Chapter 3
Research methodology and data

3.1. Introduction

This chapter summarises research methods, describes the source of data, and defines key variables. In particular, the Markov switching model (hereafter called the MS model) and its estimation method are discussed. Figure 3.1 shows research methodology; the data is first tested for unit root behaviour by using the augmented Dickey-Fuller test (hereafter called the ADF test), which is proposed by Dickey and Fuller (1979), in order to determine its stationarity. This study also selects the numbers of regimes and variables simultaneously for the MS model based on Schwarz (1978) information criterion (hereafter called SIC). Subsequently, if there are competing models, the best model will also be chosen based on SIC value.

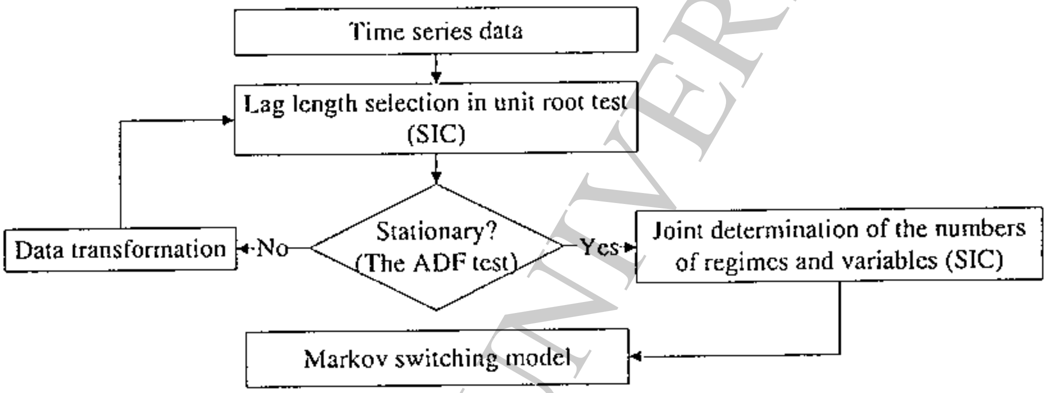

Figure 3.1: Research methodology

This diagram shows the related research methodology. It includes three methods, which are unit root test (the ADF test), model selection (Schwarz (1978) information criterion: SIC), and the Markov switching model, respectively.

3.2. Unit root test

Testing economic data for stationarity is nowadays a widespread exercise. To do this, a unit root test has to be employed to determine whether each variable has unit root behaviour (or non-stationarity). There are several tests such as the ADF test, the Phillips-Perron (1988) test (hereafter called the PP test), and the KPSS test (Kwiatkowski et al. 1992).

The ADF and PP tests are asymptotically equivalent but may differ substantially to finite samples due to a different way in which they correct for serial correlation problems. A Monte Carlo simulation of DeJong et al. (1992) suggests that in practice, the ADF test performs better than the PP test. For the autoregressive model, the ADF test also performs better (Mills 1999). This study, therefore, will employ the ADF test to determine if there is any evidence of unit root behaviour in the data. The null hypothesis of non-stationarity: H_0: $\alpha = 0$ is tested as:

$$\Delta y_t = \alpha y_{t-1} + x_t'\delta + \sum_{l=1}^{p} \beta_l \Delta y_{t-l} + u_t \, , \tag{3.1}$$

where y_t is the observed variable at time t; for $t = 1, 2, \ldots, T$,
T is the total number of observations,
x_t contains either an intercept term or an intercept term and time trend,
α, β, and δ are parameters to be estimated, and
u_t is the disturbance at time t and assumed to be white noise.

3.3. *Model selection*

An essential issue for estimating the MS model is how to determine the numbers of regimes and variables simultaneously. The simulation results of Psaradakis and Spagnolo (2006) suggest that Akaike (1974) information criterion (hereafter called AIC)[17] is the most successful criterion in terms of the frequency of correct model identification, followed by SIC and HQC (Hannan & Quinn 1979). On the other hand, Smith, Naik and Tsai (2006) derive a new information criterion called "Markov switching criterion" (hereafter called MSC). Their simulation results also show that MSC works as well as standard criteria. Recently, the simulation results of Awirothananon and Cheung (2009) show that SIC performs better than other criteria (including AIC, HQC, and MSC) in joint determination of the numbers of states and variables. As a result, this study will use SIC value to choose a suitable model and to compare with other competing MS models. SIC may be defined as:

$$\text{SIC} = -2L + k \log(T), \tag{3.2}$$

where L, T, and k is the maximised log likelihood value, the total number of observations, and the number of estimated parameters, including the intercept term, respectively.

SIC also imposes penalty for including an increasingly large number of estimated parameters (Enders 2004). For comparing two or more models, the model with the lowest SIC value is preferred. The advantage of using SIC is that it is useful for both in-sample and out-of-sample forecasting performance (Gujarati 2003). It does not explicitly suffer from nuisance parameter issues. It is, therefore, useful for both nested and non-nested models.

3.4. *Markov switching model*[18]

The result from SIC (from previous section) suggests a suitable MS model, which fits the data best. The MS model has been proposed as an alternative to a constant parameter, linear time series model of the earlier Box and Jenkins (1976) modelling, because the MS model allows for changes in regime of the process generating time series. The idea behinds this class of regime switching model is that parameters of D-dimensional vector time series process (y_t) depend upon an unobservable regime variable (s_t), which

---

[17] AIC generally selects large models in non-linear frameworks (Fenton & Gallant 1996). In addition, HQC usually suggests the same number of lags as SIC does.

[18] This section relies heavily on Krolzig (1997).

represents the probability of being in a particular state of economy. This could be specified as:

$$y_t = v(s_t) + \sum_{l=0}^{p} A_l(s_t)x_{t-l} + u_t \, , \tag{3.3}$$

where x_t is defined as exogenous variables conditional upon s_t; $s_t \in \{1, 2, \ldots, M\}$,
$v(s_t)$ and $A_l(s_t)$ are defined as estimated parameters conditional upon s_t, and
u_t is assumed to be a Gaussian innovation process conditional upon s_t: $u_t \sim$ NID(0, $\Sigma(s_t)$).

The sign of one coefficient in $A_l(s_t)$ in Equation (3.3) implies the direction of relationship between changes in money market interest rate (or the inter-bank rate: IB) and the commercial bank stock index return (hereafter called Index). For example, if the sign of one coefficient of $A_l(s_t)$ is negative, it suggests that when the IB rises, the Index tends to decline, suggesting a negative relationship. However, if the sign of one coefficient of $A_l(s_t)$ in Equation (3.3) is positive, it implies that the Index and changes in IB would move in the same direction, implies that a positive relationship.

For this model to be completed, a crucial assumption about the regime generating process is that it is a discrete state homogenous Markov chain, which is defined by the transition probabilities:[19]

$$Pr_{ij} = Pr(s_{t+1} = j \mid s_t = i) \text{ and } \sum_{i=1}^{M} \Pr_{ij} = 1; \ \forall i, j \in \{1, 2, \ldots, M\},$$

where $Prij$ is the probability that event i is followed by event j.

The probability of regime i occurring next period given that the current regime is j is fixed. These probabilities could also be represented in the transition matrix for an irreducible ergodic M state Markov process st:

$$\Pr = \begin{bmatrix} \Pr_{11} & \Pr_{12} & \cdots & \Pr_{1M} \\ \Pr_{21} & \Pr_{22} & \cdots & \Pr_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \Pr_{M1} & \Pr_{M2} & \cdots & \Pr_{MM} \end{bmatrix},$$

where $Pr_{im} = 1 - Pr_{i1} - \ldots - Pr_{i,M-1}$; for $i = 1, 2, \ldots, M$.

By inferring the probabilities of unobserved regimes conditional on available information set, it is then possible to reconstruct the regimes. In principle, all parameters of the conditional model can be made dependent on the state (s_t) of the Markov chain.

The MS model allows for a greatly wider choice of specification regarding parameters in the model. In principle, it would be possible to: (i) make all parameters regime-dependent and (ii) introduce separate regimes for each shifting parameter. In practice, only some parameters would be conditioned on the state of Markov chain, while the other parameters will be regime invariant. Enders (2004) also states that there are three important features of the MS model. First, since the transition probabilities are unknown, they need to be estimated along with coefficients. Second, the overall degree of

[19] Hence, the evolution of regimes could be inferred from the data.

persistence depends on parameters as well as transition probabilities. Third, the probabilities are all conditional probabilities. For example, if the system is in regime two the conditional probability exists whereby the system switches into regime one. Since the MS model allows for a variety of specifications, it is important to assess that which specification might be suitable for the data best.

The MS model would be referred to as the MSI model if only intercept term, $\nu(s_t)$, is regime-varying, the MSIA model if parameters, $A_i(s_t)$, also change with regime, and the MSIAH model if additionally variances, $\Sigma(s_t)$, are regime-dependent.[20] The MS model could be estimated by using a maximum likelihood procedure (hereafter called ML). As mentioned in Ehrmann, Ellison and Valla's (2003) study, "since the Markov chain is hidden, the likelihood function has a recursive nature: optimal inference in the current period depends on the optimal inference made in the previous period. Under such conditions the likelihood could not be maximised using standard techniques." The ML algorithm of this model is based on a version of the Expectation Maximisation (hereafter called EM) algorithm discussed in Hamilton (1990) and Krolzig (1997).

The EM algorithm is originally described by Dempster, Laird and Rubin (1977) as a general approach to iteratively compute the ML estimation technique. This technique is designed for general models where the observed variables are dependent on some unobserved variables, such as regimes variables (s_t). Ehrmann, Ellison and Valla (2003) state that the first expectations step optimally infers the hidden Markov-chain for given parameters. The maximisation step then re-estimates the parameters for the inferred hidden Markov-chain. This approach also continues until the likelihood converges to a local maximum of the likelihood function (Zhai 2004).

3.5. *Principle component analysis*[21]

Principle component analysis (hereafter called PCA) is statistical techniques applied to a set of variables. Variables that are correlated with another but largely independent of other subsets of variables are combined into factors. Factors could reveal underlying processes that have created the correlation among variables. PCA attempts to identify factors that explain the pattern of correlations within a set of variables. PCA is also used for data reduction to identify a small number of factors that explain most of variance in a larger number of variables. The correlation matrix can be calculated as:

$$R = V \Lambda V', \tag{3.4}$$

where Λ is the eigenvalue matrix c x c,

V is the eigenvector matrix r x c,

R is the correlation matrix r x r, and

r and c are the number of variables and factors.

[20] The MSIAH model is also able to capture the autoregressive conditional heteroskedasticity effect (Krolzig 1997).

[21] This section relies heavily on Tabachnick and Fidell (2007).

The column in V is called eigenvector and the values in the main diagonal of Λ are called eigenvalues. The first eigenvector corresponds to the first eigenvalues, and so on. Rearranging Equation (3.4) by taking the square root of the matrix of eigenvalue (Λ):

$$R = AA', \tag{3.5}$$

where R is the correlation matrix,

$A = V\sqrt{\Lambda}$ and $A' = \sqrt{\Lambda}V'$.

From Equation (3.5), the correlation matrix can be considered a product of two matrices, each combination of eignvectors and the square root of eigenvalues. It also shows the calculation of eigenvalues and eigenvectors. After the calculation has done, the factor-loading matrix (A) is found by straightforward matrix multiplication as:

$$A = V\sqrt{\Lambda}, \tag{3.6}$$

where A is factor-loading matrix r x c.

The value of eigenvector (V) from Equation (3.6) will be used to construct a global factor (hereafter called the GF) for this study as follows:

$$GF = \text{Index} \times V,$$

where GF is the global factor matrix t x c,

Index is the three developing countries' Indices matrix t x 3,

V is the eigenvector matrix 3 x c, and

c and t are the number of factors and observations.

3.6. Key variables and data

This study focuses on the money market interest rate and the commercial bank stock index return in Thailand. It further compares the results of Thailand with other two developing countries, namely Indonesia and Malaysia in the Southeast Asia region.[22] The first variable is the Index from the stock market of each country.[23] The second variable is the IB of each country.[24] All variables are collected from Thomson DataStream; all codes are provided in Table 3.1. Notice that all three countries are open economy; any shock from the international environment could influence their stock market. This study, therefore, constructs a new variable, which could be called the GF. The PCA is used to extract this factor, which measures the inter-correlation among three developing countries' Indices in the Southeast Asia region, namely Indonesia, Malaysia, and Thailand. This study will select a factor with an eigenvalue greater than one, since Tabachnick and Fidell (2007) suggest that a factor with an eigenvalue less than one is not important as it cannot explain a significant portion of data variability. The PCA result shows that there is only one factor with an eigenvalue greater than one. This factor,

[22] The coefficient on interest rate of each country will be compared. This method is also used by some prior studies such as Gultekin (1983), Chen and Chan (1989), Khil and Lee (2000), and Choudhry (2001).

[23] Indonesia: the Jakarta Stock Exchange; Malaysia: the Bursa Malaysia; Thailand: the Stock Exchange of Thailand.

[24] The money market interest rate for Thailand and Malaysia is the overnight inter-bank rate. This rate is also employed by some previous studies such as Hahm (2004) and Aggarwal, Jeon and Zhao (2005). In the case of Indonesia, this study uses the inter-bank rate. This rate is also used by some previous studies such as Faff and Howard (1999) and Faff, Hodgson and Kremmer (2005).

hence, could be interpreted as a response of all three developing countries' Indices to shocks from international environment.

Table 3.1: Code of all key variables

This table summarises the Thomson DataStream code of all key variables from three developing countries in the Southeast Asia region. These countries are Indonesia, Malaysia, and Thailand.

Countries	Commercial bank stock indices	Inter-bank rate
Indonesia	BANKSID	IDIBCAL
Malaysia	BANKSMY	MYIBKCL
Thailand	BANKSTH	THBTIBN

Recall from Chapter 2 (in Section 2.3) that there are three different monetary policy frameworks in Thailand, namely the pegged exchange rate policy, the monetary aggregates targeting policy, and the inflation targeting policy, respectively. To assess the effectiveness of the Thai monetary policy under different monetary policies, this study employs the IB and the Index from September 1993 to March 2010.[25] The same technique is also applied to other three developing countries in the Southeast Asia: Indonesia and Malaysia. This study further compares the results of these countries with Thailand. The total number of observations is 199 per variable. In addition, monthly data is used because it can mitigate the potential impact of infrequent trading effects on the statistical inference, as suggested by Harvey (1995). This data also can capture long-term movements in volatility and avoid some impacts of settlement and clearing delays that significantly affect stock return over shorter periods.

3.7. *Chapter summary*

This chapter discussed in detail the different procedures used in this study (as presented in Figure 3.1), which involve unit roots tests to determine the stationary behaviour and SIC to determine the suitable model of the MS model. It also discussed the key variables for three developing countries in the Southeast Asia region, namely Indonesia, Malaysia, and Thailand, that will be used in this study (as summarised in Table 3.1).

[25] The starting period is from September 1993 since all data for these three countries are available from this date.