilon_n, \quad \varepsilon_n | \Psi_{n-1} \sim N(0, h_n), \quad (4.52)$$

$$h_n = \gamma_{s_n} + \sum_{k=1}^K \alpha_{s_n}^k \varepsilon_{n-1}^2, \quad k = 1, 2, \dots, K, \quad (4.53)$$

where $\gamma_{s_n} \geq 0$, $\alpha_{s_n}^k \geq 0$. Ψ_{n-1} denotes the information up to time $n - 1$.

$$\text{if } s_n = 0, \quad \gamma_{s_n} = \gamma_0 \quad \text{and} \quad \alpha_{s_n}^k = \alpha_0^k, \quad (4.54)$$

$$\text{if } s_n = 1, \quad \gamma_{s_n} = \gamma_1 \quad \text{and} \quad \alpha_{s_n}^k = \alpha_1^k. \quad (4.55)$$

γ_{s_n} is the ARCH constant term, and $\alpha_{s_n}^k$ is the k -th ARCH coefficient. Both γ_{s_n} and $\alpha_{s_n}^k$ have Markov switching structures given by equations (4.54) and (4.55), and shift between different parameter values depending on the unobserved latent variable s_n . Since $\gamma_0 > \gamma_1$ is assumed, $s_n = 0$ and $s_n = 1$ represent low and high volatility regimes, respectively.

Markov Switching Models with an Exogenous Leverage Effect

The third model is the Markov switching model with an exogenous leverage effect. We propose the Markov switching heteroskedasticity model with an exogenous leverage effect (MSH-L) and the Markov switching ARCH model with an exogenous leverage effect (MSARCH-L). Let z_n be the exogenous variable that is observable at time n . Equations (4.47) and (4.53) respectively can be replaced by:

$$\sigma_{s_n} = \gamma_{s_n} + f(\lambda, z_{n-1}), \quad f(\lambda | z_{n-1}) > 0, \quad (4.56)$$

and

$$h_n = \gamma_{s_n} + \sum_{k=1}^K \alpha_{s_n}^k \varepsilon_{n-1}^2 + f(\lambda | z_{n-1}), \quad f(\lambda | z_{n-1}) > 0, \quad (4.57)$$

where $f(\cdot)$ denotes the arbitrary function with the constant unknown parameter λ . Note that $f(\cdot)$ is assumed to take non-negative values over time. The order of ARCH coefficients in MS-ARCH and MSARCH-L models is selected by the Akaike Information Criterion (AIC).

Estimations of States and Parameters

The MSH, MS-ARCH, MSH-L, and MSARCH-L models described in the previous subsection can be written in the state-space form as follows:

$$x_n = Fx_{n-1} + Gv_n, \quad (4.58)$$

$$y_n = Hx_n + w_n, \quad (4.59)$$

where y_n is the observation variable of the logarithmic index return; w_n is an observation noise, which is assumed to be zero in the MSH, MS-ARCH, MSH-L, and MSARCH-L models; $x_n \equiv (\mu_n, \varepsilon_n)^T$ is a (2×1) state vector; and $v_n \equiv (0, \varepsilon_n)^T$ is an (2×1) system noise vector in which

$$\varepsilon_n | \Psi_{n-1} \sim N(0, h_n), \quad (4.60)$$

and

$$E[v_n | \Psi_{n-1}] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad E[v_n v_n^T | \Psi_{n-1}] = \begin{bmatrix} 0 & 0 \\ 0 & h_n \end{bmatrix}, \quad (4.61)$$

where Ψ_{n-1} refers to information available up to time $n-1$, and F , G , and H are given by:

$$F = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 \end{bmatrix}. \quad (4.62)$$

Kim (1994) proposes the filter algorithm for the state-space model with Markov switching structures. Harvey, Ruiz and Sentana (1992) propose the filter algorithm for the state-space model with ARCH disturbances. We can combine these filter algorithms and apply the newly developed filter algorithm to the MSH, MS-ARCH, MSH-L, and MSARCH-L models. The filter algorithms for the state vector x_n and the unobserved variable s_n respectively are described as follows:

Filter for the state vector x_n

[Prediction]

$$x_{n|n-1}^{(i,j)} = Fx_{n-1|n-1}^{(i)}, \quad (4.63)$$

$$v_{n|n-1}^{(i,j)} = Fv_{n-1|n-1}^{(i)} F^T + GQ_n^{(j)} G^T, \quad (4.64)$$

where

$$x_{n|n-1}^{(i,j)} = E[x_n | \Psi_{n-1}, s_n = j, s_{n-1} = i], \quad (4.65)$$

$$v_{n|n-1}^{(i,j)} = E[(x_n - x_{n|n-1})(x_n - x_{n|n-1})^T | \Psi_{n-1}, s_n = j, s_{n-1} = i], \quad (4.66)$$

$$Q_n^{(j)} = E[v_n v_n^T | \Psi_{n-1}] = \begin{bmatrix} 0 & 0 \\ 0 & h_n \end{bmatrix}. \quad (4.67)$$

[Filter]

$$K_n^{(i,j)} = v_{n|n-1}^{(i,j)} H^T (H v_{n|n-1}^{(i,j)} H^T)^{-1}, \quad (4.68)$$

$$x_{n|n}^{(i,j)} = x_{n|n-1}^{(i,j)} + K_n^{(i,j)} (y_n - H x_{n|n-1}^{(i,j)}), \quad (4.69)$$

$$v_{n|n}^{(i,j)} = (I - K_n^{(i,j)} H) v_{n|n-1}^{(i,j)}, \quad (4.70)$$

where

$$x_{n|n}^{(i,j)} = E[x_n | \Psi_n, s_n = j, s_{n-1} = i], \quad (4.71)$$

$$v_{n|n}^{(i,j)} = E[(x_n - x_{n|n})(x_n - x_{n|n})^T | \Psi_n, s_n = j, s_{n-1} = i]. \quad (4.72)$$

However, to apply equations (4.68), we need to calculate $h_n = \gamma_{s_n} + \sum_{k=1}^K \alpha_{s_n}^k \varepsilon_{n-k}^2$, which is a function of the square of past unobserved shocks ε_{n-k}^2 , $k = 1, 2, \dots, K$. Thus, the Kalman filter cannot be used. Harvey, Ruiz and Sentana (1992) solve the problem of the unobserved ε_{n-k}^2 by replacing them with their conditional expectations:

$$h_n = \gamma_{s_n} + \sum_{k=1}^K \alpha_{s_n}^k E[\varepsilon_{n-k}^2 | \Psi_{n-k}]. \quad (4.73)$$

Given this specification, calculation of the terms $E[\varepsilon_{n-k}^2 | \Psi_{n-1}]$ is straightforward. Since we know that

$$\varepsilon_{n-k} = E[\varepsilon_{n-k} | \Psi_{n-k}] + (\varepsilon_{n-k} - E[\varepsilon_{n-k} | \Psi_{n-k}]), \quad (4.74)$$

it is easy to show that

$$E[\varepsilon_{n-k}^2 | \Psi_{n-k}] = E[\varepsilon_{n-k} | \Psi_{n-k}]^2 + E[(\varepsilon_{n-k} - E[\varepsilon_{n-k} | \Psi_{n-k}])^2], \quad (4.75)$$

where $E[\varepsilon_{n-k} | \Psi_{n-k}]$ is obtained from the last two elements of $x_{n-1|n-1}$, and its mean squared error $E[(\varepsilon_{n-k} - E[\varepsilon_{n-k} | \Psi_{n-k}])^2]$ is obtained from the last two diagonal elements of $v_{n-1|n-1}$. As a by-product of the filtering algorithm for the state-space x_n , we can obtain the conditional distribution y_n on s_n, s_{n-1} and Ψ_{n-1} as follows:

$$f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1}) = \frac{1}{\sqrt{2\pi r_n}} \exp \left\{ -\frac{(y_n - H x_{n|n-1}^{(i,j)})^2}{2r_n} \right\}, \quad (4.76)$$

with $r_n = H v_{n|n-1}^{(i,j)} H^T$ from the predictive distribution in equations (4.65) and (4.66).

Filter for the unobserved variable s_n **[Prediction]**

$$\Pr(s_n = j | \Psi_{n-1}) = \sum_{i=0}^1 \Pr(s_n = j, s_{n-1} = i | \Psi_{n-1}) \quad (4.77)$$

$$= \sum_{i=0}^1 \Pr(s_n = j, s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}). \quad (4.78)$$

[Filter]

$$\Pr(s_n = j | \Psi_n) = \sum_{i=0}^1 \Pr(s_n = j, s_{n-1} = i | \Psi_{n-1}, y_n) \quad (4.79)$$

$$= \sum_{i=0}^1 \frac{f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1}) \Pr(s_n = j, s_{n-1} = i | \Psi_{n-1})}{f(y_n | \Psi_{n-1})}, \quad (4.80)$$

where

$$f(y_n | \Psi_{n-1}) = \sum_{j=0}^1 f(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1}) \quad (4.81)$$

$$= \sum_{i=0}^1 \sum_{j=0}^1 f(y_n | s_n = j, s_{n-1} = i, \Psi_{n-1}) \Pr(s_n = j, s_{n-1} = i | \Psi_{n-1}). \quad (4.82)$$

[Smoother]

$$\Pr(s_n = j | \Psi_N) = \Pr(s_n = j | \Psi_n) \sum_{k=0}^1 \frac{\Pr(s_{n+1} = k | \Psi_N) \Pr(s_{n+1} = k | s_n = j)}{\Pr(s_{n+1} = k | \Psi_n)}. \quad (4.83)$$

To avoid an explosion in the number of Gaussian components, the approximation techniques proposed by Harrison and Stevens (1976) are used:

$$x_{n|n}^{(j)} = \frac{\sum_{i=0}^1 \Pr(s_{n-1} = i, s_n = j | \Psi_n) x_{n|n}^{(i,j)}}{\Pr(s_n = j | \Psi_n)}, \quad (4.84)$$

$$v_{n|n}^{(j)} = \frac{\sum_{i=0}^1 \Pr(s_{n-1} = i, s_n = j | \Psi_n) [v_{n|n}^{(i,j)} + (x_{n|n}^{(j)} - x_{n|n}^{(i,j)})(x_{n|n}^{(j)} - x_{n|n}^{(i,j)})^T]}{\Pr(s_n = j | \Psi_n)}. \quad (4.85)$$

The approximate log-likelihood is obtained as a by-product of running these filter algorithms. The approximate log-likelihood function for the unknown parameter vector θ is given by:

$$\begin{aligned} L(\theta) &= \sum_{n=1}^N \log f(y_n | \Psi_{n-1}) \\ &= \sum_{n=1}^N \left[\sum_{j=0}^1 \sum_{i=0}^1 \log f(y_n, S_n = j, S_{n-1} = i | \Psi_{n-1}) \right] \\ &= \sum_{n=1}^N \left[\sum_{j=0}^1 \sum_{i=0}^1 \log f(y_n | S_n = j, S_{n-1} = i, \Psi_{n-1}) \Pr(S_n = j, S_{n-1} = i | \Psi_{n-1}) \right] \end{aligned} \quad (4.86)$$

Several methods of estimating the unknown parameters are proposed, including the EM algorithm (Hamilton 1990) and the Gibbs sampler (Albert and Chib 1992). In this section, the quasi-maximum likelihood

method is preferred because of its computational ease. In addition, the Akaike Information Criterion (AIC) for selecting the model, the order of the ARCH coefficients, and the number of regimes, is defined by:

$$\text{AIC} = -2 \sum_{n=1}^N \log f(y_n | \Psi_{n-1}, \theta) + 2|\theta|, \quad (4.87)$$

where $|\theta|$ denotes the dimension of the vector θ (Kitagawa and Gersch 1996). Leroux and Puterman (1992) and Ryden (1995) discuss the method of model selection and identification for Markov switching models in detail.

4.2.4 Application and Model Comparison

Estimations of Markov Switching Models

First, we estimate and compare the models that do not include an exogenous leverage effect (MSH and MS-ARCH). Table 4.5 reports the estimated parameters and corresponding statistics. The orders of the ARCH coefficients are selected to minimize the AIC. ARCH coefficients in the MS-ARCH model are close to zero for both volatility regimes $s_n = 0$ and $s_n = 1$. Since there is no ARCH effect in the Markov switching structure, the MS-ARCH model is equivalent to the MSH model. Therefore, other estimated parameters and log-likelihood values are almost the same, and the MSH model can be selected as the best model by the AIC because of a reduction in the number of unknown parameters. In the MSH model, $\gamma_0 = 1.352$ and $\gamma_1 = 4.090$ represent variances of the low and high volatility regimes, respectively. Both volatility regimes have some persistence with self-loop transition probabilities $p_{00} = 0.980$ and $p_{11} = 0.928$ that the process does not change regime. Figure 4.13 plots the smoothed regime probability history for $s_n = 1$. Figure 4.13 indicates some persistence in both the low and high volatility regimes. The unconditional or stationary regime probabilities, $\Pr(s_n)$, are 0.781 and 0.219 for the low volatility regime, $s_n = 0$, and the high volatility regime, $s_n = 1$, respectively.

Secondly, we evaluate the exogenous leverage effects for the cross-sectional distribution of individual stock returns by using the Markov switching models with an exogenous leverage effect proposed in the previous subsection. We focus on the MSH-L model, since it is selected as the best model by the AIC. We examine three moments of the cross-section distribution; that is, the standard deviation, skewness, and kurtosis. The functions $f(\cdot)$ for the exogenous leverage effects in equations (4.56) and (4.57) are defined by:

$$f(\lambda | z_{n-1}) = \text{Max}(\lambda z_{n-1}, 0). \quad (4.88)$$

Here $f(\cdot) = \lambda z_{n-1}$ when $\lambda z_{n-1} > 0$. Otherwise, $f(\cdot) = 0$. Table 4.6 shows the estimated parameters and corresponding statistics for the MSH-L models. The AIC of the MSH- $L_{S,D}$, which has the cross-sectional standard deviation as the exogenous leverage factor, is less than that of the ordinary MSH model. Therefore, we can say that the cross-sectional standard deviation affects the time-series volatility dynamics with a positive correlation with $\lambda = 0.175$. In the low volatility regime, $s_n = 0$, of the MSH- $L_{S,D}$ model, the constant variance term $\gamma_0 = 0.613$ is smaller than the corresponding $\gamma_0 = 1.352$ of the MSH model. This implies that the major parts of the process in the low volatility regime can be controlled by the cross-sectional standard deviation of the previous period in the MSH- $L_{S,D}$ model. Figure 4.14 plots the smoothed regime probability history for $s_n = 1$ of the MSH- $L_{S,D}$. While the low volatility regime $s_n = 0$ with the self-loop transition probability $p_{00} = 0.986$ has almost the same persistence as the MSH model, persistence of the high volatility regime $s_n = 1$ with the self-loop transition probability $p_{11} = 0.857$ decreases in the MSH- $L_{S,D}$ model. The unconditional or stationary regime probabilities, $\Pr(s_n)$, change to 0.911 and 0.089 for the low and high volatility regimes, $s_n = 0$ and $s_n = 1$, respectively. Figure 4.13 confirms the characteristics of regime persistence. On the other hand, the MSH- $L_{S,k}$ model with cross-sectional skewness as the exogenous leverage factor hardly improves on the AIC of the MSH model with no exogenous leverage effect. In addition, the estimated $\lambda = -3.143$ is not statistically significant given a t -value of 1.456. The last MSH- $L_{K,w}$ model with cross-sectional kurtosis as the exogenous leverage factor does not have a higher log-likelihood value than that of the MSH model with no exogenous leverage effect. It can be said that the skewness and kurtosis of the cross-sectional distribution of the TOPIX500 index returns do not particularly influence the time-series volatility process.

Model Comparisons with No-switching ARCH-type Models

Finally, we compare the Markov switching models with no-switching ARCH-type models. Table 4.7 shows the log likelihood, the number of parameters, and the AIC of the alternative models; GARCH, GARCH with t distribution of the noise term (GARCH- t), exponential GARCH (EGARCH) by Nelson (1991), threshold GARCH (TGARCH) by Zakuian (1990) and GARCH in mean (GARCH-M) by Engle, Lilien and Robine (1987). The order of all ARCH and GARCH coefficients is the same for simplicity. Here, GARCH- t is selected as the best AIC model among these alternative models without the Markov switching structure. The AIC of GARCH- t is closely similar to that of the HMS model. It can be said that both models capture the characteristics of volatility — fat-tailed distribution and persistence — with different specifications, respectively.

4.2.5 Conclusions

This section introduced the Markov switching heteroskedasticity and ARCH models with exogenous leverage effects (MSH-L and MSARCH-L), which were used to describe the time-series volatility of stock market index returns and to evaluate the leverage effect of moments of the cross-sectional distribution. In the empirical analysis, we applied some standard models to the daily TOPIX500 index return series. The Markov switching heteroskedasticity model (MSH), which was selected as the best model by the AIC, was compared with MSH-L models that included either the standard deviation, the skewness or the kurtosis of the individual stock return distributions as an exogenous leverage factor. We found that the cross-sectional standard deviation affected the time-series volatility dynamics, while the higher moments (skewness and kurtosis) did not have significant impacts.

Table 4.4: Basic statistical summary of the TOPIX500 index returns.

Statistics	Overall	1999	2000	2001
Mean	-0.00028	0.19861	-0.10753	-0.08942
Std Deviation	1.40043	1.21023	1.40372	1.54965
Skewness	-0.14081	-0.17672	-0.28888	0.09637
Kurtosis	1.63498	0.90875	1.74587	1.63708
Sample Size	738	244	248	246

Table 4.5: Estimated parameters and statistics of models with no exogenous leverage effect.

	MSH		MSARCH(1)	
	Estimation	(Standard error)	Estimation	(Standard error)
μ	-0.003	(0.000)	-0.003	(0.000)
γ				
γ_0	1.352	(0.123)	1.352	(0.123)
γ_1	4.090	(0.802)	4.089	(0.803)
α				
α_0			0.000	(0.000)
α_1			0.000	(0.000)
β				
p_{00}	0.980	(0.011)	0.980	(0.011)
p_{11}	0.928	(0.041)	0.928	(0.041)
LL	-1273.47		-1273.47	
AIC	2556.94		2560.94	

Table 4.6: Estimated parameters and statistics of MSH-L models.

	MSH- $L_{S.D.}$		MSH- L_{Skw}		MSH- L_{Kurt}	
	Estimation	(Standard error)	Estimation	(Standard error)	Estimation	(Standard error)
μ	-0.017	(0.045)	-0.008	(0.031)	-0.003	(0.038)
γ_0	0.613	(0.280)	1.303	(0.113)	1.352	(0.123)
γ_1	4.485	(2.350)	4.173	(0.832)	4.089	(0.801)
λ	0.175	(0.059)	-3.143	(2.158)	0.000	(0.058)
p_{00}	0.986	(0.011)	0.981	(0.010)	0.980	(0.011)
p_{11}	0.857	(0.132)	0.923	(0.041)	0.928	(0.041)
LL	-1269.50		-1272.15		-1273.47	
AIC	2551.01		2556.29		2558.94	

Note: The AIC of the MSH model with no exogenous leverage effect is 2556.94.

Table 4.7: Log likelihood, number of parameters and AIC of no-switching models.

Models	Log likelihood	# of parameters	AIC
ARCH	-1297.40	3	2600.80
GARCH	-1283.85	4	2575.71
GARCH-t	-1274.19	4	2556.38
EGARCH	-1284.32	4	2576.64
TGARCH	-1280.73	5	2571.46
GARCH-M	-1282.63	5	2575.26

Figure 4.8: TOPIX500 index: 1999/1 – 2001/12.

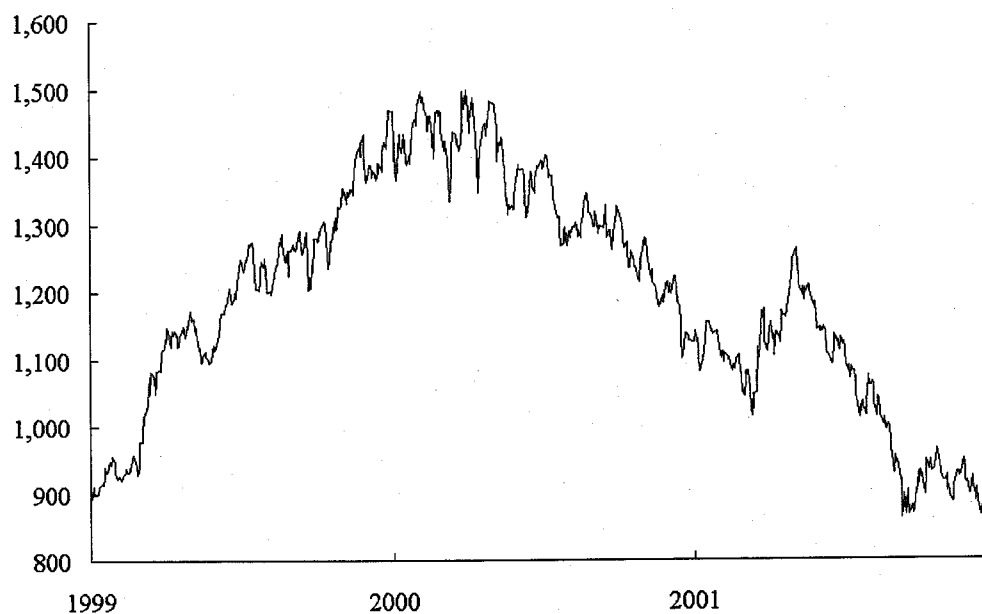


Figure 4.9: Logarithmic return of daily TOPIX500 index: 1999/1 – 2001/12.

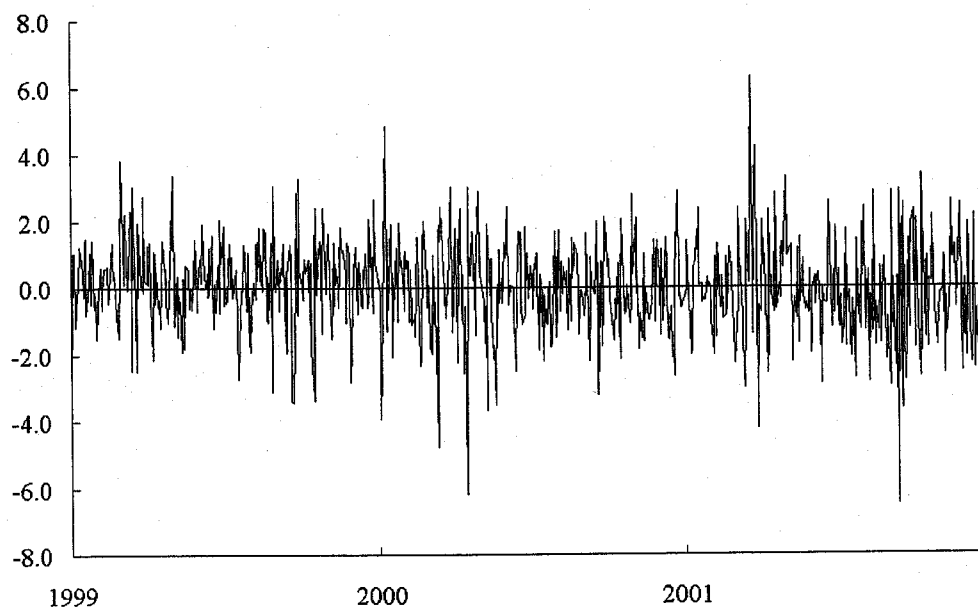


Figure 4.10: Cross-sectional standard deviation of TOPIX500.

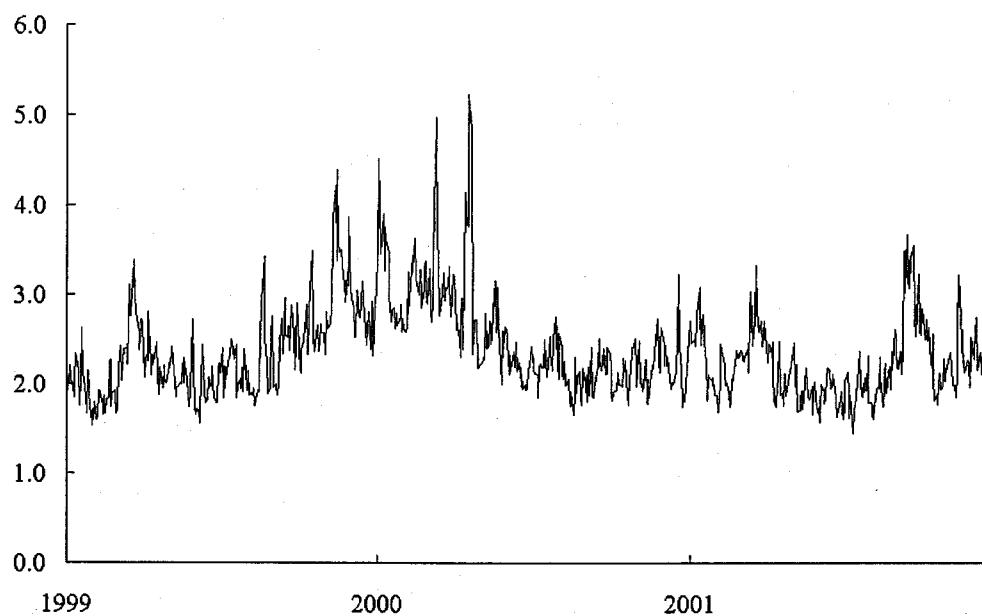


Figure 4.11: Cross-sectional skewness of TOPIX500.

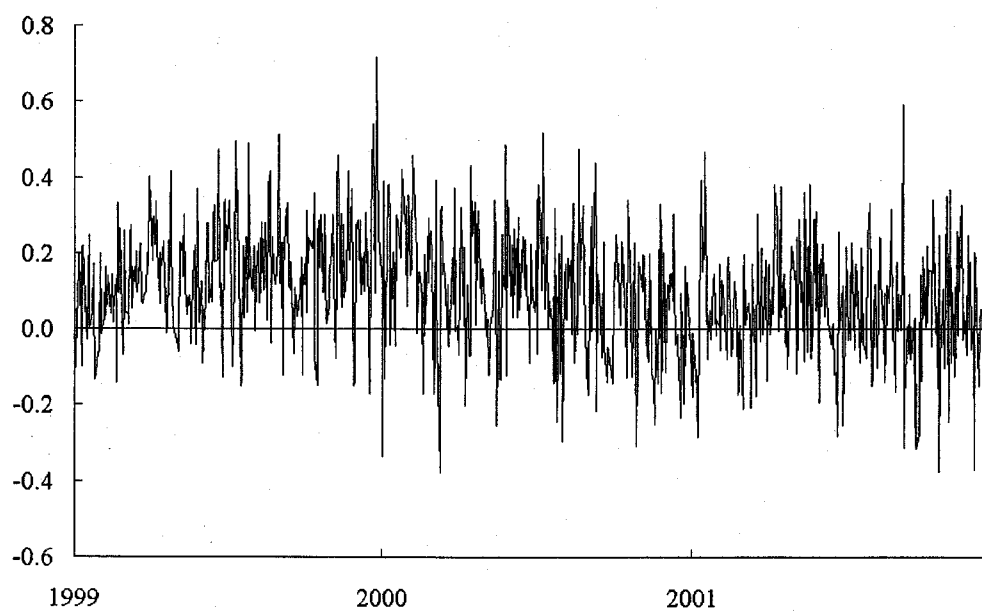


Figure 4.12: Cross-sectional kurtosis of TOPIX500.

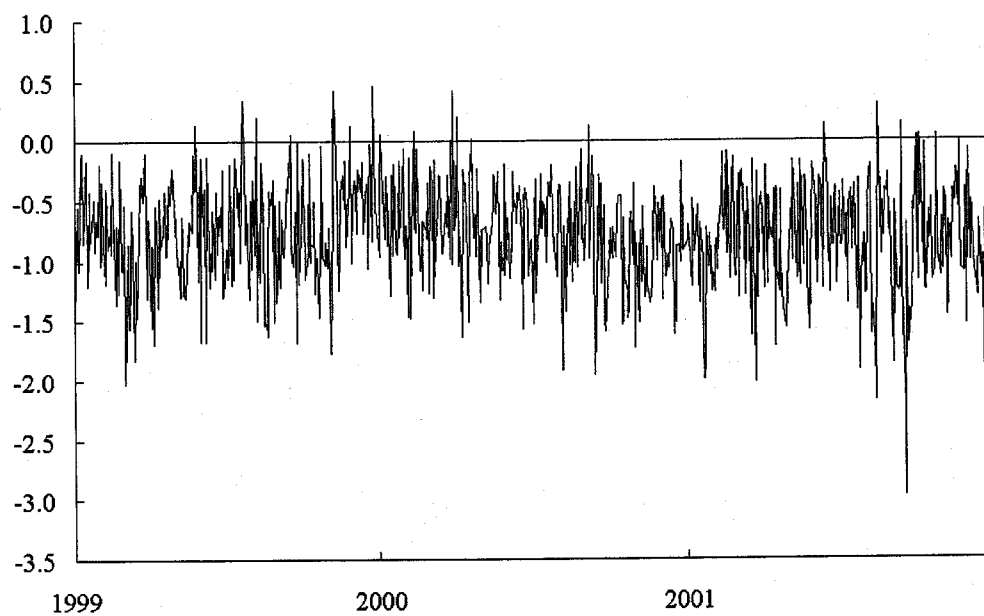
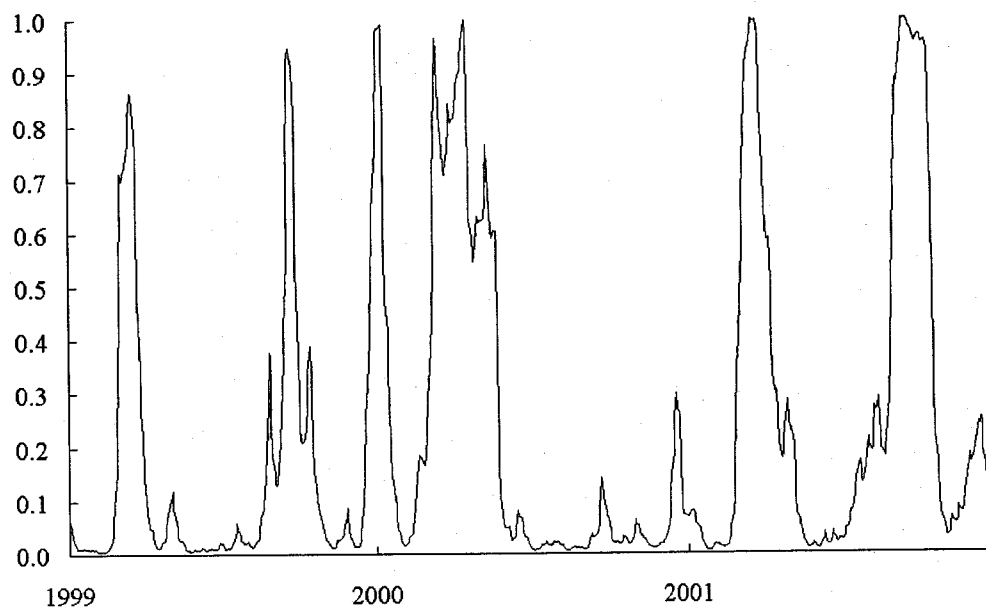
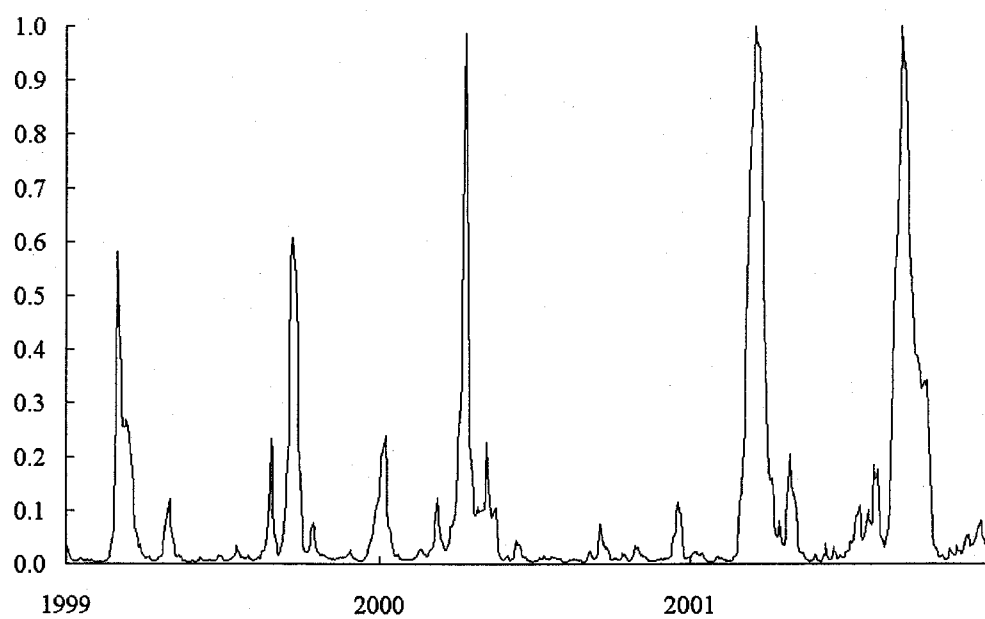
Figure 4.13: Smoothed regime probability of $S_n = 1$ (high volatility regime) of MSH model.

Figure 4.14: Smoothed regime probability of $S_n = 1$ (high volatility regime) of MSH- $L_{S,D}$ model.



4.3 Analysis of Japanese Business Cycles

— Analysis of Japanese Business Cycles Using the Multivariate Stochastic Trend Model with Simultaneous Markov Switching Drift

4.3.1 Introduction

The Japanese government releases official business cycle reference dates — the turning points between expansion and contraction of the business cycle. These dates are determined by the objective criterion of the *historical diffusion index* (Bry and Boschan 1971). To calculate this index, economic time series that represent widely differing activities or sectors of the Japanese economy are chosen to fit its business cycle chronology. The coincident indicators for the historical diffusion index are listed in Table 1 and plotted in Figure 4.15. These series were last revised in June 1996. The crucial drawback of this historical diffusion index method is that it generally takes a long time — about one year or more — to determine the turning point dates. For example, it was not until April 2000 that the Japanese government could determine “April 1999” as the end of the last economic expansion period beginning in April 1997.

Recently, types of stochastic approaches have been proposed to estimate and forecast the business cycle. The Markov switching model is one of the most popular tools used to analyze business cycles using a univariate economic model. Hamilton (1989) first proposed an autoregressive Markov switching model and applied it to U.S. quarterly GNP data. Variants of Hamilton’s model are used to capture the characteristics of business cycles and other non-linear economic time series. Filardo (1994) introduced time-varying transition probabilities into Hamilton’s model, and applied it to the U.S. monthly index of industrial production. Phillips (1991) and Kontolemis (2001) extended the univariate Markov switching model to the multivariate version, and applied it to four given time series from the U.S. composite index of coincident indicators — industrial production, real personal income, real manufacturing and trade sales, and total employee-hours in non-agricultural establishments. Also popular is Stock and Watson’s (1991) dynamic factor model, which estimates co-movement by obtaining an unobserved single common factor from a set of economic time series using the Kalman filter technique. Kim and Nelson (1998) and Kim and Yoo (1995) generalize the dynamic factor model with a Markov switching structure (DFMS model). While several recent studies of the U.S. business cycle have used multivariate stochastic models, few have focused on the Japanese business cycle.¹

In this section, we adopt the Burns and Mitchell (1946) definition of business cycles. Two key features

¹Kaufman (2000) applies the DFMS model to Japanese quarterly business cycle analysis. However, the model seems to estimate business cycle phases poorly.

of this definition are co-movement among economic time series and division into the separate phases of expansion and contraction. The DFMS model captures these two characteristics of business cycles simultaneously with a Markov switching unobserved common index. In this section, we apply the multivariate Markov switching stochastic trend model with Markov switching drift (MSST) to multivariate economic time series in order to estimate the Japanese business cycle phases. Our multivariate MSST model is a restricted special case of the DFMS model. The main difference between the DFMS model and the multivariate MSST model derives from this restriction. Our multivariate MSST model assumes that all the stochastic trends of economic time series shift simultaneously between probability distributions with high and low means. This simultaneous shift depends on a common transition probability mechanism. We view the point in time of this shift as the turning point of the business cycle. The multivariate MSST model focuses on estimating and forecasting the business cycle phases, while the DFMS model is used to estimate an unobserved common trend among economic time series with a Markov switching structure. When the amplitudes of time series clearly differ (as in Figure 4.15), it is not appropriate to use the DFMS model to estimate and forecast business cycles.

The purpose of this section is to evaluate the out-of-sample estimating accuracy of the monthly Japanese business cycle phases obtained by using the multivariate MSST model, and to compare this performance with that of the univariate MSST model. The subset of observed variables for the multivariate MSST model is selected to minimize the quadratic probability score (QPS) in the in-sample period.

The section is organized as follows. Subsection 2 describes the model specification and the state and parameter estimation methods for the multivariate MSST model. Subsection 3 presents the measure used to evaluate the accuracy of estimates of business cycle phases, and describes the variable selection method. In subsection 4, we determine the best multivariate MSST model for the in-sample period, and compare the multivariate MSST model with the univariate model according to estimating performance in the out-of-sample period. Concluding remarks are given in subsection 5.

4.3.2 Multivariate Stochastic Trend Model with Simultaneous Markov Switching Drift

Model specification

In this subsection, we describe a state-space specification of the multivariate stochastic trend model with simultaneous Markov switching drift (MSST model). Let $y_n \equiv (y_{1,n}, y_{2,n}, \dots, y_{l,n})^T$, $n = 1, 2, \dots, N$, be

a l -dimensional observed variable at time n . The observation model is given by:

$$y_n = t_n + v_n, \quad v_n \sim N(0, \Sigma), \quad (4.89)$$

where $t_n \equiv (t_{1,n}, t_{2,n}, \dots, t_{l,n})^T$ represents the l -dimensional stochastic trend component, and v_n is the observation noise, which is identically and independently distributed with zero mean and a $(l \times l)$ variance-covariance matrix $\Sigma \equiv \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_l^2)$. The system model for generating the l -dimensional stochastic trend component t_n is given by:

$$t_n = \mu + t_{n-1} + w_n, \quad w_n \sim N(0, \Omega), \quad (4.90)$$

where w_n is the system noise, which is identically and independently distributed with zero mean and a $(k \times k)$ variance-covariance matrix $\Omega \equiv \text{diag}(\tau_1^2, \tau_2^2, \dots, \tau_l^2)$. Here, a k -dimensional Markov switching drift μ is denoted as follows:

$$\mu = (1 - s_n) \begin{bmatrix} \mu_{0,1} \\ \mu_{0,2} \\ \vdots \\ \mu_{0,l} \end{bmatrix} + s_n \begin{bmatrix} \mu_{1,1} \\ \mu_{1,2} \\ \vdots \\ \mu_{1,l} \end{bmatrix}, \quad \mu_{0,m} > \mu_{1,m}, \quad m = 1, 2, \dots, l, \quad (4.91)$$

where s_n represents the unobserved Markov chain according to a discrete first-order two-regime Markov process, and takes a value of 0 or 1. s_n indicates the business cycle phases of the unobserved regime: $s_n = 0$ for an expansionary regime and $s_n = 1$ for a contractionary regime. t_n is updated by the Markov switching drift μ which is dependent on s_n . t_n follows an upward trend with a positive drift in $s_n = 0$ for the expansionary regime, and follows a downward trend with a negative drift in $s_n = 1$ for the contractionary regime. The transition probability of s_n is assumed to be time-homogeneous and is given by:

$$p_{ij} = \Pr(s_n = j | s_{n-1} = i), \quad (4.92)$$

with $\sum_{j=0}^1 p_{ij} = 1$ for all $i = 0, 1$. It is convenient to represent the transition probabilities in a (2×2) transition probability matrix:

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} p_{00} & 1 - p_{11} \\ 1 - p_{00} & p_{11} \end{pmatrix}.$$

We assume that the observed variables $y_{i,n}$, $i = 1, \dots, l$, shift between the two regimes ($s_n = 0$ and $s_n = 1$) simultaneously. This assumption allows us to extract the unobserved co-movement among the observed variables $y_{i,n}$ as binomial values of 0 and 1.

Estimations of States and Parameters

To evaluate how closely the multivariate MSST model can track the official business cycle phase, we must estimate the filtered regime probability $\Pr(s_n = j | \Psi_n)$, which is the probability that the observed variable y_n remains in regime $s_n = j$ at time n given that the information Ψ_n is known up to time n . The filter algorithm for a Markov switching state-space model was proposed by Kim (1994). Kim's filter algorithm has been derived from a Gaussian-sum filter (Sorenson and Alspach 1971) and from a non-Gaussian filter (Kitagawa, 1987).

[Prediction]

$$\begin{aligned} \Pr(s_n = j | \Psi_{n-1}) &= \sum_{i=0}^1 \Pr(s_n = j, s_{n-1} = i | \Psi_{n-1}) \\ &= \sum_{i=0}^1 \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}) \\ &= \sum_{i=0}^1 p_{ij} \Pr(s_{n-1} = i | \Psi_{n-1}). \end{aligned} \quad (4.93)$$

[Filter]

$$\begin{aligned} \Pr(s_n = j | \Psi_n) &= \Pr(s_n = j | \Psi_{n-1}, y_n) \\ &= \frac{p(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1})}{p(y_n | \Psi_{n-1})}, \end{aligned} \quad (4.94)$$

where:

$$p(y_n | s_n = j, \Psi_{n-1}) = \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{(y_n - t_n)^T \Sigma^{-1} (y_n - t_n)}{2} \right\}, \quad (4.95)$$

$$p(y_n | \Psi_{n-1}) = \sum_{j=0}^1 p(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1}). \quad (4.96)$$

As a by-product of running the above filter algorithm, we can easily obtain the log-likelihood function as follows:

$$l(\theta) = \sum_{n=1}^N \log p(y_n | \Psi_{n-1}), \quad (4.97)$$

where an unknown parameter vector θ is defined by:

$$\theta \equiv \left\{ \mu_{0,1}, \dots, \mu_{0,l}, \mu_{1,1}, \dots, \mu_{1,l}, \sigma_1, \dots, \sigma_l, \tau_1, \dots, \tau_l, p_{00}, p_{11} \right\}.$$

Several methods of estimating the unknown parameters have been proposed, such as the EM algorithm (Hamilton 1990) and the Gibbs sampler (Albert and Chib 1993). In this section, the maximum likelihood method is preferred because of its computational ease.

4.3.3 Variable Selection

Evaluating the accuracy of estimation

In this subsection, we describe the measure used to evaluate the accuracy with which business cycle phases are estimated. An important issue that needs to be defined clearly is “true business cycle reference dates”. Although the Japanese government releases the official business cycle reference date, it is not necessarily “true” in a statistical sense. However, the evaluation and comparison of different models requires a true value as a benchmark. In this section, we assume, as previous studies have, that the official business cycle reference dates published by the Japanese government are provisionally true. While this issue could be considered further, to do so is beyond the scope of the present section.

To evaluate the accuracy with which the MSST model estimates business cycle phases, we consider the quadratic probability score (QPS) proposed by Brier (1950) as follows:

$$QPS = \frac{1}{N} \sum_{n=1}^N Q_n = \frac{1}{N} \sum_{n=1}^N 2(P_n - R_n)^2. \quad (4.98)$$

Suppose that we have a time series of N probabilities $\{P_n\}_{n=1}^N$, where P_n is the *filtered* regime probability $\Pr(s_n|\Psi_n)$ in equation (4.94), which is the probability of remaining in $s_n = 1$ (the contractionary phase) at time n . Similarly, let $\{R_n\}_{n=1}^N$ be the corresponding time series of realizations released officially by the Japanese government — that is, R_n equals one if the business cycle phase is in contraction at time n and equals zero otherwise. In this section, we do not focus on forecasting, but on estimation of the business cycle phase. Therefore, when we know the information Ψ_n at time n , we calculate and evaluate the filtered regime probability $\Pr(s_n|\Psi_n)$ at time n . The QPS ranges in value from zero to two. A score of zero indicates that the model estimates the business cycle phases perfectly. The larger the score, the larger the estimating errors.

Search method

Exhaustive search is the only technique that is certain to find the variable subset with the best evaluation criteria. However, the problem is that exhaustive search is a computationally intractable technique. Since we use thirteen time series as candidates for a set of observed variables in this empirical analysis, evaluation criteria would have to be calculated for over 8,000 possible subsets. Clearly, it is not practical to conduct an exhaustive search for a subset for the best model. To cope with this computational problem, we use a forward-backward stepwise selection algorithm (Efroymson 1960, 1966), which saves much computation time.

We show step-by-step how to use the procedure for a forward-backward stepwise selection algorithm

for selecting the subset of m variables from M time series to minimize the QPS. Here, the variable dimension m is unknown.

- (1) For an initial small m_0 ($= 1, 2$ or 3), compute the QPSs of all possible m_0 -variable MSST models. Note that the number of models is ${}_MC_{m_0}$. Select L sets of models with the smallest QPSs.
- (2) Compute the QPSs of $L \times (M - m)$ sets of the $(m + 1)$ -variable MSST models by adding a new variable to each of the L sets of the m -variable models with the smallest QPSs. If the QPS of the best $(m + 1)$ -variable MSST model is larger than that of the best m -variable MSST model, stop the iteration. Otherwise, go to (3) below.
- (3) Compute the QPSs of $L \times (m + 1)$ sets of the m -variable MSST models by deleting one variable from each of the L sets of the $(m + 1)$ -variable models selected in (2). Select L sets of the m -variable models with the smallest QPSs. If the QPS of the best m -variable MSST model is smaller than that of the best $(m + 1)$ -variable MSST model obtained in (2), use L sets of the m -variable models as the current models and go to (2). Otherwise, use L sets of the $(m + 1)$ -variable models as the current models, set $m = m + 1$ and go back to (2).

4.3.4 Application and Model Comparison

Data description

The set of observed variables to be used in this empirical analysis of the Japanese business cycle are thirteen monthly time series from the composite indices of coincident indicators. These are listed in Table 4.8 and plotted in Figure 4.15. The sample period is from January 1973 to December 2000, and the sample size for each time series is 336. These historical data are available from the website of the Economic and Social Research Institute in the Cabinet Office (<http://www.esri.cao.go.jp/index-e.html>). In the parameter estimation, each time series is first standardized by subtracting the sample mean from each observation and dividing by the standard deviation, and is then multiplied by 100.

In-sample variable selection and estimation

Since the latest revisions to the group of economic time series used to calculate the historical diffusion index were made by the Japanese government in June 1996, we chose the period from January 1973 to June 1996 for in-sample variable selection and estimation. Economic time series are selected in order to fit the past business cycle chronology. For the in-sample period, we searched among the thirteen time series to find the subset of variables for the multivariate MSST model that minimizes the QPS by using

the forward-backward stepwise selection algorithm. The initial number of variables was set at $m_0 = 2$. The five-variable MSST model for LIPC, INSWH, SDS, ISWT, and ISSMSE1 was selected as the best MSST model, with a QPS=0.180. This QPS is smaller than any QPS among the univariate MSST models (0.281 – 0.806), which are listed in the second-last column of Table 4.8. Table 4.9 shows the variables, estimated parameters, and associated statistics for this QPS-minimizing five-variable MSST model in the in-sample period. Figure 4.16 shows the filtered regime probability of contraction implied by this five-variable MSST model, and its shaded areas represent contractions of the official business cycle. Figure 4.16 shows that the five-variable MSST model is successful in tracking the official business cycle reference dates, except for the period from April 1981 to March 1982.

Next, we discuss some implications of the estimated parameters for the five-variable MSST model. The drifts of the stochastic trend components are $\mu_{0,k} > 0$ and $\mu_{1,k} < 0$ for all five variables ($k = 1, 2, \dots, 5$). This implies that the five time series shift simultaneously between the stochastic trend process regimes with negative drifts (contractions) and those with positive drifts (expansions), according to the first-order Markov process. Recall that regimes $s_n = 0$ and $s_n = 1$ indicate the business cycle expansionary and contractionary phases, respectively. The estimated transition probabilities p_{00} and p_{11} are 0.954 and 0.916, respectively. The expected durations of an expansion and a contraction are 21.70 months and 11.95 months, respectively.² The fact that both transition probabilities exceed 0.90 suggests that both business cycle phases exhibit persistence during the sample period. The fact that the former transition probability exceeds the latter implies that expansions tend to continue longer than contractions.

Implication of identification errors

On the other hand, as Figure 4.15 shows, business cycles identified by the multivariate Markov switching model differ from those of the Cabinet Office in some periods. In particular, our model fails to identify business cycles with probabilities of more 50 percent, in a period that continues for eight months from 1981/6 to 1982/1 (indicating this period by drawing a circle in Figure 4.15). Here, we consider under what economic conditions such a difference between our statistical models and the Cabinet Office occurs. The period from 1981/6 to 1982/1 is included in the period of synchronized global economic slowdown. Within this period, where our model did not demonstrate goodness of fit, the second oil shock occurred due to the OPEC oil price rises. As a result, the global recession deepened, and there was little sign of an

²The expected duration of regime j can be derived as:

$$E(d) = \sum_{j=1}^{\infty} j \Pr(d = j) = \frac{1}{1 - p_{jj}}.$$

economic recovery on the horizon. It can be said that our multivariate Markov switching model identified the period of temporal economic expansion as one independent business cycle. It is not appropriate to conclude that such an identification error is due to inadequate modelling of business cycles. Paradoxically, it is possible to see that this error implies the failures of the Cabinet Office's identification of business cycles.

Out-of-sample estimation performance

In this subsection, we compare the out-of-sample estimating performance of the multivariate MSST model with that of the univariate MSST model. To evaluate the out-of-sample estimating performance of the MSST models, we calculate the out-of-sample QPSs for both the multivariate and univariate MSST models with given subsets of variables and estimated parameters in the in-sample period.

The final columns of Table 4.8 show the in-sample and out-of-sample QPSs for each economic time series for the univariate MSST model. A smaller QPS indicates a more accurate estimating performance by the univariate MSST models for the corresponding variable. A * denotes any QPS that is greater than the unconditional $QPS_0 = 0.497$, which is given by:

$$QPS_0 = \frac{1}{N} \sum_{n=1}^N Q_{n,0} = \frac{1}{N} \sum_{n=1}^N 2(P_0 - R_n)^2, \quad (4.99)$$

where $P_0 = 0.463$ is the experience probability, or the unconditional probability, which is the ordinary probability of contraction for all sample periods. A model with a QPS that is smaller than QPS_0 improves estimation performance. The final columns of Table 4.9 show the in-sample and out-of-sample QPSs of the five-variable MSST model, which is selected to minimize the in-sample QPS. This out-of-sample $QPS=0.171$ is smaller than the minimum $QPS=0.208$ of the univariate MSST model for ICRM. The business cycle phases in the out-of-sample period can be estimated more accurately with the five-variable MSST model than with the univariate models. Therefore, it can be said that our multivariate Markov switching model with the variable selection process is a more accurate method of estimating business cycle phases. Figure 4.17 plots the filtered regime probability of a contraction in the out-of-sample period from this five-variable MSST model. While the five-variable MSST model successfully tracks the business cycle reference date overall, Figure 4.17 also reveals estimating errors (1997/4–7, 1999/3–4) with regard to contractions from April 1997 to April 1999 in the out-of-sample period. One way of increasing out-of-sample estimating accuracy may be to expand the population of observed variables. However, further investigation is beyond the scope of this section, since we focus on comparisons between univariate and multivariate models.

4.3.5 Conclusions

This study estimated Japanese business cycle phases (expansion and contraction) using the multivariate stochastic trend model with Markov switching drift (MMST model). The multivariate MSST model can capture the co-movement of some economic time series. For all the time series, the common two-regime and first-order Markov chain governs the shift of each time series between stochastic processes with two different drifts. The five-variable MSST model for the time series LIPC, INSWH, SDS, ISWT, and ISSMSE1 minimizes the in-sample quadratic probability score. . Some identification errors with this model paradoxically imply the failure of business cycle detection by the Cabinet Office. In the out-of-sample period, this five-variable model performs better in estimating the business cycle than do the univariate MSST models.

Table 4.8: Coincident indicators and corresponding QPSs.

Variable Name	Symbol	Remarks	QPS ¹	QPS ²
1: Index of Industries Production (Mining and Manufacturing)	IIP	1995=100	*0.476	0.342
2: Index of Consumption of Raw Materials (Manufacturing)	ICRM	1995=100	0.248	0.208
3: Large Industrial Power Consumption	LIPC	million kwh	0.281	0.293
4: Index of Capacity Utilization Ratio (Manufacturing)	ICUR	1995=100	*0.683	*0.819
5: Index of Non-Scheduled Worked Hours	INSWH	1995=100	0.331	0.374
6: Index of Producer's Shipment of Investment Goods	IPSIG	1995=100	0.470	0.241
7: Sales at Department Stores	SDS	percent change	0.454	0.499
8: Index of Sales in Wholesale Trade	ISWT	percent change	*0.671	*0.792
9: Operating Profits (All Industries)	OP	100 million yen	*0.778	*0.888
10: Index of Sales in Small and Medium Sized Enterprises (Manufacturing)	ISSMSE1		0.379	0.357
11: Index of Shipment in Small and Medium Sized Enterprises	ISSMSE2	1995=100	*0.583	*0.607
12: Index of Wholesale Price in Small and Medium Sized Enterprises	IWPSMSE	1995=100	*0.806	*0.889
13: Effective Job Offer Rate (Excluding New School Graduates)	EJOR	times	*0.542	*0.833

Note:

1. QPS¹ is the quadratic probability score for in-sample period (1973/1 to 1996/6).
2. QPS² is the quadratic probability score for out-sample period (1996/7 to 2000/12).
3. The unconditional QPS₀¹ and QPS₀² are 0.476 and 0.481, respectively.
4. A * marks any QPS that is greater than the unconditional QPS₀.

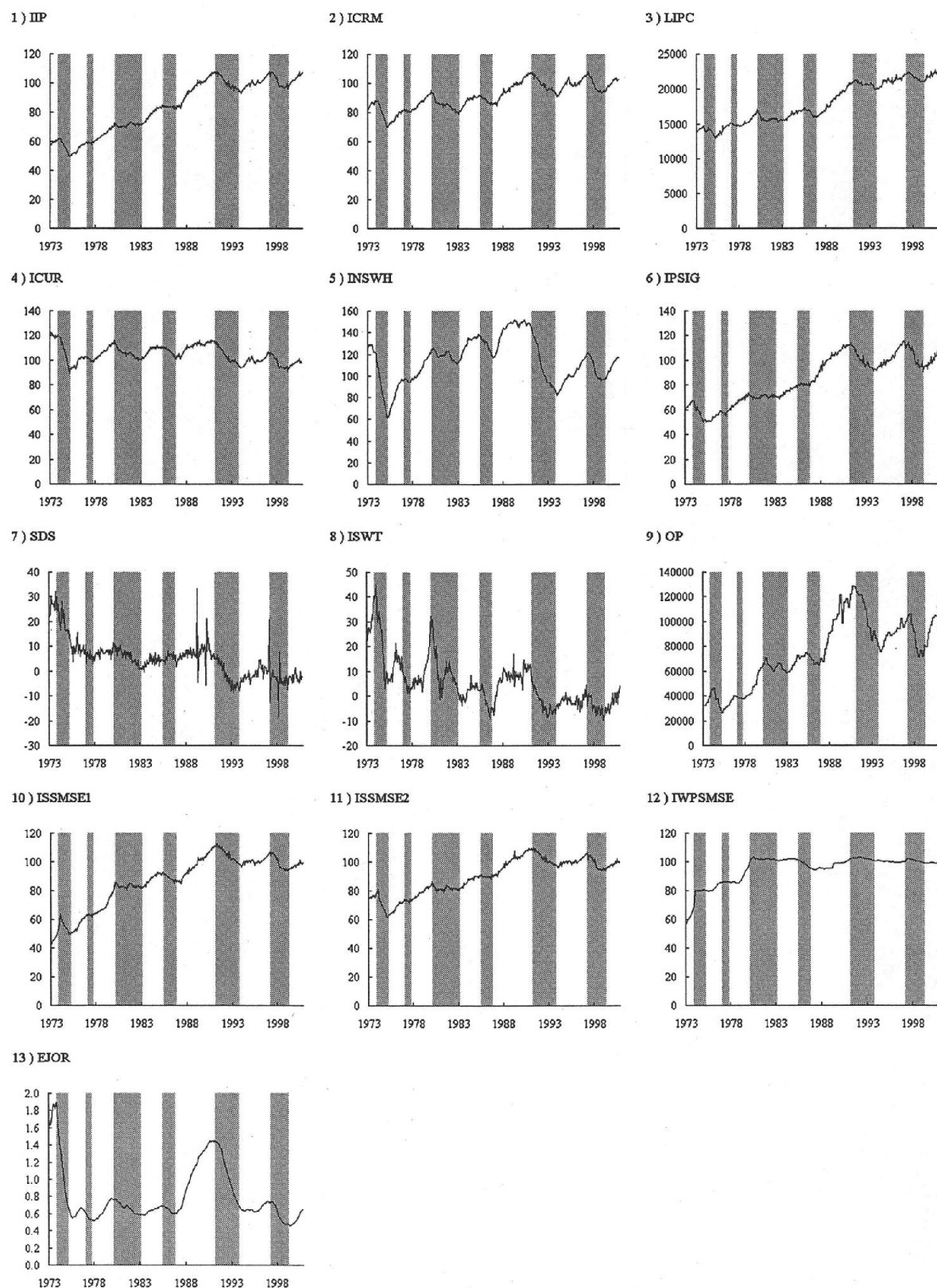
Table 4.9: Estimated parameters and statistics for the multivariate MSST model.

Symbol	μ_{k0}	μ_{k1}	τ_k	σ_k	P_{00}	P_{11}	LL	QPS^1	QPS^2
LPC	3.01 (0.40)	-2.25 (0.54)	4.83 (0.56)	5.23 (0.44)	0.854 (0.018)	0.916 (0.031)	-5753.85	0.180	0.171
INSWH	3.83 (0.62)	-7.32 (0.87)	7.65 (0.34)	0.01 (0.00)					
SDS	1.31 (0.93)	-4.83 (1.23)	10.46 (1.45)	37.55 (1.76)					
ISWT	4.34 (1.59)	-9.91 (2.06)	19.70 (1.83)	15.63 (1.60)					
ISSMSE1	2.87 (0.40)	-1.79 (0.54)	4.88 (0.44)	3.79 (0.37)					

Note:

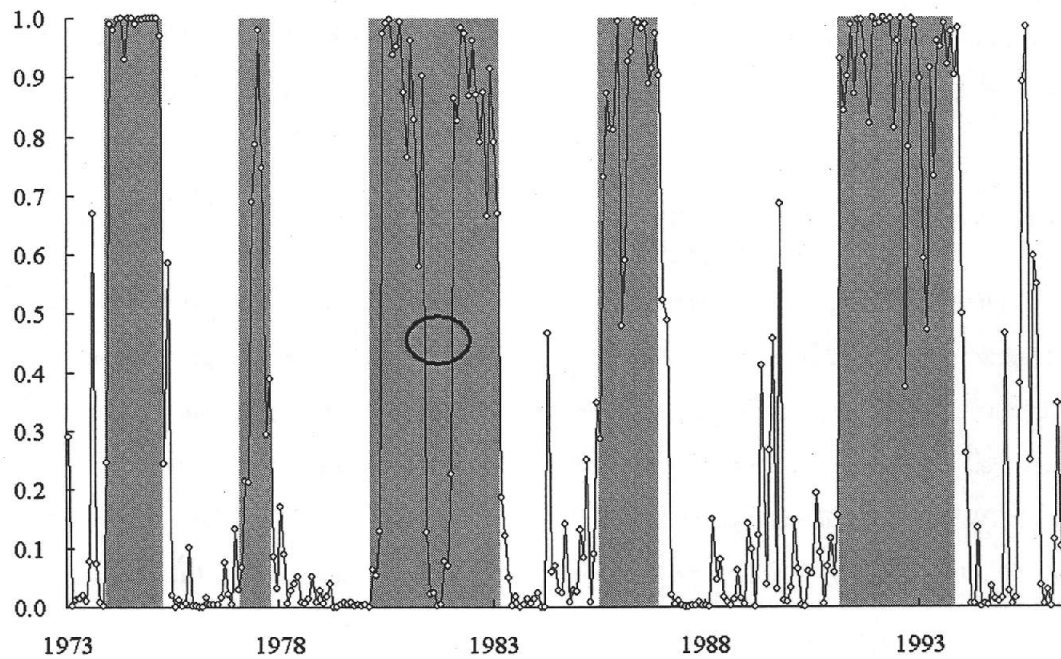
1. QPS^1 is the quadratic probability score for in-sample period (1973/1 to 1996/6).
2. QPS^2 is the quadratic probability score for out-sample period (1996/7 to 2000/12).
3. The unconditional QPS^1_0 and QPS^2_0 are 0.476 and 0.481, respectively.

Figure 4.15: Coincident indicators.



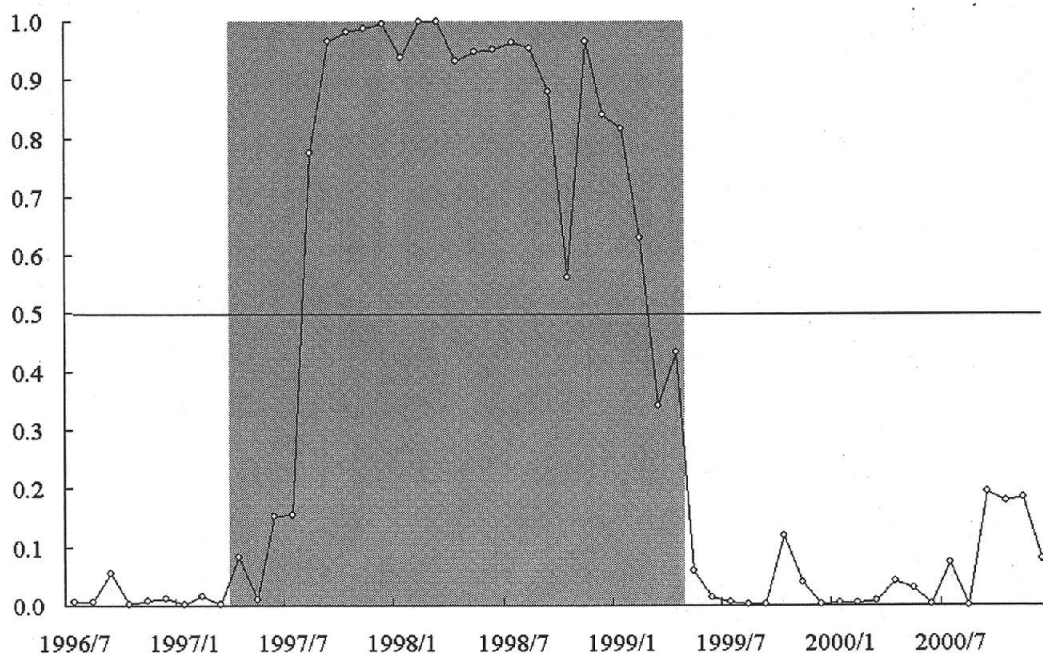
Note: shadowed periods denote economic recessions by the Cabinet Office.

Figure 4.16: Filtered regime probability of contractions in the in-sample period.



Note: shadowed periods denote economic recessions by the Cabinet Office.

Figure 4.17: Filtered regime probability of contractions in the out-of-sample period.



Note: shadowed periods denote economic recessions by the Cabinet Office.

4.4 Japan Premium and Japanese Banks' Stock Volatility

— The Japan Premium and Japanese Banks' Stock Volatility: A Bivariate Markov Switching Approach

4.4.1 Introduction

In the second half of the 1990s, Japanese banks faced a serious and embarrassing situation in two financial markets, the international money market and the domestic stock market. The first phenomenon was the so-called "Japan premium" in the international money market. The Japan premium is defined as the difference between the Eurodollar interbank borrowing rate of interest for Japanese banks and that for Western banks. When this difference expands more widely than usual regardless of the respective credit ratings (for example, ratings by Moodys and Standard & Poor's), it is highlighted as the Japan premium. The second phenomenon was the increase in the volatility of Japanese banks' stock on the Tokyo Stock Exchange. Although Japanese financial institutions had stabilized their stock prices via share cross-holdings over a period of years, resolving this share cross-holding through the second half of the 1990s was one of the factors that caused Japanese banks' stocks to become more volatile. Furthermore, the huge amount of short-selling and speculative investments by hedge funds amplified this problem. These undesirable phenomena in the financial markets set off a vicious cycle in which the increased cost of capital procurement and the destabilization of capital adequacy further weakened the Japanese banks' management bases. From an economic and political viewpoint, these problems stemmed from: 1) the vast amount of non-performing loans due to the bursting of asset bubbles in the first half of the 1990s; 2) the incomplete disclosure of bank balance sheet conditions; and 3) the lack of a Japanese government scheme to deal with near-insolvent banks. Peek and Rosengren (2000) examined the factors most responsible for movements in the Japan premium. Ito and Harada (2000) investigated how financial troubles among the Japanese banks were viewed by the international money market and by the domestic stock market.

We now turn to the relationship between the Japan premium and the increase in the volatility of Japanese banks' stock. The international money market and the domestic stock market operate with different systems and roles. There are very few mutual market participants. While international money markets are inter-bank markets, stock markets are open markets. However, recent huge capital movements and the development of financial information technologies have contributed to strengthening linkages among several financial markets. Given financial globalization, it is sensible to assume that the Japan premium and the increase in Japanese banks' stock volatility do not occur independently, but are

correlated to some degree.

The purpose of this section is to analyze the relationship between the Japan premium and the increase in Japanese banks' stock volatility in the second half of the 1990s using the bivariate Markov switching model. The two variables used are the spread of the Eurodollar borrowing rate of interest between Japanese banks and Western banks in the London Inter-Bank Offer Rates (the LIBOR spread), and the banks' stock return in the Tokyo stock exchange (the TSEBK return).

The Markov switching approach was introduced by Hamilton (1989) in the econometrics literature to capture some characteristics of nonlinear time series. Kim (1993, 1994) extended Hamilton's Markov switching model to a general state-space model. Engel and Hamilton (1990) applied the stochastic trend model with a Markov switching mean and variance to the quarterly exchange rates of developed countries. With regard to the bivariate Markov switching model, Phillips (1991) extended a univariate Markov switching model to the bivariate version, and investigated the inter-country transmission of the business cycle. Hamilton and Lin (1996) studied the relationship between stock market volatility and the business cycle. The bivariate Markov switching model used in this section is a modification of Phillips's (1991) model.

The section is organized as follows. Subsection 2 describes the analytical method within the framework of the bivariate Markov switching model. Subsection 3 presents the empirical results, including model comparisons, estimates, and the implications for the Japan premium and Japanese banks' stock volatility. Concluding remarks are presented in subsection 4.

4.4.2 The Bivariate Markov Switching Model

We describe the bivariate correlated Markov switching (BCMS) model in this subsection. We attempt to capture the mean shift effect of the LIBOR spread and the variance shift effect (heteroskedasticity) of the TSEBK return using the BCMS model. Let $y_{1,n}$ be the LIBOR spread, and $y_{2,n}$ be the TSEBK return. We assume that the observed variables $y_{1,n}$ and $y_{2,n}$ are distributed as follows:

$$y_{1,n} \sim s_{1,n}N(\mu_{1,0}, \sigma_{1,0}^2) + (1 - s_{1,n})N(\mu_{1,1}, \sigma_{1,1}^2), \quad \mu_{1,0} < \mu_{1,1}, \quad (4.100)$$

$$y_{2,n} \sim s_{2,n}N(0, \sigma_{2,0}^2) + (1 - s_{2,n})N(0, \sigma_{2,1}^2), \quad \sigma_{2,0} < \sigma_{2,1}, \quad (4.101)$$

where $s_{1,n}$ and $s_{2,n}$ are the discrete unobserved variables for the LIBOR spread and the TSEBK return, respectively. $s_{1,n}$ and $s_{2,n}$ follow the first-order and two-regime Markov switching processes taking a value of zero or one.

The unobserved variables $s_{1,n}$ and $s_{2,n}$ indicate respectively the "regime" that $y_{1,n}$ and $y_{2,n}$ are in

at time n . For the LIBOR spread, $s_{1,n} = 0$ and $s_{1,n} = 1$ indicate the narrow and wide spread regimes, respectively. The LIBOR spread follows the normal distribution $N(\mu_{1,0}, \sigma_{1,0}^2)$ in the narrow spread regime, and follows the normal distribution $N(\mu_{1,1}, \sigma_{1,1}^2)$ in the wide spread regime. For the TSEBK return, $s_{2,n} = 0$ and $s_{2,n} = 1$ indicate the low and high volatility regimes, respectively. The TSEBK return follows the normal distribution $N(0, \sigma_{2,0}^2)$ in the low volatility regime, and follows the normal distribution $N(0, \sigma_{2,1}^2)$ in the high volatility regime. We can consider four possible regimes for both the LIBOR spread and the TSEBK return. They are summarized in Table 4.10 as follows:

Table 4.10: Regime Definitions

	$y_{1,n}$: LIBOR spread	$y_{2,n}$: TSEBK return
Regime 1 ($s_{1,n} = 0, s_{2,n} = 0$)	Narrow spread	Low volatility
Regime 2 ($s_{1,n} = 1, s_{2,n} = 1$)	Wide spread	High volatility
Regime 3 ($s_{1,n} = 0, s_{2,n} = 1$)	Narrow spread	High volatility
Regime 4 ($s_{1,n} = 1, s_{2,n} = 0$)	Wide spread	Low volatility

The relationship between the LIBOR spread and the TSEBK return is characterized by the time-varying correlation among $s_{1,n}$, $s_{2,n}$, $s_{1,n-1}$ and $s_{2,n-1}$. Thus, the transition probability is defined as follows:

$$p_{ii',jj'} = \Pr(s_{1,n} = j, s_{2,n} = j' | s_{1,n-1} = i, s_{2,n-1} = i'), \quad i, j, i', j' = 0, 1, \quad (4.102)$$

and

$$\sum_{j=0}^1 \sum_{j'=0}^1 p_{ii',jj'} = 1. \quad (4.103)$$

for all $i, i' = 0, 1$. It can be written as a (4×4) transition matrix:

$$P = \begin{pmatrix} p_{00,00} & p_{11,00} & p_{01,00} & p_{10,00} \\ p_{00,11} & p_{11,11} & p_{01,11} & p_{10,11} \\ p_{00,01} & p_{11,01} & p_{01,01} & p_{10,01} \\ p_{00,10} & p_{11,10} & p_{01,10} & p_{10,10} \end{pmatrix}. \quad (4.104)$$

The BCMS model allows for various types of transition between regimes, including stationary, leading, lagging, and simultaneous regime transitions.

We now briefly describe the filter, smoother, and parameter estimation methods. Hamilton (1989) proposed the filter algorithm for the Markov switching model. As a by-product of this filter, we can easily

obtain an approximate log-likelihood function for the unknown parameter vector θ , which is given by

$$l(\theta|y) = \sum_{n=1}^N \log f(y_n|\Psi_{n-1}), \quad (4.105)$$

where

$$\begin{aligned} f(y_n|\Psi_{n-1}) &= \sum_{i=0}^1 \sum_{i'=0}^1 \sum_{j=0}^1 \sum_{j'=0}^1 f(y_n, s_{1,n} = j, s_{2,n} = j', s_{1,n-1} = i, s_{2,n-1} = i' | \Psi_{n-1}) \\ &= \sum_{i=0}^1 \sum_{i'=0}^1 \sum_{j=0}^1 \sum_{j'=0}^1 f(y_n | s_{1,n} = j, s_{2,n} = j', s_{1,n-1} = i, s_{2,n-1} = i', \Psi_{n-1}) \\ &\quad \times \Pr(s_{1,n} = j, s_{2,n} = j', s_{1,n-1} = i, s_{2,n-1} = i' | \Psi_{n-1}). \end{aligned} \quad (4.106)$$

Note that Ψ_{n-1} denotes all the information available at time $n - 1$. The smoother algorithm is also proposed by Hamilton (1989). Kim (1994) presented the filter and the smoother for the state-space model. Several methods for estimating the unknown parameters have been proposed: for example, an Expectation maximization (EM) algorithm (Hamilton 1989) and a Gibbs sampler (Albert and Chib 1992). For the purposes of this section, the quasi-maximum likelihood method is preferred because of its computational ease.

4.4.3 Application and Model Comparison

Data description

In the empirical analysis, for the observed variable on the LIBOR spread, we use the daily average spread of the three-month U.S. dollar LIBOR between two Japanese banks (Tokyo-Mitsubishi and Fuji bank) and two Western banks (Citibank and Berkeleys Capital Group). For the observed variable on the TSEBK return, we have to remove the biases against other industrial sectors, the free-float and cross-holdings, from the daily logarithmic return of the TOPIX Bank Sector Index. Fedenia *et al.* (1994) discussed the bias against the cross-holdings. To deal with these problems, we construct the TSEBK return used in the empirical analysis as follows:

$$\text{TSEBK return} = t_n f_n - (u_n - t_n w_n) \quad (4.107)$$

where t_n is the TOPIX Bank Sector Index return; f_n is the free-float adjustment factor; u_n is the S&P/TOPIX 150 Index return,³ which is adjusted for free-float due to government ownership, strategic holders, corporate cross-holdings, and foreign investment restrictions; and w_n is the index weight of the

³The S&P/TOPIX 150 Index includes 150 highly liquid securities selected from each major sector of the Tokyo market, and represents approximately 70% of the market value of the Japanese equity market. Each stock's weighting, or percentage, in the index is based on its market capitalization. The market value of each stock is adjusted to exclude the value of shares held by large shareholders.

bank stocks in the S&P/TOPIX 150 Index. Note that f_n is a constant 0.4 according to the calculation method of Standard & Poor's, and that w_n is assumed unchanged for one month.

Now, we must draw attention to the problem that the observed times for the TSEBK return and the LIBOR spread are not fully contemporaneous. While the observed time for the TSEBK return is 3:00 p.m. Tokyo time, the observed time for the LIBOR spread is 11:00 a.m. London time. Thus, on a given date, the TSEBK return is observed five hours earlier than the LIBOR spread. In our empirical analysis, we regard the TSEBK return and the LIBOR spread observed for a given date as quasi-contemporaneous observations. The financial innovations during the observed time-lag between the TSEBK return and the LIBOR spread seem to have a relatively small impact on the international money markets. This may be because this five hour time-lag does not overlap with the operating hours of the New York financial market, which seems to have a significant impact on the rest of the world. The daily data are provided by Bloomberg and Standard & Poor's. The sample period is from January 1995 to December 2000. The estimated statistics are calculated using indices obtained by multiplying the observations by 100. Figures 4.18 and 4.19 plot the LIBOR spread and the TSEBK return, respectively. It is likely that the LIBOR spread includes some structural changes in the mean, and that the TSEBK return has some heteroskedasticity.

Check for Stationarity of the Observations

To analyze the relationship of cause and effect for the multivariate stationary time series, the vector autoregressive model (VAR model) and the corresponding spectrum analysis are widely used. Ito and Harada (2000) applies the VAR model with different specification of the LIBOR spread and the TSEBK return, including news event effects. However, when the time series are not stationary, we cannot apply these approaches in a straightforward manner (Baek and Brock, 1992). As preliminary analysis, we check stationarity of the observed time series, the LIBOR spread and the TSEBK return. Figure 4.20 and 4.21 show the sample autoregressive functions of the LIBOR spread and the TSEBK return, respectively. While the TSEBK in Figure 4.21 seems to be stationary, the LIBOR spread in Figure 4.20 apparently includes a trend component and can be said to be non-stationary. Therefore, in the following part of this subsection we try to analyze the transmission mechanism between the LIBOR spread and the TSEBK return by using bivariate Markov switching models.

Model Comparison with Alternative Markov Switching Models

In this subsection, we compare the BCMS model using two null hypotheses — the bivariate “simultaneous” Markov switching (BSMS) model, and the bivariate “independent” Markov switching (BIMS) model. The null models include the different restricted transition probability forms described below.

Bivariate Simultaneous Markov Switching Model

In the null hypothesis of the BSMS model, the narrow (wide) spread regime for the LIBOR spread and the low (high) volatility regime for the TSEBK return are assumed to occur simultaneously or in synchronization. Thus, this simultaneous regime transition follows the common first-order Markov switching process across two observed variables as follows:

$$s_{1,n} = s_{2,n}, \quad (4.108)$$

$$p_{ij} = \Pr(s_{1,n} = j | s_{1,n-1} = i) = \Pr(s_{2,n} = j | s_{2,n-1} = i), \quad i, j = 0, 1, \quad (4.109)$$

and

$$\sum_{j=0}^1 p_{ij} = 1, \quad (4.110)$$

for all $i = 0, 1$.

Bivariate Independent Markov Switching Model

In the null hypothesis of the BIMS model, since $s_{1,n}$ and $s_{2,n}$ are independent, the regime transitions of the LIBOR spread and the TSEBK return are assumed to occur independently according to the independent first-order Markov switching processes as follows:

$$p_{1,ij} = \Pr(s_{1,n} = j | s_{1,n-1} = i), \quad i, j = 0, 1, \quad (4.111)$$

$$p_{2,i'j'} = \Pr(s_{2,n} = j' | s_{2,n-1} = i'), \quad i', j' = 0, 1, \quad (4.112)$$

and

$$\sum_{j=0}^1 p_{1,ij} = 1, \quad \sum_{j'=0}^1 p_{2,i'j'} = 1, \quad \sum_{j=0}^1 \sum_{j'=0}^1 p_{1,ij} p_{2,i'j'} = 1, \quad (4.113)$$

for all $i, i' = 0, 1$.

Table 4.11 shows the log-likelihood values, the Akaike information criterion (AIC) (Akaike 1973), the log-likelihood (LR) statistics, and the corresponding p -values. In this model comparison, the BCMS

model is selected to minimize the AIC. In addition, we obtain consistent results by using the AIC model selection criterion, and by testing the BCMS model on the basis of the LR statistics. Both null hypotheses of the BSMS and BIMS models can be strongly rejected against the BCMS model at the one percent significance level. It follows from this model comparison that $s_{1,n}$, $s_{2,n}$, $s_{1,n-1}$ and $s_{2,n-1}$ are time-variance correlated across both the LIBOR spread and the TSEBK return. In other words, the transition among regimes of the LIBOR spread and the TSEBK return is neither synchronized nor independent, but is time-variance correlated.

Estimation Results

Table 4.12 shows the estimated parameters and corresponding standard errors of the BCMS model. Figures 4.22 to 4.25 plot the smoothed regime probability of the BCMS model.

First, the LIBOR spread follows the normal distribution $N(4.80, 4.13^2)$ in the narrow spread regime, and follows the normal distribution $N(38.21, 19.03^2)$ in the wide spread regime. The Japan premium emerges as a serious issue for Japanese banks in those periods when the observations are in regimes 2 or 4 (Figures 4.23 and 4.25), and is normally distributed around a mean of 38.21 with a variance of 19.03.⁴ Figure 4.26 plots two joint density functions $f(y_{1,n}, s_{1,n})$, which are obtained by integrating the conditional density functions $f(y_{1,n}|s_{1,n})$ and the unconditional regime probabilities $\Pr(s_{1,n} = 0) = 0.759$ and $\Pr(s_{1,n} = 1) = 0.241$. We can see the bimodal appearance of the LIBOR spread, and can confirm its mean shift effect schematically.

Secondly, the TSEBK return follows the normal distribution $N(0, 0.59^2)$ in the low volatility regime, and follows the normal distribution $N(0, 1.29^2)$ in the high volatility regime. Figures 4.23 and 4.24 show the probabilities of periods when the volatility of the TSEBK return increases. In the same way as the LIBOR spread, the joint density functions $f(y_{2,n}, s_{2,n})$ for the TSEBK return in Figure 4.27 are given by $f(y_{2,n}, s_{2,n}) = f(y_{2,n}|s_{2,n}) \Pr(s_{2,n})$; note that $\Pr(s_{2,n} = 0) = 0.620$ and $\Pr(s_{2,n} = 1) = 0.380$. These take the form of a fat-tail distribution and suggest the variance shift effect (heteroskedasticity) of the TSEBK return.

Thirdly, we consider relationships among regimes of the LIBOR spread and the TSEBK return. Properties of the regime transition of the LIBOR spread and the TSEBK return can be obtained from the

⁴The AIC value to determine the best fitting BRMS model is given by

$$\text{AIC} = -2 \sum_{n=1}^N \log f(y_{1,n}, y_{2,n} | \Psi_{n-1}, \theta) + 2|\theta|,$$

where θ is the number of unknown parameters. Note that θ takes a value of eight, ten or eighteen for the BSMS, BIMS or BCMS model, respectively.

transition probability structure. The estimated transition probability matrix ⁵ of the BCMS model is:

$$P = \begin{pmatrix} 0.989 & 0.000 & 0.001 & 0.019 \\ (0.004) & (0.000) & (0.002) & (0.007) \\ 0.000 & 0.964 & 0.000 & 0.005 \\ (0.000) & (0.042) & (0.000) & (0.006) \\ 0.001 & 0.000 & 0.999 & 0.000 \\ (0.001) & (0.000) & (0.002) & (0.000) \\ 0.010 & 0.036 & 0.000 & 0.976 \\ (-) & (-) & (-) & (-) \end{pmatrix} \quad (4.114)$$

It is difficult to capture the relationships among four regimes from the above estimated transition probability matrix only. The diagram in Figure 4.28 helps to illustrate the relationships. In this diagram, the circles represent the regimes defined in the previous subsection. The heavy recursive arrows represent the transition towards the same regime as the previous one. The real arrows connecting two different circles represent the transition where the regimes of the LIBOR spread lead those of the TSEBK return. The dotted arrows connecting two different circles represent the transition where the regimes of the TSEBK return lead those of the LIBOR spread.

Finally, the empirical results in this subsection suggest the following characteristics of the relationships and transitions among the four regimes.

1. The smoothed probability for regime 2, in which the LIBOR spread is in the wide spread regime and the TSEBK return is in the high volatility regime, is successively greater than 50 percent from 4th September to 6th November 1998, as shown by Figure 4.22. These episodes appear to be related to the period during which the world economy faced severe financial crises, including the Russian financial crisis and the collapse of Long Term Capital Management. Also, the volatility of the TSEBK return tends to increase from September 1998, as shown by Figures 4.23 and 4.24.
2. The transition probabilities for which the regimes remain unchanged are all more than 0.95, as shown by the diagonal elements of the estimated transition probability matrix. Once the observations enter into any regime, they tend to stay in the same regime for some time. This implies that all regimes are highly persistent.
3. The regime transition with the positive correlation has never occurred simultaneously in the LIBOR spread and the TSEBK return. As shown by Figure 4.28, there is no arrow connecting regimes 1

⁵Standard errors are in parentheses.

and 2 directly. In the sample period, the observations always connect regime 1 to 2 (or regime 2 to 1) through regime 4, which represents the wide spread regime of the LIBOR spread and the high volatility regime of the TSEBK return.

4.4.4 Conclusions

In this section, we have investigated the relationship between the Japan premium and the volatility of Japanese banks' stock, using the bivariate Markov switching approach. Our BCMS model comprises the narrow and wide spread regimes for the LIBOR spread and the low and high volatility regimes for the TSEBK return. In this model, the relationships among regimes of the LIBOR spread and the TSEBK return are characterized by the time-varying correlation of the unobserved variables $s_{1,n}$ and $s_{2,n}$.

In comparisons between three variants of the Markov switching model (the BCMS, BSMS, and BIMS models), the BCMS model is selected as the best model by the AIC and by the LR test. This result suggests that the regime transition of the LIBOR spread and the TSEBK return is neither synchronized nor independent, but is time-variance correlated across the LIBOR spread and the TSEBK return. The BCMS model succeeded in capturing the characteristics of two effects: namely, the mean shift effect of the LIBOR spread, and the variance shift effect (heteroskedasticity) of the TSEBK return. The Japan premium presents problems for Japanese banks, particularly when the LIBOR spread follows the normal distribution $N(38.21, 19.03^2)$ in the wide spread regime. In addition, we identified some properties of the relationships and transitions among regimes of the LIBOR spread and the TSEBK return from the smoothed regime probability and the estimated transition probability.

Table 4.11: Model comparison.

	BCMS	BIMS	BSMS
Number of parameters	18	10	8
Log-likelihood value	-6584.82	-6604.03	-6762.07
AIC	13205.63	13228.07	13540.14
Testing for the BCMS model			
LR statistics		38.44	354.51
Degree of freedom		8	10
p-value		0.000	0.000

Table 4.12: Estimated parameters and corresponding standard errors

	BCMS	
	Estimation	(Stdard error)
(LIBOR spread)		
$\mu_{1,0}$	4.80	(0.13)
$\mu_{1,1}$	38.21	(0.95)
$\sigma_{1,0}$	4.13	(0.09)
$\sigma_{1,1}$	19.03	(0.65)
(TSEBK return)		
$\sigma_{2,0}$	0.59	(0.01)
$\sigma_{2,1}$	1.29	(0.05)

Figure 4.18: LIBOR spread.

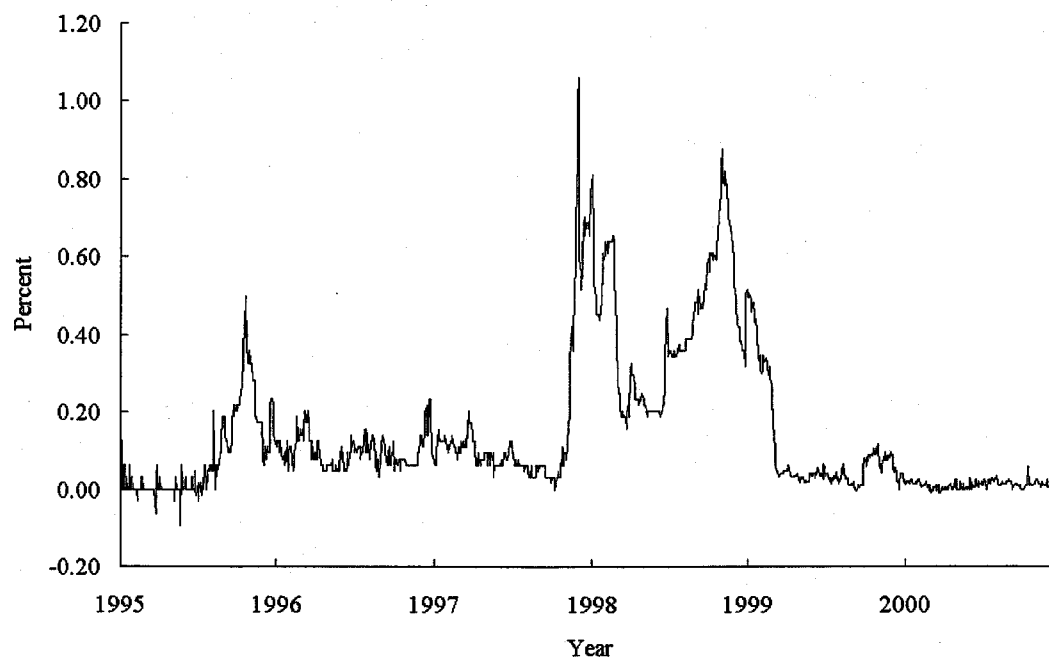


Figure 4.19: TSEBK return.

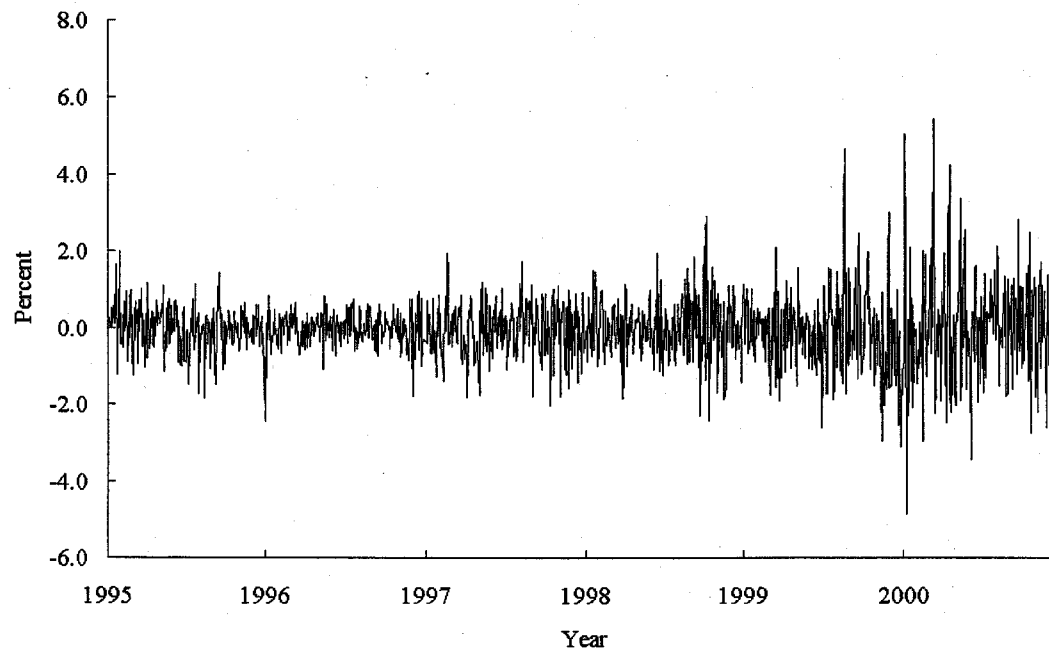


Figure 4.20: Sample autocorrelation function of the LIBOR spread.

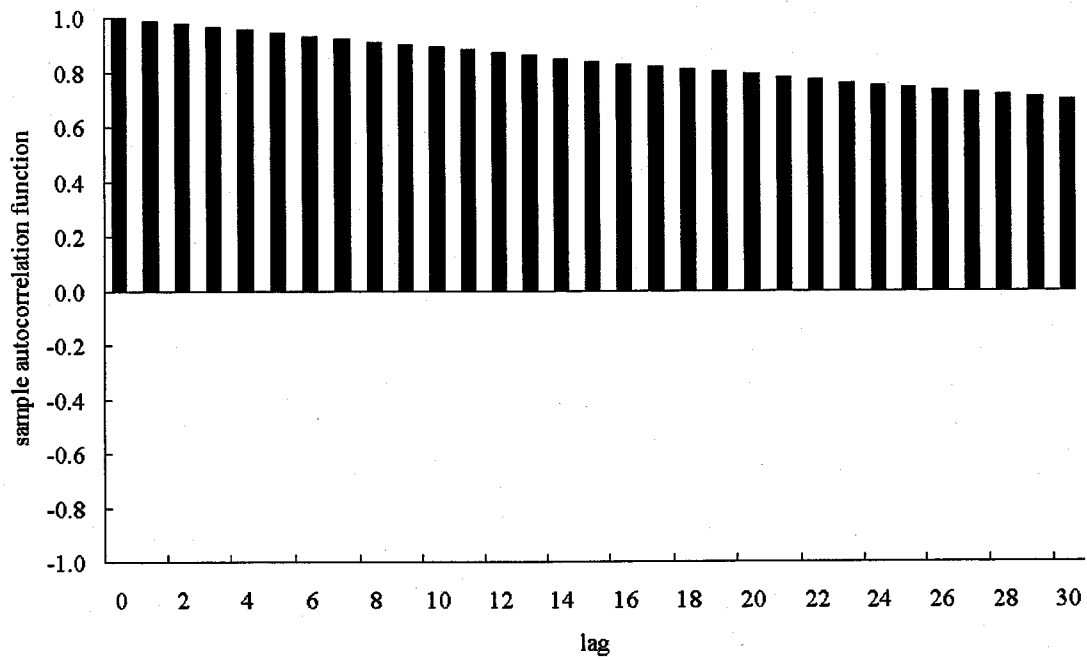


Figure 4.21: Sample autocorrelation function of the TSEBK return.

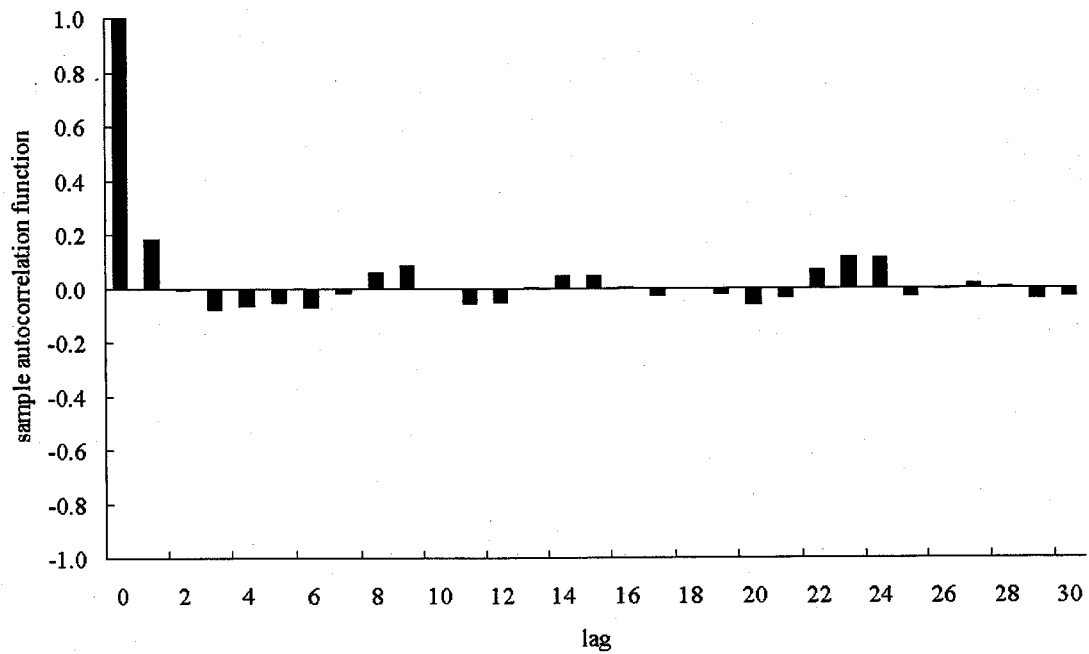


Figure 4.22: Smoothed regime probability of regime 1.

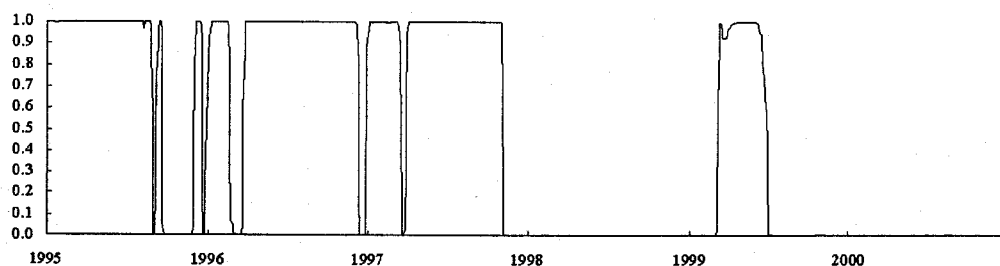


Figure 4.23: Smoothed regime probability of regime 2.

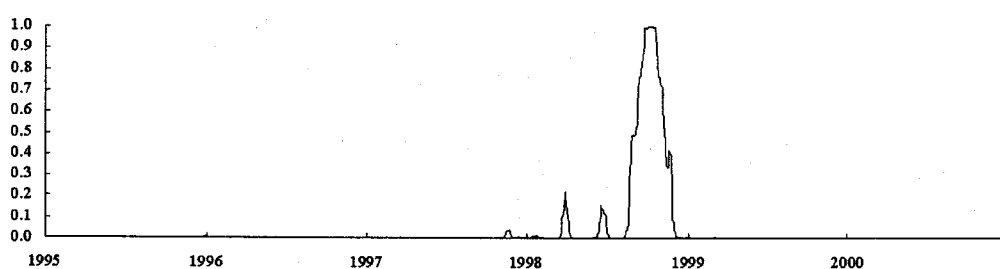
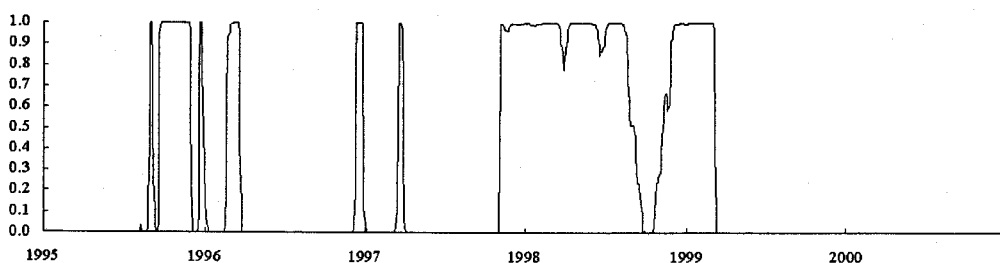


Figure 4.24: Smoothed regime probability of regime 3.

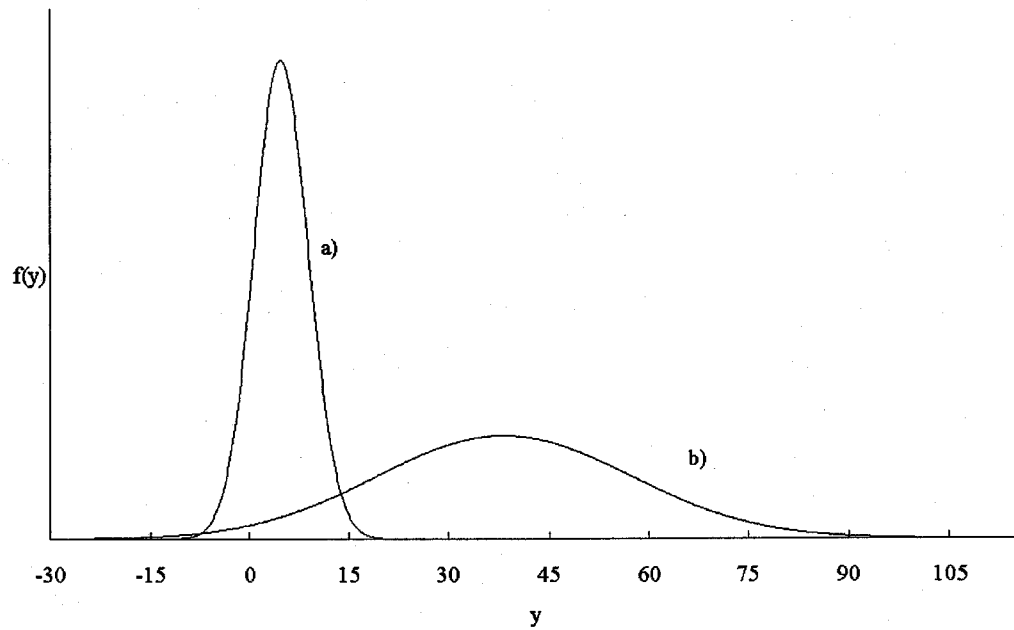


Figure 4.25: Smoothed regime probability of regime 4.



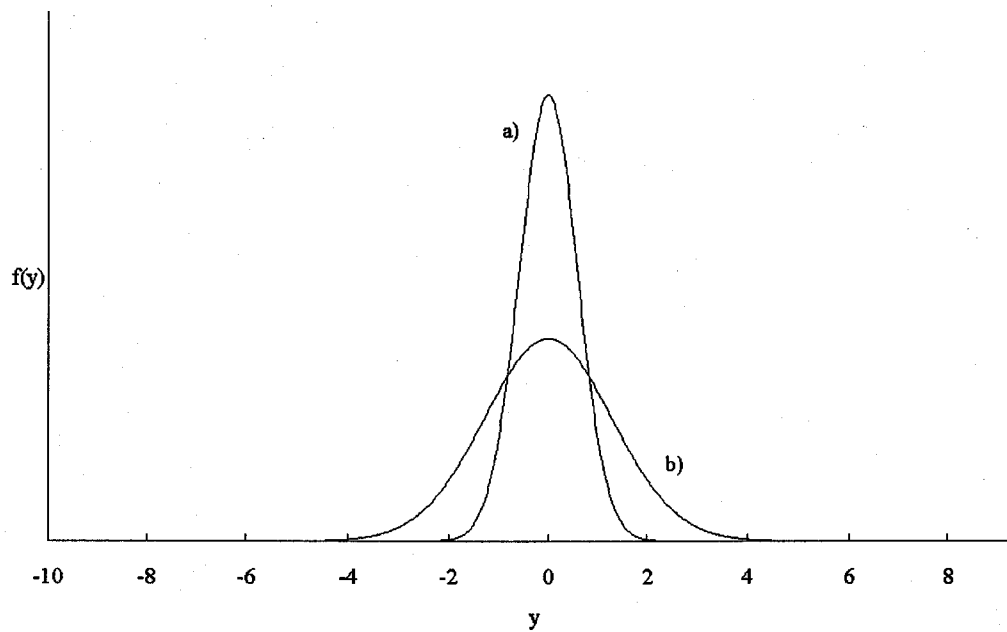
Note: Regime 1 denotes the narrow spread regime of the LIBOR spread and the low volatility regime of the TSEBK return. Regime 2 denotes the wide spread regime of the LIBOR spread and the high volatility regime of the TSEBK return. Regime 3 denotes the narrow spread regime of the LIBOR spread, the high volatility regime of the TSEBK return. Regime 4 denotes the wide spread regime of the LIBOR spread, the low volatility regime of the TSEBK return.

Figure 4.26: Two joint density functions for the LIBOR spread.



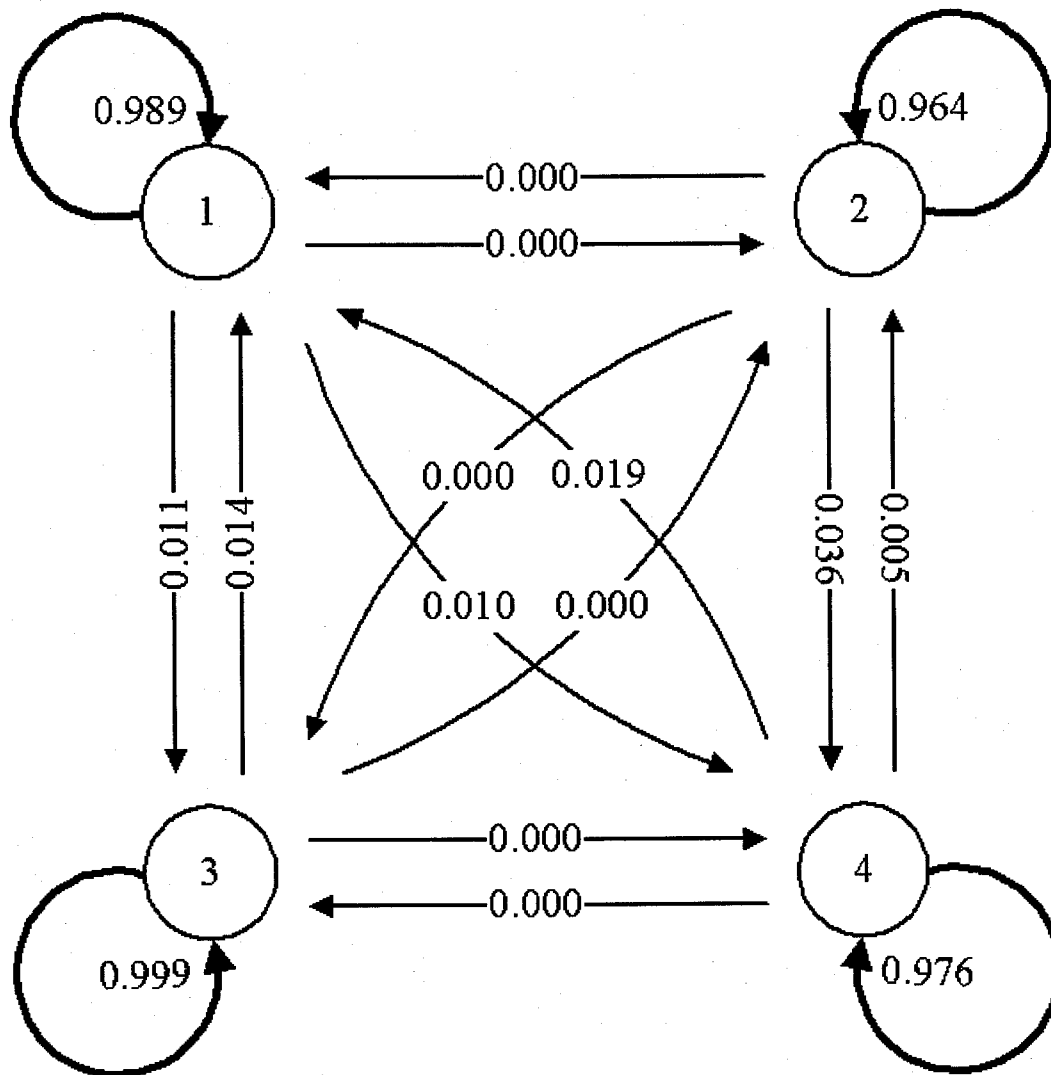
Note: $f(y_n, s_{1,n}) = f(y_n | s_{1,n}) \Pr(s_{1,n})$. a) is for the narrow spread regime; $f(y_n | s_{1,n} = 0) \sim N(4.80, 4.13)$ and $\Pr(s_{1,n} = 0) = 0.759$. b) is for the wide spread regime; $f(y_n | s_{1,n} = 1) \sim N(38.21, 19.03)$ and $\Pr(s_{1,n} = 1) = 0.241$.

Figure 4.27: Two joint density functions for the TSEBK return.



Note: $f(y_n, s_{1,n}) = f(y_n | s_{1,n}) \Pr(s_{1,n})$. a) is for the narrow spread regime; $f(y_n | s_{1,n} = 0) \sim N(0.00, 0.59)$ and $\Pr(s_{1,n} = 0) = 0.620$. b) is for the wide spread regime; $f(y_n | s_{1,n} = 1) \sim N(0.00, 1.29)$ and $\Pr(s_{1,n} = 1) = 0.380$.

Figure 4.28: Relationship among regimes of the LIBOR spread and those of the TSEBK return.



Note: Regime 1 has a narrow LIBOR spread, and low volatility of the TSEBK return. Regime 2 has a wide LIBOR spread, and high volatility of the TSEBK return. Regime 3 has a narrow LIBOR spread, and high volatility of the TSEBK return. Regime 4 has a wide LIBOR spread, and low volatility of the TSEBK return.

4.5 Transmission of Volatility

— The Transmission of Volatility between Japanese Foreign Exchange and Stock Markets: the Bivariate Markov Switching ARCH Model

4.5.1 Introduction

Foreign exchange movements tend to affect the earnings of many Japanese industrial companies. These companies have foreign exchange exposure that should be managed in order to maximize their international competitiveness and performance. Increased volatility in foreign exchange generates some uncertainty about company valuation in the stock market. In addition, advances in financial information technology have enabled rapid and global flows of large amounts of money. For example, it is said that some hedge funds take cross-asset and cross-national arbitrage positions, and execute short-term allocation changes to their asset portfolios. In these circumstances, it seems reasonable to assume that volatility of the Japanese foreign exchange and stock markets has some relevance.

However, Japanese foreign exchange and stock markets operate with different systems and roles. Each market also seems to be affected by several other factors, such as innovations in techniques, national trade balances, the political situation, and so on. Therefore, market relevance may not necessarily exhibit a constant pattern over time. It seems more reasonable to assume that the relationship between, and the transmission of, volatility in the Japanese foreign exchange and stock markets are time-varying.

The purpose of this section is to analyze the relationship between, and transmission of, volatility in Japan's foreign exchange and stock markets. In the empirical analysis, we focus on a period from the second half of the 1990s to the present, when the Japanese economy faced financial instability following the failure of Hokkaido Takushoku Bank and Yama-Ichi Securities. In this period, Japanese financial institutions suffered from having a large number of non-performing loans, the Japan premium in the international money market, and so on. To analyze this period, we propose a bivariate Markov switching ARCH model (BMSARCH), which is an extension of the univariate ARCH model with Markov switching structures proposed by Hamilton and Susmel (1994) and Cai (1994). We apply this BMSARCH model to the daily Japanese/U.S. exchange rate (JPY) and the Tokyo Stock Price Index (TOPIX) from 1996 to 2000. We also examine some diagnostic tests of model specification.

Several types of models have been proposed for capturing the time-varying variance and heteroskedasticity in financial time series. The Markov switching approach has been widely used since Hamilton (1989) first applied it to U.S. business cycle analysis. For example, foreign exchange rates (Engel and Hamilton

1990; Kaminsky 1993; Engel and Kim 1999), stock prices (Kim, Nelson and Startz 1998), and interest rates (Garcia and Perron 1996) have all been examined. However, these studies are based on a univariate framework. An alternative approach is based on the ARCH and GARCH methods introduced by Engle (1982) and Bollerslev (1986). The univariate ARCH and GARCH models have been extended to multivariate versions (Bollerslev *et al.* 1988; Engle *et al.* 1990; Bollerslev 1990). The difficulty in dealing with the multivariate ARCH and GARCH models is because of the increased number of unknown parameters that relate to the lagged conditional variance and covariance matrices. To reduce the number of parameters, several models impose restrictions on the variance and covariance matrices. These multivariate ARCH methods are summarized in Bollerslev (1994). In addition, Hamilton and Susmel (1994) and Cai (1994) have proposed ARCH models that include Markov switching structures, and have applied these models to the analysis of stock price volatility.

This section is organized as follows. Subsection 2 describes the bivariate Markov switching ARCH models. Subsection 3 presents the empirical analysis and some diagnostic tests. Concluding remarks are given in subsection 4.

4.5.2 The Bivariate Markov Switching ARCH Model

Model Specification

Let $x_n \equiv (x_{1,n}, x_{2,n})^T$ be the (2×1) logarithmic return vector of the JPY ($k = 1$) and TOPIX ($k = 2$), respectively. The BMSARCH model is given by

$$x_n = \mu + \varepsilon_n, \quad (4.115)$$

$$\varepsilon_n | \Psi_{n-1} = u_n \sqrt{\Omega_n}, \quad u_n \sim N(0, I_2), \quad (4.116)$$

where Ψ_{n-1} denotes the information available up to time $n - 1$, and $\varepsilon_n \equiv (\varepsilon_{1,n}, \varepsilon_{2,n})^T$ is the (2×2) stochastic process, I_2 is the (2×2) identity matrix, and the time-varying (2×2) covariance matrix Ω_n is with diagonal elements:

$$h_{k,n} = \gamma_k + \sum_{m=1}^M \alpha_{k,m} \varepsilon_{k,n-m}^2, \quad k = 1, 2, \quad (4.117)$$

where

$$\gamma_k = \gamma_{k,0}(1 - s_{k,n}) + \gamma_{k,1}s_{k,n}, \quad \gamma_{k,0}, \gamma_{k,1} > 0, \quad (4.118)$$

$$\alpha_{k,m} = \alpha_{k,m,0}(1 - s_{k,n}) + \alpha_{k,m,1}s_{k,n}, \quad (4.119)$$

$$\mu_k = \mu_{k,0}(1 - s_{k,n}) + \mu_{k,1}s_{k,n}. \quad (4.120)$$

γ_k and $\alpha_{k,m}$ represent the constant intercept in ARCH, and the ARCH coefficient, respectively. μ_k is the element of the (2×1) drift vector $\mu \equiv (\mu_1, \mu_2)^T$. $s_{k,n}$ denotes the unobserved Markov chain indicating

the regime or state that $x_{k,n}$ is in at time n , and takes a value of zero or one. $\gamma_k, \alpha_{k,m}$ and μ_k are all dependent on $s_{k,n}$.

Here $\varepsilon_{k,n}$ will be covariance stationary if

$$\alpha_{k,m,j_k} \leq 0 \quad \text{and} \quad \sum_{m=1}^M \alpha_{k,m,j_k} < 1, \quad (4.121)$$

for all $j_k = 0, 1$. If $\varepsilon_{k,n}$ is covariance stationary, the conditional variance and the unconditional variance respectively are given by

$$E(\varepsilon_{k,n}^2 | s_{k,n} = j_k) = \frac{\gamma_{k,j_k}}{1 - \sum_{m=1}^M \alpha_{k,m,j_k}}, \quad (4.122)$$

$$E(\varepsilon_{k,n}^2 | j_k = 0) < E(\varepsilon_{k,n}^2 | j_k = 1), \quad (4.123)$$

and

$$E(\varepsilon_{k,n}^2) = \frac{\sum_{j_k=0}^1 \gamma_{k,j_k} \Pr(s_{k,n} = j_k | \Psi_{n-1})}{1 - \sum_{m=1}^M \sum_{j_k=0}^1 \alpha_{k,m,j_k}}, \quad (4.124)$$

where $\Pr(s_{k,n} = j_k)$ is the unconditional probability of the unobserved Markov chain $s_{k,n} = j_k$. Thus if $j_k = 0$ (or $j_k = 1$), $x_{k,n}$ is in the low (or high) volatility process regime. The unobserved Markov chain $s_{k,n}$ is assumed to follow a first-order and two-regime Markov process with the time-homogeneous transition probability as follows:

$$p_{i_1 i_2, j_1 j_2} = \Pr(s_{1,n} = j_1, s_{2,n} = j_2 | s_{1,n-1} = i_1, s_{2,n-1} = i_2), \quad (4.125)$$

where

$$\sum_{j_1=0}^1 \sum_{j_2=0}^1 p_{i_1 i_2, j_1 j_2} = 1, \quad (4.126)$$

for all $i_1, i_2 = 0, 1$. The transition probabilities of the BMSARCH model can be rewritten as a (4×4) transition matrix as follows:

$$P = \begin{pmatrix} p_{00,00} & p_{11,00} & p_{01,00} & p_{10,00} \\ p_{00,11} & p_{11,11} & p_{01,11} & p_{10,11} \\ p_{00,01} & p_{11,01} & p_{01,01} & p_{10,01} \\ p_{00,10} & p_{11,10} & p_{01,10} & p_{10,10} \end{pmatrix}. \quad (4.127)$$

Here, we can redefine the four regimes of the bivariate MS-ARCH model with the latent regime indicator R_n as listed in Table 4.13. The bivariate MS-ARCH model permits many varied transmissions of volatility processes to occur. The transmissions can be classified broadly into four categories in Table 4.13:

Table 4.13: Regime definitions of the bivariate MS-ARCH model.

Regime	$x_{1,n}$: JPY	$x_{2,n}$: TOPIX
$R_n = 1$ ($s_{1,n} = 0, s_{2,n} = 0$)	Low volatility	Low volatility
$R_n = 2$ ($s_{1,n} = 1, s_{2,n} = 1$)	High volatility	High volatility
$R_n = 3$ ($s_{1,n} = 0, s_{2,n} = 1$)	Low volatility	High volatility
$R_n = 4$ ($s_{1,n} = 1, s_{2,n} = 0$)	High volatility	Low volatility

1. Self-loop

$$P_1 = \begin{pmatrix} p_{00,00} & - & - & - \\ - & p_{11,11} & - & - \\ - & - & p_{01,01} & - \\ - & - & - & p_{10,10} \end{pmatrix},$$

2. Simultaneous

$$P_2 = \begin{pmatrix} - & p_{11,00} & - & - \\ p_{00,11} & - & - & - \\ - & - & - & p_{10,01} \\ - & - & p_{01,10} & - \end{pmatrix},$$

3. Lead-relation

$$P_3 = \begin{pmatrix} - & - & p_{01,00} & - \\ - & - & - & p_{10,11} \\ - & p_{11,01} & - & - \\ p_{00,10} & - & - & - \end{pmatrix},$$

4. Lag-relation

$$P_4 = \begin{pmatrix} - & - & - & p_{10,00} \\ - & - & p_{01,11} & - \\ p_{00,01} & - & - & - \\ - & p_{11,10} & - & - \end{pmatrix}.$$

Filter, Smoother and Parameter Estimation

From the non-Gaussian filter and smoother (Kitagawa 1989), we obtain the predictive distribution, filter and smoother of the unobserved Markov chains for the Markov switching model. Those of the latent regime indicator R_n for our bivariate MS-ARCH model can be easily implemented by the state-space model with ARCH-type conditional heteroskedasticity by Harvey *et al.* (1992), and the state-space model with Markov switching heteroskedasticity by Kim (1993). The details of these filtering and smoothing algorithms are summarized in Kim (1999).

[Prediction]

$$\begin{aligned} \Pr(R_n = j^* | \Psi_{n-1}) &= \sum_{i^*=1}^4 \Pr(R_n = j^*, R_{n-1} = i^* | \Psi_{n-1}) \\ &= \sum_{i^*=1}^4 \Pr(R_n = j^* | R_{n-1} = i^*) \Pr(R_{n-1} = i^* | \Psi_{n-1}) \\ &= \sum_{i^*=1}^4 p_{i^*j^*} \Pr(R_{n-1} = i^* | \Psi_{n-1}). \end{aligned} \quad (4.128)$$

where Ψ_n denotes the information available up to time n , and i^* and j^* take values of 1, 2, 3 and 4.

[Filter]

$$\begin{aligned} \Pr(R_n = j^* | \Psi_n) &= \Pr(R_n = j^* | x_n, \Psi_{n-1}) \\ &= \frac{f(x_n | R_n = j^*, \Psi_{n-1}) \Pr(R_n = j^* | \Psi_{n-1})}{f(x_n | \Psi_{n-1})}, \end{aligned} \quad (4.129)$$

where

$$f(x_n | R_n = j^*, \Psi_{n-1}) = \frac{1}{2\pi} |\Omega_n|^{\frac{1}{2}} \exp \left\{ -\frac{(x_n - \mu)^T \Omega_n^{-1} (x_n - \mu)}{2} \right\}, \quad (4.130)$$

$$f(x_n | \Psi_{n-1}) = \sum_{j^*=1}^4 f(x_n | R_n = j^*, \Psi_{n-1}) \Pr(R_n = j^* | \Psi_{n-1}). \quad (4.131)$$

[Smoother]

$$f(R_n = j^* | \Psi_N) = f(R_n = j^* | x_n) \sum_{k^*=1}^4 \frac{f(R_{n+1} = k^* | \Psi_N) \Pr(R_{n+1} = k^* | R_n = j^*)}{f(R_{n+1} = k^* | x_n)}. \quad (4.132)$$

As a by-product of the filter algorithm above, we can easily obtain the approximate log likelihood function as follows:

$$\begin{aligned} L(\theta) &= \sum_{n=1}^N \log f(y_n | \Psi_{n-1}) \\ &= \sum_{n=1}^N \left[\sum_{j^*=1}^4 \sum_{i^*=1}^4 \log f(y_n, R_n = j^*, R_{n-1} = i^* | \Psi_{n-1}) \right] \\ &= \sum_{n=1}^N \left[\sum_{j^*=1}^4 \sum_{i^*=1}^4 \log f(y_n | R_n = j^*, R_{n-1} = i^*, \Psi_{n-1}) \Pr(R_n = j^*, R_{n-1} = i^* | \Psi_{n-1}) \right] \end{aligned} \quad (4.133)$$

where an unknown parameter vector θ is defined by

$$\theta \equiv \left\{ \mu_{k,0}, \mu_{k,1}, \gamma_{i,0}, \gamma_{i,1}, \{\alpha_{k,m,0}\}_{m=1}^M, \{\alpha_{k,m,1}\}_{m=1}^M, P^* \right\}.$$

Note that P^* includes $(k^2 - k)$ unknown parameters for the transition probability. Several methods for estimating the unknown parameters are proposed, including the Expectation maximization (EM) algorithm (Hamilton 1990 1994) and the Gibbs sampler (Albert and Chib 1992). For the purpose of this section, the maximum likelihood method is preferred because of its computational ease.

The Akaike information criterion (AIC) for the model selections, including the order of the ARCH coefficients and the number of regimes, is defined by

$$\text{AIC} = -2 \sum_{n=1}^N \log f(x_n | \Psi_{n-1}) + 2|\theta|, \quad (4.134)$$

where $|\theta|$ denotes the dimension of the vector θ (Kitagawa and Gersch 1996). Leroux and Puterman (1992) and Ryden (1995) discuss the method of model selection and identification for Markov switching model in details.

4.5.3 Application and Model Comparison

Data Description

We use the daily JPY at 5:00 p.m. in Tokyo and the TOPIX closing price at 3:00 p.m. in Tokyo. Although the JPY and the TOPIX are not fully contemporaneous, these series are the most appropriate daily data officially available. For the observed time series, we use the indices obtained by multiplying the logarithmic returns of the JPY and the TOPIX by 100. The sample period is from January 1996 to December 2000; each series has 1,232 sample data points. Figures 4.29 and 4.30 plot the original time series for the JPY and the TOPIX, respectively. Figures 4.31 and 4.32 plot the logarithmic return series for the JPY and the TOPIX, respectively.

Estimation Results

The first column of Table 4.14 shows the estimated parameters and corresponding statistics of the BM-SARCH model. Since the expectations of the conditional variances are $E(\varepsilon_{k,n}^2 | s_{k,n} = 0) < E(\varepsilon_{k,n}^2 | s_{k,n} = 1)$ for both time series in Table 4.15, we can identify the regime $s_{k,n} = 0$ as a low volatility process, and regime $s_{k,n} = 1$ as a high volatility process. Both time series exhibit ARCH effects only in the high volatility process regimes; there are no ARCH effects in the low volatility process regimes. The curves a) and b) in Figures 4.33 and 4.34 show the joint density functions $p(x_n, s_{k,n})$ for the JPY and the TOPIX, respectively. Note that the joint density function is given by

$$p(x_{k,n}, s_{k,n} = j_k) = p(x_{k,n} | s_{k,n} = j_k) \Pr(s_{k,n} = j_k), \quad (4.135)$$

where $p(x_n | s_{k,n} = j_k)$ is the conditional density of $s_{k,n}$, and $\Pr(s_{k,n} = j_k)$ is the unconditional regime probability or the stationary distribution of $s_{k,n}$. The unconditional density $p(x_{k,n})$ is a weighted sum of these two joint densities. The unconditional densities $p(x_{1,n})$ for the JPY and $p(x_{2,n})$ for the TOPIX are represented by the curves c) in Figure 4.33 and 4.34, respectively. They seem to produce fat-tailed unimodal densities. Thus, Figures 4.33 and 4.34 suggest that the BMSARCH model captures skewness rather than the characteristics of the "long swing" effect, i.e., regime shift in the conditional mean of returns (Engel and Hamilton, 1991).

The estimated transition probability matrix of Table 4.15 is summarized as follows:

$$P = \begin{pmatrix} 0.945 & 0.127 & 0.014 & 0.046 \\ (0.034) & (0.183) & (0.009) & (0.029) \\ 0.031 & 0.821 & 0.044 & 0.034 \\ (0.028) & (0.085) & (0.027) & (0.084) \\ 0.011 & 0.000 & 0.920 & 0.006 \\ (0.011) & (0.000) & (0.050) & (0.011) \\ 0.012 & 0.052 & 0.022 & 0.914 \\ (-) & (-) & (-) & (-) \end{pmatrix}, \quad (4.136)$$

where standard errors are in parentheses. The estimated unconditional regime probabilities or the stationary distributions of R_n are:

$$\begin{pmatrix} \Pr(R_n = 1) \\ \Pr(R_n = 2) \\ \Pr(R_n = 3) \\ \Pr(R_n = 4) \end{pmatrix} = \begin{pmatrix} 0.553 \\ 0.156 \\ 0.094 \\ 0.198 \end{pmatrix}. \quad (4.137)$$

Figures 4.35 to 4.38 plot the smoothed regime probabilities $\Pr(R_n = j^* | \Psi_N)$, $j^* = 1, 2, 3, 4$, of the BMSARCH model. The relationships between, and the transmission of, volatility processes among the four regimes, which are defined in Table 4.13 with $R_n = 1, 2, 3, 4$, are easily understood from Figure 4.39. In this diagram, the circles represent the regimes defined in Table 4.13. The heavy recursive arrows represent the self-loop; that is, the transition towards the same regime as the previous one. The real arrows connecting two different circles represent the transition where the regimes of the JPY lead those of the TOPIX. The broken arrows connecting two different circles represent the transition where the regimes of the TOPIX lead those of the JPY. The dotted arrows connecting two different circles represent the simultaneous transition in the same or a different direction. This diagram suggest three explanations of the regime transition among the four regimes as follows.

- (1) All the self-loop transition probabilities for which the regimes remain unchanged are greater than or close to 0.9. This implies that all the regimes ($R_n = 1, 2, 3, 4$) have some persistence. That is, once the observations (JPY and TOPIX) enter into any regime, they tend to remain in the same regime for some time.
- (2) There is not enough evidence to conclude that the JPY always tends to lead the TOPIX, or that the TOPIX always tends to lead the JPY. Thus, in the volatility processes of the observations, lead or lag relations between the JPY and the TOPIX are mixed.

- (3) The regime shift from the high-high volatility process regime ($R_n = 1$)⁶ to the low-low volatility process regime ($R_n = 1$)⁷ occurs with a relatively high transition probability in the sample period. This phenomenon means that there is a simultaneous transmission of volatility on the JPY and TOPIX.

Japanese financial crisis

From Figure 4.36 for the high-high volatility process regime ($R_n = 2$), we identify an interesting characteristic of volatility processes. The episodes of the high-high volatility process regime ($R_n = 2$) appear to have some relation to the instability of the Japanese financial system in the second half of the 1990s. However, it is difficult to evaluate and identify the periods when the Japanese financial system experienced crises. To deal with this problem, we consider the "Japan premium" as a proxy variable. The Japan premium is defined as the difference between the Eurodollar interbank borrowing rate of interest for Japanese banks and that for Western banks. In the second half of the 1990s, Japanese banks had to pay an extra funding cost compared to leading U.S. and European banks owing to a decline in their creditworthiness in international money markets. It is possible to regard an increase in the Japan premium as evidence of increased instability in the Japanese banking and financial system. Figure 4.40 plots the daily average spread of the three-month U.S. dollar LIBOR (London Interbank Offered Rate) between two Japanese banks (Tokyo-Mitsubishi and Fuji Bank) and two Western banks (Citibank and Berkeley's Capital Group). The shaded bars in Figures 4.36 and 4.40 show the periods when this spread is more than 40 basis points (0.004). The episodes of the high-high volatility process regime ($R_n = 2$) appear to be related to those of an increase in the Japan premium in the second half of the 1990s — that is, the periods of increased instability in the Japanese financial and banking system.

Diagnostic Tests for Independence and Simultaneity of Markov Switching

First, we check whether regime shifts occur independently and simultaneously across the JPY and the TOPIX. The null hypothesis of independence of Markov switching is given by

$$H_0^1 : p_{i_1 j_1} = \Pr(s_{1,n} = j_1 | s_{1,n-1} = i_1), \quad p_{i_2 j_2} = \Pr(s_{2,n} = j_2 | s_{2,n-1} = i_2), \quad (4.138)$$

⁶The high-high volatility process regime ($R_n = 2$) represents a regime where the JPY is in the high volatility process regime and the TOPIX is in the high volatility regime

⁷The low-low volatility process regime ($R_n = 1$) represents the regime where the JPY is in the low volatility process regime and the TOPIX is in the low volatility regime.

where $\sum_{j_1=0}^1 p_{i_1 j_1} = 1$ and $\sum_{j_2=0}^1 p_{i_2 j_2} = 1$ for all $i_1, i_2 = 0, 1$. The transition probability matrix can be represented by using the Kronecker product as follows:

$$P = P^1 \otimes P^2 = \begin{pmatrix} p_{00}^1 & 1 - p_{11}^1 \\ 1 - p_{00}^1 & p_{11}^1 \end{pmatrix} \otimes \begin{pmatrix} p_{00}^2 & 1 - p_{11}^2 \\ 1 - p_{00}^2 & p_{11}^2 \end{pmatrix}. \quad (4.139)$$

Here regime shifts of the JPY and the TOPIX occur independently without reference to the previous regime of another observation. Secondly, the null hypothesis of simultaneity of Markov switching is given by

$$H_0^2 : s_{1,n} = s_{2,n} = s_n, \quad (4.140)$$

and the transition probability is denoted as follows:

$$p_{ij} = \Pr(s_n = j | s_{n-1} = i), \quad (4.141)$$

where $\sum_{j=0}^1 p_{ij} = 1$ for all $i = 0, 1$. This transition probability can be written as the (2×2) transition probability matrix

$$P = \begin{pmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{pmatrix} = \begin{pmatrix} p_{00} & 1 - p_{11} \\ 1 - p_{00} & p_{11} \end{pmatrix}. \quad (4.142)$$

Here all the regime shifts are perfectly correlated under this simultaneous hypothesis.

The second and third columns of Tables 4.14 to 4.16 show the estimated parameters and corresponding statistics for the null hypotheses. Most values of the estimated parameters — the means of returns, ARCH intercepts and coefficients — are similar to those of the original BMSARCH model. The estimated transition probabilities imply that both low-low and high-high volatility process regimes have some tendency to remain where they were in the previous period.

It is straightforward to test for simultaneity of Markov switching using the likelihood ratio (LR) test and the AIC. The null hypotheses of independence and simultaneity include eight and ten restricted parameters in the transition probabilities, respectively. Both the LR statistics are distributed $\chi^2(8)$ and $\chi^2(10)$. The LR statistic of independence is $2 \times (3931.01 - 3920.21) = 21.60$, and the corresponding $\chi^2(8)$ p -value is 0.006. That of simultaneity is $2 \times (3955.50 - 3920.21) = 70.58$, and the corresponding $\chi^2(10)$ p -value is 0.000. Both the null hypotheses of independence and simultaneity of Markov switching can be rejected at the one percent significance level. In addition, the AIC of the original BMSARCH model is smaller than for the null models. Therefore, we find highly significant evidence that the regime shift of volatility processes does not necessarily occur independently or simultaneously across the JPY and the TOPIX, and that the transmission of volatility processes have some relevance to each other.

Model Comparison with No-switching Models

In this part, we compare the BMSARCH model with some bivariate ARCH-type no-switching models. Table 4.17 shows the log likelihood, the number of parameters and AICs for alternative models, including a diagonal VEC model (Bollerslev, Engle and Woldridge 1988), BEKK model (Engle and Kroner 1995), Matrix diagonal model (Ding 1994, Bollerslev, Engle and Nelson 1995), vector diagonal model, scalar diagonal model, and conditional constant correlation model (Bollerslev 1987). All the parameters are estimated using the S-Plus program with S+GARCH module. The diagonal VEC model is selected as the best AIC model among alternative no-switching ARCH-type models. However, the AIC of BMSARCH model is even smaller than that of the diagonal VEC model.

4.5.4 Conclusions

In this section, we examined the transmission of volatility processes between the Japanese foreign exchange and stock markets. In order to capture the time-varying transmission structure of the two financial time series, we extended the univariate Markov switching ARCH model to the bivariate case. Our bivariate Markov switching ARCH model (BMSARCH) assumes that observations shift among some regimes not just simultaneously but also with a time-varying transmission structure.

In the empirical analysis, we analyzed the daily Japanese/U.S. exchange rate (JPY) and the Tokyo Stock Exchange Price Index (TOPIX) in the period of the second half of the 1990s, when the Japanese economy faced secular stagnation and financial turmoil. We identified the following properties in relation to the transmission of volatility processes between the JPY and the TOPIX: 1) all regimes have some persistence; 2) lead-lag relations are mixed; and 3) a direct regime shift from the high-high volatility process regime ($R_n = 1$) to the low-low volatility process regime ($R_n = 2$) occurs with relatively high probability. We also found that a characteristic of volatility process regimes is that the episodes of the high-high volatility process regime correspond to periods of increased instability in the Japanese financial and banking system. For the specification tests and the AIC model selections, the hypotheses of independent and simultaneous regime shifts across the JPY and the TOPIX were both rejected.

Table 4.14: Estimated parameters and corresponding statistics.

	BMSARCH		Independent switching		Simultaneous switching	
	Estimation	(Stand. error)	Estimation	(Stand. error)	Estimation	(Stand. error)
$\mu_{1,0}$	0.065	(0.042)	0.062	(0.018)	0.065	(0.019)
$\mu_{2,0}$	-0.018	(0.091)	-0.022	(0.034)	-0.033	(0.041)
$\mu_{1,1}$	-0.074	(0.000)	-0.091	(0.069)	-0.085	(0.062)
$\mu_{2,1}$	0.002	(0.146)	0.009	(0.000)	0.042	(0.095)
μ_1						
μ_2						
$\gamma_{1,0}$	0.244	(0.029)	0.262	(0.022)	0.248	(0.018)
$\gamma_{2,0}$	0.804	(0.053)	0.727	(0.049)	0.809	(0.053)
$\gamma_{1,1}$	1.138	(0.311)	1.256	(0.165)	1.176	(0.127)
$\gamma_{2,1}$	3.581	(0.713)	2.732	(0.305)	2.569	(0.300)
$\alpha_{1,0}$	0.000	(0.000)	0.000	(0.000)	0.018	(0.015)
$\alpha_{2,0}$	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
$\alpha_{1,1}$	0.165	(0.073)	0.158	(0.063)	0.204	(0.075)
$\alpha_{2,1}$	0.044	(0.253)	0.100	(0.063)	0.186	(0.100)
$ \theta $		24		16		14
LL		-3920.21		-3931.01		-3955.50
AIC		7888.41		7894.01		7939.01
LR statistics				21.60		70.58
p -value				0.006		0.000

Table 4.15: Estimated transition probability.

	BMSARCH		Independent switching		Simultaneous switching	
	Estimation	(Stand. error)	Estimation	(Stand. error)	Estimation	(Stand. error)
$p_{00,00}$	0.945	(0.034)				
$p_{01,00}$	0.031	(0.028)				
$p_{10,00}$	0.011	(0.011)				
$p_{11,00}$	0.012	—				
$p_{00,01}$	0.127	(0.183)				
$p_{01,01}$	0.821	(0.085)				
$p_{10,01}$	0.000	(0.000)				
$p_{11,01}$	0.052	—				
$p_{00,10}$	0.014	(0.009)				
$p_{01,10}$	0.044	(0.027)				
$p_{10,10}$	0.920	(0.050)				
$p_{11,10}$	0.022	—				
$p_{00,11}$	0.046	(0.029)				
$p_{01,11}$	0.034	(0.084)				
$p_{10,11}$	0.006	(0.011)				
$p_{11,11}$	0.914	—				
p_{00}					0.907	(0.233)
p_{11}					0.808	(0.296)
$p_{1,00}$			0.955	(0.011)		
$p_{1,11}$			0.899	(0.033)		
$p_{2,00}$			0.981	(0.009)		
$p_{2,11}$			0.948	(0.020)		

Table 4.16: Estimated parameters and corresponding statistics.

	BMSARCH	Independent switching	Simultaneous switching
$E(\varepsilon_{1,n}^2 s_{1,n} = 0)$	0.237	0.250	0.235
$E(\varepsilon_{2,n}^2 s_{2,n} = 0)$	0.865	0.856	0.871
$E(\varepsilon_{1,n}^2 s_{1,n} = 1)$	1.242	1.321	1.213
$E(\varepsilon_{2,n}^2 s_{2,n} = 1)$	3.618	3.000	3.616
$E(\varepsilon_{1,n}^2)$	0.567	0.565	0.571
$E(\varepsilon_{2,n}^2)$	1.539	1.493	1.535

Table 4.17: Log likelihood, number of parameters and AIC of no-switching models.

Models	Log likelihood	# of parameters	AIC
Diagonal VECH GARCH	-3956.77	11	7935.53
BEKKH GARCH	-3963.43	13	7952.86
Matrix diagonal GARCH	-3978.58	11	7979.17
Vector diagonal GARCH	-3968.86	9	7955.71
Scalar diagonal GARCH	-3969.15	7	7952.29
Conditional constant correlation GARCH	-3986.01	8	7988.01

Figure 4.29: Daily Japan/U.S. exchange rate: 1996/1 – 2000/12.

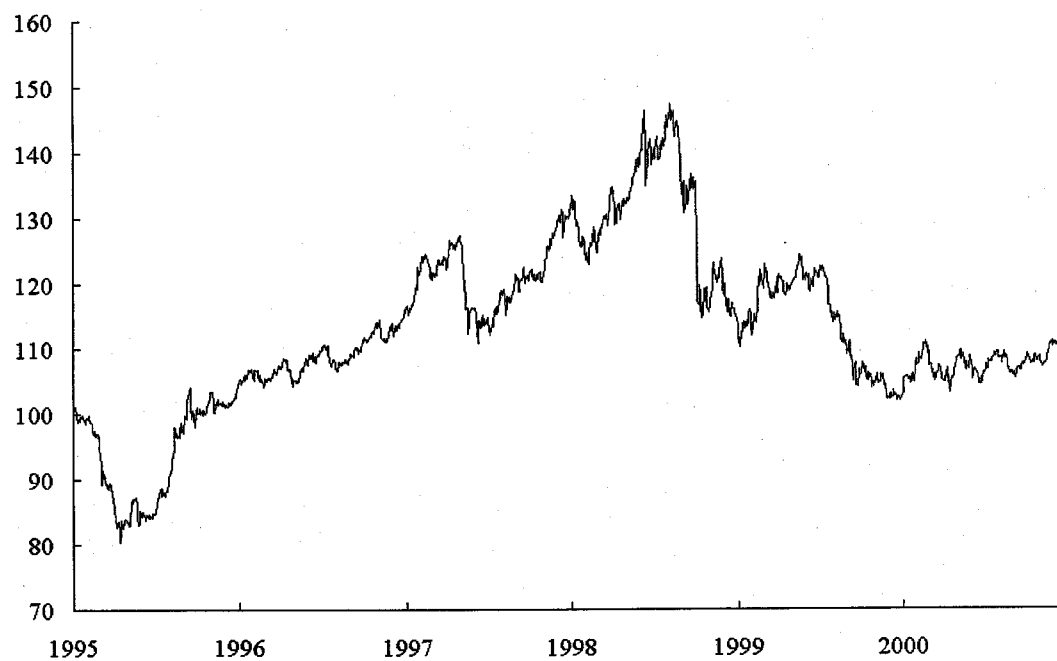


Figure 4.30: Daily TOPIX: 1996/1 – 2000/12.

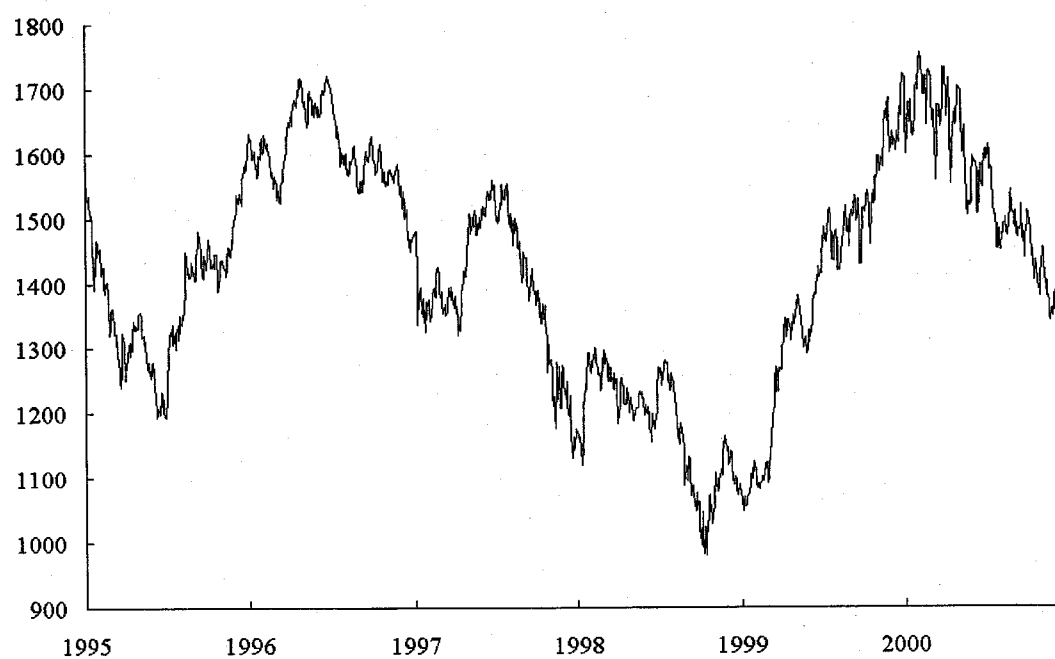


Figure 4.31: Logarithmic return of the daily Japanese/U.S. exchange rate: 1996/1 – 2000/12.

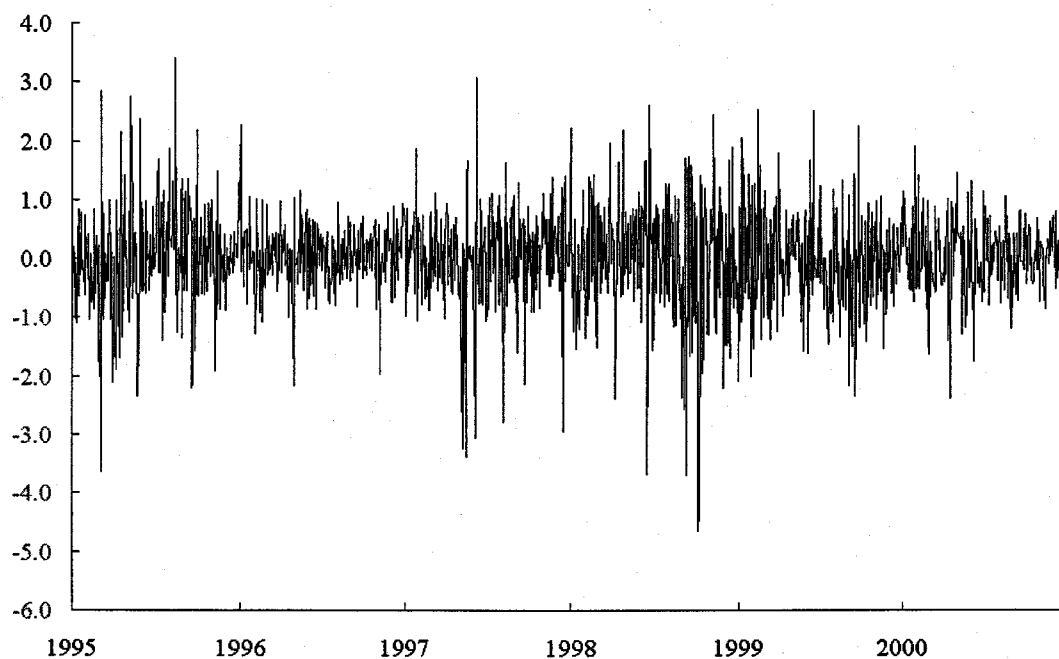


Figure 4.32: Logarithmic return of the daily TOPIX: 1996/1 – 2000/12.

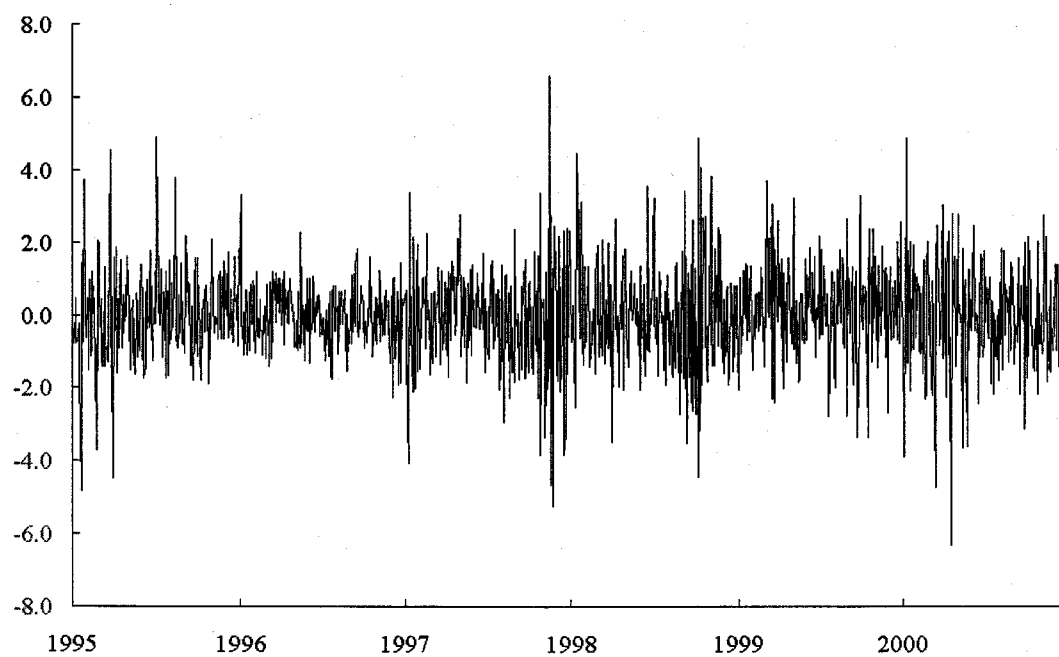
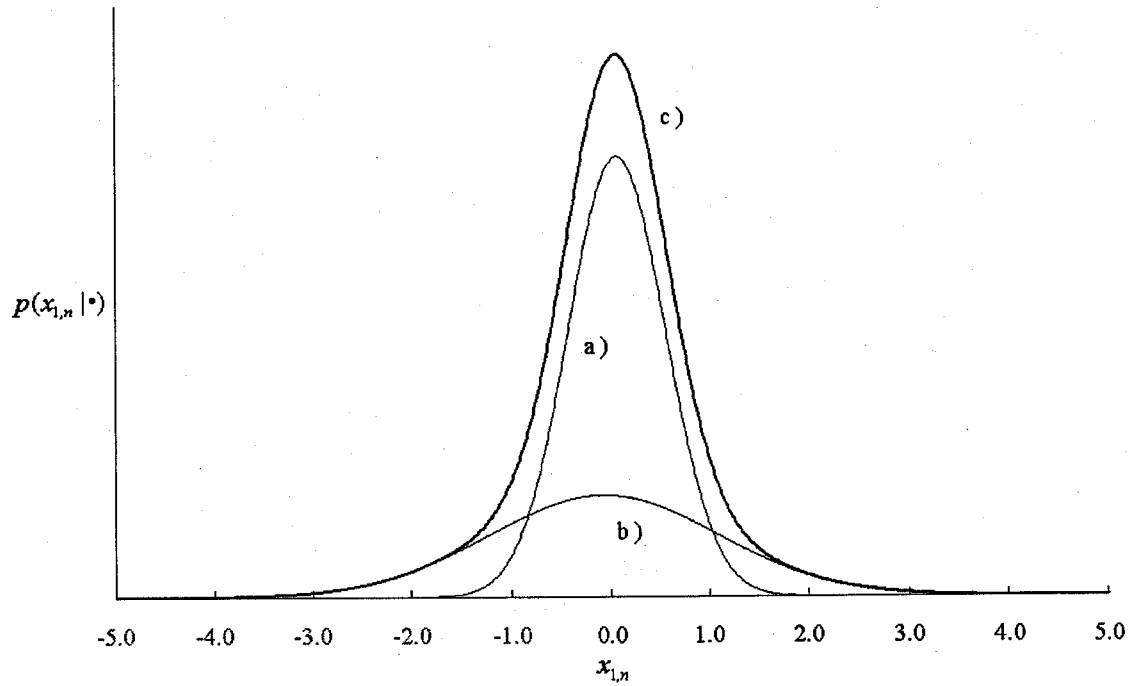


Figure 4.33: Two joint densities and the unconditional density for the JPY.

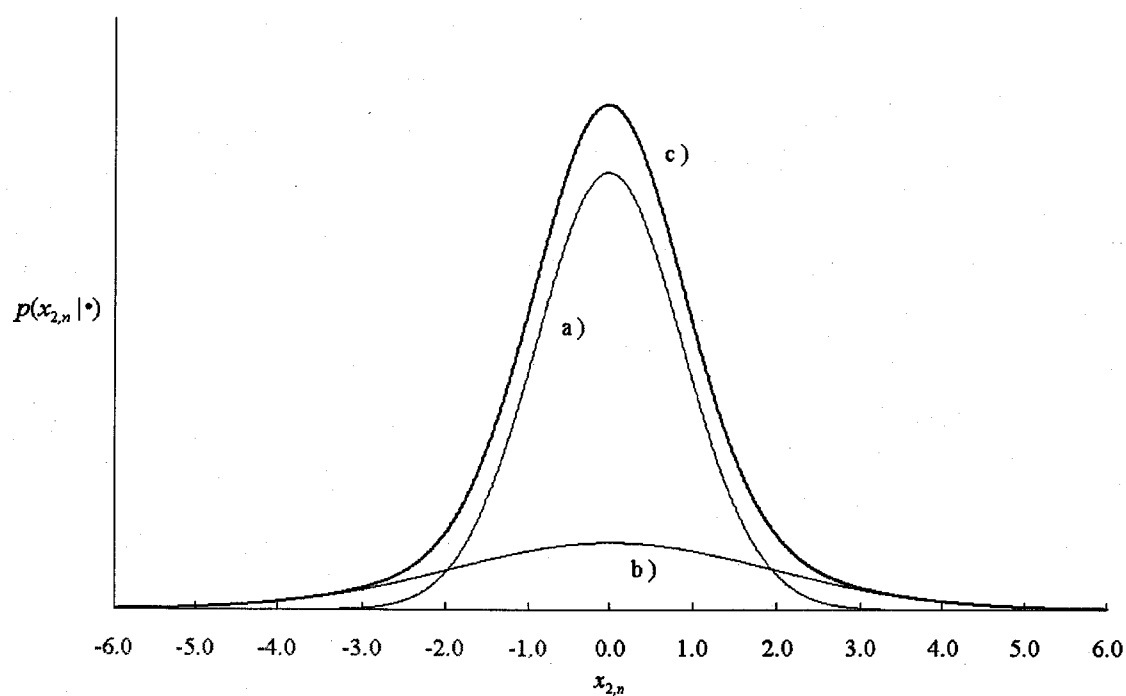


Note: $p(x_{1,n}, s_{1,n}) = p(x_{1,n}|s_{1,n}) \Pr(s_{1,n})$: a) is for the low volatility process regime;

$p(x_{1,n}|s_{1,n} = 0) \sim N(0.065, 0.494)$ and $\Pr(s_{1,n} = 0) = 0.646$, b) is for the high volatility process regime;

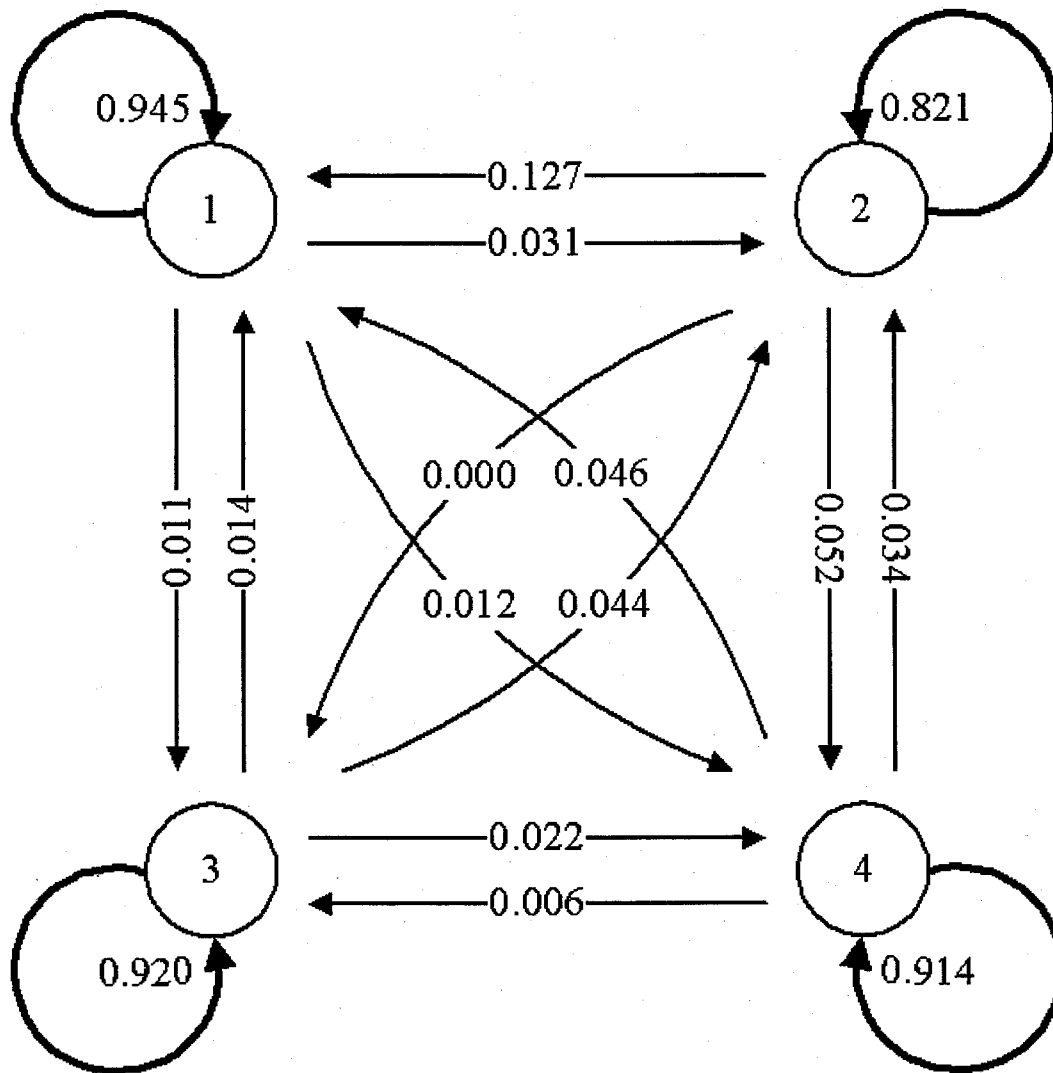
$p(x_{1,n}|s_{1,n} = 1) \sim N(-0.074, 1.168)$ and $\Pr(s_{1,n} = 1) = 0.354$. $f(x_{1,n})$: c) is a weighted sum of a) and b).

Figure 4.34: Two joint densities and the unconditional density for TOPIX.

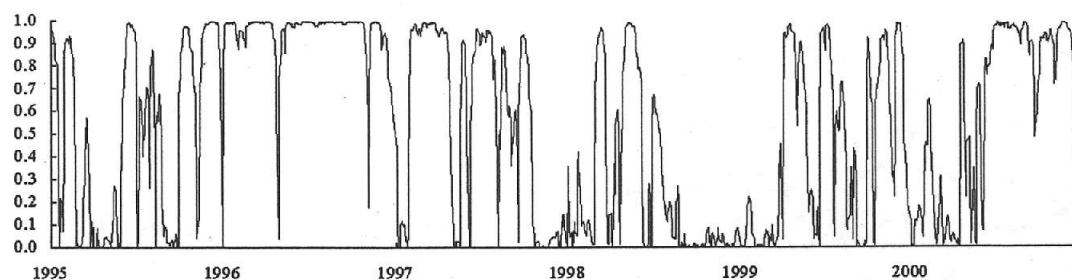
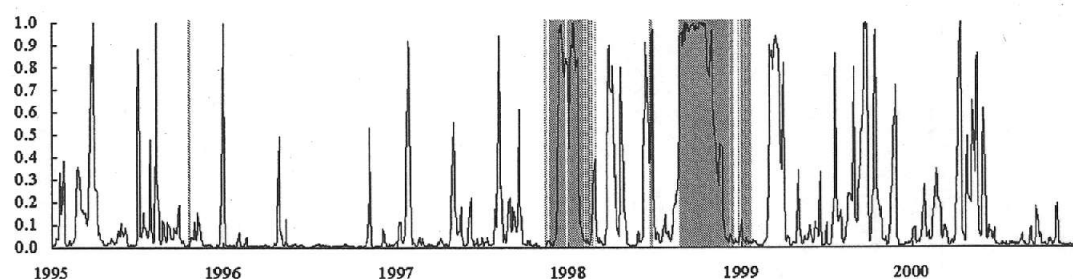
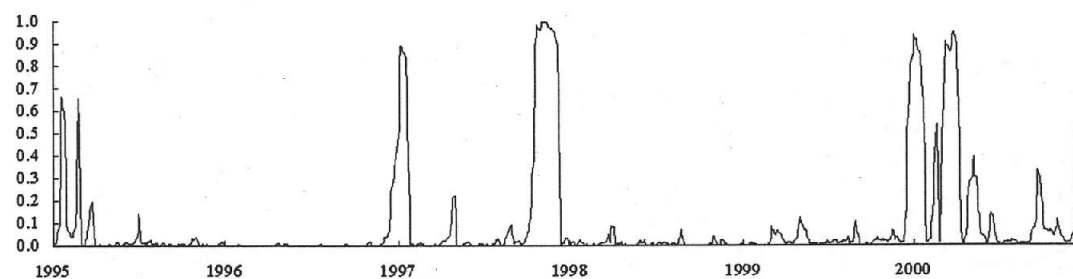
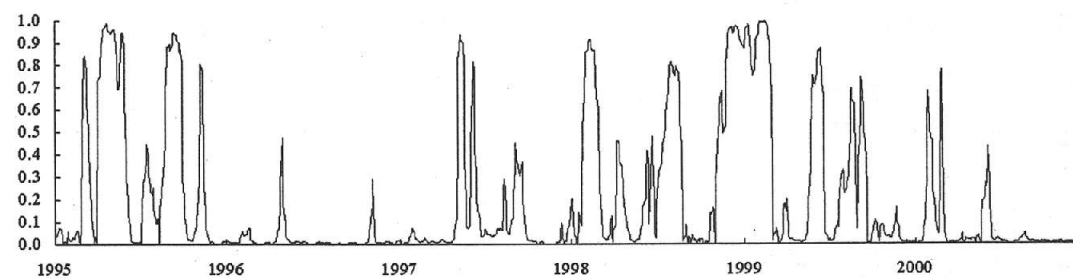


Note: $p(x_{1,n}, s_{1,n}) = p(x_{1,n}|s_{1,n}) \Pr(s_{1,n})$: a) is for the low volatility process regime; $p(x_{2,n}|s_{2,n} = 0) \sim N(-0.018, 0.897)$ and $\Pr(s_{1,n} = 0) = 0.751$, b) is for the high volatility process regime; $p(x_{2,n}|s_{2,n} = 1) \sim N(0.002, 1.935)$ and $\Pr(s_{1,n} = 1) = 0.250$. $p(x_{1,n})$: c) is a weighted sum of a) and b).

Figure 4.39: Transmission of volatility process between the JPY and the TOPIX.

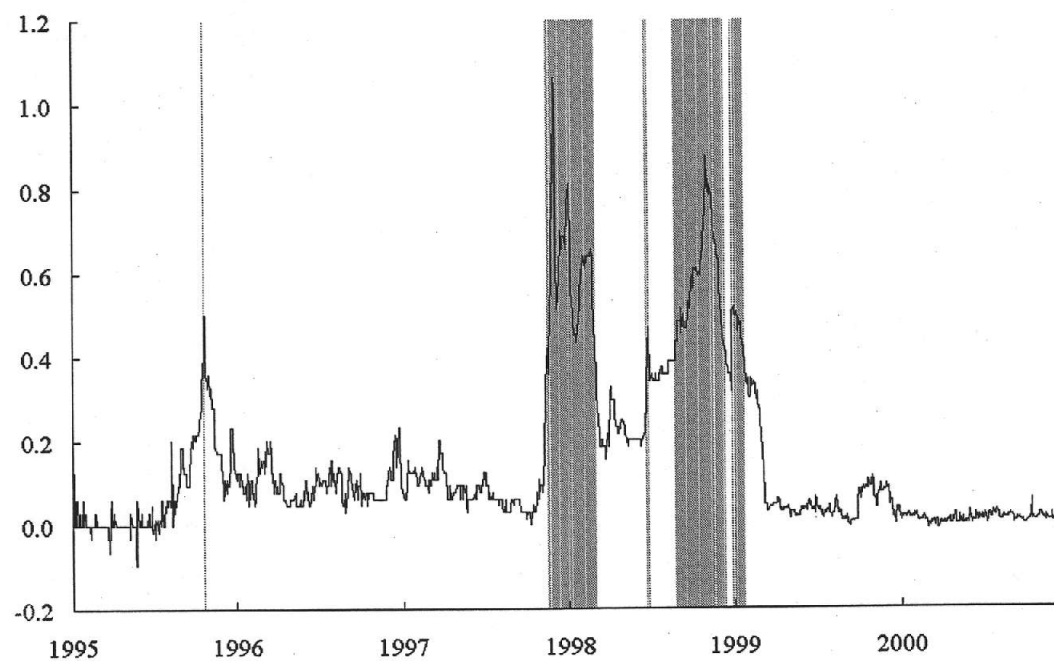


Note: Regime 1 ($R_n = 1$) is for the low-low volatility process. Regime 2 ($R_n = 2$) is for the high-high volatility process. Regime 3 ($R_n = 3$) is for the low volatility process of the JPY and the high volatility process of the TOPIX. Regime 4 ($R_n = 4$) is for the high volatility process of the JPY and the low volatility process of the TOPIX.

Figure 4.35: Smoothed probability of regime $R_n = 1$.Figure 4.36: Smoothed probability of regime $R_n = 2$.Figure 4.37: Smoothed probability of regime $R_n = 3$.Figure 4.38: Smoothed probability of regime $R_n = 4$.

Note: Regime 1 ($R_n = 1$) is for the low-low volatility process. Regime 2 ($R_n = 2$) is for the high-high volatility process. Regime 3 ($R_n = 3$) is for the low volatility process of the JPY and the high volatility process of the TOPIX. Regime 4 ($R_n = 4$) is for the high volatility process of the JPY and the low volatility process of the TOPIX.

Figure 4.40: Japan Premium.



Note: Shadowed bars show the periods when Japan premium is more than 40 basis points.

4.6 Self-organizing Markov Switching State Space Model

- Self-organizing Markov Switching State Space Model with Time-varying Transition Probability
- International Transmission of Business Cycles: a Self-organizing Markov-switching State Space Model

4.6.1 Introduction

Although many important problems in time series analysis can be solved using an ordinary state space model, more complex systems such as a Markov-switching model are sometimes required in general nonlinear or non-Gaussian situations (Kitagawa 1987). Recent improvements in computing power and the development of Monte Carlo-based algorithms (Kitagawa 1996) have made the use of nonlinear and non-Gaussian time series models feasible. Kitagawa (1998) proposed the self-organizing state space model to deal with the model identification problem associated with the use of the recursive Monte Carlo method. In this approach, the state vector is augmented by the unknown parameters of the model and the state and the parameters are estimated simultaneously by the recursive filter and smoother. In the method used in this section, accurate approximations of the marginal posterior densities of the state and the parameters are obtained by the Monte Carlo filter (Kitagawa 1996).

The Markov-switching model is a useful tool for capturing the dynamic structures of time series that change under the unknown discrete Markov chain regimes. Hamilton (1989) first applied a Markov-switching autoregressive model to the analysis of the U.S. business cycle using quarterly data. Hamilton's (1989) model assumes that GNP data shifts between low and high growth states or regimes according to a first-order Markov process. Variants of Hamilton's (1989) model have been used in single country analyses of the business cycle. Kim and Yoo (1995) and Kontolemis (2001) extended the univariate Markov-switching model to a multivariate version to capture the co-movement of some economic indicators. Filardo (1994) and Filardo *et al.* (1998) presented a Markov switching model with time-varying transition probabilities. Filardo's model changes the transition probability between states depending on the exogenous variables.

In this section, we propose a self-organizing Markov switching state space (SOMS) model, which incorporates the ordinary Markov switching model with the self-organizing state space model. Two types of SOMS models — a model with time-varying transition probability and a bivariate model — are described. These models are applied firstly to volatility analysis of the Japan/U.S. foreign exchange rate

and its relationship with interventions by the Bank of Japan, and secondly, to the transmission of the business cycle between the U.S. and Germany.

This section is organized as follows. In subsection 2 and 3 respectively, we present the self-organizing Markov switching state space modeling for time-varying transition probability and the bivariate version. An analysis of the Japan/U.S. exchange rate and the Bank of Japan interventions is presented in subsection 4, and an analysis of the international transmission of the business cycle is shown in subsection 5. Subsection 6 contains the conclusion.

4.6.2 The Self-organizing Markov Switching State Space Model with Time-varying Transition Probability

We introduce the self-organizing Markov switching state space model (SOMS) by incorporating the self-organizing method into the ordinary Markov switching model. In this section, we deal with two types of SOMS models — first, SOMS stochastic trend models with constant transition probability (Kitagawa and Hakamata 2001) and second, with exogenous time-varying transition probability (Kitagawa and Hakamata 2002). Let y_n , $n = 1, 2, \dots, N$, be the observed time series. The two-state first-order Markov switching state space model in the linear state space form is given by

$$x_n = x_{n-1} + v_n \quad (\text{System model}), \quad (4.143)$$

$$y_n = x_n + w_n \quad (\text{Observation model}), \quad (4.144)$$

where x_n is the k -dimensional state vector. w_n is the one-dimensional observation noise according to the density function $r(w)$, which does not necessarily follow a Gaussian distribution. The system noise v_n is the l -dimensional system noise according to the mixture density function including a Markov switching structure as follows:

$$q(v) = q_0(v)(1 - s_n) + q_1(v)s_n, \quad (4.145)$$

where $q_0(\cdot)$ and $q_1(\cdot)$ are arbitrary density functions; for example, Gaussian, Student, Cauchy and others. Here s_n is a latent two-state first-order Markov chain taking values of zero or one. s_n is updated via the self-loop transition probability as follows:

[Constant transition probability]

$$p_{ii,n} = \Pr(s_n = i | s_{n-1} = i, z_n) = \frac{\exp(\alpha)}{1 + \exp(\alpha)}, \quad \text{for } i = 0, 1, \quad (4.146)$$

where α take a constant value over time,

[Exogenous time-varying transition probability]

$$p_{ii,n} = \Pr(s_n = i | s_{n-1} = i, z_n) = \frac{\exp[g(z_n)]}{1 + \exp[g(z_n)]}, \quad \text{for } i = 0, 1, \quad (4.147)$$

where z_n is the exogenous information available up to n , and $g(\cdot)$ denotes an arbitrary linear or non-linear function. Equation (4.147) guarantees that $p_{ii,n}$ varies from zero to one. The transition probability matrix of the state $s_n = i (i = 0, 1)$ can be expressed by using the self-loop transition probabilities as follows:

$$P_n = \begin{pmatrix} p_{00,n} & 1 - p_{11,n} \\ 1 - p_{00,n} & p_{11,n} \end{pmatrix}. \quad (4.148)$$

Under the assumptions above, the joint conditional distribution of x_n and s_n given the previous values x_{n-1} and s_{n-1} is obtained by

$$\begin{aligned} p(x_n, s_n | x_{n-1}, s_{n-1}, z_n) &= p(s_n | x_{n-1}, s_{n-1}, z_n) p(x_n | s_n, x_{n-1}, s_{n-1}, z_n), \\ &= \Pr(s_n | s_{n-1}, z_n) p(x_n | x_{n-1}), \end{aligned} \quad (4.149)$$

where the two terms on the right-hand side of the above equation are specified by equations (4.143) and (4.147), respectively. Therefore, our SOMS models in equations (4.143) to (4.149) can be expressed in the general state space model form (Kitagawa 1996) as:

$$x_n \sim Q(\cdot | x_{n-1}, s_{n-1}) \quad (\text{System model}), \quad (4.150)$$

$$y_n \sim R(\cdot | x_n, s_n) \quad (\text{Observation model}), \quad (4.151)$$

Here, $Q(x_n, s_n | x_{n-1}, s_{n-1})$ is equivalent to the conditional distribution defined in equation (6), and $R(y_n | x_n, s_n)$ is the conditional density of the observation given the state.

We introduce the self-organizing state space formula (Kitagawa 1998) into the Markov switching model, and consider the self-organizing SOMS models for simultaneous estimation of the state and the parameters. For this purpose, we define an augmented state vector as follows:

$$x_n^* = (x_n, s_n, \theta_n)^T, \quad (4.152)$$

where x_n is the original state vector, s_n is the latent Markov chain and θ_n is the parameter vector defined in section 3. We assume a (vector) random walk model

$$\theta_n = \theta_{n-1} + u_n, \quad (4.153)$$

where u_n is a Gaussian white noise process with mean zero and covariance $\text{diag}\{\varepsilon_1, \dots, \varepsilon_k\}$. Here, all of the structural parameters are assumed to be time-varying. The SOMS model for this augmented state vector x_n^* in the general state space model form is immediately given by

$$x_n^* \sim Q^*(\cdot | x_{n-1}^*), \quad (4.154)$$

$$y_n \sim R^*(\cdot | x_n^*), \quad (4.155)$$

where $Q^*(x_n^*|x_{n-1}^*)$ is the conditional density of the augmented state given the previous one, and $R^*(y_n^*|x_n^*)$ is that of the observation given the augmented state. By applying the Monte Carlo filter to this self-organizing state space model, we can estimate the state vector and parameters simultaneously (Kitagawa 1998). The SOMS model has no need to estimate the parameter θ with a maximum likelihood method.

4.6.3 The Bivariate Self-organizing Markov Switching State Space Model

In this section, we extend the univariate SOMS model to a multivariate version (Hakamata and Kitagawa 2002). For simplicity, we consider a bivariate model for the time series $y_{k,n}$,

$$y_{k,n} = t_{k,n} + w_{k,n}, \quad (k = 1, 2) \quad (4.156)$$

where $t_{k,n}$ is an unknown individual trend component, $w_{k,n}$ is Gaussian observation noise with zero mean and variance σ_k^2 for each time series. For simplicity, we assume that the trend component $t_{k,n}$ follows the first-order stochastic trend model with individual drift as follows

$$t_{k,n} = t_{k,n-1} + v_{k,n}, \quad (4.157)$$

where $v_{k,n}$ is Gaussian system noise depending on a $\{0, 1\}$ -valued Markov chain $s_{k,n}$; that is,

$$p(t_{k,n}|t_{k,n-1}, s_{k,n} = j_k) = \begin{cases} N(\mu_{k,0}, \tau_k^2), & \text{if } j_k = 0, \\ N(\mu_{k,1}, \tau_k^2), & \text{if } j_k = 1. \end{cases} \quad (4.158)$$

The Markov chain $s_{k,n}$ indicates the "regime" of economic activity. If $j_k = 0$, the economy of the k -th country is expanding. If $j_k = 1$, it is contracting. The transition probability of $s_{k,n}$ is given by

$$p_{i_1 i_2, j_1 j_2} = \Pr(s_{1,n} = j_1, s_{2,n} = j_2 | s_{1,n-1} = i_1, s_{2,n-1} = i_2) \quad (4.159)$$

with $\sum_{j_1=0}^1 \sum_{j_2=0}^1 p_{i_1 i_2, j_1 j_2} = 1$ for all $i_1, i_2 = 0, 1$. This can be rewritten in transition matrix form as follows

$$P = \begin{pmatrix} p_{00,00} & p_{11,00} & p_{01,00} & p_{10,00} \\ p_{00,11} & p_{11,11} & p_{01,11} & p_{10,11} \\ p_{00,01} & p_{11,01} & p_{01,01} & p_{10,01} \\ p_{00,10} & p_{11,10} & p_{01,10} & p_{10,10} \end{pmatrix}. \quad (4.160)$$

The Markov switching model allows for several types of transmission of the business cycle between two countries, such as precedent, positive, and negative simultaneous regime shifts. Given the assumption

of this model specification, the joint conditional distribution of $t_{k,n}$ and $s_{k,n}$ given the previous values $t_{k,n-1}$, $s_{1,n-1}$ and $s_{2,n-1}$ is given by

$$\begin{aligned} p(t_{1,n}, t_{2,n}, s_{1,n}, s_{2,n} | t_{1,n-1}, t_{2,n-1}, s_{1,n-1}, s_{2,n-1}) \\ = \Pr(s_{1,n}, s_{2,n} | s_{1,n-1}, s_{2,n-1}) p(t_{1,n} | t_{1,n-1}, s_{1,n}) p(t_{2,n} | t_{2,n-1}, s_{2,n}), \end{aligned} \quad (4.161)$$

The terms on the right-hand side of the above equation are specified by equations (4.158) and (4.159), respectively.

Then, by defining the state vector x_n as

$$x_n = (t_{1,n}, t_{2,n}, s_{1,n}, s_{2,n})^T, \quad (4.162)$$

our bivariate Markov-switching model can be expressed in terms of the general state space model. In addition, we can use the self-organizing state space model for simultaneous estimation of the state and the parameters. Here, we obtain the SOMS model by incorporating the self-organizing state space form into the Markov switching model. For that purpose, we define an augmented state vector by $z_n \equiv (x_n^T, \theta_n^T)^T$ where x_n is the original state vector defined in equation (4.162) and θ is the parameter vector defined in the previous section. We assume a (vector) random walk model

$$\theta_n = \theta_{n-1} + u_n, \quad (4.163)$$

where u_n is Gaussian white noise with zero mean and covariance $\text{diag}\{\xi_1, \dots, \xi_{20}\}$. All of the structural parameters, σ_k^2 , τ_k^2 , $\mu_{k,0}$, $\mu_{k,1}$ and $p_{i_1 i_2, j_1 j_2}$ ($i_1, i_2, j_1 \in \{0, 1\}, j_2 = 0$) are assumed to be time-varying, and the parameter vector is defined by

$$\theta_n = (\sigma_{k,n}^2, \tau_{k,n}^2, \mu_{k,0,n}, \mu_{k,1,n}, p_{i_1 i_2, j_1 j_2, n})^T. \quad (4.164)$$

4.6.4 Application I: Foreign Exchange Volatility and Intervention

Data Description

The Bank of Japan has continued to intervene in foreign exchange markets. One of its goals is to reduce the increasing exchange rate volatility. There are several reasons why the Bank of Japan wants to reduce exchange rate volatility (Bonser-Neal 1996). First, volatility may impede international investment flows. Second, it may adversely affect international trade. Third, volatility could spread throughout domestic financial markets. Figure 4.41 shows the daily Japan/U.S. exchange rate, the logarithmic returns and the Bank of Japan's interventions, respectively. From Figure 4.42 it is difficult to see significant evidence of a relationship between the Japan/U.S. foreign exchange volatility and the Bank of Japan's interventions.

We apply the SOMS model with exogenous time-varying transition probability to the daily Japan/U.S. exchange rate. Intervention in the Japan/U.S. exchange rate market by the Bank of Japan is used for the exogenous variable. Each sample size is 985, using data from January 1997 to December 2000.

Model Specification and Estimation Results

To identify the periods when the Japan/U.S. exchange rate is in a high volatility state, and to investigate the influence of intervention on foreign exchange volatility, we use the first-order stochastic trend model (random walk model) with Markov switching mean and variance in the trend components as follows:

$$x_n = \mu + x_{n-1} + v_n \quad v_n \sim N(0, \tau^2), \quad (4.165)$$

$$y_n = x_n + w_n, \quad w_n \sim N(0, \sigma^2), \quad (4.166)$$

where

$$\mu = \mu_0(1 - s_n) + \mu_1 s_n, \quad (4.167)$$

$$\tau = \tau_0(1 - s_n) + \tau_1 s_n, \quad \tau_0 < \tau_1. \quad (4.168)$$

Here s_n is the latent first-order two-state Markov chain taking values of zero or one. Since it is assumed that $\tau_0 < \tau_1$, $s_n = 0$ and $s_n = 1$ can be identified as the low and high volatility states, respectively. $g(\cdot)$ in equation (5) is defined by

$$g(z_n) = \alpha_i + \beta_i z_n, \quad \text{for } i = 0, 1, \quad (4.169)$$

where z_n is the absolute value of the intervention. A greater amount of intervention enhances the volatility state persistence for both low and high volatility states. The parameter vector is defined by

$$\theta_n \equiv \{ \log \tau_{0,n}, \log \tau_{1,n}, \log \sigma_n, \mu_{0,n}, \mu_{1,n}, \alpha_{0,n}, \alpha_{1,n}, \beta_{0,n}, \beta_{1,n} \}^T. \quad (4.170)$$

Figure 4.45 shows the results of the proposed model with $\varepsilon_1 = \dots = \varepsilon_9 = 0.01$. The top three plots show the marginal posterior means of the probability of the state $s_n = 0$. The second and third plots show the evolution of the nine parameters. From the top they are τ_1 , τ_0 , μ_0 and μ_1 in the second plot; and α_0 , α_1 , β_1 , σ and β_0 in the third plot, respectively. It can be seen that even though it was assumed that these parameters were time-varying, they are rather stable and satisfy the assumption of $\tau_0 < \tau_1$ over time. Because $\beta_{0,n} < 0$ in the low volatility state of $s_n = 0$, and $\beta_{1,n} > 0$ in the high volatility state of $s_n = 1$, it can be concluded that the Bank of Japan's interventions tend to increase the volatility of the Japan/U.S. exchange rate.

4.6.5 Application II: International Business Cycle Transmission

Data Description

In this section, we apply the SOMS model to analyze transmission of the business cycle between the U.S. and Germany. Because each country's business cycle is not observed, we must use the observable economic indicators as proxy variables. We use monthly industrial production indices (seasonally adjusted, and equal to 100 in the base year, 1995) for the U.S. and Germany from January 1961 to December 2000. The indices are provided by the OECD and Figure 4.47 shows the original time series.

Estimation Results

Before moving on to the results of the SOMS model, it is desirable to describe the maximum likelihood estimate (MLE) of the Markov switching model in equations (4.156) and (4.157) with fixed-value parameters. Figure 4.48 illustrates the estimated transition probabilities between the four regimes. These approximate MLE are obtained by Hamilton's filter (Hamilton 1989) and the maximum likelihood method.

We then turn to an application of the SOMS model. Because the computational difficulties increase as the dimension of the model becomes higher, we attempt to self-turn the parameters with the SOMS model. Figure 4.49 shows the smoothed joint probabilities $\Pr(s_{1,n} = j_1, s_{2,n} = j_2)$ for $j_1, j_2 = 0, 1$ from the SOMS model with $\xi_1 = \dots = \xi_{20} = 0.01$. The shadowed bars (in Figure 4.49) denote those from Kim's smoothing (Kim 1994) algorithm given the approximate MLE above. Note that the number of particles is $m = 20,000$, and the initial mean values of the parameter vector are $\sigma_{1,0}^2 = 0.03, \sigma_{2,0}^2 = 0.75, \tau_{1,0}^2 = 0.45^2, \tau_{2,0}^2 = 0.60^2, \mu_{1,0,0} = 0.30, \mu_{2,0,0} = 0.30, \mu_{1,1,0} = -0.50, \mu_{2,1,0} = -0.50$ and

$$P_0 = \begin{pmatrix} 0.97 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.97 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.97 & 0.01 \\ - & - & - & - \end{pmatrix}. \quad (4.171)$$

Although we set the initial distributions of the parameter vector in equation (4.164) approximately, the smoothed joint probabilities from the SOMS model are quite similar to the results from the approximate MLE. We obtained the results from a single run of the Monte Carlo filter and smoother. From Figure 4.49, it is apparent that all regimes are highly persistent, and that the SOMS model succeeds in classifying the whole period into four regimes.

Therefore, it can be said that the SOMS model allows the estimate of the unobserved regimes of the business cycles between two countries without an MLE, and captures the international transmission and

relationships of business cycles.

4.6.6 Conclusions

In this section, we proposed a self-organizing Markov-switching state space (SOMS) model to evaluate the international transmission of the business cycle. Our SOMS model is an effective analytical tool in understanding complex systems. Although the application of the Monte Carlo filter to the Markov-switching model may make parameter estimation difficult, we dealt with this by incorporating a self-organizing state space model into the Markov-switching model. In the SOMS model, the state and unknown parameters are estimated simultaneously. We applied our model to an analysis of business cycles in the U.S. and Germany, and successfully captured the characteristics of the transmission of, and relationship between, business cycles in different countries.

Figure 4.41: Japan/U.S. exchange rate.

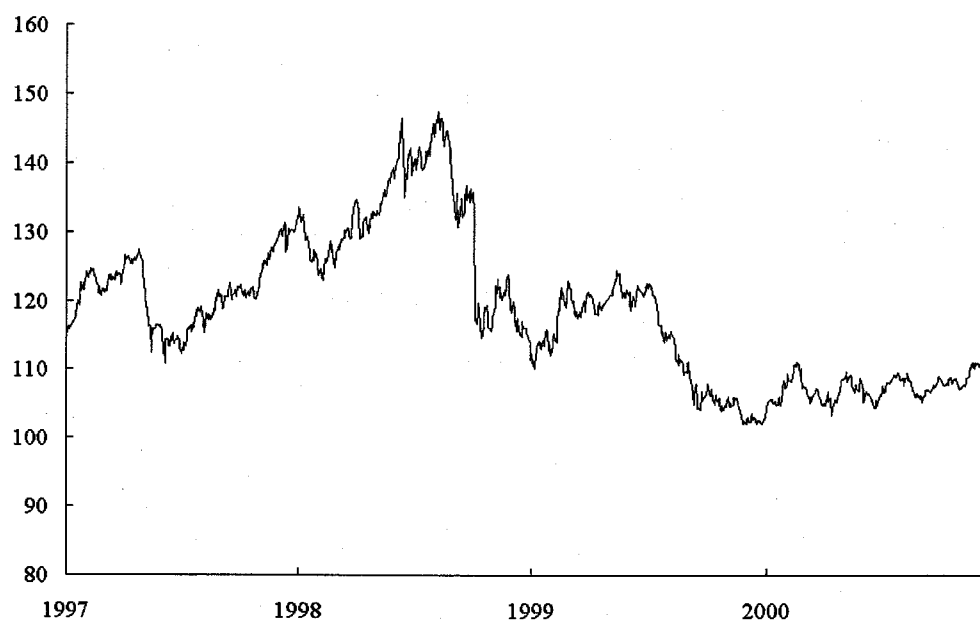


Figure 4.42: Logarithmic return of the Japan/U.S. exchange rate.

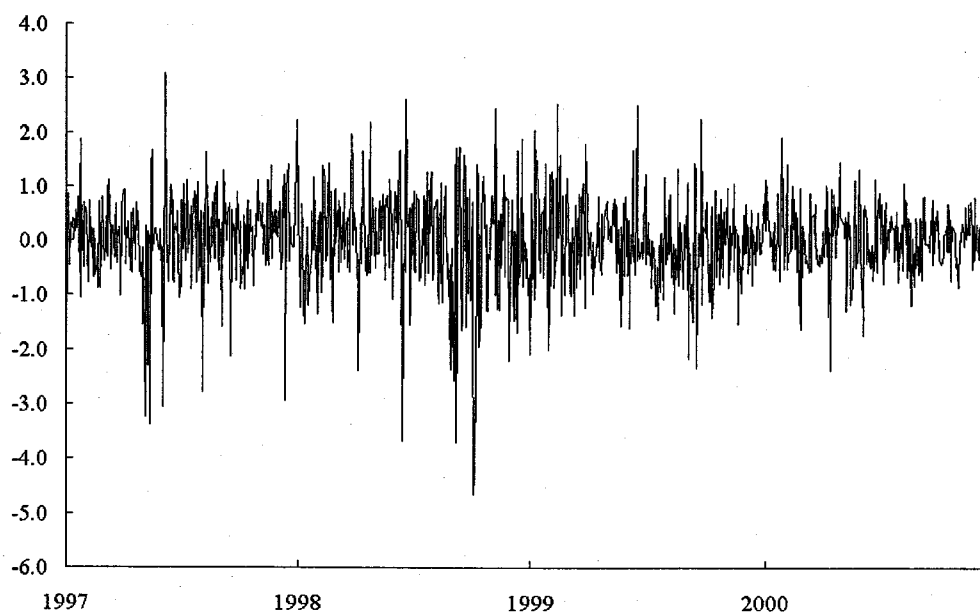
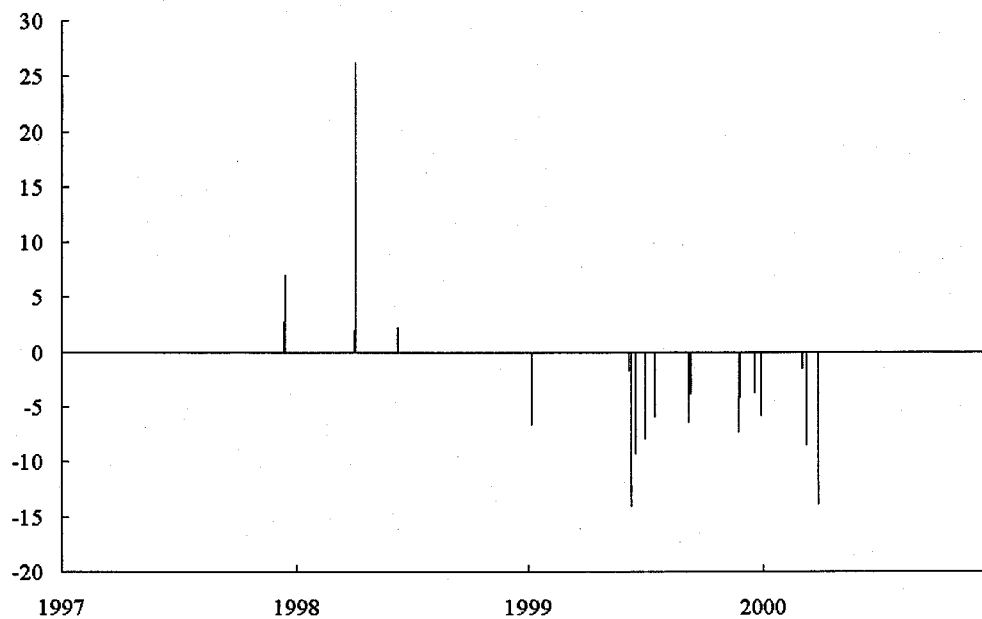
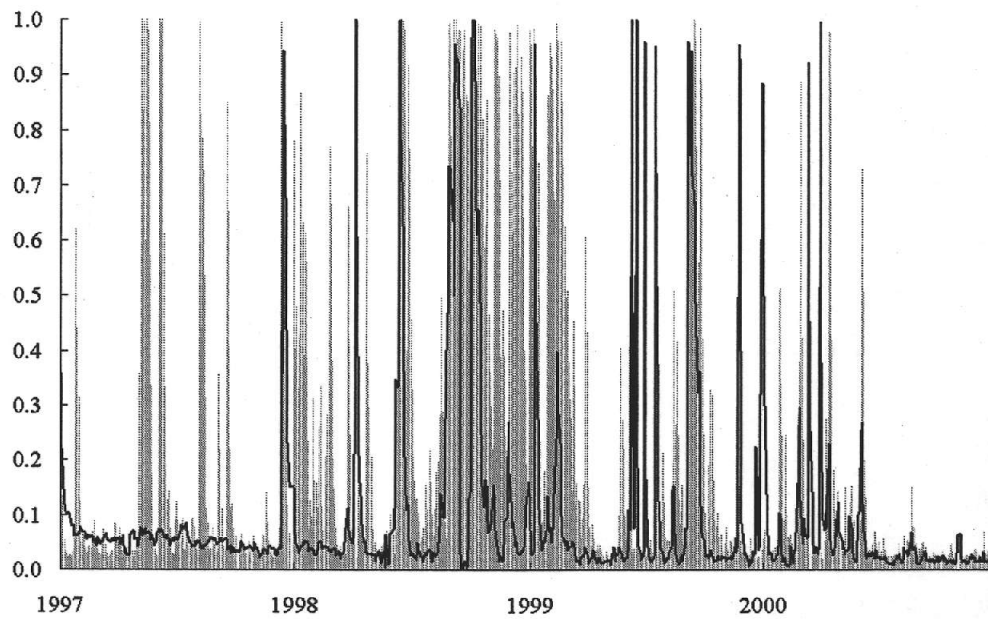


Figure 4.43: Amount of intervention (100 billion yen) by the Bank of Japan.



Note: Japanese Yen purchases by the Bank of Japan are denoted as positive values, and Japanese Yen sales are denoted as negative values.

Figure 4.44: Filtered probability of high volatility regime.



Note: Shaded bars show the filtered probability of a high volatility regime based on a maximum likelihood estimate.

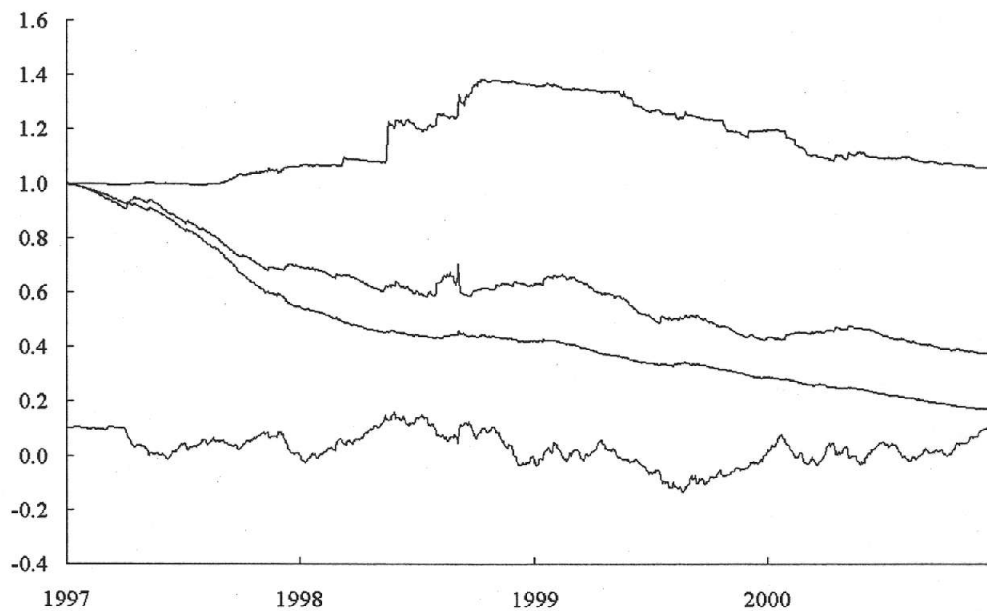
Figure 4.45: Time-varying parameter: τ_1 , τ_0 , μ_0 and μ_1 (from top to bottom).

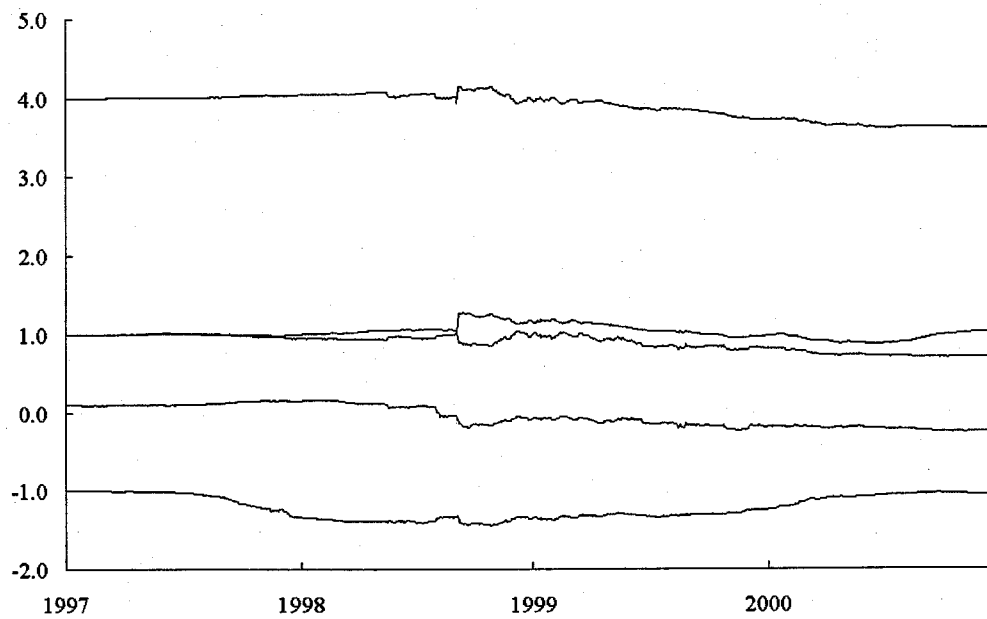
Figure 4.46: Time-varying parameter: α_0 , α_1 , β_1 , σ and β_0 (from top to bottom).

Figure 4.47: Industrial production indices for the U.S. and Germany.

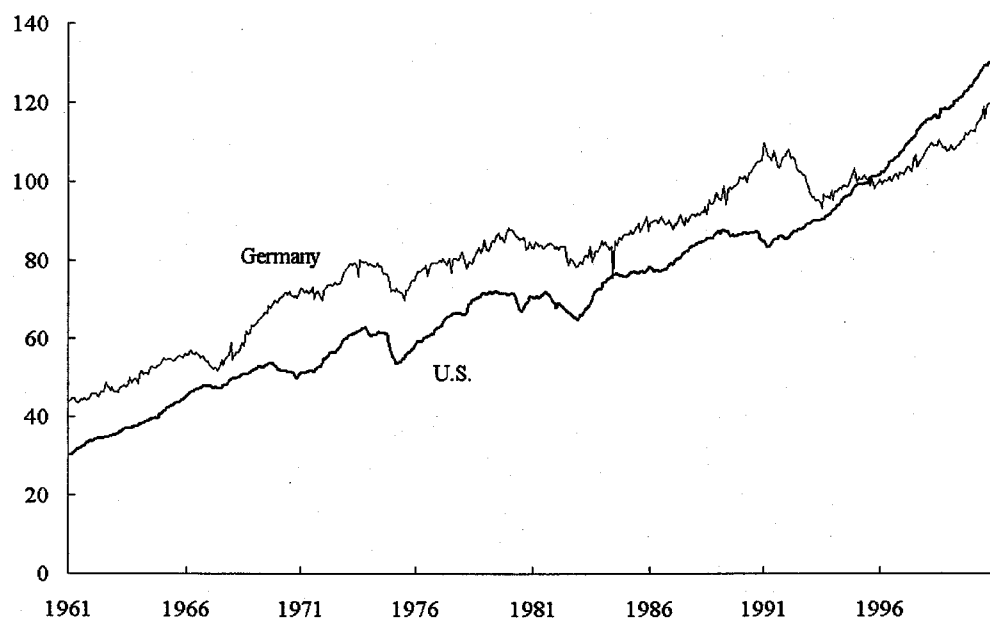
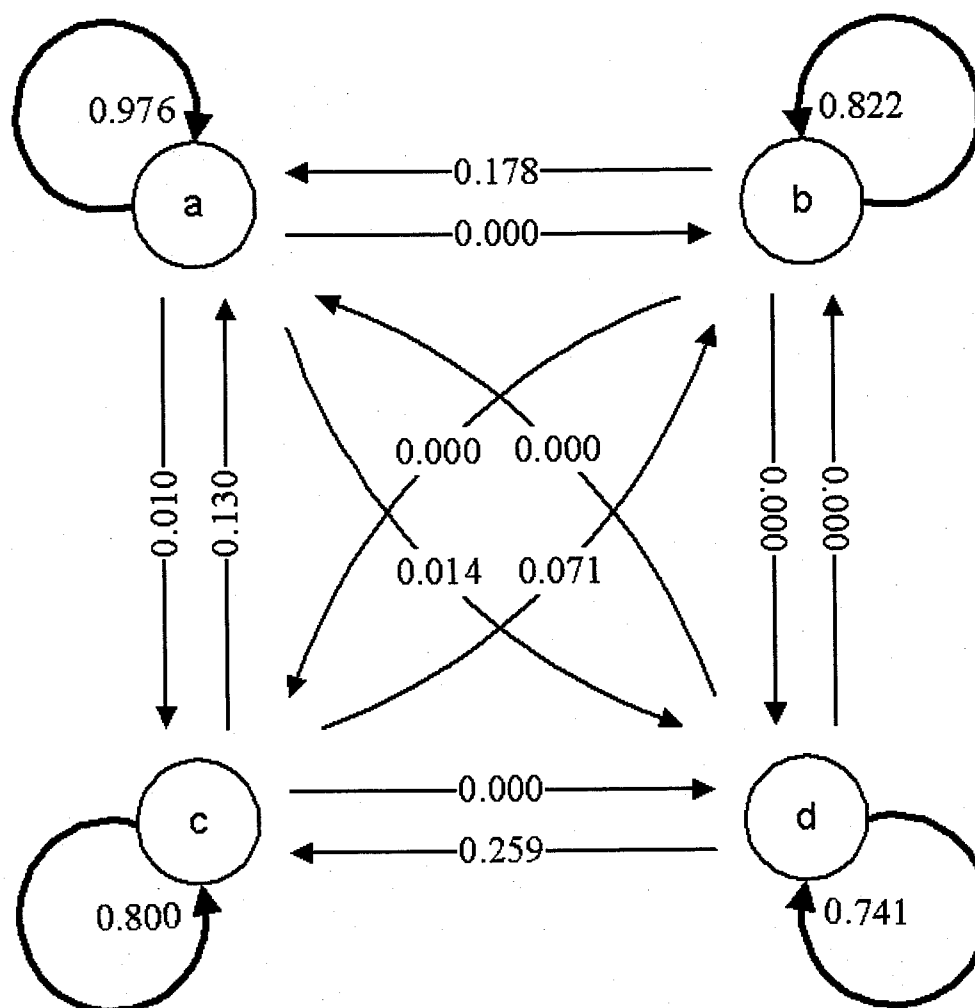
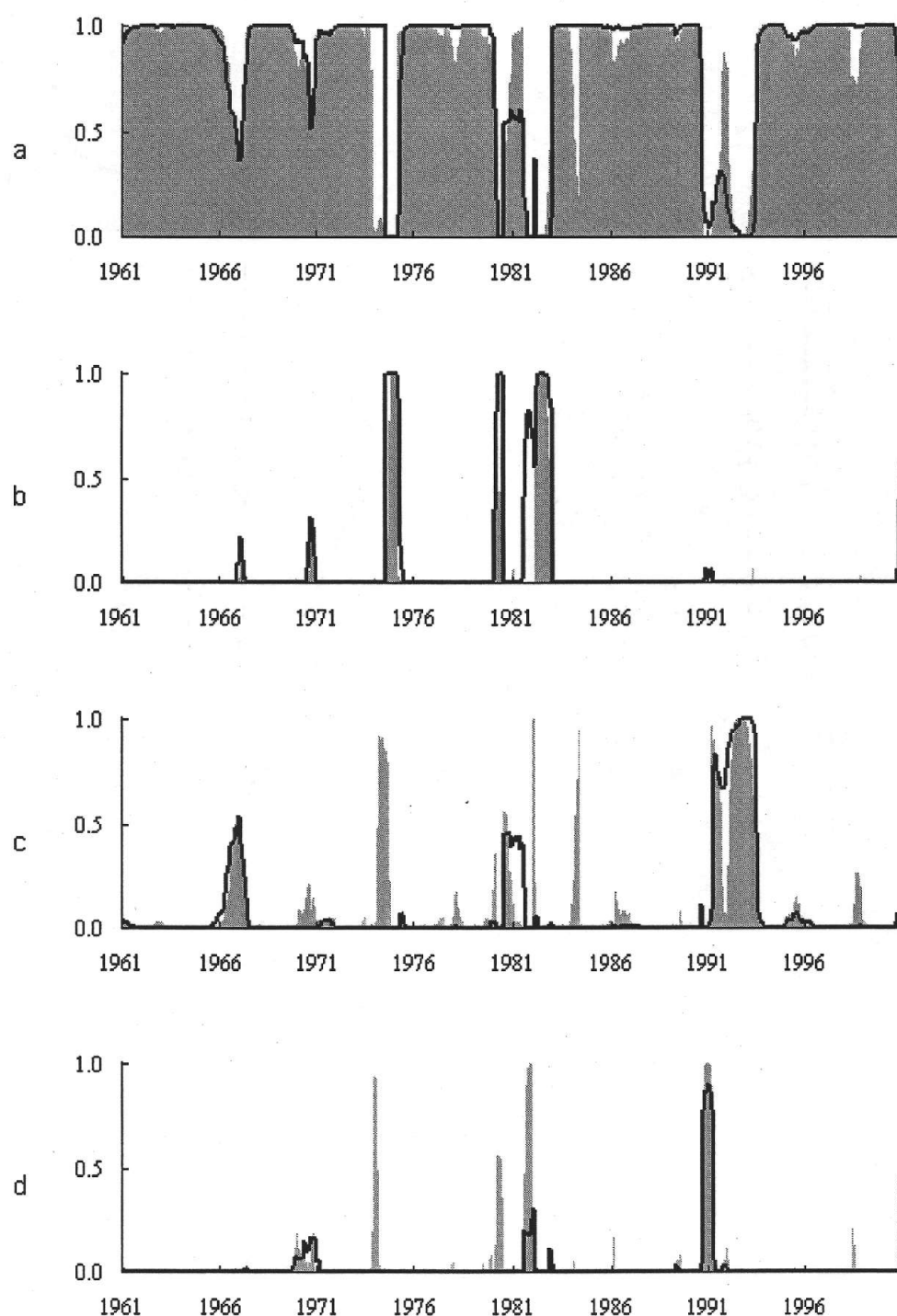


Figure 4.48: Transition probabilities of the approximate maximum likelihood estimate.



Note: Circle a is for $s_{1,n} = 0$ and $s_{2,n} = 0$, circle b is for $s_{1,n} = 1$ and $s_{2,n} = 1$, circle c is for $s_{1,n} = 0$ and $s_{2,n} = 1$, circle d is for $s_{1,n} = 1$ and $s_{2,n} = 0$.

Figure 4.49: Smoothed joint probabilities of the SOMS model.



Note: a is for $\Pr(s_{1,n} = 0, s_{2,n} = 0)$, b is for $\Pr(s_{1,n} = 1, s_{2,n} = 1)$, c is for $\Pr(s_{1,n} = 0, s_{2,n} = 1)$, d is for $\Pr(s_{1,n} = 1, s_{2,n} = 0)$. The shadowed bars denote the smoothed joint probabilities of the approximate maximum likelihood estimate.

Chapter 5

Summary of Extended Markov Switching Models

We describe some extended Markov switching models used in the empirical analysis of chapter 4, by the gross. The new models proposed in this thesis are based on modified ordinary Markov switching models. In addition, we briefly discuss the specifications of semi-Markov switching models, which we intend to study in future. Table 5.1 lists the empirical analysis topics corresponding with the extended Markov switching models described in this chapter.

Table 5.1: Corresponding Empirical Analysis Topics.

Extended Markov Switching Model	Empirical Analysis Topics in the Chapter 4
5.1 The modified Markov switching ARCH model	4.5 Volatility transmission
5.2 The modified multivariate Markov switching ARCH model	4.5 Volatility transmission
5.3 The Markov switching slope change and ARCH model	4.1 Trend identification and trading strategy
5.4 The Markov switching models with exogenous leverage effects	4.2 Time-series and cross-sectional volatility
5.5 The self-organizing Markov switching state-space modeling	4.6 Foreign exchange volatility and International business cycle transmission
5.6 The semi-Markov switching model	—

5.1 Modified Markov Switching ARCH Model

In this section, we modify the SWARCH model (Hamilton and Susmel 1994, Cai 1994) and the switching-regime ARCH model (Cai 1994), and introduce a modified Markov switching autoregressive heteroskedasticity (MS-ARCH) model. Let y_n be an observation, such as the logarithmic returns of stock price, foreign exchange and so on. The univariate MS-ARCH model is given by

$$y_n = \mu + \varepsilon_n, \quad \varepsilon_n | \Psi_{n-1} \sim N(0, h_{s_n, n}), \quad (5.1)$$

where Ψ_{n-1} is the information up to time $n-1$, and μ is a drift term, ε_n is a stochastic process according to the time-varying variance h_{s_n} given by

$$h_{s_n, n} = \gamma_{s_n} + \sum_{k=1}^K \alpha_{k, s_n} \varepsilon_{n-k}^2, \quad (5.2)$$

where γ_{s_n} and α_{k, s_n} represent the constant intercept in ARCH, and the k -lagged ARCH coefficient, respectively. Both γ_{s_n} and α_{k, s_n} change depending on the m -state latent discrete Markov chain $s_n \in \{1, 2, \dots, m\}$.

Our specification of the ARCH process with Markov switching structure is a general case of the prior studies (Hamilton and Susmel, 1994 Cai 1994). In the SWARCH model of Hamilton and Susmel (1994), the magnitude of the regime shift in the constant intercept in ARCH and the ARCH coefficients are the same through the common scale parameter g_{s_n} . In the regime-switching ARCH model of Cai (1994), only the constant intercept in ARCH has a Markov switching process. On the other hand, our MS-ARCH model allows for the Markov switching shifts of γ_{s_n} and α_{k, s_n} without common scale parameters.

The unobserved Markov chain s_n is assumed to follow a first-order and m -state Markov process with a time-homogeneous transition probability as follows:

$$p_{ij} = \Pr(s_n = j | s_{n-1} = i), \quad i, j = 1, 2, \dots, m, \quad (5.3)$$

where

$$\sum_{j=1}^m p_{ij} = 1 \quad \text{for all } i = 1, 2, \dots, m. \quad (5.4)$$

The density of y_n , conditional on s_n taking the value j , is

$$f(y_n | s_n = i, \Psi_{n-1}) = \frac{1}{\sqrt{2\pi h_{s_n}}} \exp \left\{ -\frac{(y_n - \mu)^2}{2h_{s_n}} \right\}. \quad (5.5)$$

To determine the log likelihood function, we first consider the joint density of y_n and s_n , which is the product of the conditional and marginal densities

$$f(y_n, s_n = j | \Psi_{n-1}) = f(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1}), \quad (5.6)$$

where Ψ_{n-1} refers to information up to time $t-1$. Then, to obtain the marginal density of y_n , we integrate s_n out of the above joint density by summing all possible values of s_n

$$\begin{aligned} f(y_n|\Psi_{n-1}) &= \sum_{j=1}^m f(y_n, s_n = j|\Psi_{n-1}) \\ &= \sum_{j=1}^m f(y_n|s_n = j, \Psi_{n-1}) \Pr(s_n = j|\Psi_{n-1}) \\ &= \sum_{j=1}^m \frac{1}{\sqrt{2\pi h_{j,n}}} \exp \left\{ -\frac{(y_n - \mu)^2}{2h_{j,n}} \right\} \Pr(s_n = j|\Psi_{n-1}), \end{aligned} \quad (5.7)$$

where $\Pr(s_n = j|\Psi_{n-1})$ can be obtained recursively by using a non-Gaussian filter (Kitagawa 1987) as follows

$$\Pr(s_n = j|\Psi_n) = \frac{f(y_n|s_n = j, \Psi_{n-1}) \Pr(s_n = j|\Psi_{n-1})}{f(y_n|\Psi_{n-1})}, \quad (5.8)$$

and

$$\Pr(s_n = j|\Psi_{n-1}) = \sum_{i=1}^m \Pr(s_n = j|s_{n-1} = i) \Pr(s_{n-1} = i|\Psi_{n-1}). \quad (5.9)$$

The log likelihood function is then given by

$$\begin{aligned} L(\theta) &= \sum_{n=1}^N \log f(y_n|\Psi_{n-1}). \\ &= \sum_{n=1}^N \log \left[\sum_{j=1}^m f(y_n, s_n = j, s_{n-1} = i|\Psi_{n-1}) \right] \\ &= \sum_{n=1}^N \log \left[\sum_{j=1}^m \sum_{i=1}^m \Pr(s_n = j|s_{n-1} = i) \Pr(s_{n-1} = i|\Psi_{n-1}) \right. \\ &\quad \left. \times \frac{1}{\sqrt{2\pi h_{j,n}}} \exp \left\{ -\frac{(y_n - \mu)^2}{2h_{j,n}} \right\} \right]. \end{aligned} \quad (5.10)$$

Here, for ε_n to be variance stationary, we need to have ARCH constants and coefficients in equation (5.2) as follows

$$\gamma_{s_n} > 0, \quad \alpha_{k,j} \geq 0 \quad \text{and} \quad \sum_{k=1}^K \alpha_{k,j} < 1, \quad (5.11)$$

for all $j = 1, 2, \dots, m$, and $k = 1, 2, \dots, k$. If ε_n is variance stationary, the conditional variance and the unconditional variance respectively are given by

$$E(\varepsilon_n^2|s_n = j) = \frac{\gamma_j}{1 - \sum_{k=1}^K \alpha_{k,j}}, \quad j = 1, 2, \dots, m, \quad (5.12)$$

and

$$E(\varepsilon_n^2) = \frac{\sum_{j=1}^m \gamma_j \Pr(s_n = j|\Psi_{n-1})}{1 - \sum_{k=1}^K \sum_{j=1}^m \alpha_{k,j}}. \quad (5.13)$$

5.2 Modified Multivariate Markov Switching ARCH Model

In this subsection, we extend the univariate MS-ARCH model described in the previous subsection to a multivariate version. This multivariate MS-ARCH model is a modification of the multivariate SWARCH model by Hamilton and Lin (1996) and Susmel (2000). Let $y_n \equiv (y_{1,n}, y_{2,n}, \dots, y_{k,n})^T$ be the $(k \times 1)$ observation vector. The multivariate Markov switching autoregressive conditional heteroskedasticity (MS-ARCH) model is given by

$$y_n = \mu + \varepsilon_n, \quad \varepsilon_n | \Psi_{n-1} \sim N(0, \Omega_{s_n, n}), \quad (5.14)$$

where Ψ_{n-1} is the information up to time $n-1$, and the $(k \times 1)$ drift vector is

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix}, \quad (5.15)$$

and the time-varying $(k \times k)$ variance-covariance matrix is

$$\Omega_{s_n, n} = \text{diag} \begin{pmatrix} h_1(s_{1,n}) \\ h_2(s_{2,n}) \\ \vdots \\ h_k(s_{k,n}) \end{pmatrix} = \text{diag} \begin{pmatrix} \gamma_{1,s_n} + \sum_{l=1}^L \alpha_{l,1,s_n} \varepsilon_{1,n-l}^2 \\ \gamma_{2,s_n} + \sum_{l=1}^L \alpha_{l,2,s_n} \varepsilon_{2,n-l}^2 \\ \vdots \\ \gamma_{k,s_n} + \sum_{l=1}^L \alpha_{l,k,s_n} \varepsilon_{k,n-l}^2 \end{pmatrix}, \quad (5.16)$$

The first and second terms of the right-hand side of equation (5.16), $\gamma_{1,s_n}, \dots, \gamma_{k,s_n}$, and, $\alpha_{l,1,s_n}, \dots, \alpha_{l,k,s_n}$, represent the constant intercepts in ARCH, and the ARCH coefficients, respectively.

In the same manner as the multivariate Markov switching stochastic trend model, k -dimensional and m -state Markov switching $s_{l,n}$, $l = 1, 2, \dots, k$, can be re-defined with the univariate m^k -dimensional Markov chain $s_n^* \in \{1, 2, \dots, m^k\}$. Depending on s_n^* , $\Omega_{s_{l,n}}$ can also be replaced by $\Omega_{s_n^*}^*$ as follows

$$\begin{aligned} \text{if } s_n^* = 1, & \quad \Omega_{1,n}^* = \text{diag}(h_1(1), h_2(1), \dots, h_k(1))^T, \\ \text{if } s_n^* = 2, & \quad \Omega_{1,n}^* = \text{diag}(h_1(1), h_2(2), \dots, h_k(1))^T, \\ & \quad \vdots \\ \text{if } s_n^* = m^k - 1, & \quad \Omega_{1,n}^* = \text{diag}(h_1(m), h_2(m), \dots, h_k(m-1))^T, \\ \text{if } s_n^* = m^k, & \quad \Omega_{1,n}^* = \text{diag}(h_1(m), h_2(m), \dots, h_k(m))^T. \end{aligned}$$

With the newly defined Markov chain s_n^* , the joint conditional distribution of y_n , s_n^* and s_{n-1}^* on Ψ_{n-1} is given by

$$\begin{aligned} f(y_n, s_n^* = j^*, s_{n-1}^* = i^* | \Psi_{n-1}) &= f(y_n | s_n^* = j^*, s_{n-1}^* = i^*, \Psi_{n-1}) \\ &\quad \times \Pr(s_n^* = j^*, s_{n-1}^* = i^* | \Psi_{n-1}), \end{aligned} \quad (5.17)$$

where Ψ_{n-1} denotes the information available up to time $n - 1$. The first and second terms of the right-hand side of equation (5.17) are expressed respectively by

$$f(y_n | s_n^* = j^*, s_{n-1}^* = i^*, \Psi_{n-1}) = (2\pi)^{-\frac{k}{2}} |\Omega_{s_n^*, n}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \bar{y}_n^T \Omega_{s_n^*, n}^{-1} \bar{y}_n \right\}, \quad (5.18)$$

and

$$\Pr(s_n^* = j^*, s_{n-1}^* = i^* | \Psi_{n-1}) = \Pr(s_n^* = j^* | s_{n-1}^* = i^*) \Pr(s_{n-1}^* = i^* | \Psi_{n-1}), \quad (5.19)$$

where s_n^* follows the first-order m^k -state Markov chain with the transition probability $\Pr(s_n^* = j^* | s_{n-1}^* = i^*)$ in a general case. Note that it is possible to consider several types of transitions (general, independent and simultaneous) as $\Pr(s_n^* = j^* | s_{n-1}^* = i^*)$.

The conditional density of y_n can be obtained over all possible states:

$$\begin{aligned} f(y_n | \Psi_{n-1}) &= \sum_{j^*=1}^{m^k} \sum_{i^*=1}^{m^k} f(y_n, s_n^* = j^*, s_{n-1}^* = i^* | \Psi_{n-1}) \\ &= \sum_{j^*=1}^{m^k} \sum_{i^*=1}^{m^k} \Pr(s_n^* = j^* | s_{n-1}^* = i^*) \Pr(s_{n-1}^* = i^*) \end{aligned} \quad (5.20)$$

$$\times (2\pi)^{-\frac{k}{2}} |\Omega_{s_n^*, n}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \bar{y}^T \Omega_{s_n^*, n}^{-1} \bar{y} \right\}. \quad (5.21)$$

The log likelihood of the multivariate MS-ARCH model is given by

$$\begin{aligned} L(\theta) &= \sum_{n=1}^N \log f(y_n | \Psi_{n-1}) \\ &= \sum_{n=1}^N \log \left[\sum_{s_n^*=1}^{m^k} \sum_{s_{n-1}^*=1}^{m^k} f(y_n, s_n^* = j^*, s_{n-1}^* = i^* | \Psi_{n-1}) \right] \\ &= \sum_{n=1}^N \log \left[\sum_{j^*=1}^{m^k} \sum_{i^*=1}^{m^k} \Pr(s_n^* = j^* | s_{n-1}^* = i^*) \Pr(s_{n-1}^* = i^* | \Psi_{n-1}) \right. \\ &\quad \times (2\pi)^{-\frac{k}{2}} |\Omega_{s_n^*, n}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \bar{y}^T \Omega_{s_n^*, n}^{-1} \bar{y} \right\} \left. \right] \\ &= \sum_{n=1}^N \log \left[\sum_{j^*=1}^{m^k} \Pr(s_n^* = j^* | \Psi_{n-1}) (2\pi)^{-\frac{k}{2}} |\Omega_{s_n^*, n}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \bar{y}^T \Omega_{s_n^*, n}^{-1} \bar{y} \right\} \right]. \end{aligned} \quad (5.22)$$

where θ denotes the unknown parameters.

For $\varepsilon_n = \text{diag}(\varepsilon_{1,n}, \varepsilon_{2,n}, \dots, \varepsilon_{k,n})^T$ to be variance stationary,

$$\gamma_{1,s_n^*}, \dots, \gamma_{k,s_n^*} > 0, \quad \alpha_{l,1,s_n^*}, \dots, \alpha_{l,k,s_n^*} \geq 0,$$

and

$$\sum_{l=1}^L \alpha_{l,1,s_n^*}, \dots, \sum_{l=1}^L \alpha_{l,k,s_n^*} < 1,$$

for all $l = 1, 2, \dots, L$ and $s_n^* = 1, 2, \dots, m^k$. If ε_n is variance stationary, the conditional variance and the unconditional variance respectively are given by

$$E(\varepsilon_n^T \varepsilon_n | s_n^* = j^*) = \begin{pmatrix} \frac{\gamma_{1,s_n^*}}{1 - \sum_{l=1}^L \alpha_{l,1,s_n^*}} \\ \frac{\gamma_{2,s_n^*}}{1 - \sum_{l=1}^L \alpha_{l,2,s_n^*}} \\ \vdots \\ \frac{\gamma_{k,s_n^*}}{1 - \sum_{l=1}^L \alpha_{l,k,s_n^*}} \end{pmatrix} \otimes I_k, \quad (5.23)$$

and

$$E(\varepsilon_n^T \varepsilon_n) = \begin{pmatrix} \frac{\sum_{j^*=1}^{m^k} \gamma_{1,s_n^*} \Pr(s_n^* = j^* | \Psi_{n-1})}{1 - \sum_{l=1}^L \sum_{j^*=1}^{m^k} \alpha_{l,1,j^*}} \\ \frac{\sum_{j^*=1}^{m^k} \gamma_{2,s_n^*} \Pr(s_n^* = j^* | \Psi_{n-1})}{1 - \sum_{l=1}^L \sum_{j^*=1}^{m^k} \alpha_{l,2,j^*}} \\ \vdots \\ \frac{\sum_{j^*=1}^{m^k} \gamma_{k,s_n^*} \Pr(s_n^* = j^* | \Psi_{n-1})}{1 - \sum_{l=1}^L \sum_{j^*=1}^{m^k} \alpha_{l,k,j^*}} \end{pmatrix} \otimes I_k. \quad (5.24)$$

5.3 Markov Switching Slope Change and ARCH Model

By incorporating the structure of slope changes into the Markov switching ARCH model, we introduced the Markov Switching Slope change and ARCH model (MS-SC-ARCH). In this model, two components of the system simultaneously shift between discrete-valued states. Let y_n , $n = 1, 2, \dots, N$, be the observed time series. The Markov switching slope change and ARCH model is given by

$$y_n = t_n + \varepsilon_n, \quad \varepsilon_n | \Psi_{n-1} \sim N(0, h_{s_n,n}), \quad (5.25)$$

$$h_{s_n,n} = \gamma_0(1 - s_n) + \gamma_1 s_n + \alpha \varepsilon_{n-1}^2, \quad (5.26)$$

where $\gamma_1 > \gamma_0 > 0$, $\alpha \geq 0$. t_n is an unobserved trend component. ε_n shows the fluctuation around the trend component and follows the ARCH process. Latent variable $s_n \in \{0, 1\}$ follows the first-order and two-state Markov process. Depending on this Markov chain s_n , the time series process shifts between two different states. When $s_n = 0$, the time series process is in a low volatility state with ARCH constant γ_0 . When $s_n = 1$, it is in a high volatility state with ARCH constant γ_1 . Further, the system model of MS-SC/ARCH model is given by

$$t_n = t_{n-1} + \Delta t_n + v_n, \quad v_n \sim N(0, \tau^2), \quad (5.27)$$

$$\Delta t_n = \Delta t_{n-1} + w_n, \quad w_n \sim N(0, v_1^2 s_n). \quad (5.28)$$

The unobserved trend component t_n follows a random walk process with a drift term Δt_n and Gaussian white noise v_n . Δt_n can be considered the slope component of the unobserved trend component t_n . In addition, depending on the Markov chain s_n , the process of this slope component shifts between two

different states. $s_n = 0$ shows the state where there is no slope change in the trend component, and follows a random walk process with a Gaussian white noise w_n . $s_n = 1$ shows the state where slope changes occur. In the MS-SC/ARCH model, the state shifts of volatility and trend slope are assumed to occur simultaneously according to the following common Markov chain s_n and its transition probability

$$p_{ij} = \Pr(s_n = j | s_{n-1} = i), \quad i, j = 0, 1, \quad (5.29)$$

where

$$\sum_{j=0}^1 p_{ij} = 1, \quad \text{for all } i = 0, 1, \quad (5.30)$$

This transition probability p_{ij} is constant over time. It can be written as the following transition probability matrix.

$$P = \begin{pmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{pmatrix} = \begin{pmatrix} p_{00} & 1 - p_{11} \\ 1 - p_{00} & p_{11} \end{pmatrix}. \quad (5.31)$$

The density of y_n , conditional on s_n taking on the value j , is

$$f(y_n | s_n = j, \Psi_{n-1}) = \frac{1}{\sqrt{2\pi h_{s_n}}} \exp \left\{ -\frac{(y_n - t_{n|n-1})^2}{2h_{s_n}} \right\}. \quad (5.32)$$

where Ψ_{n-1} refers to information up to time $t - 1$. To determine the log likelihood function, we first consider the joint density of y_n and s_n , which is the product of the conditional and marginal densities given by

$$f(y_n, s_n = j | \Psi_{n-1}) = f(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1}), \quad (5.33)$$

Then, to obtain the marginal density of y_n , we integrate s_n out of the above joint density by summing all possible values of s_n

$$\begin{aligned} f(y_n | \Psi_{n-1}) &= \sum_{j=1}^m f(y_n, s_n = j | \Psi_{n-1}) \\ &= \sum_{j=1}^m f(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1}) \\ &= \sum_{j=1}^m \frac{1}{\sqrt{2\pi h_{j,n}}} \exp \left\{ -\frac{(y_n - t_{n|n-1})^2}{2h_{j,n}} \right\} \Pr(s_n = j | \Psi_{n-1}). \end{aligned} \quad (5.34)$$

The log likelihood function is then given by

$$\begin{aligned}
 L(\theta) &= \sum_{n=1}^N \log f(y_n) \\
 &= \sum_{n=1}^N \log \left[\sum_{j=1}^m f(y_n, s_n = j, s_{n-1} = i) \right] \\
 &= \sum_{n=1}^N \log \left[\sum_{j=1}^m \sum_{i=1}^m \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}) \frac{1}{\sqrt{2\pi h_{j,n}}} \right. \\
 &\quad \left. \times \exp \left\{ -\frac{(y_n - t_{n|n-1})^2}{2h_{j,n}} \right\} \right]. \tag{5.35}
 \end{aligned}$$

where θ denotes the unknown parameters. The maximum likelihood estimation of θ is obtained by maximizing equation (5.35). Here, since the Markov chain s_n is not observed, we can obtain $\Pr(s_{n-1} | \Phi_{n-1})$ in equation (5.35) by using the same method as for equations (2.27) and (2.28) of the univariate Markov switching stochastic trend model.

5.4 Markov Switching Models with Exogenous Leverage Effects

We expect that other variables may contain relevant information for the volatility of a series. Such evidence has been found by Bollerslev and Melvin (1994), Engle and Mezrich (1996) and Engle *et al.* (1990a, b). In this subsection, we describe the Markov switching model with an exogenous leverage effect. We introduce the Markov switching heteroskedasticity model with an exogenous leverage effect (MSH-L) and the Markov switching ARCH model with an exogenous leverage effect (MSARCH-L). Let y_n , $n = 1, 2, \dots, N$, be the observed time series, such as, for example, stock, foreign exchange rate and bond price returns.

5.4.1 Markov switching heteroskedasticity model with an exogenous leverage effect

First, the MSH-L model is given by

$$y_n = \mu + \varepsilon_n, \quad \varepsilon_n \sim N(0, \sigma_{s_n}^2), \tag{5.36}$$

where

$$\text{if } s_n = 0, \quad \sigma_{s_n} = \gamma_0 + f(\lambda, z_{n-1}), \quad f(\lambda, z_{n-1}) > 0, \tag{5.37}$$

$$\text{if } s_n = 1, \quad \sigma_{s_n} = \gamma_1 + f(\lambda, z_{n-1}), \quad f(\lambda, z_{n-1}) > 0, \tag{5.38}$$

where $0 < \sigma_0 < \sigma_1$. Note that $f(\cdot)$ denotes an arbitrary function with the constant unknown parameter λ , and z_n is the exogenous variable that is observable at time n . Note that $f(\cdot)$ is assumed to take

non-negative values over time.

Here s_n is a latent variable indicating the volatility regime that takes a value of zero or one. When $s_n = 0$, y_n follows a normal distribution with mean μ and variance σ_0^2 . When $s_n = 1$, y_n follows a normal distribution with mean μ and variance σ_1^2 . Since $\sigma_0 < \sigma_1$, $s_n = 0$ indicates a low volatility regime and $s_n = 1$ indicates a high volatility regime. We assume that the latent variable s_n shifts between zero and one according to a first-order Markov process with constant transition probability as follows:

$$p_{ij} = \Pr(s_n = j | s_{n-1} = i), \quad i, j = 0, 1, \quad (5.39)$$

where

$$\sum_{j=0}^1 p_{ij} = 1, \quad \text{for all } i = 0, 1, \quad (5.40)$$

This transition probability p_{ij} is assumed to be homogeneous over time, and can be written in the form of the following transition probability matrix:

$$P = \begin{pmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{pmatrix} = \begin{pmatrix} p_{00} & 1 - p_{11} \\ 1 - p_{00} & p_{11} \end{pmatrix}. \quad (5.41)$$

The density of y_n , conditional on s_n taking on the value j , is

$$f(y_n | s_n = j, \Psi_{n-1}) = \frac{1}{\sqrt{2\pi\sigma_{j,n}^2}} \exp \left\{ -\frac{(y_n - \mu)^2}{2\sigma_{j,n}^2} \right\}. \quad (5.42)$$

where Ψ_{n-1} refers to information up to time $t-1$. To determine the log likelihood function, we first consider the joint density of y_n and s_n , which is the product of the conditional and marginal densities

$$f(y_n, s_n = j | \Psi_{n-1}) = f(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1}), \quad (5.43)$$

Then, to obtain the marginal density of y_n , we integrate s_n out of the above joint density by summing all possible values of s_n

$$\begin{aligned} f(y_n | \Psi_{n-1}) &= \sum_{j=0}^1 f(y_n, s_n = j | \Psi_{n-1}) \\ &= \sum_{j=0}^1 f(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1}) \\ &= \sum_{j=0}^1 \frac{1}{\sqrt{2\pi\sigma_{j,n}^2}} \exp \left\{ -\frac{(y_n - \mu)^2}{2\sigma_{j,n}^2} \right\} \Pr(s_n = j | \Psi_{n-1}). \end{aligned} \quad (5.44)$$

The log likelihood function is then given by

$$\begin{aligned}
 L(\theta) &= \sum_{n=1}^N \log f(y_n). \\
 &= \sum_{n=1}^N \log \left[\sum_{j=0}^1 f(y_n, s_n = j, s_{n-1} = i) \right] \\
 &= \sum_{n=1}^N \log \left[\sum_{j=0}^1 \sum_{i=0}^1 \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}) \right. \\
 &\quad \left. \times \frac{1}{\sqrt{2\pi\sigma_{j,n}^2}} \exp \left\{ -\frac{(y_n - \mu)^2}{2\sigma_{j,n}^2} \right\} \right]. \tag{5.45}
 \end{aligned}$$

where θ denotes the unknown parameters. The maximum likelihood estimation of θ is obtained by maximizing equation (5.45). Here, since the Markov chain s_n is not observed, we can obtain $\Pr(s_{n-1} | \Phi_{n-1})$ in equation (5.45) by using the same method as for equations (2.27) and (2.28) of the univariate Markov switching stochastic trend model.

5.4.2 Markov switching ARCH model with an exogenous leverage effect

Second, the MSARCH-L model is given by

$$y_n = \mu + \varepsilon_n, \quad \varepsilon_n | \Psi_{n-1} \sim N(0, h_{s_n, n}), \tag{5.46}$$

$$h_{s_n, n} = \gamma_{s_n} + g(\lambda, z_{n-1}) + \sum_{k=1}^K \alpha_{s_n}^k \varepsilon_{n-1}^2, \tag{5.47}$$

where $\gamma_{s_n} \geq 0$, $\alpha_{s_n}^k \geq 0$, $g(\lambda, z_{n-1}) > 0$, and

$$\text{if } s_n = 0, \quad \gamma_{s_n} = \gamma_0 \quad \text{and} \quad \alpha_{s_n}^k = \alpha_0^k, \quad k = 1, 2, \dots, K, \tag{5.48}$$

$$\text{if } s_n = 1, \quad \gamma_{s_n} = \gamma_1 \quad \text{and} \quad \alpha_{s_n}^k = \alpha_1^k, \tag{5.49}$$

where Ψ_{n-1} is the information up to time $n-1$, and $g(\cdot)$ denotes an arbitrary function with the constant unknown parameter λ . Note that $g(\cdot)$ is assumed to take non-negative values over time. γ_{s_n} is the ARCH constant term, and $\alpha_{s_n}^k$ is the k -th ARCH coefficient. Both γ_{s_n} and $\alpha_{s_n}^k$ shift between different parameter values depending on the unobserved latent variable s_n . Since $\gamma_0 > \gamma_1$ is assumed, $s_n = 0$ and $s_n = 1$ represent low and high volatility regimes, respectively.

$$h_{s_n, n} = \gamma_{s_n} + g(\lambda, z_{n-1}) + \sum_{k=1}^K \alpha_{s_n}^k \varepsilon_{n-1}^2, \quad g(\lambda, z_{n-1}) > 0, \tag{5.50}$$

The density of y_n , conditional on s_n taking on the value j , is

$$f(y_n | s_n = i, \Psi_{n-1}) = \frac{1}{\sqrt{2\pi h_{s_n, n}}} \exp \left\{ -\frac{(y_n - \mu)^2}{2h_{s_n, n}} \right\}. \tag{5.51}$$

where Ψ_{n-1} refers to information up to time $t - 1$. To determine the log likelihood function, we first consider the joint density of y_n and s_n , which is the product of the conditional and marginal densities

$$f(y_n, s_n = j | \Psi_{n-1}) = f(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1}), \quad (5.52)$$

Then, to obtain the marginal density of y_n , we integrate s_n out of the above joint density by summing all possible values of s_n

$$\begin{aligned} f(y_n | \Psi_{n-1}) &= \sum_{j=0}^1 f(y_n, s_n = j | \Psi_{n-1}) \\ &= \sum_{j=0}^1 f(y_n | s_n = j, \Psi_{n-1}) \Pr(s_n = j | \Psi_{n-1}) \\ &= \sum_{j=0}^1 \frac{1}{\sqrt{2\pi}h_{j,n}} \exp \left\{ -\frac{(y_n - \mu)^2}{2h_{j,n}} \right\} \Pr(s_n = j | \Psi_{n-1}). \end{aligned} \quad (5.53)$$

The log likelihood function is then given by

$$\begin{aligned} L(\theta) &= \sum_{n=1}^N \log f(y_n). \\ &= \sum_{n=1}^N \log \left[\sum_{j=0}^1 f(y_n, s_n = j, s_{n-1} = i) \right] \\ &= \sum_{n=1}^N \log \left[\sum_{j=0}^1 \sum_{i=0}^1 \Pr(s_n = j | s_{n-1} = i) \Pr(s_{n-1} = i | \Psi_{n-1}) \right. \\ &\quad \left. \times \frac{1}{\sqrt{2\pi}h_{j,n}} \exp \left\{ -\frac{(y_n - \mu)^2}{2h_{j,n}} \right\} \right]. \end{aligned} \quad (5.54)$$

where θ denotes the unknown parameters. The maximum likelihood estimation of θ is obtained by maximizing equation (5.54). Here, since the Markov chain s_n is not observed, we can obtain $\Pr(s_{n-1} | \Phi_{n-1})$ in equation (5.54) by using the same method as for equations (2.27) and (2.28) of the univariate Markov switching stochastic trend model.

5.5 Self-organizing Markov Switching State-Space Modeling

In this section, we introduce the self-organizing state-space formula (Kitagawa 1998) into the Markov switching model, and describe the self-organizing Markov switching model (Kitagawa and Hakamata 2001, Higuchi 2001, Hakamata and Kitagawa 2002). Let y_n be an observation variable. We first consider the univariate Markov switching stochastic trend model as follows

$$t_n = \mu + t_{n-1} + v_n, \quad v_n \sim N(0, \tau^2), \quad (5.55)$$

$$y_n = t_n + w_n, \quad w_n \sim N(0, \sigma^2), \quad (5.56)$$

where t_n is the trend component according to a first-order random walk process, and μ is a drift term

$$\mu = \mu_0(1 - s_n) + \mu_1 s_n. \quad (5.57)$$

Here, s_n denotes a two-state and first-order Markov chain taking a value of zero or one. The transition probability of s_n is given by

$$p_{ij} = \Pr(s_n = j | s_{n-1} = i), \quad i, j = 0, 1, \quad (5.58)$$

where $\sum_{j=0}^1 p_{ij} = 1$ for all $i = 0, 1$. Under the above assumptions, the joint conditional distribution of t_n and s_n given the previous values t_{n-1} and s_{n-1} is obtained by

$$p(t_n, s_n = j | t_{n-1}, s_{n-1} = i) = \Pr(s_n = j | s_{n-1} = i) p(t_n | t_{n-1}, s_n = j). \quad (5.59)$$

In the ordinary Markov switching modeling, the state vector x_n is given by

$$x_n = (t_n, s_n)^T, \quad (5.60)$$

and its general state-space form (Kitagawa and Gersch 1996) can be expressed as follows

$$x_n \sim Q(x_n | x_{n-1}), \quad (5.61)$$

$$y_n \sim R(y_n | x_n). \quad (5.62)$$

The unknown parameter θ is obtained by the maximum likelihood method via the Kalman filter algorithm described in section 4.

For simultaneous estimation of the state x_n and the unknown parameter θ , we consider Bayesian estimation by augmenting the state vector as follows:

$$x_n^* = (t_n, s_n, \theta_n)^T, \quad (5.63)$$

where t_n is the original state vector, s_n is the Markov chain and θ_n is the parameter vector. Here, we assume a random walk model

$$\theta_n = \theta_{n-1} + \varepsilon_n, \quad (5.64)$$

where ε_n is a Gaussian white noise process with zero mean and covariance $\text{diag}\{\epsilon_1^2, \epsilon_2^2, \dots, \epsilon_m^2\}$. All the structural parameters are assumed to be time-varying. The self-organizing Markov switching model for this augmented state vector x_n^* in the general state-space form is immediately given by

$$x_n^* \sim Q^*(x_n^* | x_{n-1}^*), \quad (5.65)$$

$$y_n \sim R^*(y_n | x_n^*). \quad (5.66)$$

where $Q^*(x_n^*|x_{n-1}^*)$ is the conditional density of the augmented state given the previous one, and $R^*(y_n|x_n^*)$ is that of the observation given the augmented state. By applying the Monte Carlo filter to this self-organizing state-space model, we can estimate the state vector and parameters simultaneously (Kitagawa 1998). Using this self-organizing Markov switching model there is no need to estimate the parameter θ with a maximum likelihood method.

5.6 Semi-Markov Switching Model

5.6.1 Univariate Semi-Markov Switching Stochastic Trend Model

Durland and McCurdy (1994) extended Hamilton's (1989) Markov switching model to allow the transition to be duration dependent. Hamilton's model specification can be said to be a particular parameterization of a semi-Markov process. Xianping and Padhraic (2000) and Kitagawa and Hakamata (2001) used a semi-Markov chain as the unobserved variable governing the process shifts. In this subsection, we consider the first-order stochastic trend model with semi-Markov switching structure. Let y_n be an observed time series at time n . The observation model of the semi-Markov switching stochastic trend model is given by

$$y_n = y_{n-1} + \varepsilon_n, \quad \varepsilon_n \sim N(\mu_{s_n}, \sigma_{s_n}^2), \quad (5.67)$$

where ε_n is the innovation term depending on the latent discrete semi-Markov chain $s_n \in \{1, 2, \dots, m\}$ as follows

$$\begin{aligned} \text{if } s_n = 1, & \quad \varepsilon_n \sim N(\mu_1, \sigma_1^2) \\ \text{if } s_n = 2, & \quad \varepsilon_n \sim N(\mu_2, \sigma_2^2) \\ & \quad \vdots \\ \text{if } s_n = M, & \quad \varepsilon_n \sim N(\mu_m, \sigma_m^2). \end{aligned}$$

The Markov switching model presented in the previous section can be extended to a semi-Markov switching model where the switching (transition) occurs according to a probability distribution P_d (Ross, 1996). Here d is the duration of the new regime. If d is the realization of the distribution P_d at time n , then we have $s_n = s_{n+1} = \dots = s_{n-d}$ and

$$\begin{aligned} p_{ij} &= \Pr(s_n = j | s_{n-1} = \dots = s_{n-d} = i) \\ &= \Pr(s_n = j | s_{n-1} = i, D_{n-1} = d), \end{aligned} \quad (5.68)$$

where D_n denotes the time that the semi-Markov chain has remained in the current regime since the previous switching. The transition of D_n is given by

$$D_n = \begin{cases} 0 & \text{with probability } \beta^{-1}P_{D_{n-1}+1} \\ D_{n-1} + 1 & \text{with probability } 1 - \beta^{-1}P_{D_{n-1}+1} \end{cases} \quad (5.69)$$

where

$$\beta = \sum_{j=D_{n-1}+1}^{\infty} P_j. \quad (5.71)$$

As a parametric model for the duration probability, we shall use the negative binomial distribution,

$$P_{d+e}(e, p) = \binom{d+e-1}{d} p^d (1-p)^e. \quad (5.72)$$

Note that $P_k(l, p) = 0$ for $k = 0, \dots, l$. In the ordinary Markov switching model, the probability of duration time k is given by $p_{ii}^k(1 - p_{ii})$ if $s_n = i$. Therefore, the probability of duration is a monotone decreasing function of time k . By contrast, the duration time of the semi-Markov process with the negative binomial distribution attains its maximum at $k = P_0^{-1}(l - 1)$. The Markov switching model can be further extended by using a different duration probability for a different state, $P_k^{(l)}$, $l=1, 2, \dots, m$. In this case, we need to assume that the self-loop transition probability $p_{ij} = 0$, $i = j$, in order to avoid redundancy of the parameters. This means that when the duration time has passed, the probability of switching occurring is one.

In particular, this parameterization of the conditional probabilities ensures that they lie in the interval $(0, 1)$, sum to 1, and, if $d = 0$ for all i , the $\Pr(s_n = j | s_{n-1} = i, D_{n-1} = d) = \Pr(s_n = j | s_{n-1} = i)$ and the process collapses to a first-order Markov process identical to that assumed by Hamilton (1989).

The conditional joint density of y_n , s_n and D_{n-1} is given by

$$f(y_n, s_n = j, D_{n-1} = d | \Psi_{n-1}) = f(y_n | s_n = j, D_{n-1} = d, \Psi_{n-1}) \Pr(s_n = j, D_{n-1} = d | \Psi_{n-1}). \quad (5.73)$$

Here, the density of y_n , conditional on s_n and D_{n-1} , is

$$f(y_n | s_n = j, D_{n-1} = d | \Psi_{n-1}) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left\{ -\frac{(y_n - \mu_j)^2}{2\sigma_j^2} \right\}, \quad (5.74)$$

for all $j = 1, 2, \dots, m$. Since the semi-Markov chain s_n is not observed, we can calculate it recursively as follows

$$\begin{aligned} \Pr(s_n = j, D_{n-1} = d | \Psi_{n-1}) &= \frac{f(y_n, s_n = j, D_{n-1} = d | \Psi_{n-1})}{f(y_n | \Psi_{n-1})} \\ &= \frac{f(y_n | s_n = j, D_{n-1} = d, \Psi_{n-1}) \Pr(s_n = j, D_{n-1} = d | \Psi_{n-1})}{f(y_n | \Psi_{n-1})}, \end{aligned} \quad (5.75)$$

where

$$\Pr(s_n = j, D_n = d | \Psi_{n-1}) = \sum_{i=1}^m \sum_{d=1}^{n-1} \Pr(s_n = j | s_{n-1} = i, D_{n-1} = d) \Pr(s_{n-1} = i, D_{n-1} = d | \Psi_{n-1}). \quad (5.76)$$

The conditional density of y_n is given by

$$\begin{aligned} f(y_n | \Psi_{n-1}) &= \sum_{i=1}^m \sum_{j=1}^m \sum_{d=1}^{n-1} f(y_n, s_n = j, s_{n-1} = i, D_{n-1} = d | \Psi_{n-1}), \\ &= \sum_{i=1}^m \sum_{j=1}^m \sum_{d=1}^{n-1} \Pr(s_n = j, s_{n-d} = i, D_{n-1} = d | \Psi_{n-1}) \\ &\quad \times \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left\{ -\frac{(y_n - \mu_j)^2}{2\sigma_j^2} \right\} \end{aligned} \quad (5.77)$$

from which the log-likelihood function is formed as

$$\begin{aligned} L(\theta) &= \sum_{n=1}^N \log f(y_n | \Psi_{n-1}) \\ &= \sum_{n=1}^N \log \left[\sum_{i=1}^m \sum_{j=1}^m \sum_{d=1}^{n-1} f(y_n, s_n = j, s_{n-1} = i, D_{n-1} = d | \Psi_{n-1}) \right] \\ &= \sum_{n=1}^N \log \left[\sum_{i=1}^m \sum_{j=1}^m \sum_{d=1}^{n-1} \Pr(s_n = j, s_{n-1} = i, D_{n-1} = d | \Psi_{n-1}) \right. \\ &\quad \left. \times \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left\{ -\frac{(y_n - \mu_j)^2}{2\sigma_j^2} \right\} \right], \end{aligned} \quad (5.78)$$

where θ denotes the unknown parameters.

5.6.2 Multivariate Semi-Markov Switching Slope Change Model

Let $y_{k,n}$ be the K -dimensional time series. We consider the multivariate semi-Markov switching slope change model (SMS-SC) for $y_{k,n}$ as follows

$$y_{k,n} = t_{k,n} + w_{k,n}, \quad w_{k,n} \sim N(0, \Sigma), \quad (5.79)$$

where $t_{k,n}$ is a K -dimensional unobserved trend component, and $w_{k,n}$ is a K -dimensional Gaussian observation noise with zero mean and variance-covariance matrix Σ . Note that the diagonal elements of Σ are $\{\sigma_1^2, \dots, \sigma_K^2\}$. The trend component in equation (5.79) assumes that

$$t_{k,n} = t_{k,n-1} + \Delta t_{k,n}. \quad (5.80)$$

The k -dimensional distribution of the slope $\Delta t_{k,n}$ depends on its previous value $\Delta t_{k,n-1}$ and a semi-Markov chain $s_n \in \{1, 2, 3, 4\}$ as follows

$$p(\Delta t_n | \Delta t_{n-1}, s_n = j) = \begin{cases} N(\Delta t_{n-1}, \Omega) & \text{if } j = 1 \\ \text{Unif}(-c, 0) & \text{if } j = 2 \\ \text{Unif}(c, 0) & \text{if } j = 3 \end{cases} \quad (5.81)$$

Here $N(v, \Omega)$ and $\text{Unif}(a, b)$ denote the Gaussian distribution with mean v and k -dimensional variance-covariance matrix $\Omega = \text{diag}\{\tau_1^2, \dots, \tau_K^2\}$ and the uniform distribution over (a, b) , respectively. The uniform distributions in equation (5.81) can be replaced by other distributions such as Gaussian distributions, $N(\mu_0^2, \xi_0^2)$ and $N(\mu_0^2, \xi_0^2)$. The transition probability of the semi-Markov chain s_n is given by

$$p_{ij} = \Pr(s_n = j | s_{n-1} = \dots = s_{n-d} = i) \quad (5.82)$$

$$= \Pr(s_n = j | s_{n-1} = i, D_{n-1} = d), \quad (5.83)$$

where

$$p_{23} = \Pr(s_n = 3 | s_{n-1} = 2) = 0 \quad \text{and} \quad p_{32} = \Pr(s_n = 2 | s_{n-1} = 3) = 0, \quad (5.84)$$

where D_n denotes the function of the duration time d , and the transition of D_n is given in equation (5.70). We assume that the regime shifts occur in all the trend components of $t_{k,n}$ for $k = 1, 2, \dots, K$ simultaneously.

In the multivariate SMS-SC model, the joint conditional distribution of $t_{k,n}, \Delta t_{k,n}$ and s_n given the previous values $t_{k,n-1}, \Delta t_{k,n-1}$ and s_{n-1} is obtained by

$$\begin{aligned} & f(t_{k,n}, \Delta t_{k,n}, s_n = j | t_{k,n-1}, \Delta t_{k,n-1}, s_{n-1} = i) \\ &= f(s_n = j | t_{k,n-1}, \Delta t_{k,n-1}, s_{n-1} = i) f(\Delta t_{k,n} | s_n = j, s_{n-1} = i, t_{k,n-1}, \Delta t_{k,n-1}) \\ & \quad \times f(t_{k,n} | s_n = j, s_{n-1} = i, t_{k,n-1}, \Delta t_{k,n-1}) \\ &= \Pr(s_n = j | s_{n-1} = i) f(t_{k,n} | \Delta t_{k,n}, t_{k,n-1}), \end{aligned} \quad (5.85)$$

where the two terms on the right-hand side are specified by equations (5.80) and (5.81), respectively.

Chapter 6

Conclusions

In this thesis, we focused on the Markov switching approach, and applied the ordinary and newly extended Markov switching models to the empirical analysis of economic and financial time series as follows:

- (1) Trend identification and trading strategy,
- (2) Time-series and cross-sectional volatility analysis,
- (3) Japanese business cycle analysis,
- (4) Japan premium and Japanese banks' stock volatility,
- (5) Transmission of volatility,
- (6) Foreign exchange volatility and intervention,
- (7) International business cycle transmission.

Our applications concentrated on the unique issues and scope of economic and financial markets. In these empirical analysis, we proposed five main extended Markov switching models with constant parameters: the univariate and multivariate modified Markov switching ARCH models, the Markov switching slope change and ARCH model, the Markov switching models with exogenous leverage effects, and the semi-Markov switching models. In addition, we introduced the self-organizing state space Markov switching model, which does not need to obtain the maximum likelihood estimate of parameters. Therefore, it is more useful in the high dimensional situation. To evaluate our extended Markov switching models, we compared them with the ordinary stochastic models and analytical methods.

Bibliography

- [1] Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle, in: B.N. Petroc and F. Caski, eds., *Second International Symposium in Information Theory*, (Akademiai Kiado, Budapest) 267-281.
- [2] Albert, J. and Chib, S. (1993) Bayes Inference via Gibbs Sampling of autoregressive time series subject to Markov mean and variance shifts, *Journal of Business and Economic Statistics* 11, 1-15.
- [3] Alspach, D.L. and Sorenson, H.W. (1972) Nonlinear Bayesian Estimation Using Gaussian Sum Approximations, *IEEE Transactions on Automatic Control*, AC-17-4, 439-448.
- [4] Andersen, T.G. and Bollerslev, T. (1998) Deutschmark-Dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies, *Journal of Finance* 53, 219-265.
- [5] Ball, C.A, Roma, A. (1993) A jump diffusion model for the European monetary system, *Journal of International Money Finance* 12, 475-492.
- [6] Ball, C.A, Torous, W.N. (1985) On jumps in common stock prices and their impact on call option pricing, *Journal of Finance* 40, 155-173.
- [7] Bates, D. (1996) Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options, *Review of Financial Studies* 9, 69-107.
- [8] Bates, D. (2000) Post-'87 Crash Fear in S&P 500 Futures Options, *Journal of Econometrics* 94, 181-238.
- [9] Birchenhall, C., Jessen, H., Osborn, D., Simpson, P. (1999) Predicting U.S. Business-Cycle Regimes, *Journal of Business and Economic Statistics* 17-3, 313-323.
- [10] Boldin, M. (1994) Dating Turning Points in the Business Cycle, *Journal of Business* 67-1, 97-131.

- [11] Boldin, M. (1996) A Check on the Robustness of Hamilton's Markov Switching Model Approach to the Economic Analysis of the Business Cycle, *Studies in Nonlinear Dynamics and Economics* 1-1, 35-46.
- [12] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307-327.
- [13] Bollerslev, T. (1990) Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach, *Review of Economics and Statistics* 72, 498-505.
- [14] Bollerslev, T., R.Y. Chou and K.F. Kroner (1992) ARCH modeling in finance: a review of the theory and empirical evidence, *Journal of Econometrics* 52, 5-59.
- [15] Bollerslev, T., R.F. Engle and D.B. Nelson (1994) ARCH models, in: Engle, R. F. and D. McFadden, eds, *Handbook of Econometrics* vol IV. Amsterdam: North-Holland, 2959-3038.
- [16] Bollerslev, T., R.F. Engle and J.M. Woldridge (1988) A capital asset pricing model with time varying covariance, *Journal of Political Economy* 96, 116-131.
- [17] Bollerslev, T. and M. Melvin (1994) Bid-ask spreads and the volatility in the foreign exchange market: an empirical analysis, *Journal of International Econometrics* 36, 355-372.
- [18] Bonser-Neal, C (1996) Does Central Bank intervention stabilize foreign exchange rate ?, *Economic Review First Quarter 1996 - Federal Reserve Bank of Kansas City*, 43-57.
- [19] Brier, G.W. (1950) Verification of forecasts expressed in terms of probability, *Monthly Weather Review* 75, 1-3.
- [20] Bry, G. and Boschan, C. (1971) Cyclical Analysis of Time Series: Selected Procedures and Computer Programs, *NBER Technical Paper* 20.
- [21] Burns, A.F. and Michell, W.C. (1946) Measuring Business Cycles, *National Bureau of Economic Research, New York*.
- [22] Cai, J. (1994) A Markov model of switching-regime ARCH, *Journal of Business and Economic Statistics* 12, 309-31.
- [23] Carlin, B.P., Polson, N.G. and Stoffer, D.S. (1992) A Monte Carlo approach to nonnormal and nonlinear state-space modeling, *Journal of the American Statistical Association* 87, 495-500.

- [24] Carter, C.K. and Kohn, R. (1994) On Gibbs sampling for state space models, *Biometrika* 81, 541-553.
- [25] Dacco, R., Satchell, S. (1999) Why do Regime-switching Models Forecast so Badly?, *Journal of Forecasting*, 19, 1-16.
- [26] Doucet, A., de Freitas, N. and Gordon, N. (2001) Sequential Monte Carlo Methods in Practice, *Springer-Verlag, New York*.
- [27] Dueker, M. (1997) Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility, *Journal of Business and Economic Statistics* 15-1, 26-34.
- [28] Durland, T., McCurdy, T. (1994) Duration-Dependent Transitions in a Markov Model of U.S. GNP Growth, *Journal of Business and Economic Statistics* 12-3, 279-288.
- [29] Efroymson, M.A. (1960) Multiple regression analysis, in: A. Ralston and H.S. Wilf, eds, *Mathematical Methods for Digital Computers Vol.1. Wiley & Sons, New York*, 191-203.
- [30] Efroymson, M.A. (1966) Stepwise Regression — a Backward and Forward Look. *Presented at the Easter Regional Meetings of the Inst. of Meth. Statist. Florham Park, New Jersey*.
- [31] Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50, 987-1007.
- [32] Engle, R.F., T. Ito and W.L. Lin (1990) Meteor showers or heat waves? Heteroscedastic intra-daily volatility in the foreign exchange market, *Econometrica* 58, 525-542.
- [33] Engle, R.F., D.M. Lilien and R.P. Robins (1987) Estimating time varying risk premia in the term structure: The ARCH-M model, *Econometrica* 55, 391-407.
- [34] Engle, R.F. and J. Mezrich (1996) GARCH for groups, *Risk* 9, 36-40.
- [35] Engle, R.F., V.K. Ng and M. Rothschild (1990) Asset pricing with a factor-ARCH covariance structure, *Journal of Econometrics* 45, 213-238.
- [36] Engel, C. and J.D. Hamilton (1990) Long swings in the dollar: Are they in the data and do markets know it? *The American Economic Review* 80, 689-713.
- [37] Engel, C. and C.J. Kim (1999) The Long-Run U.S./U.K. Real Exchange Rate, *Journal of Money, Credit, and Banking* 31-3, 335-356.

- [38] Fedenia, M., Hodder, J.E. and Triantis, A.J (1994) Cross-Holdings: Estimation Issues, Biases, and Distortions, *Review of Financial Studies* 7-1, 61-96.
- [39] Filardo, A.J. (1994) Business-cycle phases and their transitional dynamics, *Journal of Business and Economic Statistics* 12, 299-308.
- [40] Filardo, A.J. and Gordon, S. (1998), Business Cycle Durations, *Journal of Econometrics* 85-1, 99-123.
- [41] Garcia, R. and Perron, P. (1996) An Analysis of the Real Interest Rate under Regime Shifts *The Review of Economics and Statistics* 78, 111-125.
- [42] Fong, W. (1997), Volatility Persistence and Switching ARCH in Japanese Stock Returns, *Financial Engineering and the Japanese Markets* 4, 37-57.
- [43] Franses, P., Dijk, D. (2000), Nonlinear Time Series Models, *Cambridge University Press*.
- [44] Fruhwirth-Schnatter (2001), Fully Bayesian analysis of switching Gaussian state space models, *Annals of the Institute of Statistical Mathematics* 53-1, 31-49.
- [45] Garcia, R. (1998) Asymptotic null distribution of the likelihood ratio test in Markov switching models, *International Economic Review* 39-3, 763-788.
- [46] Goldfred, S., Quandt, R. (1973) A Markov model for switching regression, *Journal of Econometrics* 1, 3-16.
- [47] Glosten, L.R., R. Jagannathan and D.E. Runkle (1993) On the relation between the expected value and the volatility of the nominal excess returns on stocks, *Journal of Finance* 48, 1779-1801.
- [48] Gordon, N.J., Salmond, D.J. and Smith, A.F.M. (1993) A Novel Approach to Nonlinear/non-Gaussian Bayesian State Estimation *IEE Proceedings-F* 140-2, 107-113.
- [49] Hakamata, M. (2002) International Transmission of Business Cycles: a Markov-switching Self-organizing State-space Model, *Proceedings in 6th Joint Conference on Information Sciences* 1106-1110.
- [50] Hakamata, M., Kitagawa, G. (2002) Self-organizing Markov Switching State Space Model with Time-varying Transition Probability, *ISM Research Memorandum*
- [51] Hamilton, J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle, *Journal of Econometrics* 57-2, 357-384.

- [52] Hamilton, J.D. (1990) Analysis of Time Series Subject to Changes in Regime, *Journal of Econometrics* 45, 39-70.
- [53] Hamilton, J.D. (1994) Time Series Analysis, *Princeton University Press, Princeton, N.J.*
- [54] Hamilton, J.D. (1991) A quasi-Bayesian Approach to Estimating Parameters for Mixtures of Normal Distributions, *Journal of Business and Economic Statistics* 9-1, 27-39.
- [55] Hamilton, J.D. (1996) Specification testing in Markov-switching time-series models, *Journal of Econometrics* 70, 127-157.
- [56] Hamilton, J.D. and G. Lin (1996) Stock market volatility and the business cycle, *Journal of Applied Econometrics* 11, 573-593.
- [57] Hamilton, J. D. and Susmel, R. (1994) Autoregressive conditional heteroscedasticity and changes in regime, *Journal of Econometrics* 64, 307-333.
- [58] Hall, S., Psaradakis, Z., Sola, M. (1999) Detecting periodically collapsing bubbles: a Markov-switching unit root test, *Journal of Applied Econometrics* 14, 143-154.
- [59] Hansen, B. (1992) The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GDP, *Journal of Applied Econometrics* 7, 61-82.
- [60] Hansen, B. (1996) Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* 64-2, 413-430.
- [61] Harrison, P.J. and Stevens, C.F. (1976) Bayesian Forecasting, *Journal of the Royal Statistical Society, Series B* 38, 205-247.
- [62] Harvey, A., Ruiz, E. and Sentana, E. (1992) Unobserved component time series models with ARCH disturbances, *Journal of Econometrics* 52, 129-157.
- [63] Harvey, A.C. and Todd, P.H.J. (1983) Forecasting economic time series with structural and Box-Jenkins models: A case study, *Journal of Business and Economic Statistics* 1, 299-315.
- [64] Higuchi, T. (2001) Evolutionary Time Series Model with Parallel Computing, *Proceedings of The Third Japan-US Joint Seminar on Statistical Time Series Analysis* 183-190.
- [65] Iino, M., Ozaki, T. (1999) Filtering and Prediction of Non-Gaussian Time Series via Jump Diffusion Process (in Japanese), *Proceedings of the Institute of Statistical Mathematics* 47, 327-342.

- [66] Ito, T. and K. Harada (2000) Japan premium and stock prices: two mirrors of Japanese banking crises. *NBER working paper series*.
- [67] Jorion, P. (1998) On jump processes in the foreign exchange and stock markets. *Review of Financial Studies* 1, 427-445.
- [68] Kaminsky, G. (1993) Is there a peso problem? Evidence from the dollar/pound exchange rate, 1976-1987, *The American Economic Review* 83, 456-472.
- [69] Kaufmann, S. (2000) Measuring business cycles with a dynamic Markov switching factor model: an assessment using Bayesian simulation methods, *Econometrics Journal* 3, 39-65.
- [70] Kim, C.J. (1993) Unobserved-component time series models with Markov-switching heteroscedasticity: Changes in regime and the link between inflation uncertainty, *Journal of Business and Economic Statistics* 11, 341-349.
- [71] Kim, C.J. (1994) Dynamic linear models with Markov-switching, *Journal of Econometrics* 60, 1-22.
- [72] Kim, C.J. and Nelson, C.R. (1999) State-Space Models with Regime Switching, *MIT Press, Cambridge, Mass.*
- [73] Kim, M.J. and Yoo, J.S. (1995) New index of coincident indicators: A multivariate Markov switching factor model approach, *Journal of Monetary Economics* 36, 607-630.
- [74] Kitagawa, G. (1987) Non-Gaussian state-space modeling of nonstationary time series, *Journal of the American Statistical Association* 82, 1032-1063.
- [75] Kitagawa, G. (1994) The Two-filter Formula for Smoothing and an Implementation of the Gaussian-sum Smoother *Annals of the Institute of Statistical Mathematics* 46-4, 605-623.
- [76] Kitagawa, G. (1994) *Markov switching autoregressive model* (in Japanese), *Proceedings of the Institute of Statistical Mathematics* 42, 75-82.
- [77] Kitagawa, G. (1996) Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Model, *Journal of Computational and Graphical Statistics* 5, 1-25.
- [78] Kitagawa, G. (1998) Self-Organizing State Space Model, *Journal of the American Statistical Association* 93-443, 1203-1215.
- [79] Kitagawa G. and Gersch W. (1996) Smoothness Priors Analysis of Time Series, *Springer-Verlag, N.Y.*

- [80] Kitagawa, G. and Hakamata, M. (2001) Semi-Markov switching slope model, *ISM Research Memorandum*
- [81] Komaki, F. (1993) State-space modeling of time series sampled from continuous process with pulses., *Biometrika* 80, 417-429.
- [82] Kontolemis, Z.G. (2001) Analysis of the US business cycle with a Vector-Markov-switching model., *Journal of Forecasting* 20, 47-61.
- [83] Kontorovich, V. and Lyandres, V. (1996) Impulsive noise: A nontraditional approach., *Signal Processing* 51-2, 121-132.
- [84] Krolzig, M. (1997) Markov-Switching Vector Autoregressions: Modeling, Statistical Inference, and Application to Business Cycle Analysis, *Springer*.
- [85] Kwan, J., Lam, L., So, M. (2000) Forecasting and Trading Strategies Based on a Price Trend Model, *Journal of Forecasting* 19, 485-498.
- [86] Lahiri, K. Wang, J. (1994) Predicting Cyclical Turning Points with Leading Index in a Markov Switching Model, *Journal of Forecasting* 13, 245-263.
- [87] Lam, P. (1990) The Hamilton model with a general autoregressive component, *Journal of Monetary Economics* 26, 409-432.
- [88] Leroux, B.G. and M.L. Puterman (1992) Maximum-penalized-likelihood estimation for independent and Markov dependent mixture models, *Biometrics* 48, 545-558.
- [89] Lindgren, G. (1978) Markov Regime Models for Mixed Distributions and Switching Regressions, *Scandinavia Journal of Statistics* 5, 81-91.
- [90] Luginbuhl, R. and Vos, A.D. (1999) Bayesian Analysis of an Unobserved-Component Time Series Model of GDP with Markov-Switching and Time-Varying Growths, *Journal of Business & Economic Statistics* 17, 456-465
- [91] Ljung, L. (1979) Asymptotic Behavior of the Extended Kalman Filter as a Parameter Estimator for Linear Systems, *IEEE Transactions on Automatic Control* AC-24-1, 36-50.
- [92] MacDonald, J.A. and H. Shawky (1995) On estimating stock market volatility: An exploratory approach, *Journal of Financial Research* 18-4, 449-463.

- [93] McCulloch, R. Tsy, R. (1994) Statistical analysis of economic time series via Markov switching models, *Journal of Time Series Analysis* 15-5, 523-539.
- [94] Maheu, J., McCurdy, T. (2000) Volatility dynamics under duration-dependent mixing, *Journal of Empirical Finance* 7, 345-372.
- [95] Murphy, J.J. (1986) Technical Analysis of the Futures Markets, New York Institute of Finance, N.Y.
- [96] Nelson, D.B. (1991) Conditional heteroscedasticity in asset return: a new approach, *Econometrica* 59, 347-370.
- [97] Peek, J. and E.S. Rosengren (2000) Determinants of the Japan premium: Actions speak louder than words, *Journal of International Economics* 53, 283-305.
- [98] Phillips, K.L. (1991) A two-country model of stochastic output with changes in regime, *Journal of International Economics* 31, 121-142.
- [99] Ravn, M., Sola, M. (1995) Stylized facts and regime changes: Are prices procyclical?, *Journal of Monetary Economics* 36, 497-526.
- [100] Ryden, T. (1995) Estimating the order of hidden Markov models, *Statistics* 26, 345-354.
- [101] Sakamoto, Y., M. Ishiguro and G. Kitagawa (1986) Akaike Information Criterion Statistics, D-Reidel Pub., Dordrecht.
- [102] Shumway, R., Stoffer, D. (1991) Dynamic Linear Models With Switching, *Journal of the American Statistical Association* 86-415, 763-769.
- [103] Snyder, D.L. and Miller, M.I. (1991) Random Point Processes in Time and Space, Springer, New York, Chapter 5, 215-286.
- [104] So, M., Lam, K., Li, W.K. (1998) A Stochastic Volatility Model With Markov Switching, *Journal of Business and Economic Statistics* 16-2, 244-253.
- [105] Sorenson, H.W. and Alspach, D.L. (1971) Recursive Bayesian estimation using Gaussian sums, *Automatica* 7, 465-479.
- [106] Schnatter, S. (2001) Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models, *Journal of Business and Economic Statistics* 96, 194-209.

- [107] Susmel, R. (2000) Switching Volatility in Private International Equity Markets, *International Journal of Financial and Economics* 5, 265-283.
- [108] Taylor, M. P. (1992) The use of technical analysis in the foreign exchange market, *Journal of International Money and Finance* 11, 304-314.
- [109] Tyssedal, J., Tjostheim, D. (1988) An Autoregressive Model with Suddenly Changing Parameters and an Application to Stock Market Prices, *Applied Statistics* 37-3, 353-369.
- [110] Xianping Ge and Padhraic Smyth (2000) Deformable Markov Model Templates for Time-Series Pattern Matching, *KDD-2000: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining* August 20-23, 2000.