國 立 交 通 大 學

電 機 與 控 制 工 程 研 究 所

碩 士 論 文

應用貝氏定理與鑑別分析建立金融危機預警系統

Building Financial Crisis Early Warning Systems Based on Bayesian Theorem and Discriminant Analysis

研 究 生 : 李建威

指導教授 : 周志成 博士

中華民國 九十五 年 九 月

應用貝氏定理與鑑別分析建立
金融危機預警系統

Building Financial Crisis Early Warning Systems Based on Bayesian Theorem and Discriminant Analysis

研 究 生 ：李建威　　Student：Chien-Wei Lee

指導教授：周志成　　Advisor：Dr. Chi-Cheng Jou

國 立 交 通 大 學

電 機 與 控 制 工 程 研 究 所

碩 士 論 文

A Thesis
Submitted to Department of Electrical and Control Engineering
College of Electrical Engineering and Computer Science
National Chiao-Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master
in
Electrical and Control Engineering
September 2006
Hsinchu, Taiwan, Republic of China

中華民國 九十五 年 九 月

應用貝氏定理與鑑別分析建立
金融危機預警系統

學　生：李建威　　　　　　　　　　　指導教授：周志成　博士

國立交通大學電機與控制工程研究所

中 文 摘 要

本論文的目標在於開發二個預警系統用以預測金融危機的發生。從決策者的觀點, 預警系統之所以舉足輕重在於它能夠偵測出潛藏的經濟疲弱, 進而使決策者能夠在金融危機發生之前提早擬出對策, 進而減輕金融危機所帶來的衝擊, 甚至能夠制止此金融危機的發生。就預測能力而言, 欲建立一個健全系統, 需倚賴於鑑定出的領先指標是否可靠。

根據以往的經驗可以察覺, 一個金融體系在金融危機發生之前, 必有經濟疲弱的發生。因此, 我們首先利用貝氏理論偵測金融變數是否發生體系變遷, 並利用 F 統計量和斯皮爾曼相關分析去評估經由貝氏分析所得到的事後機率值和危機前導變數在時間序列的相似度, 如此便可篩選出關鍵變數。爾後我們繼續利用因素分析以及鑑別分析, 分別從這些關鍵變數之中鑑定出兩組不同的領先指標。

最後, 我們再利用這兩組領先指標建立二個預警系統, 較諸於之前已存在的預警系統模型, 期能達成預測表現之增進與改善。

Building Financial Crisis Early Warning Systems Based on Bayesian Theorem and Discriminant Analysis

Student : Chien-Wei Lee Advisor : Dr. Chi-Cheng Jou

Department of Electrical and Control Engineering
National Chiao Tung University

Abstract

The objective of this thesis is to develop two new Early Warning Systems (EWS) for predicting financial crises. From the policy-makers' perspectives, an EWS plays an important role by allowing them to detect potential economic weaknesses and take pre-emptive strategies to lessen the impact of a crisis or even to prevent it. In terms of predictive power, building a robust system relies on identifying reliable leading indicators.

According to an observation of the past experience, before a financial crisis takes place in an economic system, there must be economic weaknesses. Therefore, we first used Bayesian Theorem to detect regime shifts of financial variables, and then used F statistics and Spearman correlation to estimate the degree of similarity between the obtained posterior probability and a forward crisis variable in a time series in order to extract key variables. We proceeded to identified two sets of leading indicators among the key variables with factor analysis and discriminant analysis.

Finally we employed the leading indicators to build two EWS, which achieve the expectancy of improving the performance upon existing models.

Acknowledgement

I would like to thank all the people who have made writing this thesis a more pleasant task. In particular, I'd like to thank my principal advisor, Dr. Chi-Cheng Jou, who introduced me to this topic and gave me broad support and guidance throughout my time at Chiao-Tung. In addition, I'd like to thank Professor Yon-Ping Chen and Professor Kou-Yang Young for their valuable comments and advice given while serving on my reading committee.

There are many people whom I would like to thank. First, I'd like to thank Grandfather inside my heart, Father, Mother, Grandmother and Sister in my family whose unconditional support made this thesis possible. Second, I'd like to thank my girl friend Cheng-Wei, my intimates Huei-Cheng, Shih-Ting, Syue-Fen, Ming-Jin, Yi-Fei, Ya-Li, Pei-Wun, Pei-Ru, Miho-Adachi and my dear friends Jia-Sian, Jyun-Fu, Yu-Ren, who had been always standing for me. Without them, my life will not be so colorful and lovely.

Last, but certainly not the last, I'd like to thank Ta-Wen, Cheng-Shih, Sing-Rong, Ren-Hong and many colleagues who encouraged and supported me on studying. For them, I can make writing this thesis with no fear of disturbance in the rear. At that moment my heart is full of joy and thankfulness. I wish all of us being always good and having a great future.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Problem Statement and Motivations

In the recent years, we saw a large number of financial crises in emerging market economies (EMEs). These crises often occurred with devastating economics, external debts, social and political consequences. Furthermore, instead of being confined to individual economies or a single country, these financial crises were contagious in spreading across markets through trade linkages in many cases. Take crises of the last decade for example, the Mexican currency crisis of 1994-95, the Asian financial crisis of 1997-98, the Russian debt crisis of 1998, and the following Argentine external debt crisis affected a wide group of countries and the international financial system as well [1].

According to an observation of the past experience, before a financial crisis takes place in an economic system, there must be economic weaknesses and vulnerabilities. Owing to a financial crisis can cause great damage to social security, macro-economics, development of industries, trade of stock market, and social cost, international organizations and many academic institutions have developed Early Warning System (EWS) models with the aim of anticipating *whether* and *when* individual countries may suffer from a financial crisis. From the policy-makers' perspectives, EWS models play an important role to them by allowing them to detect potential economic weaknesses and vulnerabilities, and taking both pre-emptive and appropriate strategies to lessen the impact of a crisis or to reduce the probability of experiencing a crisis, or even to prevent it.

Up to now, some influential approaches have been proposed to develop EWS models by Kaminsky, Lizondo and Reinhart in 1998 [2] and by Berg and Pattillo in 1999 [3] since the International Monetary Fund (IMF) took a lead in making research into developing EWS models for EMEs. But also many central banks and various private institutions have put significant efforts to develop models [1]. However, the central concern is that, these models only can predict about seventy percentage of crises as a whole so far.

The objective of this thesis is to develop a new EWS model that can identify economic weaknesses and vulnerabilities to anticipate a financial crisis and can improve the performance upon existing models.

1.2 Features of the Problem

To our knowledge, there must be economic weaknesses and vulnerabilities before a financial crisis occur. By detecting the related critical financial variables, we could predict reliably whether a crisis will occur or not in a country instead of when a crisis will happen. The occurrence of a crisis is possible to be prevented as long as we send early warning signals to the policy-makers. In other words, what we are trying to predict is the occurrence, not the timing of a crisis. Therefore, the aim of our EWS is to predict whether a crisis will occur within a specific time horizon.

A sound and reliable EWS model is dependent on the extraction of leading indicators. Because the source data is likely to be a huge two-dimensional matrix (column of the matrix as financial variable, row of the matrix as sample), it will take a great load no matter which we proceed to analyze or make operations with too many variables. So it is necessary to make applicable extraction from the input variables. That is to say, choosing the important input variables as key variables which have significant influence on the goodness-of-fit of an EWS model of is essential. The five goodness-of-fit criteria are percentage of observations correctly estimated, percentage of crises correctly estimated, percentage of false alarms of total alarms, probability of a crisis given an alarm, and probability of a crisis given no alarm. With the variable extraction, we can not only improve the anticipating performance and enhance the efficiency of operations and analyses, but also find out which variables are more important for the consequence.

We note that the input variables are usually highly-correlated with each other. It means that there exist some redundancies, the co-varying or highly coupling input variables. If an EWS model is developed with the redundancies within our input variables, it may fail to extract the correct leading indicators. Seriously it will cause unacceptable errors in predicting the financial crisis. Because the occurrence of a crisis refers to many factors, and a factor might reflect on different input variables which are co-varying with each other. In order to develop a robust system, we should give consideration to all the factors in connection with the occurrence of a crisis. However, if we incorporate all of those co-varying variables into a leading indicator set, the same factor will be considered more than once. It means the factor is emphasized and weighted. Then it will increase systematic load, and seriously, it may cause the system to miss other important factors. Thus, it is important to remove redundancy from the key variables, so as to identify the appropriate leading indicators.

Recently, IMF had proposed the Macro Prudential Indicators (MPIs) to develop an EWS model to estimate whether a financial system is stable or not. However, the range of MPIs is so extensive that they cannot express the economic characteristics of all countries in the world; they can only provide a rough framework to explain and anticipate crises. Actually, each country is supposed to find out the leading indicators of its own according to its particular situation.

1.3 Related Literatures

Previous early warning systems of financial crises have used methods that fall into three broad categories. The first approach extracts early signals directly from a range of leading indicators, the second approach builds logit models by a set of independent variables, and the third uses regime shift detection methods to determine whether there is abnormality or not.

The leading indicator approach considers the key variables which reflect the economic weaknesses, and they are independent variables. If an indicator crosses a critical threshold, it will send a signal or flash a red light corresponding to the particular indicator to warn the policy-maker that a crisis might occur. The first problem of this approach is that the lower the threshold, the more signals this indicator will send. By contrast, raising the threshold level will reduce the number of wrong signals, but at the cost of increasing the number of missing crisis signals. Second, these various indicators do not provide a synthetic framework to predict the economic vulnerability of a given country. That is to say, if only indicators A and B cross the critical threshold with another situation where indicators C and D are in the critical zone, it will be difficult to rank [4].

The logit and probit model approach provides a synthetic picture in view of all the variables. In the logit and probit models the probability of a crisis is a non-linear function. We create two variables that indicate a crisis occur or not and stays within the 0-1 bounds to indicate the probability respectively. The aim of the model is to estimate the effect of the variables on the probability of experiencing a crisis. For this approach, an issue is that the logit and probit models do not extract leading indicators from the variables and this may cause system errors. Because there are redundancies in the variables, they will reduce the anticipating performance [5].

In the regime shift detection approach, there are various methods to detect the regime shifts by mean, variance, frequency structure, and system. For the methods of detecting shifts in the mean, student t -test method is the most commonly used techniques for testing a hypothesis on the basis of a difference between sample means. It determines a probability that two groups are the same with respect to the variable tested and it can be applied sequentially for each data point. The position of the greatest t value exceeding the given threshold indicates the location of abnormality [6]. The student t -test method is robust to the assumption of normality and equality of variances, but has a problem in testing for change occurring at a specified time [7]. For the methods of detecting shifts in the variance, the Downton-Katz test method uses a non-parametric bootstrap technique to compute confidence intervals to decide a change-point. However, it requires a reference time series with no change-points and the change-points of the sample data have to be separated at least ten years apart [8]. As for the methods of detecting shifts in the system, the principal component analysis (PCA) method is widely used to identify coherent patterns of variability among large sets of time series because it is designed to solve large-sized problems. The feature of PCA is that it can reduce the dimensionality of the data matrix and it requires additional time series analysis method to assess the character of temporal changes [9].

1.4 Concepts and Techniques of the Proposed Method

After considering the advantages and drawbacks of the aforementioned approaches, we find that the extraction of leading indicators is a necessary step to improve anticipating performance. The logit and probit model is a type of wrapper models which implement a machine learning approach. It selects the features systematically used by a classifier with an aim to achieve the highest predictive accuracy as the optimal features and allows an explicit modeling of the input variables. As for the regime shift detection method, the loss of information is questionable. That is, in the regime shift detection method, it focuses on the change of values whereas values of input variables are meaningless. Even so, the regime shift detection is still a popular and reliable method to detect the occurrence of crises because our purpose is to predict a crisis will occur or not alternatively. According to the above-mentioned

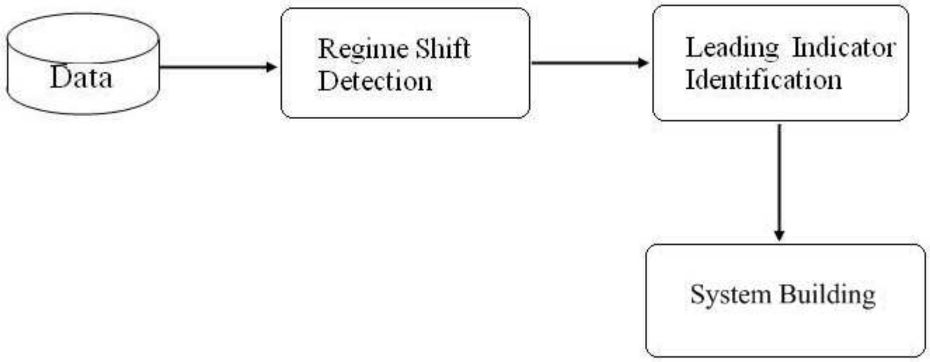

Figure 1.1: The flow chart of proposed systems

consideration, we develop new EWS that focuses on the improvement of the prediction performance and the systematic stability. The proposed systems (Figure 1.1) include three procedure: regime shift detection, leading indicator identification, and system building.

In the procedure of regime shift detection, we build a mathematical model with Bayesian Theorem to detect the training data. The results of regime shift detection have significant influence on the anticipating performance of an EWS model. The method is based on the approach that detects shifts in the mean. In this mathematical model, we take the training data as input variables and transform them into series of probability in the range of 0 to 1 as output variables to estimate a crisis may occur in the future or not. If the time point with probability close to 1 is earlier than a crisis that really occurs for a variable, we will regard the variable as a key variable. With the Bayesian analysis method, it will efficiently extract key variables that reflect the coming of a crisis from a large number of input variables and reduce the number of wrong alarms. Generally speaking, there are tuning parameters for each model. The difference of tuning parameters can decide the complexity of a model and result in different anticipating errors. Therefore, we compare the simulation data to decide the values of those tuning parameters in order to optimize the model.

In the procedure of leading indicator identification, we have to identify the leading indicators from the key variables that are extracted from those input variables. First we have to set up a binary forward crisis variable Y_{t-j}, which is defined as

$$Y_{t-j} = \begin{cases} 1 & \text{if } \exists j = 0, 1, 2, 3, 4 \text{ such that } CC_t = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1.1}$$

The variable CC_t indicates whether a crisis really occurs or not at time t and our model attempts to predict whether a crisis will occur within the subsequent 4 months by Y_{t-j}. Details can be found in Appendix A. The reason why choose the length of this period is to consider it can represent a sign of a coming crisis more accurately than others and reduce the number of false alarms. Second, we use F statistics and Spearman correlation to estimate the relationship between input variables and the forward crisis variable. In this way, we can extract key variables from all of the input variables and remove insignificant variables. Finally, we propose two methods to extract leading indicators. The first method is factor analysis and the other is discriminant analysis.

In the procedure of system building, we design two frameworks to build systems for each country. The two frameworks are basically based on the theories: weighted Bayesian analysis and discriminant analysis. Discriminant analysis is a category of wrapper models, and weighted Bayesian analysis is a category of filter models. Because of the characteristic of a wrapper model, the system built with discriminant analysis is more robust. On the other hand, the system built with weighted Bayesian analysis has the advantage of showing better performance in terms of prediction, but the performance is likely to drop faster as the time goes by.

1.5 Outline of the Thesis

This thesis proceeds as follows. Chapter 2 starts by describing the identification of regime shifts and the methods of regime shift detection. Chapter 3 introduces four methods of leading indicator identification. Chapter 4 shows the idea that how to combine regime shift detection with the methods discussed in chapter 3 to build early warning systems. Chapter 5 presents the simulation results of anticipating fianacial crises with our data samples, and chapter 6 concludes.

Chapter 2

Regime Shift Detection

Regime shifts are defined as rapid reorganizations of a system from one relatively stable state to another. A regime shift occurs when a statistically significant difference exists between the mean values of the variable before and after a certain point. As our knowledge, there must be economic weaknesses and vulnerabilities before a crisis occurs. Therefore, the disturbances and changes will occur throughout the normal operation and the corresponding regime shifts will be also found in many financial variables. There are a number of methods designed for a detection of regime shifts in both the individual time series and entire systems [11]. Bayesian Analysis is a method we propose to address this problem.

2.1 Bayesian Analysis

Bayesian analysis is applied to detect change points in a time series. The approach based on Bayes theorem provides uncertainty estimates of change points and means for shift detection. Specifically, the Bayesian analysis approach is formulated to demonstrate the posterior probability of the shifts at time t. Let X_t be the time series under investigation and y_t be the regime shift indicator. We denote $y_t = 1$ if shift occurs at t, and $y_t = 0$ otherwise. The posterior probability can be computed by Bayesian Theorem as

$$P(y_t = 1|\mathcal{X}) = \frac{P(\mathcal{X}|y_t = 1)P(y_t = 1)}{P(\mathcal{X}|y_t = 1)P(y_t = 1) + P(\mathcal{X}|y_t = 0)(1 - P(y_t = 1))} \quad (2.1)$$

where $\mathcal{X} = \{X_k,\, k = 1, 2, \ldots, N\}$ in the observed time series. We use a prior probability for the parameter $P(y_t = 1)$ in the inference procedure to reflect our information about shifts before we see the data, and a sampling model $P(\mathcal{X}|y_t = 1)$ for our data about the likelihood of observing the data in which shifts occur. We then compute the posterior probability $P(y_t = 1|\mathcal{X})$ which represents our updated information about shifts after we see the data. Thus the prior probability is modified by the data to obtain the posterior probability.

In the following we would like to specify the posterior probability calculation. The size of pre-change window W_1 and post-change window W_2 have important effect on the predicting performance and on the delay of the regime shift detection because there will be delay of $W_2 - 1$ in detection of the shifts. Figure 2.1 shows an example of regime shift with the pre-change and the post-change windows. We begin by using Gaussian distribution to approximate the true probability density for normal data in the expression

$$P_1(X_k) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{X_k - m_1}{\sigma}\right)^2\right\} \qquad k < t \quad (2.2)$$

$$P_2(X_k) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{X_k - m_2}{\sigma}\right)^2\right\} \qquad k \geq t \quad (2.3)$$

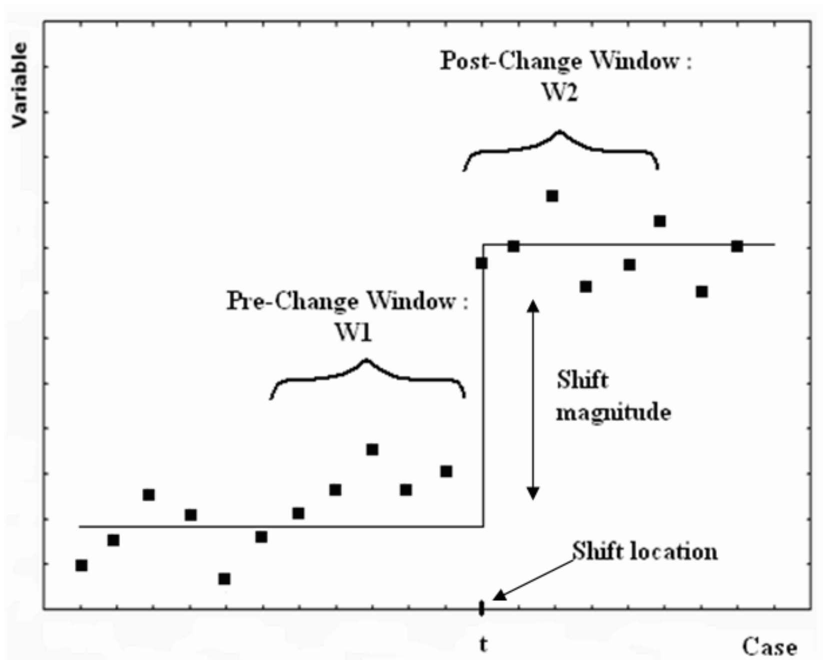

Figure 2.1: Using pre- and post-change window with regime shift detection as an example

where $P_1(X_k)$ is the Gaussian distribution before time t, $P_2(X_k)$ is the Gaussian distribution after time t, m_1 is the mean of the pre-change window and m_2 is the mean of the post-change window. The estimation of the level m_1 and m_2 can be manipulated to make the EWS model more sensitive to regime shifts. σ is the deviation and its value can't be too small because it will cause $P_1(X_k)$ and $P_2(X_k)$ to tend towards zero. It can lead to the system inaccuracy as a whole. Then it proceeds to give the likelihood function:

$$P(\mathcal{X}|y_t = 1) = \prod_{k=t-1}^{t-W_1} P_1(X_k) \prod_{k=t}^{t+W_2-1} P_2(X_k) \tag{2.4}$$

$$P(\mathcal{X}|y_t = 0) = \prod_{k=t-1}^{t-W_1} P_1(X_k) \prod_{k=t}^{t+W_2-1} P_1(X_k) \tag{2.5}$$

The strategy of the proposed algorithm of Bayesian analysis is to estimate the magnitude of the disturbance using measurement before and after the present month. The posterior probability $P_step(t)$ that a regime shift did occur will be updated with measurement for each month t, so that we can consider the pattern of the shift probability in the posterior probability sequence. Substituting (2-4) and (2-5) into (2-1) yields

$$\begin{aligned} P_{step}(t) &= P(y_t = 1|\mathcal{X}) \\ &= \frac{\prod_{k=t}^{t+W_2-1} P_2(X_k)P(y_t = 1)}{\prod_{k=t}^{t+W_2-1} P_2(X_k)P_0 + \prod_{k=t}^{t+W_2-1} P_1(X_k)(1 - P(y_t = 1))}. \end{aligned} \tag{2.6}$$

In order to enhance the model performance and solve the problem of system inaccuracy caused by the extremely small deviation, we can rewrite the formula. Substituting (2-2) and (2-3) into (2-6) yields

$$P_{step}(t) = \frac{1}{1 + \frac{(1 - P(y_t = 1))}{P(y_t = 1)} K(t)} \tag{2.7}$$

where

$$K(t) = \prod_{k=t}^{t+W_2-1} \exp\left\{ -\frac{1}{2}\left(\frac{X_k - m_1(t)}{\sigma_t}\right)^2 + \frac{1}{2}\left(\frac{X_k - m_2(t)}{\sigma_t}\right)^2 \right\}. \tag{2.8}$$

To facilitate computation, let $K(t) = e^{Q(t)}$, hence

$$Q(t) = \log K(t) = \sum_{k=t}^{t+W_2-1} \left[\frac{(X_k - m_1(t))^2 - (X_k - m_2(t))^2}{-2\sigma_t^2} \right]. \tag{2.9}$$

With the algorithm we can overcome the defects against the original model to obtain the posterior probability $P_{step}(t)$ via $Q(t)$. As for $m_1(t)$ and $m_2(t)$ are the means of the pre-change window and the post-change window of a specific time t, and σ_t is the corresponding deviation which can be captured in various methods [12].

The Bayesian analysis differs from the classical method for inference in its use of a prior probability and a posterior probability to express the uncertainty before and after seeing the data respectively. The numerical probability is regarded as representing a degree of reasonable belief rather than a frequency, and it can be associated with degrees of confidence that we experience a crisis attack. As long as new information is available, the probability will be always updated according to the present state of information.

2.2 Deviation Estimation

2.2.1 Sample Standard Deviation

In the method of regime shift detection, we had proposed an algorithm based the Bayesian theorem to compute a series of posterior probability. The formula for our algorithm has a very important parameter σ_t which has significant influence on the performance about anticipating crises. The sample standard deviation (SSD) is a commonly used measure of variation. It is also called simple standard deviation (STD) or estimated standard deviation (ESD). For each single deviation σ_t, it has different sampling population according to time t and pre-change window size W_1. Then the sample standard deviation is computed as:

$$\sigma_t = \sqrt{\frac{1}{W_1 - 1} \sum_{k=t-W_1}^{t-1} (X_k - m_1(t))^2} \tag{2.10}$$

2.2.2 Half-Mean-Square Successive Deviation

Half-mean-square successive deviation (HMSSD) is another measure that we use to compute the deviation and it is given by

$$\sigma_t = \sqrt{\frac{1}{2W_1} \sum_{k=t-W_1+1}^{t} (X_k - X_{k-1})^2} \tag{2.11}$$

2.2.3 Recursive Deviation

When there is a shift in the data set, we want our model to be able to adapt to the trend and continuously learn the new information. The purpose of doing this is to achieve adaptive modeling. The method is to update our model when a new data sample arrives. The new data is added into the modeling data set and a new model is built. The ways of updating a model differ in terms of the nature of a data set. In order to learn the trend quickly and remove the serious inertia related to a large data set, we should assign different weighting to samples to forget the old information gradually. Here we use the recursive deviation (RD) as a measure of variation to achieve model updating recursively. Let μ be the forgetting factor, we can write the recursive mean b_t as:

$$
\begin{aligned}
&b_t = \mu b_{t-1} + (1-\mu)X_t, \quad t \geq 2 \\
&b_1 = X_1
\end{aligned}
\tag{2.12}
$$

Then the recursive deviation is expressed in this form:

$$
\begin{aligned}
&\sigma_t = \sqrt{\mu(\sigma_{t-1}^2 + \Delta b_t^2) + (1-\mu)(X_t - b_t)^2}, \quad t \geq 2 \\
&\sigma_1 = 0
\end{aligned}
\tag{2.13}
$$

When implementing the method, we can decide how frequently to update the model and how fast to forget the old information by tuning the parameter μ. Besides, because of the recursive nature, the update calculation needs much less computation and is suitable for online model updating [13].

2.3 Parameter Optimization

Generally speaking, there are tuning parameters for each model. In this thesis, the EWS model which use Bayesian analysis to detect regime shifts has four tuning parameters. They are pre-change window (W_1), post-change window (W_2), forgetting factor (μ), and prior probability $P(y_t = 1)$. The difference of tuning parameters can decide the complexity of a model and result in different anticipating errors. Therefore, we wish that we can search out the optimized parameters in order to minimize the anticipating errors.

The window W_1 can be manipulated to modulate the sensitivity to shifts. When W_1 increases, the model becomes more sensitive. But it may fail to detect a shift occurring in early periods. As for W_2, it causes delay of W_2-1 in shift detection because we have to use the data after the change point. So the value of W_2 cannot be too large. The value of forgetting factor μ represents the degree we give thought to the old information. With a small value of μ, a model updates more frequently and it means the old information is forgotten faster. Except of that, the increase of false alarms is also a serious problem. The prior probability $P(y_t = 1)$ is decided in reference to the experience, which means the frequency of shifts occurred in the past.

The simplest and most widely used method for parameter optimization is K-fold cross validation. However, it relies on enough data samples and that's what we are unable to attain. What we do is comparing the simulation data to optimize those parameters, and then improve the anticipating performance.

Chapter 3

Leading Indicators Identification

After introducing the regime shift detection, we would like to describe several methods we use to identify leading indicators which are highly-related to the occurrence of crises. This chapter proceeds as follows. Section 3.1 starts by introducing a kind of statistics for testing population variance, the F statistics. Section 3.2 discusses the Spearman correlation as a method to express a relationship between two variables. Section 3.3 then discusses the factor analysis as a way of uncovering relationships among variables as well, and section 3.4 describes the concept of discriminant analysis.

3.1 F Statistics

An F statistic is a statistic for a test concerning the differences among means. It is the ratio of two estimates of population variance based on the information in two or more samples. When an F statistic is employed in the procedure entitled ANOVA which is an acronym for *analysis of variance*, the obtained value of F provides a test for the statistical significance of the differences among the means of two or more samples from a given population.

When we want to determine whether an input variable is a key variable among a lot of input variables, we can estimate the degree of how this input variable affect the response variable by means of the F statistics [14]. Supposing an input variable with N data samples is given by $\mathbf{x} = [x_1\ x_2\ \cdots\ x_N]^T$, and the response variable is given by $\mathbf{y} = [y_1\ y_2\ \cdots\ y_N]^T$, where x_i and y_i are the i th samples of $\mathbf{x}$ and $\mathbf{y}$ respectively. In the first place, the response variable $\mathbf{y}$ is manipulated to divide samples into k classes as Figure 3.1, and each class is labeled as C_j, $j = 1, 2, \ldots, k$.

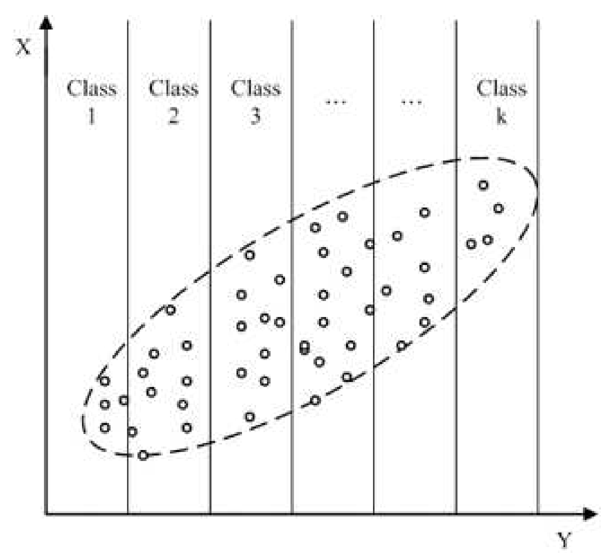

Figure 3.1: Samples are divided into k classes according to the values of $\mathbf{y}$

Then the F statistic is formulated as follows:

$$F = \frac{S_B/df_B}{S_W/df_W} \tag{3.1}$$

where S_B is the between-class variance, S_W is the within-class variance, $df_B = k-1$ is the number of degrees of freedom associated with S_B, and $df_W = N-k$ is the number of degrees of freedom associated with S_W. Among them, the between-class variance and the within-class variance are defined by:

$$S_B = \sum_{j=1}^{k} N_j (m_j - m)^2 \tag{3.2}$$

$$S_W = \sum_{j=1}^{k} \sum_{x_n \in C_j} (x_n - m_j)^2 \tag{3.3}$$

where N_j represents the number of samples for class C_j, m_j represents the sample mean of class C_j, and m is the mean of total samples:

$$m_j = \frac{1}{N_j} \sum_{x_n \in C_j} x_n \tag{3.4}$$

$$m = \frac{1}{N} \sum_{n=1}^{N} x_n = \frac{1}{N} \sum_{j=1}^{k} N_j m_j \tag{3.5}$$

From (3.1) and (3.2), we realize that the larger the F value is, the larger the differences in class means. It means that the input variable will affect the response variable more and more significantly while the F value is getting larger. It implies that the response variable will vary with the input variable. Thus, we can calculate the F values of all the input variables to obtain the order of importance of them. Besides, we can also use the F values to calculate the significance level (p-level):

$$
\begin{aligned}
p(F) = &\Gamma\left[(df_B + df_W)/2\right] / \left[\Gamma(df_B/2)\Gamma(df_W/2)\right] * (df_B/df_W)^{df_B/2} \\
&* F^{[(df_B/2)-1]} * \left[1 + (df_B/df_W)\, F\right]^{[-(df_B+df_W)/2]}
\end{aligned}
\tag{3.6}
$$

where Γ is the Gamma function. From (3.6), it is seen that the p-level decreases while the F value increases. The p-level is the significance level for the F value, and it represents a degree of importance for a variable. It is also a measure of confidence degree of a result. For example, if the p-level for an input variable is given by 0.02, the confidence degree to believe that variable is a key variable is equal to 98%. The lower the p-level for a variable, the more we can believe that it is a key variable. In many areas of research, the p-level of 0.05 is customarily regarded as a borderline acceptable error level, and variables which are significant at the $p \leq 0.01$ level are commonly considered highly significant. In the procedure of leading indicators identification, using the F statistics is the first step to make extraction. When there are a large number of input variables, we can use the F statistics and the p-level to remove those with p-level greater than 0.05 and obtain the importance sequence, and then proceed to the follow-up analysis procedure with the selected key variables.

3.2 Spearman Correlation

Correlation coefficient is usually computed to express a relationship between two variables. Pearson correlation calculations are based on the assumption that two variables x and y are sampled from populations that follow a normal (Gaussian) distribution, at least approximately. However, sometimes the variables are not normally distributed, such as the posterior probability

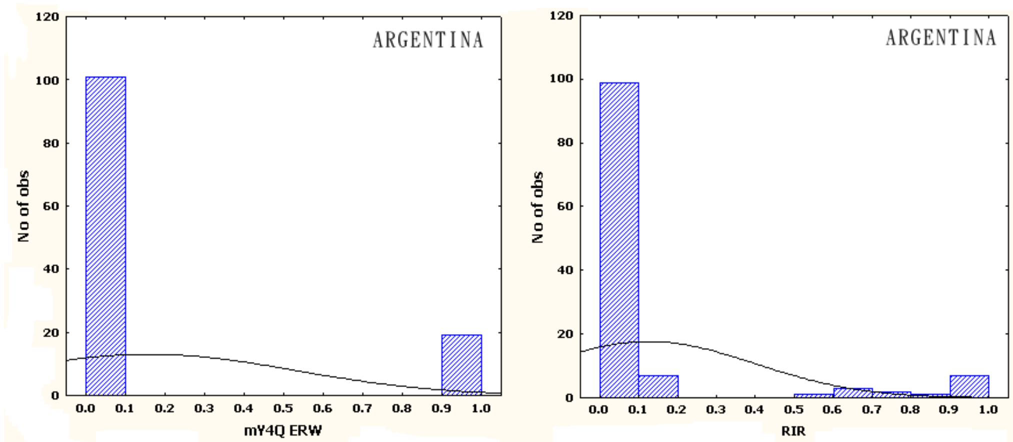

Figure 3.2: Histograms of a forward crisis variable mY4QERW and the posterior probability of RIR

$P_{step}(t)$ of (2.6) and the forward crisis variable Y_{t-j} of (1.1). Figure 3.2 shows their histograms as examples. The left plot shows a forward crisis variable called mY4QERW and the right plot shows the posterior probability of a variable called RIR. Calculating a Pearson correlation coefficient on data that are non-Gaussian in distribution may lead to false conclusions. It is then appropriate to use nonparametric Spearman correlation.

The Spearman correlation calculation is based on ranking two variables from highest to lowest rather than the original values. It makes no assumption about the distribution of the values. In addition, Spearman correlation can detect both linear and non-linear correlations while Pearson correlation is only appropriate for finding linear correlations. Spearman correlation assumes that variables under consideration are measured on an ordinal scale. The ordinal scale of measurement represents the ranks of a variable's values. Values measured on an ordinal scale contain information about their relationship to other values only in terms of whether they are greater than or less than other values but not in terms of how much greater or how much smaller. That is, the individual cases can be ranked into two ordered series. Let $\mathbf{x} = [x_1\ x_2\ \cdots\ x_N]^T$ and $\mathbf{y} = [y_1\ y_2\ \cdots\ y_N]^T$ are two variables with N data samples, where x_i and y_i are the i th samples of $\mathbf{x}$ and $\mathbf{y}$ respectively. They are ranked separately and the differences in rank are calculated at each position i. Then the correlation coefficient R_s between $\mathbf{x}$ and $\mathbf{y}$ is computed with the following formula:

$$R_s = 1 - \frac{6 \sum_{i=1}^{N} \left(\text{rank}\left(x_i\right) - \text{rank}\left(y_i\right)\right)^2}{N\left(N^2 - 1\right)} \tag{3.7}$$

The Spearman correlation coefficient ranges from -1 to +1 and it quantifies the direction and magnitude of correlation. The sign tells us whether the correlation is positive or negative where negative values indicate an inverse relationship and positive values indicate a direct relationship. But when we decide whether a result is significant or not, the sign can be ignored. Modulus of a correlation coefficient reflects the strength of the relationship between two variables [15]. Table 3.1 provides an overview about the degrees of strength in correlation with various regions of R_s. Besides, we can use a

Table 3.1: Regions vs degrees

Value of R_s	Degree of Correlation
$R_s = 1$	Perfect
$0.75 \le R_s < 1$	Strong
$0.5 \le R_s < 0.75$	Moderate
$0.25 \le R_s < 0.5$	Fair
$0 < R_s < 0.25$	Little
$R_s = 0$	No proof of correlation

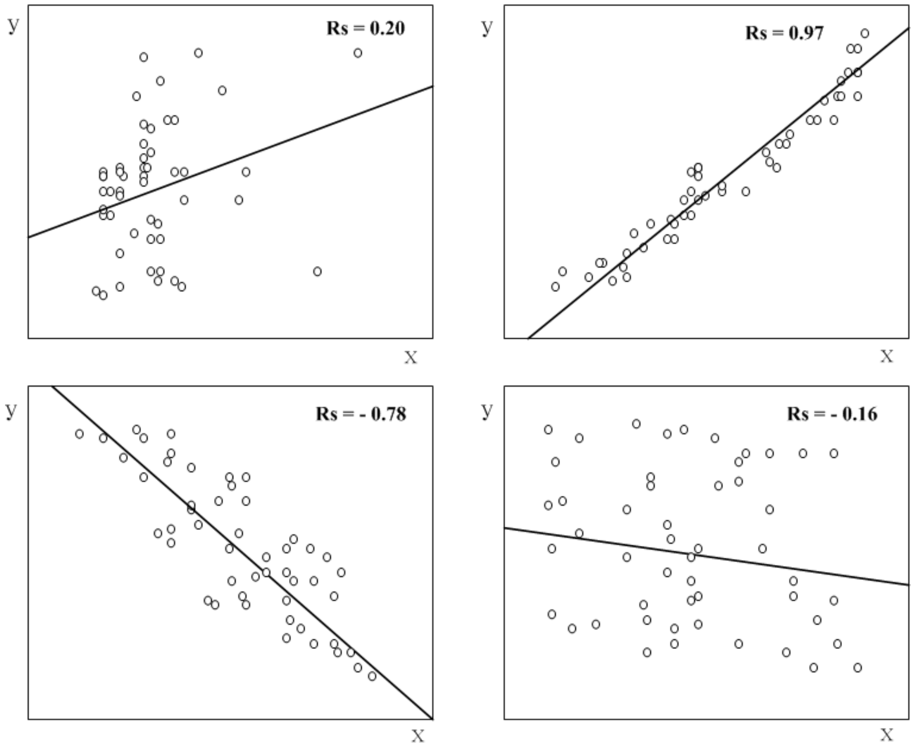

Figure 3.3: Scatterplots with different correlation coefficients

two-dimensional scatterplot to visualize a correlation between two variables **x** and **y**. A two-dimensional scatterplot is a plot of points that represent the values of the two continuous variables for each case in the data set. The two coordinates (x and y) determine the location of each point corresponding to its specific values on the two variables. Figure 3.3 is an illustration showing some scatterplots with different correlation coefficients between two variables **x** and **y**.

From Figure 3.3 we know that a scatterplot can make us realize how the two variables vary together, and correlation coefficients can reflect the closeness to a linear relationship between two variables. Besides, we still have to consider the statistical significance, p-level [16]. The p-level mentioned before is a measure of reliable degree of a result. It represents the probability of error that is involved in accepting our result as valid. In other words, the value of a p -level represents a probability that the relation between the variables found in our sample is a coincidence. The lower the p -level, the more we can believe that the observed relation between variables is reliable. Oppositely, the higher the p-level, the more we doubt that the correlation is a coincidence.

In the procedure of leading indicators identification, using the Spearman correlation is the second step to make extraction. After implementing the Bayesian analysis, we can obtain a posterior probability series, $P_{step}(t)$ of each variable. For each variable, we continue to use the Spearman correlation to estimate the relationship between the $P_{step}(t)$ and the forward crisis variable Y_{t-j}. Then we compare the results so as to find out key variables. There is a matter needing attention is that those key variables are not leading indicators yet because some redundancies may exist among them. The problem can be solved with factor analysis and discriminant analysis which will be discussed in the following sections.

3.3 Factor Analysis

Factor analysis is a statistical procedure used to uncover relationships among many variables. Factor analysis finds relationships or natural connections where variables are maximally correlated with one another and minimally correlated with other variables, and then groups the variables accordingly. This allows numerous intercorrelated variables to be condensed into fewer dimensions, called factors. In the procedure, it reduces a large number of variables to a smaller number of factors and a factor is actually a linear combination of multiple variables. The factor analysis technique is mainly applied to reduce the number of variables or to classify variables [17].

A correlation table can tell us the interrelationships among all the variables. We summarize the correlation between two variables with a correlation coefficient in a scatterplot and a regression line is fitted to represent the summary of the linear relationship between the variables. Given a high correlation between the two variables, we can conclude that they are quite redundant. Note that redundancy can cause some errors of results in a system because it will increase systematic load and miss other important information. If we could define a factor that approximates the regression line in such a plot, then the factor would capture most of the essence of the two items. In a sense we have reduced the two variables to one factor. Combining two correlated variables into one factor, illustrates the basic idea of factor analysis. If we extend the two-variable example to multiple variables, then the computations become more involved, but the basic principle of condensing the information contained in a number of original variables into a smaller set of factors remains the same.

While performing the factor analysis, there are three decisions to be made: the method of factor extraction, the type of factor rotation, and the number of factors to be used [18]. First, principle components analysis (PCA) is the method we use in factor extraction. Basically, the extraction of principal components amounts to a variance maximizing rotation of the original variable space. After we have found the line on which the variance is maximal, that is the first factor has been extracted, there remains some variance around this line. We continue to define another line that maximizes the remaining variance, and so on. In this manner, consecutive factors are extracted. Because each consecutive factor is defined to maximize the variance that is not captured by the preceding factor, factors are independent of each other. In other words, factors are orthogonal to each other. Hence we arrive at a classification of the variables. Mathematically, consider a data set of p variables and n cases, then PCA performs analyses in the p-dimensional space defined by n cases. The data can be arranged in a matrix $\mathbf{X}$ of n rows and p columns

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \tag{3.8}$$

where X_{ij} represents the value of the j th variable for the i th case. Singular value decomposition (SVD) is a computationally more robust procedure than the PCA which expresses the principal components of the variables in $\mathbf{X}$. From SVD, the $n \times p$ matrix $\mathbf{X}$ has the form

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T \tag{3.9}$$

where $\mathbf{U}$ and $\mathbf{V}$ are $n \times p$ and $p \times p$ orthogonal matrices, with the columns of $\mathbf{U}$ spanning the column of $\mathbf{X}$, and the columns of $\mathbf{V}$ spanning the row space. $\mathbf{D}$ is a $p \times p$ diagonal matrix with diagonal entries $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$ called the singular values of $\mathbf{X}$. Given the sample covariance matrix $\mathbf{\Sigma} = \mathbf{X}^T\mathbf{X}/n$, the eigen decomposition of $\mathbf{\Sigma}$ is

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{D}^2\mathbf{V}^T \tag{3.10}$$

where the singular values are the square root of eigenvalues, λ_j of $\mathbf{X}^T\mathbf{X}$ and the columns of $\mathbf{V}$ are the corresponding unit eigenvectors. The eigenvectors $\mathbf{v}_j$ are also called the principal components or factors of the variables in $\mathbf{X}$.

The first eigenvector is extracted in a manner which causes it to account for a maximum amount of variance in the data. After each eigenvector is extracted, a residual data matrix is calculated and the procedure is repeated until there are no significant eigenvectors left. The variance accounted for by each eigenvector is measured by its eigenvalue. The variance is equal to the square of the eigenvalue. In this way, the factors are linear combinations of the original variables and result in the maximum sums of squares for the orthogonal projections. Consequently, a lower dimensional factor space is generated onto which the cases and the variables can be projected and classified into categories [19, 20]. Figure 3.4 illustrates the principal components of some data points in two dimensions. The largest principal component is the direction that maximizes the variance of the projected data, and the smallest principal component minimizes that variance.

In factor analysis, the factor coordinates are referred to factor loadings and the α th factor coordinate of the i th variable is given by

$$Z_{\alpha i} = \sqrt{\lambda_\alpha} v_{\alpha i} \tag{3.11}$$

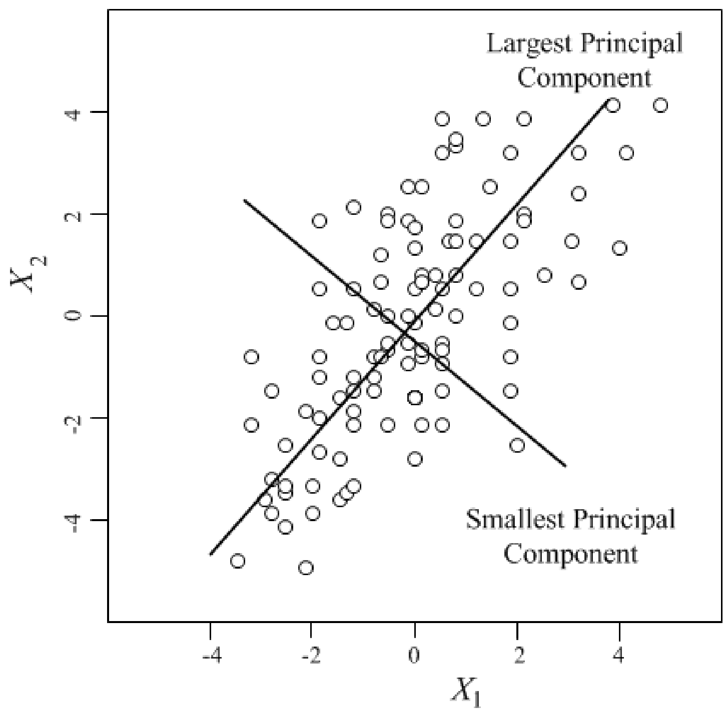

Figure 3.4: Principal components of some data points in two dimensions

As we know, a principal component is a linear combination of the variables that are most correlated with it. This implies that the factor coordinates of a variable are the correlations between the variable and the factor axes. Therefore, we can infer that the higher the absolute value of the factor loading of a variable on a particular factor, the more strongly is the variable related to that factor. In other words, the higher the magnitude of the factor coordinate of a variable, the more indicative is the variable of the concept represented by that factor.

Factor rotation serves to make the output more understandable and is usually necessary to facilitate the interpretation of factors. The sum of eigenvalues is not affected by rotation, but rotation will alter the eigenvalues of particular factors and will change the factor loadings. Since alternative rotations have different factor loadings and factor loadings are used to intuit the meaning of factors, this means that different meanings may be ascribed to the factors depending on the different rotation methods. Therefore, when we proceed the factor analysis, we have to decide a rotation method which leads to the most interpretable factor structure among all the oblique rotations, including no rotation, varimax rotation, quartimax rotation, equimax rotation, direct oblimin rotation, and promax rotation [21, 22]. According to the essence of the PCA which we use in factor extraction, the extraction of factors amounts to a variance maximizing rotation of the original variable space. This type of rotation is called variance maximizing abbreviated as varimax because the goal of the rotation is to maximize the variance of a new factor. Varimax rotation is an orthogonal rotation of the factor axes to maximize the variance of the squared loadings of a factor on all the variables in a factor matrix, which has the effect of differentiating the original variables by extracted factor. A varimax solution yields results which make it as easy as possible to identify each variable with a single factor. This is the most common rotation option.

Except of the issues of factor extraction and factor rotation, the number of factors is the third question that has to be determined. Kaiser criterion, scree plot, parallel analysis, minimum average partial criterion, variance explained criteria, Joliffe criterion, and comprehensibility are several different criteria for determining the number of factors [23]. A common rule of thumb for dropping the least important factors from the analysis is the Kaiser criterion proposed in 1960. It can only retain factors with eigenvalues greater than

1 and drops all factors with eigenvalues under 1. In essence this is like saying that, unless a factor extracts at least as much as the equivalent of one original variable, we drop it. It may sometimes overestimate the true number of factors, but it is still the most commonly used criterion [24].

3.4 Discriminant Analysis

Discriminant analysis (DA) is a technique used to build a predictive model of group membership based on observed characteristics of each case, and determine which variables discriminate between two or more naturally occurring groups. Specifically, DA generates functions from a sample of cases for which group membership is known; the functions can then be applied to new cases with measurements for the predictor variables but unknown group membership. The functions are known as discriminant functions. A variable will be called a predictor variable if the means for this variable are significantly different in different groups, and we can say that this variable discriminates between the groups [25].

Discriminant analysis is used in machine learning to find the linear combination of features which best separate two or more classes of object or event. The resulting combinations may be used as a linear classifier. DA is closely related to ANOVA and regression analysis, which also attempt to express one dependent variable as a linear combination of other features or measurements. In the other two methods however, the dependent variable is a numerical quantity, while for DA it is a categorical variable. According to the above-mentioned, DA allows us to group cases into two groups of occurring a crisis or not occurring a crisis, and proceed to build a predictive model based on predictor variables. That is, knowing the predictor variables of a new case, we can perform the discriminant analysis to determine whether a crisis will occur or not.

There are basically three types of DA: direct, hierarchical and stepwise. In direct DA, all the variables enter at once; in hierarchical DA, the order of variable entry is determined by the researcher; and in stepwise DA, statistical criteria alone determine the order of entry. In our model we concentrate on stepwise DA and there are both forward and backward stepwise versions. In forward stepwise discriminant function analysis, a model of discrimination is built step-by-step. Specifically, at each step all variables are reviewed and evaluated to determine which one will contribute most to the discrimination between groups. That variable will then be included in the model, and the process starts again. Oppositely, in backward stepwise analysis, all variables are included in the model in the beginning. At each step, the variable that contributes least to the prediction of group membership is eliminated. No matter what kind of stepwise analysis is performed, we would only keep the important variables as predictors in order to contribute the most to the discrimination between groups in the model. The stepwise procedure is guided by the respective F -to-enter and F -to-remove values. The F value for a variable indicates its statistical significance in the discrimination between groups. That is, it is a measure of the extent to determine whether a variable is a predictor.

Discriminant analysis is a technique for classifying a set of observations into predefined classes. We use the training samples which group membership is known to build a linear classification model. Based on the training samples, DA constructs a set of linear functions of the predictors, known as discriminant functions. The equation is given by:

$$f(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_d x_d \tag{3.12}$$

where β_0 through β_d are discriminant coefficients and x_1 through x_d are predictors. Let $\mathbf{c} = [\beta_0\ \beta_1 \cdots \beta_d]^T$ and $\mathbf{x} = [1\ x_1 \cdots x_d]^T$, the equation (3.12) can be written as:

$$f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}. \tag{3.13}$$

We use the training data $(\mathbf{x}^1,\ y^1), (\mathbf{x}^2,\ y^2), \ldots, (\mathbf{x}^N,\ y^N)$ to determine the parameter $\mathbf{c}$ to minimize the error rate of classification in our model. Each $\mathbf{x}^n$ for $n = 1,\ 2, \ldots, N$ is a vector of feature of the form $\mathbf{x}^n = [1\ x_1^n\ x_2^n \cdots x_d^n]^T$. Then we use least squares (LS) method to minimize the residual sum of squares:

$$\mathrm{E}(\mathbf{c}) = \sum_{n=1}^{N} (y^n - f(\mathbf{x}^n))^2 = \sum_{n=1}^{N} \left(y^n - \mathbf{c}^T\mathbf{x}^n\right)^2. \tag{3.14}$$

Differentiating the equation (3.14) with respect to $\mathbf{c}$ we obtain

$$\begin{aligned}\frac{\partial \mathrm{E}}{\partial \mathbf{c}} &= -2\sum_{n=1}^{N}(y^n - \mathbf{c}^T\mathbf{x}^n)\mathbf{x}^n \\ &= -2\left[\sum_{n=1}^{N}\mathbf{x}^n y^n - \sum_{n=1}^{N}\mathbf{x}^n(\mathbf{x}^n)^T\mathbf{c}\right] \\ &= -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\mathbf{c})\end{aligned} \tag{3.15}$$

where

$$\mathbf{X} = \begin{bmatrix}(\mathbf{x}^1)^T \\ (\mathbf{x}^2)^T \\ \vdots \\ (\mathbf{x}^N)^T\end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix}y^1 \\ y^2 \\ \vdots \\ y^N\end{bmatrix}.$$

Setting the derivative to zero yields the linear optimality condition of the form:

$$\mathbf{X}^T(\mathbf{y} - \mathbf{X}\mathbf{c}) = 0. \tag{3.16}$$

Assuming that $\mathbf{X}$ is nonsingular and hence $\mathbf{X}^T\mathbf{X}$ is positive define, we can obtain the unique solution:

$$\mathbf{c} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{3.17}$$

Therefore the discriminant functions can obtain the best fitted values in this form:

$$f(\mathbf{x}) = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{3.18}$$

These discriminant functions are used to predict the class of a new observation with unknown class. For a k class problem k discriminant functions are constructed. Given a new observation, all the k discriminant functions are evaluated and the observation is assigned to class j if the j th discriminant function has the highest value.

Chapter 4

The Early Warning Systems

4.1 System I of BEWS-A

Figure 4.1 is the flow chart of a Bayesian early warning system, named BEWS-A. The system takes advantage of the filter model to make feature selection. The research effort is devoted to the investigation of indirect performance measures, such as noise removal, data reduction, and information measure, in selecting features. The system, BEWS-A is characterized by the nature of the filter model. It does not rely on a particular classifier's bias, but on the intrinsic properties of the data, so the selected features can be used to learn different classifiers.

In step A, we first collect monthly data over period 1994M01 to 2003M12 for a sample of 28 countries and each country has 48 different financial variables. A key assumption of our thesis is that countries in the sample don't share common characteristics as far as openness to capital flows. Therefore, for each country, we will find out the leading indicators of its own and build a Bayesian early warning system, BEWS-A according to its particular characteristic. Thus the following steps from B to H will be described for a single

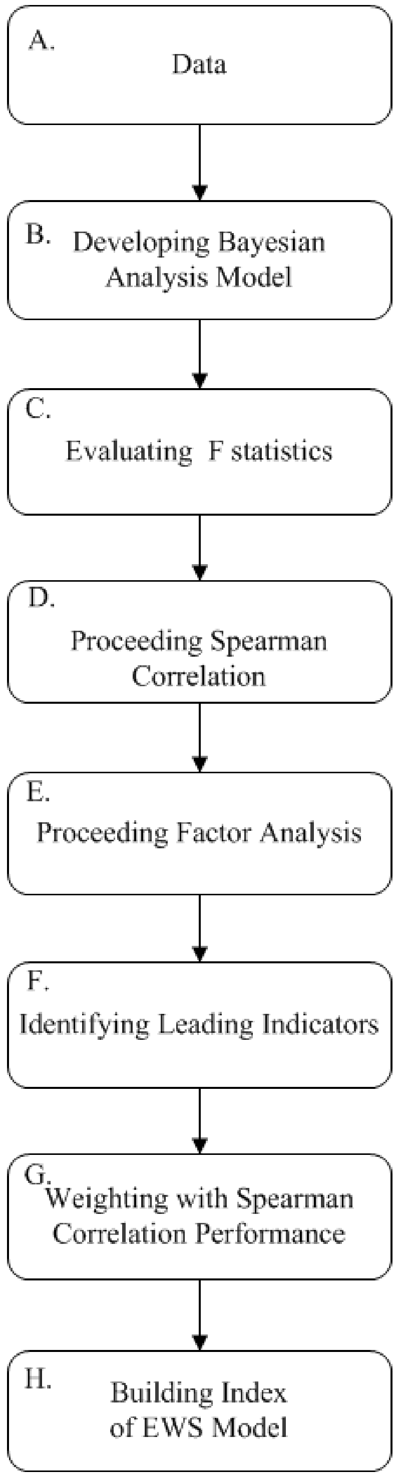

Figure 4.1: Flow chart of BEWS-A

country. In step B, we develop a Bayesian analysis model which is applied to detect regime shifts in a time series. A regime shift occurs when a significant difference exists between the mean value of the variable before and after a certain point. It implies that there are disturbance and change throughout the normal operation. So the detected regime shifts show signs of financial crises. As Figure 4.2 shows, the inputs of a Bayesian analysis model are the original values of all variables. They are transformed into posterior probabilities as the outputs which are representative of the probability that shifts occur.

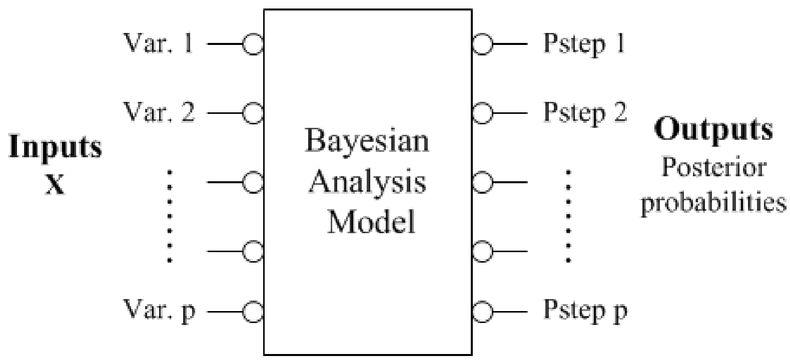

Figure 4.2: Inputs and Outputs for Bayesian analysis model

After building the Bayesian analysis model, the obtained posterior probability series, $P_{step}(t)$ will be used to evaluate both of the F statistic and the Spearman correlation coefficient for each of the variables. In step C and step D, the purpose is to extract key variables from all of the variables. The difference is that step C makes the first stage of extraction, and step D makes the second stage of extraction. Recall that, in chapter 1, we had introduced a forward crisis variable Y_{t-j} which is given with the aim of predicting four months ahead of time. In step C, it is treated as a response variable and the

F values could tell us the degree of how a variable affect the response variable Y_{t-j}. The higher the F values for a variable, the more we believe that it is a key variable. In step D, the Spearman correlation coefficients make it possible to find out the relationship between variables and Y_{t-j}. Modulus of a correlation coefficient reflects the strength of the relationship. This helps us to determine whether a variable is important or not as well. Figure 4.3 illustrates the procedure of variable extraction in step C and step D, where the columns of the matrix represent variables and the rows represent samples. Among these variables, the bold ones are key variables extracted by means of the F statistics and the shaded ones are key variables extracted by means of the Spearman correlation.

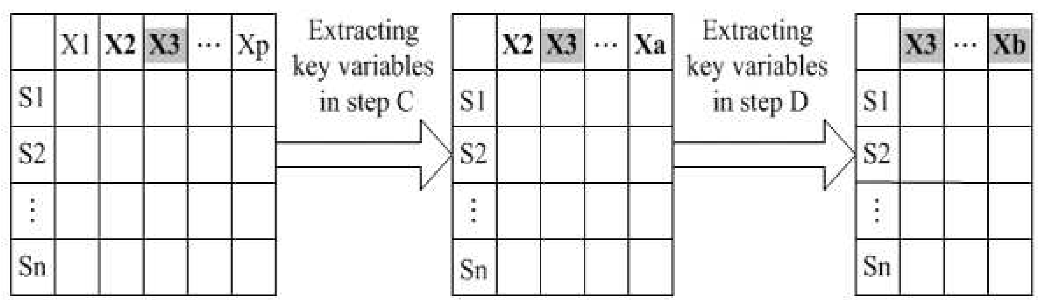

Figure 4.3: variable extraction in step C and D

In BEWS-A, We use both of the F statistics and the Spearman correlation to compare the results so as to find out key variables. However, those key variables are not leading indicators yet because they are intercorrelated. Thus there must be some redundancy existing among those key variables. In step E, we continue to employ factor analysis to classify key variables. Factor analysis is capable of finding relationships between key variables and separating those key variables into a smaller number of factors. In each factor, the factor loading of each key variable is a measure of significance. That is, the higher the factor loading is, the more significant the associated variable is. In this way, a variable with the highest Spearman correlation coefficient in a specific factor will be regarded as a leading indicator, and consequently the number of leading indicators we pick is as the same as the number of extracted factors. Therefore, the step F is achieved through step C, step D, amd step E. Table 4.1 is an example of leading indicator identification. Assuming that there are 8 key variables extracted from step C and step D. In the procedure of factor analysis, the 8 key variables are condensed into 3 factors and the associated factor loadings for each key variable are also calculated. As the table shows, the 8th key variable has the highest factor loading in factor 1. The 7th and the 3th key variables have the highest factor loadings in factor 2 and factor 3 respectively. Therefore, the 3th, 7th, and 8th variables will be identified as leading indicators.

Table 4.1: An example of leading indicator identification

Key Variables	Factor 1	Factor 2	Factor 3
Key Var.1	0.14	0.59	0.35
Key Var.2	0.85	0.06	0.21
Key Var.3	0.18	0.01	**0.91***
Key Var.4	0.14	0.80	0.17
Key Var.5	0.89	0.16	0.04
Key Var.6	0.23	0.56	0.11
Key Var.7	0.16	**0.91***	0.05
Key Var.8	**0.93***	0.14	0.05

Owing to the property of factor analysis, the identified leading indicators are minimally correlated with each other. They are different in significance according to the performance of predicting crises. So a weighting procedure is necessary to building a robust system. As we have mentioned before, the Spearman correlation coefficient quantifies the direction and magnitude of correlation, and the modulus of a correlation coefficient can reflect the strength of the relationship. The higher the modulus of a correlation coefficient is, the more important the variable is. Therefore, in step G, the leading indicators will be weighted with their Spearman correlation coefficients. In step H, we use the posterior probabilities of the leading indicators obtained from the Bayesian analysis model to create an index probability variable with their Spearmon correlation coefficients. Suppose $\boldsymbol{\mathcal{Z}} = \{Z_i(t),\, i = 1, 2, \ldots, d\}$ are the posterior probabilities of the leading indicators. $\boldsymbol{\mathcal{R}} = \{R_i,\, i = 1, 2, \ldots, d\}$ are their correlation coefficients and $\mathcal{P}(t)$ is the index probability variable. Then the equation is defined as follows:

$$\mathcal{P}(t) = \frac{Z_1(t)R_1 + Z_2(t)R_2 + \cdots + Z_d(t)R_d}{R_1 + R_2 + \cdots + R_d} \tag{4.1}$$

In addition, a threshold value $\mathcal{T}$ is determined based on the training samples. For BEWS-A, the calculated index variable $\mathcal{P}(t)$ will be ultimately put to use in anticipating crises in this form:

$$\mathcal{Y}(t) = \begin{cases} 1 & \text{if } \mathcal{P}(t) \geq \mathcal{T} \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

where $\mathcal{Y}(t) = 1$ represents that crises may occur within future four months and then an alarm will be sent. This system has accorded with our expectation to give an early warning. As long as new information is available, the new posterior probabilities of leading indicators will be calculated, and then $\mathcal{P}(t)$ will be updated according to the present state of information.

4.2 System II of BEWS-B

Figure 4.4 is the flow chart of the other Bayesian early warning system, named BEWS-B. We use a wrapper model to make feature selection in this system. The major concept of the wrapper model is the machine learning approach. The major concern of machine learning is to minimize the classifier error rate, and equal the measurement cost for all the features. The system, BEWS-A is characterized by the nature of the wrapper model. It builds a classifier with an aim to achieve the highest predictive accuracy as possible, and selects the features used by the classifier as the optimal features. When feature subsets are systematically generated, for each subset of features, a classifier is generated from the training data with chosen features. Its accuracy is recorded and the feature subset with the highest accuracy is kept. When the selection process terminates, the subset with the best accuracy is chosen.

In BEWS-B, step A, step B, step C, and step D are all the same as in BEW-A. We still have to evaluate the F statistics and the Spearman correlation coefficients to make variable extraction as usual. It allows to make primary selection from all of the variables, and to avoid proceeding discriminant analysis with too large data size. After variable extraction, the key variables will be taken as inputs in step E. The discriminant analysis

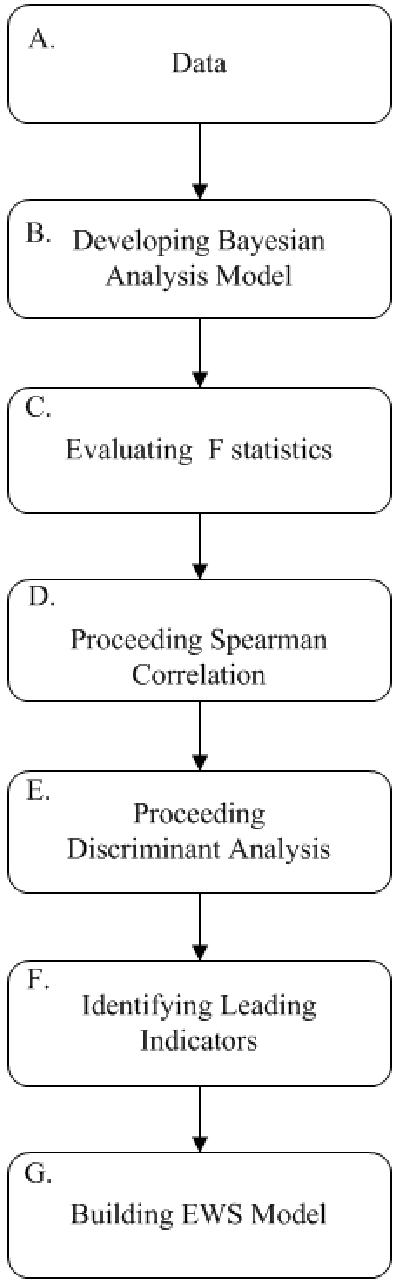

Figure 4.4: Flow chart of BEWS-B

is the most important process for BEWS-B. Owing to we focus on forward stepwise discriminant analysis, the input variables will be taken into account step-by-step. The stepwise procedure is guided by the F-to-enter and F-to-remove values which indicate the significance in the discrimination between classes. We would only keep the important variables as leading indicators in order to contribute the most to the discrimination between classes, and then the identification of leading indicators of step F is achieved. Finally, in step G, the discriminant analysis constructs a set of discriminant functions of the leading indicators. These discriminant functions will be used to build an EWS model. When there are new observations with unknown class, we can take advantage of the model to predict the class of the new observations. The BEWS-B can achieve the purpose of selecting the optimal features and improving predictive accuracy.

Chapter 5

Simulation Results and Discussions

5.1 Data Introduction and Simulation Process

In the recent years, we saw a large number of financial crises in emerging market economies with devastating economic consequences. These financial crises caused great damage to macro-economics, development of industries, and trade of stock market. Seriously the financial crises can bring about great depressions. In the financial futures market, there are various financial variables such as current account, trade balance, public debts, inflation rate, and so on. By detecting underlying economic weaknesses and vulnerabilities among these financial variables, we can offer an early warning before a crisis comes, and then take some pre-emptive steps to reduce the risks of experiencing a crisis. As a result, many international organizations have begun to develop Early Warning Systems (EWS) for the purpose of anticipating *whether* and *when* individual countries may be affected by a financial crisis. The aim of our thesis is to develop a new EWS that improves the anticipating performance with the financial variables we collected.

In our thesis, we use monthly data over the period Jan. 1994 to Dec. 2003 for a sample of 28 countries. The data include 48 different financial variables and a financial crisis variable for each country. The financial crisis variable is named mCC ERW, and it is a binary response variable which indicates whether a crisis really occurs or not. Table 5.1 lists the 48 finaicial variables, and they are the input variables which are used to predict financial crises. The response variable and input variables are both time series, and there are 120 samples. Besides, in order to achieve the purpose of predicting financial

Table 5.1: financial variables tested in our system

RER	mEX G	12mEX G	mIM G
12mIM G	CA/GDP	FR G	12mFR G
M1/FR	m M1/FR G	12m M1/FR G	M2/FR
m M2/FR G	12m M2/FR G	SD/FR	TD/FR
FR/IM	FR/EX	mM1 G	12mM1 G
mM2 G	12mM2 G	MMM(M2/RM)	mMMM G
12mMMM G	DC/GDP	m DC/GDP G	12m DC/GDP G
EMR	RIR	CBD	m CBD G
12m CBD G	RRA	FBY	SD/GDP
TD/GDP	INR	12mINR	GDPC
mGDPC G	12mGDPC G	NSR	mNSR G
12mNSR G	mGDP G	12mGDP G	(EX - IM)/GDP

crises, we set up another forward crisis variable, named mY4Q ERW to take place of original financial crisis variable in process of the following analysis. Figure 5.1 is an example of how mCC ERW changes into mY4Q ERW for ARGENTINA, the value of 1 means a crisis occurs, and 0 otherwise. As for the simulation procedures are illustrated as Figure 5.2. First, we use the Bayesian analysis to detect regime shifts of all input variables. Second, we use the obtained posterior probability from Bayesian analysis model to evaluate both of the F statitic and the Spearman correlation coefficient for each of the input variables, and then proceed to identify leading indicators for each country. Finally, we take advantage of the extracted leading indicators to build EWS models for each country.

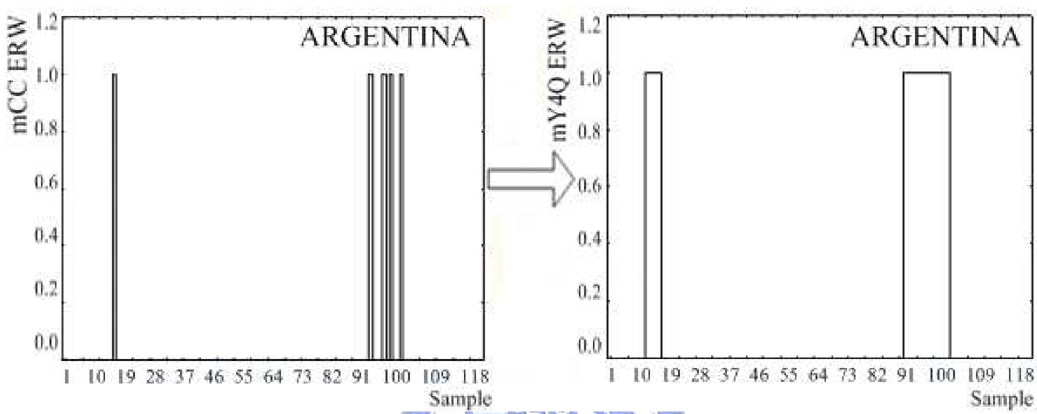

Figure 5.1: Financial crises for ARGENTINA

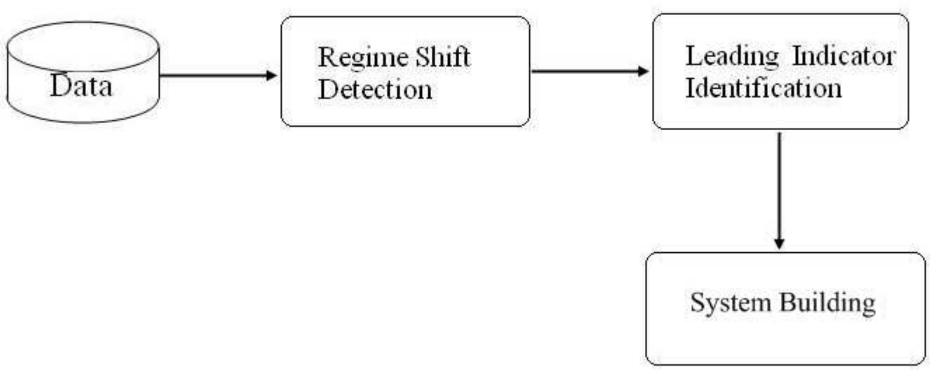

Figure 5.2: The flow chart of simulation

5.2 Collection of Simulation Results for Regime Shift Detection

In the procedure of regime shift detection, Bayesian analysis is applied to detect mean shifts in a time series. The inputs of a Bayesian analysis model are the original values of all variables. The outputs are posterior probability series. Before we implement the Bayesian analysis, there are four tuning parameters have to be determined. They are pre-change window (W_1), post-change window (W_2), forgetting factor (μ), and prior probability $P(Y_t = 1)$. The difference of tuning parameters can decide the complexity of a model and result in different anticipating errors. From Figure 5.3 we can see an output comparison with different values of W_1. For a single variable, when W_1 gradually increases, we detect more regime shifts and the posterior probability gets higher. The advantage is allowing us to predict a financial crisis more easily. However, the drawback is that it may fail to detect shifts occurring in early periods and the number of false alarms will increase oppositely. So we have to seek a balance and $W_1 = 15$ is an appropriate value. As for W_2, it causes delay of $W_2 - 1$ in shift detection because we have to use the data samples after the change point. So the value of W_2 cannot be too large. Owing to we use the forward crisis variable to take place of the crisis variable in order to predict crises in 4 months ahead of time. The values 1 to 5 are acceptable for W_2. From Figure 5.4 we realize that $W_2 = 3$ is the most applicable value. The value of forgetting factor μ represents the degree we give thought to the old information. As the value of μ gets smaller, a model will update more frequently and it means that the old information will be forgot faster. Basically, we detect regime shifts by estimating the difference

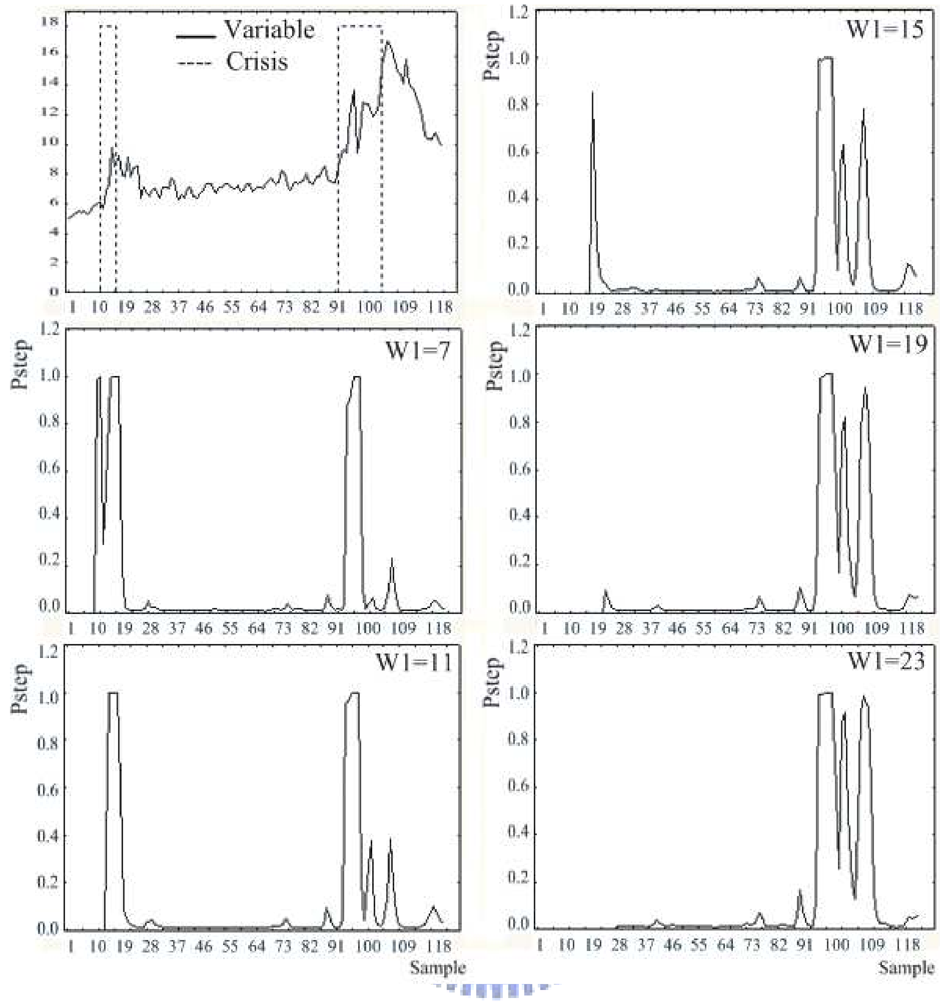

Figure 5.3: Comparison with defferent values of W1

between the old data and the newest data samples. The old data play a very important role in the Bayesian analysis model. So the value of μ is given by 0.95 in our model. The last tuning parameter needed to be determined is the prior probability $P(Y_t = 1)$. we decide its value in reference to the past experience. From Jan. 1994 to Dec. 2003, there are 120 data samples, and most countries suffer from financial crises less than 2 times. So the value of $P(Y_t = 1)$ is given by 0.01 in our model. After determining the four tuning parameters, we will detect the regime shifts for all of the input variables in

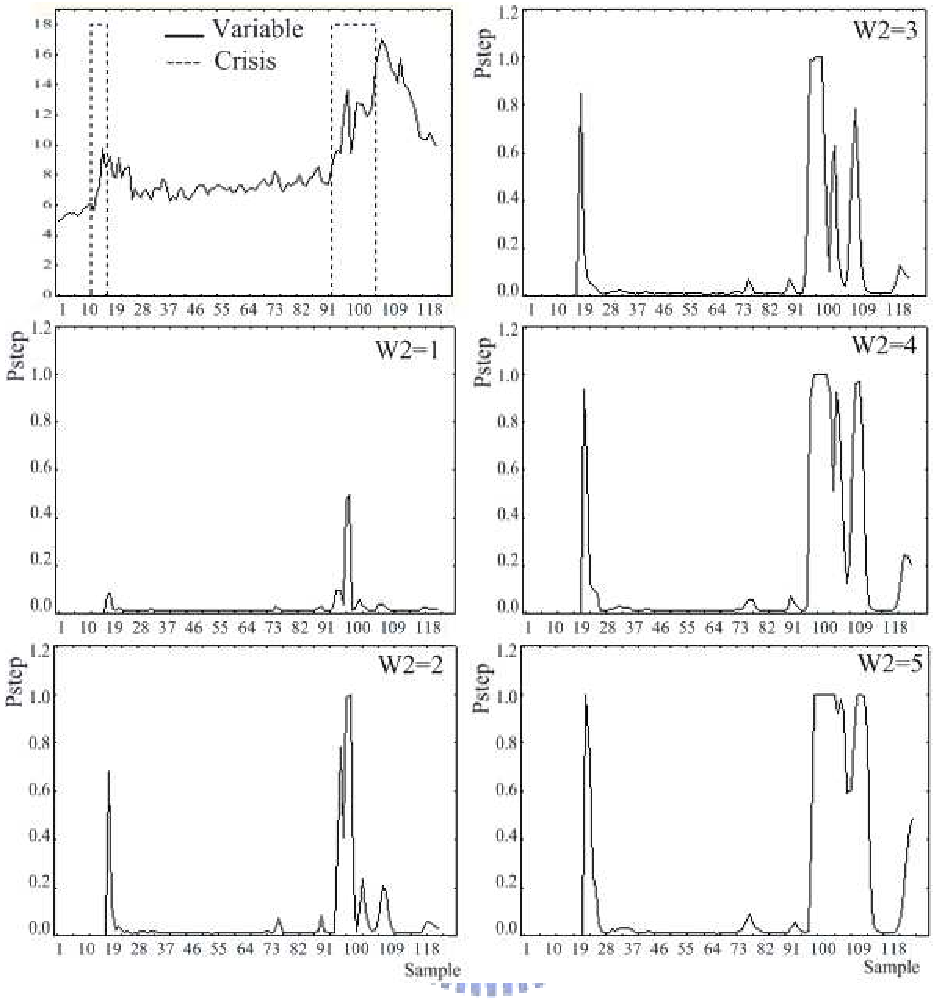

Figure 5.4: Comparison with defferent values of W2

this Bayesian analysis model.

Figure 5.5 shows the simulation results of regime shift detection for ARGENTINA. It also displays the forward crisis variable, mY4Q ERW in comparison with the simulation results. we can notice that when a crisis occurs, lots of regime shifts will be detected among the variables. But, It is quite obvious that the detected regime shifts do not match financial crises. This is because the input variables are too many, and we have not made extraction. Therefore, making variable extraction and pick out key variables are necessary. We will use the simulation results to calculate F statistics and Spearman correlation coefficients in order to identify leading indicators in next procedure.

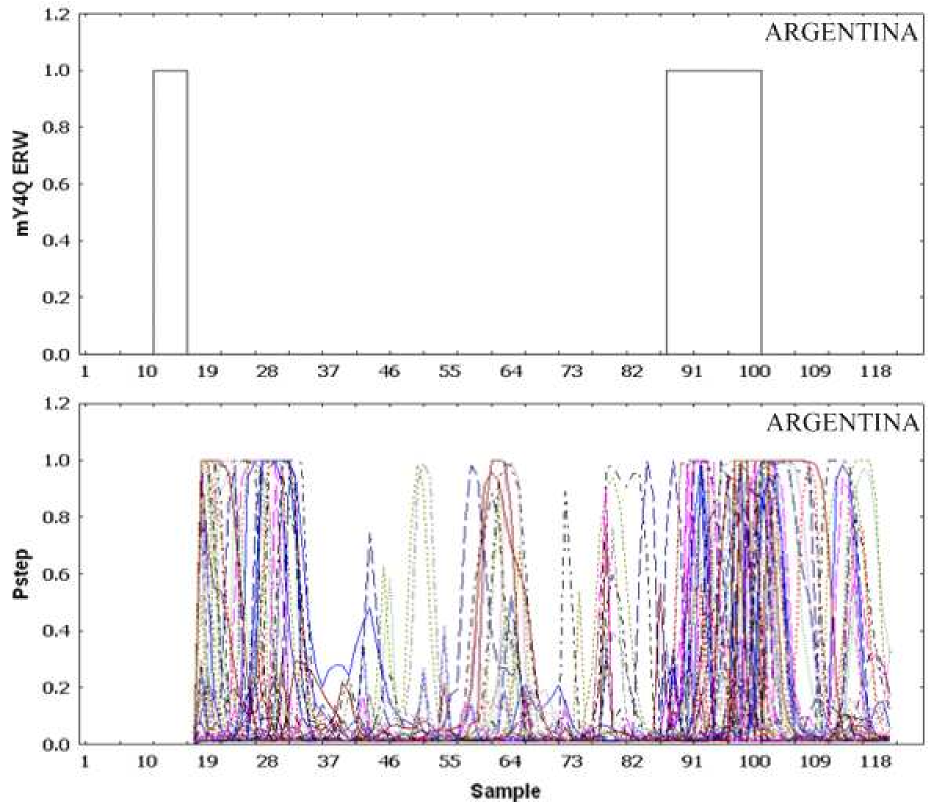

Figure 5.5: Simulation results of regime shift detection for ARGENTINA

5.3 Leading Indicator Identification

For 48 different input variables with 120 samples, we first use F statistics to make feature selection. Table 5.2 presents the results and sequencing of significance using ANOVA to evaluate the F statistics and the p-level of all input variables. The F value provides a test for the statistical significance of the differences among means of classified samples.

Table 5.2: Key variables extracted by F statistics

ARGENTINA					
Variable	F statistic	p-level	Variable	F statistic	p-level
TD/FR	175.46345	0.00000	DC/GDP	10.97127	0.00129
M2/FR	90.87242	0.00000	FR/IM	10.35467	0.00175
(EX - IM)/GDP	82.22062	0.00000	12m M1/FR G	8.55043	0.00428
RIR	81.08383	0.00000	12mIM G	8.22778	0.00504
CA/GDP	78.96180	0.00000	TD/GDP	8.07963	0.00544
INR	71.51459	0.00000	12mM1 G	7.61407	0.00690
mM2 G	55.57029	0.00000	M1/FR	7.01714	0.00940
FR/EX	54.72298	0.00000	mMMM G	6.58445	0.01179
mIM G	53.12155	0.00000	FR G	6.49024	0.01239
MMM(M2/RM)	51.80418	0.00000	RER	6.08802	0.01533
mM1 G	44.70306	0.00000	SD/GDP	4.39464	0.03860
mGDP G	42.54848	0.00000	EMR	4.07456	0.04624
mGDPC G	42.52536	0.00000	12mEX G	3.15495	0.07877
m DC/GDP G	36.86279	0.00000	GDPC	1.95798	0.16486
12mMMM G	30.82983	0.00000	FBY	1.23840	0.26848
m M2/FR G	30.57065	0.00000	12mGDPC G	1.05769	0.30625
m CBD G	28.65698	0.00000	12mGDP G	1.04226	0.30979
m M1/FR G	25.39791	0.00000	12mINR	0.96480	0.32838
RRA	24.18581	0.00000	NSR	0.86332	0.35507
mNSR G	22.89918	0.00001	12mNSR G	0.68292	0.41057
12m CBD G	21.40104	0.00001	SD/FR	0.42338	0.51676
12m M2/FR G	18.52652	0.00004	mEX G	0.30513	0.58193
12mFR G	16.00145	0.00012	12mM2 G	0.13088	0.71829
12m DC/GDP G	13.84032	0.00033	CBD	0.07593	0.78346

For a variable, the larger the F value is, the more significantly the variable affects the response variable. Comparing with other variables, TD/FR is the most significant variable by F statistics. Variables with p-level > 0.05 are relatively insignificant ones. So we remove 12mEX G, GDPC, FBY, 12mGDPC G, 12mGDP G, 12mINR, NSR, 12mNSR G, SD/FR, mEX G, 12mM2 G and CBD these 12 variables. Afterwards we select the remained 36 variables as key variables at first stage of variable eatraction.

For the key variables extracted by F statistics, we continue to use Spearman correlation and to make further feature selection. Modulus of a correlation coefficient represents the strength of the relationship between two variables. Here we take advantage of the Spearman correlation to find out which variables vary with the forward crisis variable. Table 5.3 presents the results of implementing the Spearman correlation for the 36 key variables and arranges in an order according to the important degree. Variables with coeffi-

Table 5.3: Key variables extracted by Spearman correlation

ARGENTINA					
Variable	Correlation Coefficient	p-level	Variable	Correlation Coefficient	p-level
TD/FR	0.57104	0.00000	m M2/FR G	0.33710	0.00057
(EX - IM)/GDP	0.54215	0.00000	INR	0.33026	0.00074
CA/GDP	0.53188	0.00000	mM1 G	0.32630	0.00087
RIR	0.53126	0.00000	m M1/FR G	0.32531	0.00090
FR/EX	0.52875	0.00000	12m M1/FR G	0.30174	0.00217
M2/FR	0.51016	0.00000	mNSR G	0.28502	0.00387
MMM(M2/RM)	0.49730	0.00000	12mM1 G	0.28433	0.00396
12mMMM G	0.42268	0.00001	12mIM G	0.27646	0.00513
mGDP G	0.41769	0.00001	DC/GDP	0.26588	0.00720
mGDPC G	0.41671	0.00002	RER	0.23755	0.01676
RRA	0.41580	0.00002	m DC/GDP G	0.23293	0.01907
12m DC/GDP G	0.41165	0.00002	FR/IM	0.22951	0.02096
12m M2/FR G	0.38952	0.00006	FR G	0.22114	0.02626
mIM G	0.38919	0.00006	M1/FR	0.21327	0.03225
mM2 G	0.36757	0.00016	EMR	0.20541	0.03934
m CBD G	0.36659	0.00016	SD/GDP	0.20344	0.04130
12m CBD G	0.36562	0.00017	mMMM G	0.17298	0.08366
12mFR G	0.36180	0.00020	TD/GDP	0.12188	0.22470

cients greater than 0.2 will be thought acceptable. Among them, TD/FR has the highest correlation coefficient with the forward crisis variable. Besides, (EX - IM)/GDP, CA/GDP, RIR, FR/EX and M2/FR also have quite high correlation coefficients greater than 0.5 as well. These six variables certainly have strong relationships with the forward crisis variable. After the second stage of variable extraction, variables with p-level > 0.05, such as mMMM G and TD/GDP, will be removed. Therefore, the remained 34 variables will be regarded as the final key variables. After the final key variables are extracted, we will find out leading indicators among them in two kinds of methods. One of the methods is factor analysis, which is employed for system BEWS-A. The other method is discriminant analysis, which is employed for system BEWS-B.

For BEWS-A, factor analysis is proceeded to uncover relationships among the key variables, and classify them. The purpose is to reduce the number of key variables and remove the redundancy. While performing the factor analysis, we use PCA to extract consecutive factors, and the extracted factors are orthogonal to each other. Figure 5.6 is the plot of eigenvalues evaluated in PCA. Given an eigenvalue threshold of 1.0 to determine the number of factors according to the Kaiser criterion, so we only retain the first 8 factors. In each factor, the factor loading of a variable represents the correlation between the variable and the factor axis. If the modulus of a factor loading for a variable is high on a specific factor, we can infer that the variable is strongly correlated with the factor. Figure 5.7 lists the factor loadings obtained from the factor analysis for all of the key variables and factors. It indicates that the correlation is high enough for the fields which are marked. Because the factors are independent of each other, only one field will be marked in one

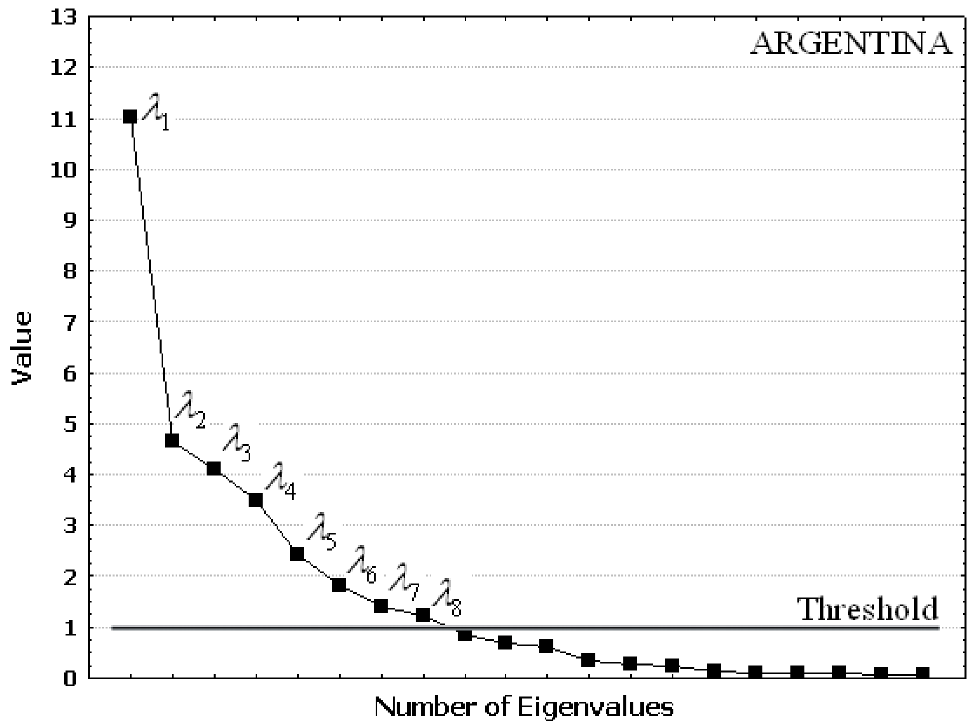

Figure 5.6: Plot of eigenvalues evaluated in factor extraction

row. Therefore, each variable will only belong to a specific factor. In this way, we succeed in dividing the key variables into 8 classes. In each factor, the number of variables which have high correlations with this factor are more than one, so we realize that there must exist some redundancy. Taking factor 1 for example, there are many marked variables which have high correlations with it, including RER, CA/GDP, M2/FR, TD/FR, FR/EX, 12mM1 G, MMM(M2/RM), 12mMMM G, DC/GDP and (EX-IM)GDP. These variables are belong to the same class and highly related with each other. Therefore, we will only choose one variable as leading indicator in one factor, and the variable must have the highest Spearman correlation coefficient with the forward crisis variable. Table 5.4 summarize the identified leading indicators for all of the factors.

ARGENTINA								
Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
RER	0.935907*	0.086785	0.007438	0.141852	0.151113	0.048308	-0.020801	-0.233475
mIM G	0.168957	0.128020	0.544796*	0.215629	-0.262663	-0.052715	0.105964	0.158536
12mIM G	0.035043	-0.251905	0.073042	0.023672	-0.889642*	0.032738	0.091501	0.192701
CA/GDP	0.937611*	0.001216	-0.009456	0.164019	0.207637	-0.112171	0.035130	-0.078538
FR G	0.006423	0.036014	-0.888970*	0.083908	-0.204563	-0.146091	0.159112	0.073242
12mFR G	-0.246571	0.244011	-0.092978	-0.036357	-0.829076*	0.113220	0.139199	0.047203
M1/FR	-0.201602	-0.384192	0.100877	-0.029019	-0.009029	0.092304	0.071841	0.830655*
m M1/FR G	0.038290	-0.219191	0.841666*	0.054686	0.010874	0.012839	-0.093485	0.076546
12m M1/FR G	0.106186	-0.936128*	0.118274	-0.054722	-0.001116	-0.044694	0.029686	0.189327
M2/FR	-0.673710*	-0.347604	0.234636	-0.171110	0.215045	-0.232109	-0.009909	0.379065
m M2/FR G	-0.037108	-0.043243	0.921378*	-0.034414	-0.069329	-0.152931	-0.143371	0.140955
12m M2/FR G	-0.360644	-0.865519*	0.133865	-0.091394	-0.117153	-0.112311	-0.020182	0.158473
TD/FR	0.805696*	-0.030226	0.162651	0.063767	0.416025	-0.228476	-0.015027	-0.198121
FR/IM	0.324897	0.039409	-0.241849	-0.030195	0.377052	0.093875	-0.060287	-0.776807*
FR/EX	-0.724316*	0.141617	-0.105086	-0.263336	-0.155946	0.117901	-0.055952	-0.388913
mM1 G	0.176851	-0.261368	-0.306372	0.144562	0.089464	0.336257	0.506461*	-0.101083
12mM1 G	0.858723*	0.212911	-0.014362	0.057220	-0.244342	0.185226	0.127327	0.077764
mM2 G	0.145283	0.029502	-0.253740	0.061795	-0.135118	0.104323	0.912502*	0.044589
MMM(M2/RM)	-0.657755*	0.422577	0.119262	-0.043816	0.013860	-0.242468	0.016158	-0.507809
12mMMM G	-0.709579*	0.292451	0.037367	-0.016395	-0.254572	0.091220	0.046384	0.431145
DC/GDP	0.845012*	0.056085	0.023438	0.054481	0.262636	-0.002149	0.012264	-0.440615
m DC/GDP G	-0.033788	-0.037889	0.055527	-0.604261*	0.357513	0.372824	0.444356	-0.165746
12m DC/GDP G	0.361884	0.500537	-0.013048	0.067824	0.701214*	0.170465	0.011313	-0.165025
EMR	-0.397217	0.130617	-0.022845	0.178163	-0.287038	0.708880*	0.038548	0.361539
RIR	0.294655	0.145526	0.210809	0.237218	0.643610*	0.271927	-0.243540	-0.157637
m CBD G	-0.116821	0.019499	-0.062519	-0.148634	-0.274145	-0.202743	0.847709*	0.135603
12m CBD G	-0.511584	0.144380	-0.026340	-0.062604	-0.731462*	0.225707	0.170796	0.045946
RRA	0.606645	-0.674233*	-0.073221	-0.026658	-0.102720	0.121307	-0.043402	0.140142
SD/GDP	0.394848	0.253556	0.060131	0.071543	0.240570	-0.092845	-0.061036	-0.768770*
INR	0.396473	0.172404	-0.001064	0.377883	0.447672	0.518373*	-0.176186	-0.176032
mGDPC G	0.174649	0.031938	-0.021652	0.968518*	0.015996	0.028768	0.028701	0.000649
mNSR G	0.120013	0.068176	0.028425	0.905317*	0.179675	0.162438	-0.052494	-0.079404
mGDP G	0.173663	0.032096	-0.021737	0.968667*	0.015958	0.029021	0.028691	0.001631
(EX - IM)/GDP	0.925821*	0.029550	-0.012893	0.188981	0.205403	-0.058022	-0.008987	-0.191818

Figure 5.7: Factor loadings of the key variables and the factors extracted in factor analysis

Table 5.4: Leading indicators identified in BEWS-A for ARGENTINA

ARGENTINA	
Factor	Leading indicator
Factor1	TD/FR
Factor2	RRA
Factor3	mIM G
Factor4	mGDP G
Factor5	RIR
Factor6	INR
Factor7	mM2 G
Factor8	FR/IM

Table 5.5 persents the correlation matrix for the 8 leading indicators and the forward crisis variable. we can notice that the highest correlation coefficient between the leading indicators is smaller than 0.60. so it is believable that there are not redundancy existing among the leading indicators. Besides, from the sactterplots of Figure 5.8, we can know these leading indicators truly have high correlations with the forward crisis variavle, mY4Q ERW (which is also called the response variable). As for the Figure 5.9 and 5.10, they show the posterior probability series using regime shift detection for the 8 leading indicators in comparison with the forward crisis variable. We can find the posterior probabilities for the samples with crises are obviously greater than the posterior probabilities for the samples without crises.

For BEWS-B, discriminant analysis is the technique used to identify leading indicators. In this thesis, we adopt a forward stepwise analysis. The model of discriminanation is built step by step relying on including the key variables one by one to examine the F-to-enter and F-to-remove values. After finishing all of the steps, the variables remained in the model are exactly

Table 5.5: Correlation matrix for leading indicators and the forward crisis variable

ARGENTINA									
Variable	mY4Q ERW	TD/ FR	RRA	m IM G	m GDPG	RIR	INR	m M2 G	FR/ IM
mY4Q ERW	1.00	0.57	0.42	0.39	0.42	0.53	0.33	0.37	0.23
TD/FR	0.57	1.00	0.56	0.13	0.17	0.59	0.25	0.30	-0.07
RRA	0.42	0.56	1.00	0.01	-0.03	0.54	0.41	0.43	0.12
mIM G	0.39	0.13	0.01	1.00	0.43	0.21	0.22	0.02	0.23
mGDP G	0.42	0.17	-0.03	0.43	1.00	0.20	0.03	0.08	0.21
RIR	0.53	0.59	0.54	0.21	0.20	1.00	0.42	0.37	0.14
INR	0.33	0.25	0.41	0.22	0.03	0.42	1.00	0.23	0.57
mM2 G	0.37	0.30	0.43	0.02	0.08	0.37	0.23	1.00	0.25
FR/IM	0.23	-0.07	0.12	0.23	0.21	0.14	0.57	0.25	1.00

the leading indicators for BEWS-B. These leading indicators will be used to construct discriminant functions, and to contribute the most to the discrimination between groups in samples. Table 5.6 presents the results using the discriminant model with identified leading indicators. The Wilks' lambda is generally used to denote the statistical significance of the discriminatory power of the current model. Its value ranges from 1.0 (no discriminatory power) to 0.0 (perfect discriminatory power). Each value in the second column listed in Table 5.6 denotes the Wilks' lambda after the respective variable is entered into the model. Partial lambda is the Wilks' lambda for the unique contribution of the respective variable to the discrimination between groups. Because a lambda of 0.0 denotes perfect discriminatory power, the smaller the Partical lambda, the greater is the contribution to the overall discrimination of the respective variable. As we can see, the Partial lambda in-

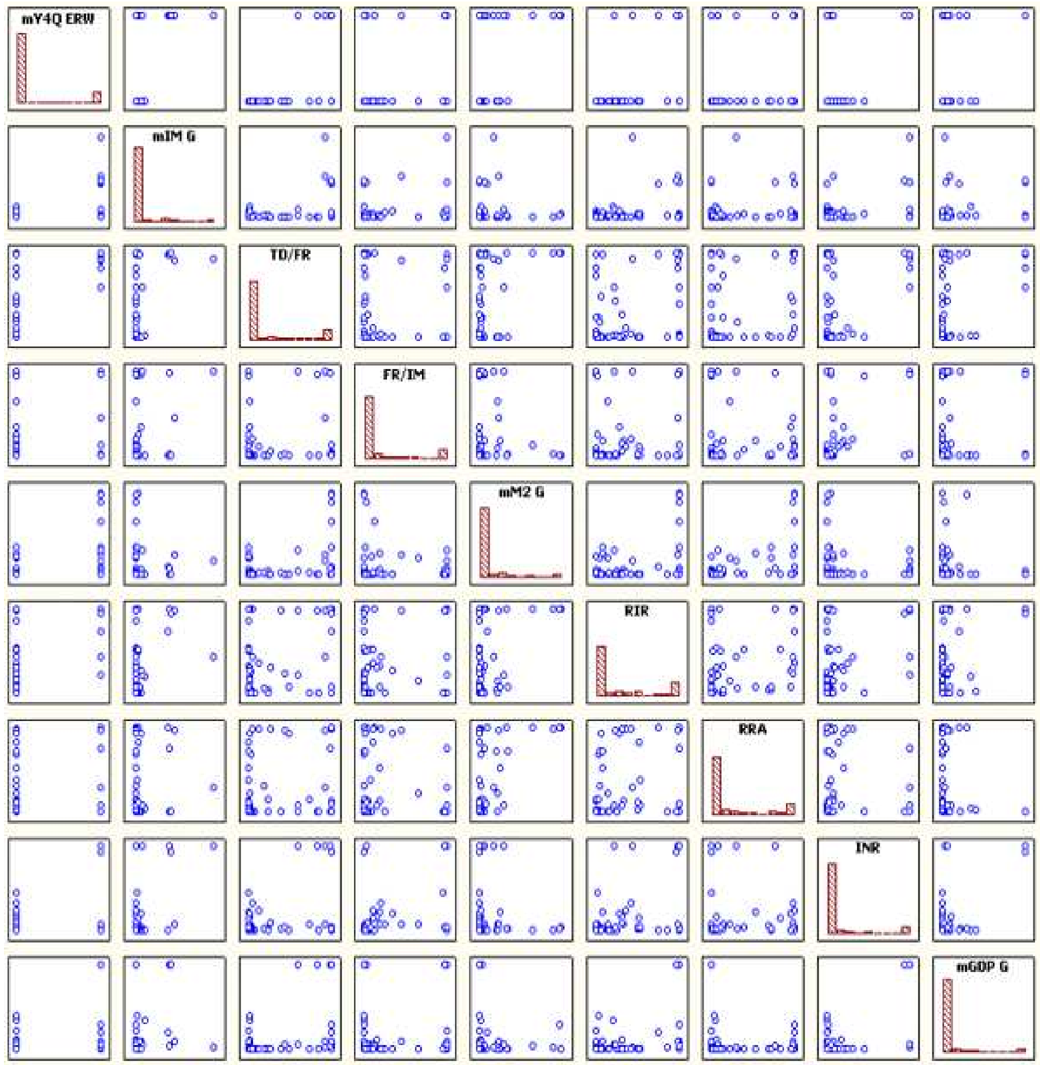

Figure 5.8: Scatterplots of posterior probabilities of extracted leading indicators and the forward crisis variable for ARGENTINA

dicates that variable M2/FR contributes most, variable M1/FR second most, variable 12mIM G third most, and variable 12mFR G contributes least to the overall discrimination. Thus, we can conclude that M2/FR and M1/FR are the major variables that allow us to discriminate between samples with and without crisis. As for the Tolerance, it is defined as 1 minus R-square of the respective variable with all other variables in the model, and this value

Table 5.6: Leading indicators identified in BEWS-B and the discriminant model for ARGENTINA

ARGENTINA						
Leading Indicator	Wilks' Lambda	Partial Lambda	F-remove (1,106)	p-level	Tolerence	1-Toler. (R-Sqr.)
12mFR G	0.30226	0.99983	0.01831	0.89262	0.34330	0.65677
M2/FR	0.52296	0.57789	77.42519	0.00000	0.06484	0.93516
FR/IM	0.31784	0.95083	5.48172	0.02109	0.21375	0.78625
m CBD G	0.31630	0.95547	4.94027	0.02836	0.12466	0.87534
RRA	0.30269	0.99841	0.16917	0.68168	0.06766	0.93234
12m DC/GDP G	0.31665	0.95439	5.06542	0.02647	0.11063	0.88937
12mIM G	0.32733	0.92327	8.80913	0.00371	0.15825	0.84176
M1/FR	0.34929	0.86521	16.51314	0.00009	0.05886	0.94115
mM2 G	0.30401	0.99407	0.63207	0.42837	0.12691	0.87309
MMM(M2/RM)	0.31515	0.95895	4.53707	0.03548	0.03842	0.96158
INR	0.31026	0.97406	2.82338	0.09585	0.27147	0.72853
m M2/FR G	0.31290	0.96584	3.74873	0.05551	0.18320	0.81680
FR G	0.30703	0.98431	1.68988	0.19644	0.19231	0.80769

gives an indication of the redundancy of the respective variable. For example, when the variable INR is about to enter into the model, it has a tolerance value of 0.27, then INR can be considered to be 73% redundant with the variables already included. In this system, we set the Tolerance threshold at its default value of 0.01. If a variable is included in the model that is more than 99% redundant with other variables, it means that its practical contribution to the improvement of the discriminatory power is dubious and the variable will be removed.

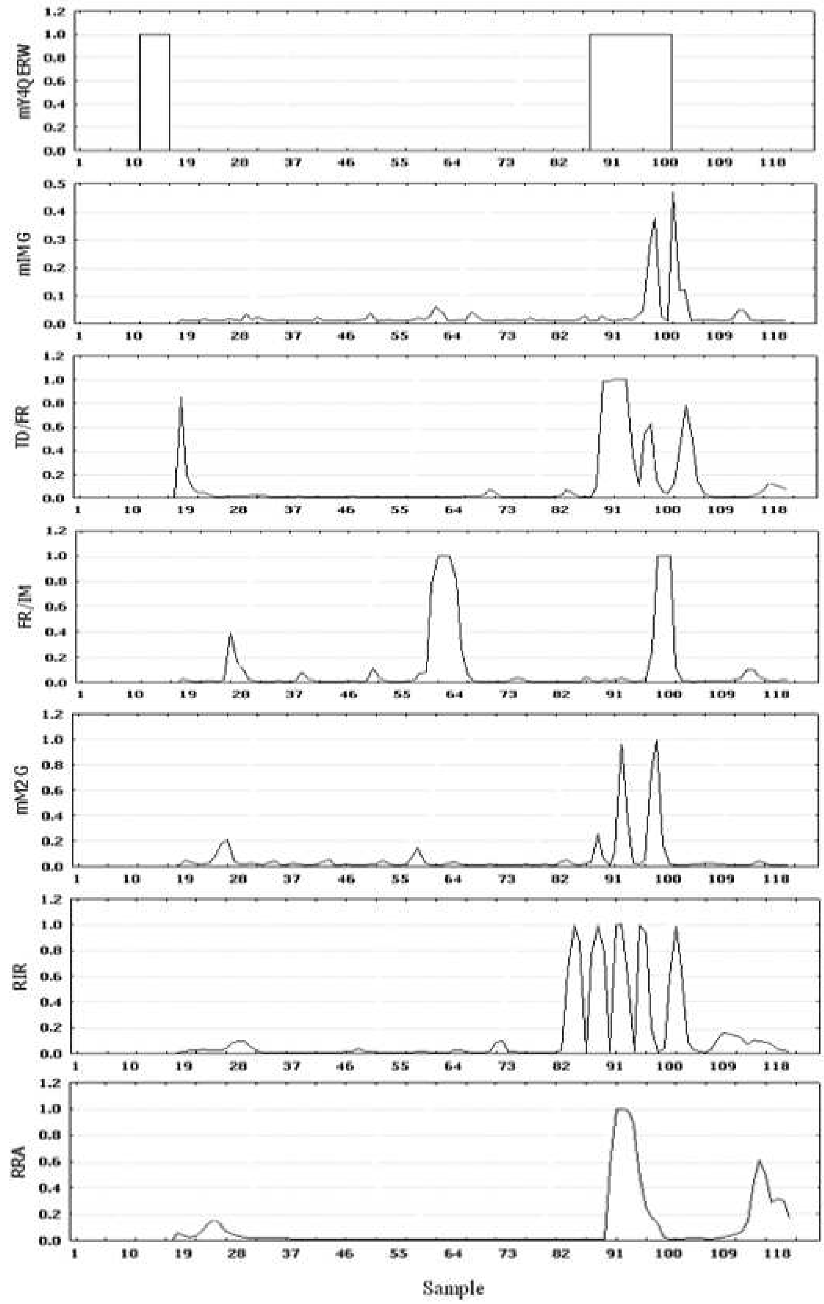

Figure 5.9: Comparison between posterior probabilities of each leading indicator and the forward crisis variable

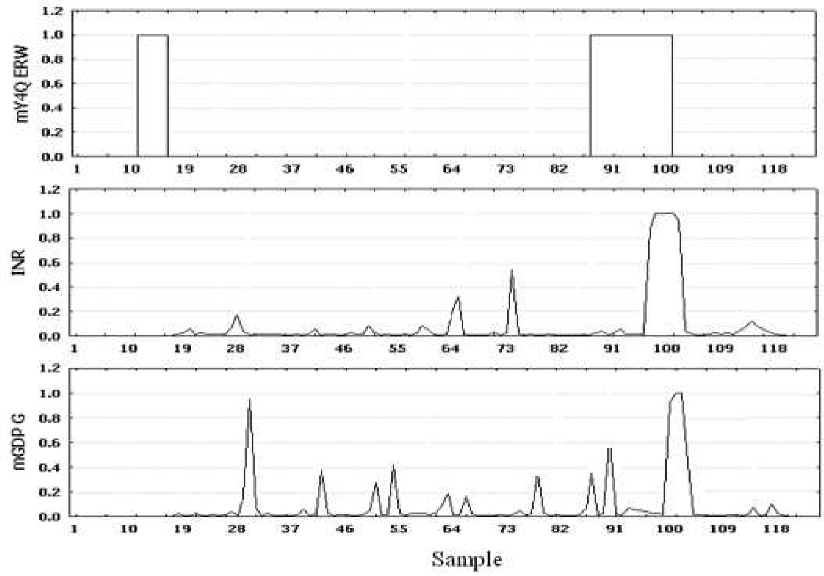

Figure 5.10: Comparison between posterior probabilities of each leading indicator and the forward crisis variable

5.4 Building the Early Warning Systems

After identifying the two sets of leading indicators for BEWS-A and BEWS-B, we can proceed to build the early warning systems accordingly for each country. Take ARGENTINA for example, By weighing with the Spearman correlation coefficients, we create an index variable for BEWS-A with the leading indicators extracted in factor analysis. As Figure 5.11 shows, we create an index variable to compare between the posterior probability and the forward crisis variable. Give an probability threshold of 0.2 for the index variable, as long as the posterior probability of the idex variable is greater than 0.2, then the system BEWS-A will send an alarm signal that financial crises may occur within four months in the future. On the other hand, BEWS-B will be built with the other set of leading indicators in the process of discriminant analysis. Except of BRAZIL, CHINA, POLAND and SLOVENIA

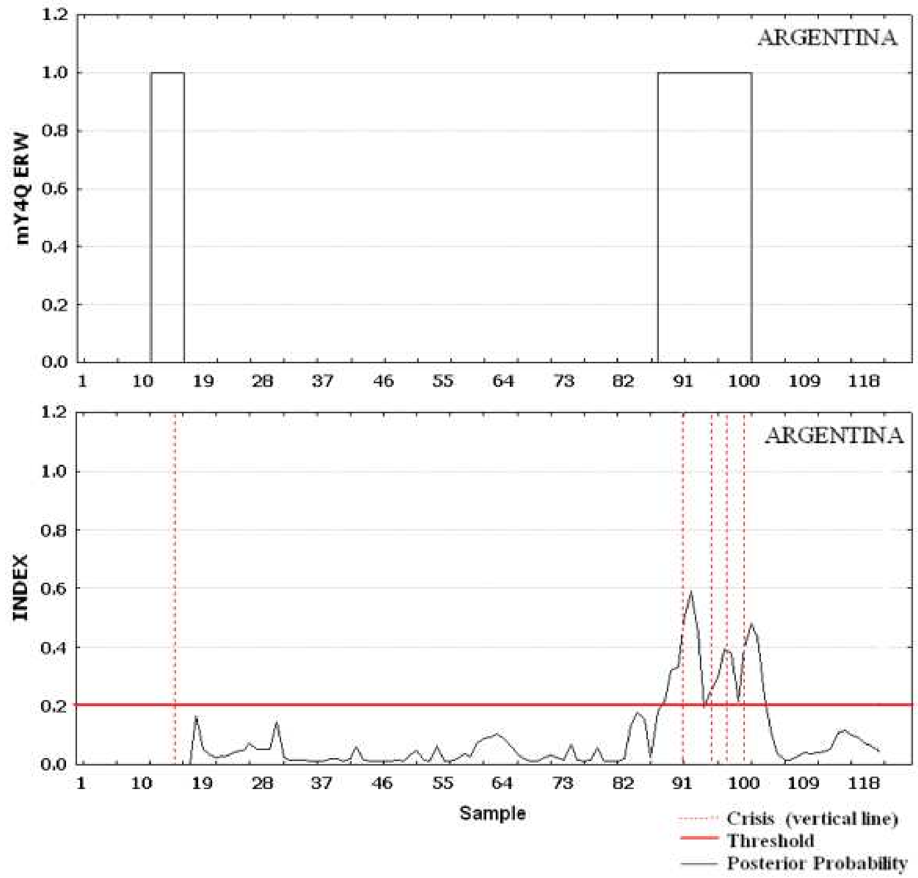

Figure 5.11: Given an probability threshold of 20% for the index variable to identify crisis signals for BEWS-A and compare with the forward crisis variable

which didn't suffer any crises during 1994 to 2003, all the goodness-of-fit analyses of 24 countries for BEWS-A and BEWS-B are summarized in Table 5.7, where $S_t = 0$ represents that the system didn't send an alarm signal at time t, $S_t = 1$ represents that the system sent an alarm signal at time t, $Y_t = 1$ represents that a crisis truly occurred within the next four months at time t, and $Y_t = 0$ represents that no crisis occurred within the next four months at time t. We compare the the two systems in terms of the five goodness-of-fit criteria, that are probability of observations correctly called, probability of

Table 5.7: Simulation results for BEWS-A and BEWS-B

BEWS-A:

	$S_t = 0$	$S_t = 1$	Total
$Y_t = 0$	2483	62	2545
$Y_t = 1$	55	280	335
Total	2538	342	2880

Prob. of obs. correctly called:	95.9
Prob. of crises correctly called:	83.6
Prob. of false alarms:	18.1
Prob. of crisis given an alarm:	81.9
Prob. of crisis given no alarm:	2.2

BEWS-B:

	$S_t = 0$	$S_t = 1$	Total
$Y_t = 0$	2499	46	2545
$Y_t = 1$	78	257	335
Total	2577	303	2880

Prob. of obs. correctly called:	95.7
Prob. of crises correctly called:	76.7
Prob. of false alarms:	15.2
Prob. of crisis given an alarm:	84.8
Prob. of crisis given no alarm:	3.0

crises correctly called, probability of false alarms of total alarms, probability of crisis given an alarm, and probability of crisis given no alarm. Table 5.8 to 5.9 provide a comparison of the goodness-od-fit of our systems with those models of previous studies on the subject. Our systems perform better in each of the five goodness-of-fit criteria than the Pooled logit BF model (on the left of Table 5.8), the IMF-DCSD model (on the right of Table 5.8), the Kaminsky-Lizondo-Reinhart model (on the left of Table 5.9) and the Goldman-Sachs model (on the right of Table 5.9).

The system BEWS-A correctly calls the highest ratio of observations (95.9%) and of crises months (83.6%), and the system BEWS-B gives the fewest false alarms of any of the model above-mentioned. Besides, the conditional probabilities of having a crisis if an alarm occurred are more than 80% for both BEWS-A and BEWS-B, which are much higher than any of the other models. The better performance of our systems means that the leading indicators we have been using are more reliable, and that the country sample and time period are more appropriate.

Table 5.8: Pooled logit BF model and IMF-DCSD model

	$S_t = 0$	$S_t = 1$	Total
$Y_t = 0$	1140	164	1304
$Y_t = 1$	82	164	246
Total	1222	328	1550

Prob. of obs. correctly called: 84.1
Prob. of crises correctly called: 66.7
Prob. of false alarms: 50.0
Prob. of crisis given an alarm: 50.0
Prob. of crisis given no alarm: 6.7

	$S_t = 0$	$S_t = 1$	Total
$Y_t = 0$	1965	525	2490
$Y_t = 1$	167	311	478
Total	2132	836	2968

Prob. of obs. correctly called: 76.7
Prob. of crises correctly called: 65.1
Prob. of false alarms: 62.8
Prob. of crisis given an alarm: 37.2
Prob. of crisis given no alarm: 7.8

Table 5.9: IMF-KLR model and Goldman-Sachs model

	$S_t = 0$	$S_t = 1$	Total
$Y_t = 0$	1834	704	2538
$Y_t = 1$	200	298	498
Total	2034	1002	3036

Prob. of obs. correctly called: 70.2
Prob. of crises correctly called: 59.8
Prob. of false alarms: 70.3
Prob. of crisis given an alarm: 29.7
Prob. of crisis given no alarm: 9.8

	$S_t = 0$	$S_t = 1$	Total
$Y_t = 0$	543	279	822
$Y_t = 1$	50	98	148
Total	593	377	970

Prob. of obs. correctly called: 66.1
Prob. of crises correctly called: 66.2
Prob. of false alarms: 74.0
Prob. of crisis given an alarm: 26.0
Prob. of crisis given no alarm: 8.4

Chapter 6

Conclusions

In this thesis, two new Early Warning Systems (EWS) for predicting financial crises had been developed. The main difference to existing EWS models and the intended contribution of this thesis focused on three areas. First, it proposed a systematic framework by combining Bayesian Theorem with four kinds of feature selection methods, including F statistics, Spearman correlation, factor analysis, and discriminant analysis, to identify two different sets of leading indicators. The second area is that our EWS offered each of the leading indicators a weighting to create an index variable. By monitoring the posterior probability of the index variable, the early warning signals will be determined whether to send out or not. And third, we found out different leading indicators for different countries in terms of their own characteristics.

From a policy perspective, developing EWS that help to reliably anticipate financial crises could become a more important tool for policy-makers in the future. Many financial crises over the past few decades had caused damage to social security, economics and development of industries. Developing reliable EWS there can be of substantial value by allowing the policy-makers to obtain clear signals *when* and *how* to take pre-emptive measures in order to mitigate or even prevent financial turmoil. It should be stressed that EWS can not replace the sound judgment of the policy-maker to guide policy, but it can play an important role as a neutral and objective measure of vulnerability.

This thesis showed that detecting regime shifts of financial variables and estimating the degree of similarity between the obtained posterior probability and a forward crisis variable can indeed identify more reliable leading indicators and improve the performance of EWS substantially. Besides, applying factor analysis and discriminant analysis to distinguish different classes of variables avoids including the leading indicators with the same characteristic in a system. It would be beneficial to keep the predictive stability for a system. From the simulation results of section 5.4, it is fair to say that both of the systems developed in this thesis performed much better than any EWS models of previous studies on the subject in terms of predictive power. However, because the collected data of financial variables and crisis variable in this thesis only have 120 samples, it is insufficient for making cross validation in the process of parameter optimization. It may cause an issue of overfitting for the simulation results. In addition, because we don't have enough data to regard as testing samples, we only can provide simulation results, instead of experiment results.

In conclusion, the simulation results accords with our expectancy. But it should be emphasized that the EWS developed in this thesis doesn't constitute the final step towards a comprehensive EWS of financial crises. Future research on EWS may focus on the following points:

1) Considering an overlap between pre and post windows or the magnitude of the regime shifts is helpful to eliminate false detection. Besides,

a further method is to combine the posterior probability calculated through different methods.

2) Regime shifts could be classified into three kinds of types, including the smooth regime shifts, the abrupt regime shifts and the discontinuous regime shifts [26]. It is possible to use different sets of window definition for classification of different types of regime shifts so as to reduce classification delay.

3) Considering that a post-crisis bias will sometimes appear in an EWS. If we fail to distinguish between pre-crisis and post-crisis periods, it may bring about a bias in the estimation results. So, suggesting a new method to solve this problem could be a further step towards developing an EWS that is more powerful in terms of prediction.

4) Developing a framework that allows the policy-makers to design the features of their EWS according to their preferences and degree of risk-aversion. It includes adding dynamic components to EWS, or providing choices of the timing and the length of different regimes.

References

[1] M. Bussiere and M. Fratzscher, *Towards a New Early Warning Systen of Financial Crises*. European Central Bank, Working Paper Series NO. 145, (May 2002).

[2] G. L. Kaminsky and C. M. Reinhart, 1998, *Financial Crises in Asia and Latin America: Then and Now*. American Economic Review, Papers and Proceedings; 88(2), pp. 444-48, (May 1998).

[3] A. Berg and C. Pattillo, 1999b, Predicting Currency Crises: The Indicators Approach and an Alternative, *Journal of International Money and Finance*. 18(4), pp. 561-86, (August 1999).

[4] G. L. Kaminsky, S. Lizondo and C. M. Reinhart, *Leading Indicators of Currency Crises*. IMF Staff Papers, 45, NO. 1, pp. 1-48 (1998).

[5] G. S. Maddala, *Limited-dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs, Cambridge University Press, (1989).

[6] J. F. Ducre-Robitaille, L. A. Vincent and G. Boulet, *Comparison of Techniques for Detection of Discontinuities in Temperature Series*. International Journal of Climatology, 1087-1101, (2003).

[7] E. S. Epstein, *Detecting Climate Change*. J. Appl. Meteorol. 21, 1172, (1982).

[8] T. R. Karl and C. N. Williams, *An Approach to Adjusting Climatological Time Series for Discontinuous Inhomogeneities: Journal of Climate and Applied Meteorology*. 26, 1744-1763, (1987).

[9] N. J. Mantua, *Methods for Detecting Regime Shifts in Large Marine Ecosystems: a Review with Approaches Applied to North Pacific Data*. Progress in Oceanography, 60, 165-182, (2004).

[10] G. L. Kaminsky and C. M. Reinhart, *The Twin Crises: The Causes of Banking and Balance-of-payments Problems*. International Finance Discussion Papers 544, Board of Governors of the Federal Reserve System (U.S.), (1996).

[11] S. N. Rodionov, *A Brief Overview of the Regime Shift Detection Methods*. Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, WA 98195, (2005).

[12] J. Wang, and Q. P. He, *A Pattern Matching Approach for Fast Disturbance Detection and Classification Using Bayesian Statistics*. AMD (Sep. 2005).

[13] H. H. Yue and M. Tomoyasu, *Weighted Principal Component Analysis and its Applications to Improve FDC Performance*. Tokyo Electron AT Ltd, Nirasaki, Yamanashi, Japan, (2004).

[14] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers, London, (1998).

[15] Janet, E. A. Forrester, Ph. D. Fall, *CMH110 Epidemiology and Biostatistics*. (2004/2005).

[16] K. A. Brownlee. *Statistical Theory and Methodology in Science and Engineering*. JohnWiley, New York. (1960).

[17] J. F. Hair, R. E. Anderson, R. L. Tatham, and W. C. Black, *Multivariate Data Analysis (3rd ed.)*. New York, Macmillan, (1992).

[18] J. Kim and C. W. Mueller, *Introduction to Factor Analysis: What It Is and How to Do It*. Newbury Park, CA: Sage Publications (1978).

[19] Bryant and Yarnold, *Principal Components Analysis and Exploratory and Confirmatory Factor Analysis*. In Grimm and Yarnold, Reading and understanding Multivariate Analysis. American Psychological Association Books (1995).

[20] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. (2001).

[21] J. O. Kim and C. W. Mueller, *Factor Analysis: Statistical Methods and Practical Issues*. Thousand Oaks, CA: Sage Publications, (1978).

[22] L. L. Thurtstone, *Multiple-Factor Analysis*. Chicago: University of Chicago Press, (1947).

[23] G. H. Dunteman, *Principal Components Analysis*. Thousand Oaks, CA: Sage Publications, (1989).

[24] C. E. Lance, M. M. Butts, and L. C. Michels, *The sources of four commonly reported cutoff criteria: What did they really say? Organizational*

Research Methods. 9(2): 202-220. Discusses Kaiser and other criteria for selecting number of factors, (2006).

[25] P. A. Lachenbruch, *Discriminant analysis*. NY: Hafner, (1975).

[26] *Collection of Definitions of Regime Shift*, `http://biocore.ecolab.bas.bg/events/past/unesco-ws/ws-workgr/regshift.htm`

Appendix A

A Summary of Collected Data

There are various types of financial crises: currency crises, banking crises, sovereign debt crises, private sector debt crises, and equity market crises. Currency crises are believed to be the most important type among them because they often coincide or occur in succession with other types of financial crises [10]. For this reason, our EWS model in this thesis focuses mainly on currency crises as well as most EWS models in the literature. We proceed to define a currency crisis by employing a variable named exchange market pressure ($EMP_{i,t}$). The $EMP_{i,t}$ variable for each country i and period t is defined as

$$\begin{aligned} EMP_{i,t} = \omega_{RER} \left(\frac{RER_{i,t} - RER_{i,t-1}}{RER_{i,t-1}} \right) + \omega_r \left(r_{i,t} - r_{i,t-1} \right) \\ - \omega_{res} \left(\frac{res_{i,t} - res_{i,t-1}}{res_{i,t-1}} \right). \end{aligned} \tag{A.1}$$

$EMP_{i,t}$ is a weighted average of the change of the real effective exchange rate (RER), the change in the interest rate (r), and the change in foreign exchange reserves (res). ω_{RER}, ω_r, and ω_{res} are the relative precision of each variable for all countries over the full sample period 1994-2003.

The rationale for using $EMP_{i,t}$ is expounded as follows. If investors consider there are some underlying vulnerable economic factors attacking a currency of a specific country, the government essentially has two options to deal with this kind of situation. One is to abstain from defending the currency either by abandoning a fixed exchange rate regime or by avoiding intervening in foreign exchange markets, and to let the currency devalue. The other is to defend the currency regime by raising interest rates and running down foreign exchange reserves. We can give consideration to both sides with $EMP_{i,t}$.

The next step is to define a currency crisis ($CC_{i,t}$) which is given by

$$CC_{i,t} = \begin{cases} 1 & \text{if } EMP_{i,t} > \overline{EMP_i} + 2SD(EMP_i), \\ 0 & \text{if otherwise,} \end{cases} \tag{A.2}$$

as the event when $EMP_{i,t}$ is two standard deviations (SD) or more above its country average EMP_i [1]. Besides, the thesis uses monthly data over period from 1994 to 2003 for a sample of 28 countries and each country has 48 different financial variables. These 48 variables concerned with external competitiveness, external exposure, domestic real and public sector, domestic financial sector, global factors and the contagion through trade channel, are expected to be able to express the situation that crises will occur in the future. And we will take them as the training data in our model.