Reconstruction of World Bank Classification of Countries and Moody's Rating System

Nima Mirzaei

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the Degree of

Doctor of Philosophy
in
Industrial Engineering

Eastern Mediterranean University
Jun 2013
Gazimağusa, North Cyprus

ABSTRACT

This thesis has two main objectives. The first objective is to analyze whether the classification of countries provided by the World Bank (WB) can be reconstructed with a linear and/or integer-programming model known as Multi-Group Hierarchical Discrimination method, using only data published by the WB. The model's parameters were determined for a collection of 44 countries, and the model was verified using another 39 countries. Moreover, the study examines the relative importance of factors in classification of countries.

The second purpose of this study is applying Logical Analysis of Data for country risk rating to provide an approximate rating method. The employed data is available in World Bank and International Monetary Fund and the results are compared with Moody's rating scale on year 2010. The country risk rating model was established for a collection of 71 countries, and the model was verified using another 34 countries. Furthermore, the study examines the relative importance of economical, environmental, educational, and infrastructure criteria in determining countries risk rating.

Keywords: Multi-Group Hierarchical Discrimination, Classification of countries, Country Risk Rating, Logical Analysis of Data

ÖZ

Bu tezin iki temel amacı bulunmaktadır. Birinci amacı; Dünya Bankası (DB) tarafından sağlanan ülkeler sınıflandırmasının, sadece DB tarafından yayınlanan verilerle Çoklu Grup Hiyerarşik Ayrıştırma yöntemi olarak bilinen doğrusal ve/veya tamsayı programlamla ile tekrar yapılandırılmasını sağlamaktır. Model parametreleri 44 ülkeden toplamıştr ve 39 başka ülke üzerinden model doğrulandırılması gerçekleştirilmiştir. Ayrıca, bu çalışma ülke sınıflandırılmasında kullanılan faktörlerin göreli etkenlerini incelemektedir.

Bu çalışmanın ikinci amacı ise ülke risk derecelendirilmesi için Mantıksal Veri Analizi kullanarak yaklaşık derecelendirme yöntemi elde etmektir. Kullanılan veri Dünya Bankası'nda ve Uluslarası Para Fonu'nda bulunmaktadır ve elde edilen sonuçlar 2010 yılı için Moody's derecelendirme ölçeğiyle kıyaslanmıştır. Ülke risk derecelendirme modeli 71 ülkenin toplanmasıyla oluşturulmuştur ve geliştirilen model 34 başka ülke kullanılarak doğrulanmıştır. Bundan başka, bu çalışma ekonomik, çevresel, eğitimsel ve altyapısal kriterlerin göreli önemini inceleyerek, ülkelerin risk değerlendirilmesinin oluşturulmasını içermektedir.

Anahtar Kelimeler: Çoklu Grup Hiyerarşik Ayrıştırma, Ülke Sınıflandırılması, Ülke Risk Derecelendirmesi, Mantıksal Veri Analizi

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Dr. Béla Vizvári for the continuous support of my PhD study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study. Honestly, he is the salt of the earth.

Besides my supervisor, I would like to thank the rest of my thesis committee: Assoc. Prof. Dr. Oya Ekin Karaşan, Assoc. Prof. Dr. Tonguç Ünlüyurt, Asst. Prof. Dr. Sahand Daneshvar, for their encouragement, insightful comments, and hard questions. I would like to thank Asst. Prof. Dr. Orhan Korhan for Turkish translation of my thesis abstract.

My sincere thanks also goes to the department chair Asst. Prof. Dr. Gökhan İzbırak for their support, and help. I am thankful that in the midst of all his activities. Also I thank all faulty members and my friends who helped me during last five years for research and work.

Last but not the least, I would like to thank my gorgeous wife for unconditional support and my parents for giving birth to me at the first place and supporting me spiritually throughout my life.

To My

Beloved Wife *and* ***Dear Family***

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

In the last decades, country risk rating also known as sovereign risk rating is a popular topic in the fields of economics, operations research, and statistics. In the past, only some of well-known organizations or agencies such as Standard & Poor (S&P), Moody's and Fitch published reports which in general include credit and non-credit rating related information.

Credit rating examples are country risk rating, bank deposit rating, and insurance financial strength rating. Non-credit rating involves different aspects of risk such as investment quality rating or market risk rating. Generally, country risk rating reports contain three separate lists, which are long term obligation ratings, medium note ratings, and short term ratings. These lists which are published yearly, semiannually, and/or quarterly, provide with information about future of country default risk.

The published reports affect economical and political future of countries in many different ways. Many investors, companies, financial institutions, and banks make decisions based on these reports for potential investment or lending money to a specific country.

This research developed mathematical methods and employed optimization techniques to propose models for classification of countries and rate country risk default. The aim of this thesis is two folds: classification of countries, and reconstruction of country risk rating.

The first part of this study which is described in Chapter 2, includes literature review on the subject of classification of countries and country risk rating. Several methodologies and techniques that appear to be significant in country classification and rating are described in this chapter. Also some models which are applied in country classification and rating are presented here.

Chapter 3 contains information about data collection and filtering. The input data in this research was collected from a number of databases such as World Bank (WB) and International Monetary Fund (IMF). The data that utilized in this research were collected in two stages. First a large set of data is collected and then filtered according to some specific conditions, and is denoted as training set.

The training set is employed to modify and calibrate the mathematical model which is used in this study. After that, a set of new data called the test set was collected. The test data and training data are completely different and the test data is used for verification of the proposed model. The characteristic of criteria (indicators) has differ as economical, environmental, educational, and infrastructure. The dataset includes information about numerous criteria that a country can be evaluated accordingly. For example, economical criteria are important in the field of microeconomics. Those factors that have an effect directly or indirectly on unemployment rate, inflation rate, and productivity growth were nominated for this area.

In the classification of countries for each country criteria the averages of observations from 1990-2008 were used for input. However, the input data for country risk rating is only from 2010. Furthermore, the choice of the countries is dictated by data availability. More information and descriptions about data collection is discussed in Chapter 3.

After finalizing the dataset, reconstruction of classification of World Bank was constructed using mathematical models as detailed in the Chapter 5. Classification of World Bank (which is based on Gross National Income per Capita) was reconstructed by a methodology called Multi-Criteria Decision Aid (MCDA) and Multi-Group Hierarchical Discrimination (MHDIS). The model which was proposed by Doumpos and Zopunidis (2001) is modified and improved in order to use for classifying countries into four classes similar to the work which is done by World Bank.

Meanwhile, the most important criteria (indicators) that have the main affect in the classification of countries were identified and finally the new classification of the developed model is compared with the previous classification of World Bank. Up to this point, the aim of this research is accomplished, and the developed model is able to classify a country based on specific indicators in different periods of time. LINGO 12 package is used to solve the developed model in the classification of countries.

Subsequently, in Chapters 5 and 6we go one step further and start to rate the countries by using Logical Analysis of Data (LAD). The rating scale is similar to Moody's rating scale and it is used for evaluation and comparison of new scaling system for countries. The Moody's sovereign credit rating scales and rating procedures were described in detail. Similar to the classification procedure, in country risk rating process, essential indicators were identified and afterward a test set is used to demonstrate the accuracy of the model. The proposed model is not an exact or perfect prototype for country risk rating, b ut it is a valuable model compared to other country risk rating model. Its advantages are that, it is easy to calculate the rating of any country suggested by the model and it almost perfectly reconstructs Moody's rating even on the test set. The model can be used by a person without having any serious mathematical background. Similar to previous chapter, LINGO 12.0 package and Xpress optimization suit are employed to solve the developed mod el.

In addition to those three sections which cover in the main body of this research, Chapter 7 describes some efficient mathematical and statistical methods that are used in this study for classification and rating countries. In particular, it is explained why those mathematical and statistical methods are impotent or incompetent to classify or rate countries. For example, it is explained why it is not possible to classify or rate countries by using a linear regression model

Finally, the last chapter makes a summary of conclusions and provides some directions for future study for country risk rating methodologies and indicators (factors) which contributes as input for rating or classifying country.

Chapter 2

LITERATURE REVIEW

This chapter consists of two main sections reviewing the literature: classification of country and sovereign risk rating. In the first section, some clustering and classification techniques are introduced. In classification or clustering data, the aim is to identify a group or number of groups of similar objects which have the same aspect. In this section some of the popular methods for classification and clustering are discussed. In the second section, we discuss country risk rating models that are valuable in literature and some efficient factors utilized in rating model are introduced.

2.1 Clustering and Classification Techniques

2.1.1 Clustering

One of the helpful tasks in data mining is clustering, which attempts to identify interesting patterns and distribution in the set of data (Halkidi *et al.,* 2001). In other words, separating a set of data points into groups (clusters) having similar properties is called clustering (Guha *et al.,* 1998). In a clustering technique, there are no predefined classes at the beginning of analysis and they are defined based on the nature of dataset or criterion of clustering. There are many clustering algorithms which are proposed in the literature for clustering data. Some of those prominent algorithms are K-mean clustering, partitional clustering, hierarchical clustering, density-based clustering, and grid-based clustering (Jain *et al.*, 1999). Halkidi *et al.* (2001) classified clustering algorithms according to:

- Type of input data,
- Similarity between data point and clustering criterion,
- Clustering techniques.

As a result, different methods can be used in different situations. For example K-mean method can be used if the input data are numerical and it is desirable to apply partitional clustering algorithm, or K-mode can be used if the input data is categorical (Halkidi *et al.*, 2001). In another example, if it is desirable to apply hierarchical clustering algorithm and input data is numerical, then BIRCH or CURE are suitable methods for clustering (Halkidi *et al.*, 2001).

2.1.2 Classification

Unlike clustering, in classification method each data in dataset is assigned to predefined classes or groups (Fayyad *et al.*, 1996). It is better to say that clustering is the initial step for classifying data and it can be used in classification procedure (Halkidi *et al.*, 2001). Based on data type and criterion, different techniques can be used for classifying data. Some of these methods are statistics-based, and some other are based on mathematical models. In this study MGHDIS method is applied to classify countries into classes. Multi Group Hierarchical Discrimination (MGHDIS) is classifying alternatives into predefined classes based on hierarchical discrimination method (Zopounidis and Doumpos 2000). The method generates a set of utility functions and those functions determine which alternative belongs to which class based on defined criteria (Doumpos and Zopounidis 2001). Multi Group Hierarchical Discrimination methods and techniques are discussed in detail in Chapter 4.

2.1.3 Logical Analysis of Data (LAD)

Logical analysis of data is a methodology that is employed for finding out hidden structural information in data sets, and it is developed by using Boolean functions to evaluate binary (0,1) data (Boros *et al.*, 1997). A numerical data set can be classified by using logical analysis of data through a process that is called "binarization". In this method each observation in a numerical data set is transformed into a binary vector. Boros *et al.*, (1997) utilized a binarization process to study the combinatorial optimization problems and to minimize the number of binary variables. They developed polynomial time algorithms for those problems. Logical analysis of data is applicable in the wide area of research such as productivity (Hammer *et al.*, 1996), oil exploration (Boros *et al.*, 2000), economic analysis (Hammer *et al.*, 2007), and many other fields.

Hammer *et al.*, (2007) used logical analysis of data to develop a consistent and stable country risk rating model which can be competitive with or closely approximate the model provided by major rating agency (Standard & Poor, Moody's, and Institutional Investors). Their proposed model has high accuracy and it has 95.5% correlation level between predicted and actual rating (with the Standard & Poor rating and high-quality correlation with Institutional Investors and Moody's). One of the significant advantages of their study is non-recursiveness of the proposed model and then it can be applied to not yet rated countries. Furthermore, both economic and political variables are utilized in the model.

2.2 Understanding Country Risk and Country Risk Rating

There are many valuable works in global finance and economics related to country risk rating in the literature. The purpose of country risk rating is that, instead of comparing from a diversity of information about a country, a single metric rate to evaluate or compare the country with others is used (Hammer *et al.*, 2007). Consequently, the rating system was constructed to make it easy and understandable for categorizing countries according to their risk level. In the financial market credit rating agencies play an important role in rating and updating information about country risk default (Cantor and Packer, 1994). Some leading rating agencies such as Standard & Poor, Moody's, Fitch and *etc*, develop standard rating scales to rate the countries and many investors, big companies, and bankers refer those ratings before investment in a country. On the other hand, after some failure in the past by these agencies in anticipating a number of crises such as Asian crisis (1997-1999), Russia and Brazil crisis (1998), and Argentina (2001), they have been criticized by some specialists (Altman, Rijken, 2004). For instance, Ferri *et al.*,(1999) provided evidence which proves that credit rating agencies failed for anticipating East Asia crisis in 1999 and become excessively conservative after that. Furthermore, they start downgrading some of East Asia countries more than what they deserve and it worsens the situation for those countries. For that reason, many researches in this field started to develop models for country risk rating at the beginning of 1990's. Some examples are Cosset and Roy (1991), Cosset *et al.,* (1992), Oral *et al.,* (1992) Lee (1993), Dahl *et al.,* (1993), and Moon and Stotsky (1993). At that time many researchers tried to develop mathematical, statistical, stochastic, and/or probabilistic models and use them in determining sovereign risk. But their models were not accurate enough or not applicable in some area.

2.2.1 Importance of Variable Selection in Country Risk Rating

One of the important issues in the country risk rating is the selection of variables. There are many variables (qualitative or quantitative) that might have an effect in country risk rating, but some of them have played crucial role in the rating procedure. In a study by Haque *et al.,* (1998), the relative importance of political and economic variables in country risk rating was described. In this study a list of variables which are used by commercial rating agencies such as Institutional Investor, Euromoney, and Economist Intelligence Unit was introduced. One of the main results in this study shows that the political variables do not have much effect on the rating of a country, therefore they focused on economical variables and the objective is to remove political variables from the regression model. In an earlier study, the importance of economic determinants of country risk rating was examined, and it was shown that economic fundamentals were key factors in the rating system (Haque *et al.*, 1996). As a result, excluding political variables from the model does not severely bias the factor estimates for the economic variables. However, Brewer and Rivoli (1990) claim that both economic and political variables affect country risk rating. Until now, because of lack of clarity, still there are controversy and disagreement between experts to decide that which one of the economic or political variables is more important. However, in many studies, researchers provided evidence that political factors can influence country rating in different ways.

Edward (1984), Cantor and Packer (1996) discuss a similar topic about variables and their relations in sovereign credit rating. Besides, Alfonso (2003) employed a large set of explanatory variables, and found correlation of several variables with key variable and then identified that six variables appear to be important in deciding about a sovereign credit rating. Those six variables are Gross Domestic Product (GDP) per capita, external debt, level of economic development, default history, real growth rate, and inflation rate. In another study, Cantor and Packer (1996) mentioned that per capita income, GDP Growth, inflation, external debt, level of economic development and default history are the most important factors in country risk rating. Moreover, they could not find any systematic relationship between the rating and current deficit, and this may have happened because of the endogeneity of financial strategy and international capital flows. It is clear that, mostly economic variables, or it is better to say macroeconomic factors have major influence in country risk rating.

Haque *et al.*, (1997) shows that rating agencies have different policies and use different criteria for credit rating. Table 2.1 that is developed by Haque *et al.*,(1997) gives details about criteria for assessing country risk of three rating agencies Institutional Investor, Euromoney, and Economist Intelligence Unit. As it is illustrated in the Table 2.1 each agency has a specific strategy for selecting indicators, and even the same indictor might have a different weight in the rating system of agencies.

Table 2.1. Criteria for Assessing Country Risk (Haque et al., 1997)

Rating agency	Criteria for ratings
Institutional investor	Information provided by 75–100 leading banks that grade each country on a scale of 0–100, with 100 representing least chance of default. Individual responses are weighted using a formula that gives more importance to responses from banks with greater worldwide exposure. Criteria used by the individual banks are not specified.
Euromoney	Assessment based on three main indicators; **Analytical indicators** (40 percent): ▪ Political risk (15 percent) ▪ Economic risk (10 percent) ▪ Economic indicators (15 percent) (debt service/exports, external debt/GNP, balance of payments/GNP) **Credit indicators** (20 percent): ▪ Payment record (15 percent) ▪ Rescheduling (5 percent) **Market indicators** (40 percent): ▪ Access to bond markets (15 percent) ▪ Selldown on short-term paper (10 percent) ▪ Access to discount available on forfeiting (15 percent)
Economist Intelligent Unit	Medium-term lending risk (45 percent): Total external debt/GDP, total debt-service ratio, interest-payment ratio, current account/GDP, savings/investment ratio, and arrears on international bank loans, recourse to IMF credit, and the degree of reliance on a single export. Political and policy risk (40 percent) Short-term trade risk (15 percent)

As Haque *et al.,* (1997) mentioned in their paper, some agencies such as Euromoney are more willing to assign higher rating to Asia and Europe compared to Institutional Investor, plus Euromoney and Economist Intelligence Unit are more optimistic compared to Institutional Investor and it is confirmed that regional considerations have a strong influence on assessing risk rating. Altman and Saunders (1998) indicated that agencies slowly but surely will downgrade the rating of a country which is in crisis, although they do not want to harm the country, which is also their client and they do not want damage their relationship with that country. As it is clears from the country risk rating literature, most of the rating agencies possibly will consider other criteria except than economic and political factors while rating the country. Even the application of subjective elements cannot be excluded.

2.2.2 Country Risk Rating Methodologies and Techniques

There are many techniques and methodologies which are used in country risk rating, however only some of these techniques are able to generate a useful outcome. Some of the studies use regression analysis to set up a model. For instance, Alesina *et al.,* (1992) utilized some simple regression methods for assessing default risk on government debt in Organization for Economic Cooperation and Development (OECD) countries. They have accomplished the study by collecting a sample of 12 OECD countries over the period 1974-89. They tried to measure default risk on government debt by the ratio of the public interest rate over the private interest rate or by the differential between the two of them. Cantor and Packer (1996) employ multi regression model to quantify the correlation between rating and their determinants. Furthermore, they explored how dollar bond spreads responding to rating announcement of agencies. Haque *et al.,* (1998) did the similar analysis.

Alfonso (2003) applied the combination of linear, logistic, and exponential transformation models to rate the sovereign debt and compare the result with both Standards & Poor's and Moody's rating scale in June 2001. The following Equation 2.1 was generated by linear/nonlinear transformation.

$$\text{RATING}_i = \alpha_0 + \alpha_1\text{GDPPC}_i + \alpha_2\text{INFL}_i + \alpha_3\text{GDPGR}_i + \alpha_4\text{DEVELOP}_i + \alpha_5\text{DEBTX}_i + \alpha_6\text{DEF}_i + \alpha_7\text{BUDGET}_i \quad (2.1)$$

In the (2.1) α_i's (α_1-α_7) are coefficients of each variable, and those variables are per capita GDP, inflation, GDP real growth, developed country indicators, external debt to-export ratio, default indicator, and budget balance, respectively. At the end, results of estimation prove that logistic transformation gives better assessment compared to other methods (linear) by concerning the collected sample.

2.2.3 Properties of Probabilistic Models in the Rating Procedure

In some other methodologies authors employ probabilistic and stochastic models for sovereign credit ratings. Lando and Skødeberg (2002) utilize continuous time method to calculate transition matrices. They estimated discrete time transition matrices for using in Markov chain method. Hu *et al.,* (2002) employed a model which is called "probit Model". Hu *et al.,* (2002) used ordered probit model to generate rating transition matrices for countries which is used in credit portfolio modeling. Rating transition matrices are generally used in determining future loss distribution for pricing purpose. They move toward an empirical or quasi-Bayesian procedure of combining information, and selected those variables which were highly statistically significant. Rating matrixes were created by Hu *et al.,* (2002) for both Standard & Poor and Moody's based on rating categories. On the other hand, sovereign rating matrix shows an estimation of the changes (upgrading, downgrading or not change) of the scale of countries in the future.

Lando and Skødeberg (2002) applied hidden Markov chain to estimate on year transition probability matrix, and the study were followed by Christensen *et al.,* (2004). Both studies proved the importance of the Markov assumption in the method of estimating transition matrices. Based on estimated transition probability matrix, it is obvious to verify that a newly downgraded or "excited" state has higher tendency to be downgraded in the following rating processes, compared to those non-excited states (Christensen *et al.*, 2004).

Sohn and Choi (2006) used Data Envelopment Analysis (DEA) and Decision Making Unit (DMU) for estimating efficiency of a new technology group for rating system. The analysis is based on company instead of country, and they incorporate random effect logistic model and DEA to show the significance and correlation between and within group of factors in rating system. Afterward, they employed similar methodology and utilized random effects multinomial regression model to compute credit rating transition probability matrices. They claim that random effect model affords a transition matrix that is less diagonally dominant. In another world, when a matrix is to be diagonally dominant, it shows that the probability mass is located in the diagonal (Kim and Sohn 2008). The study utilized seven variables. Four of them are rating specific variable and three of them are economical variable, and in the correlation analysis Kim and Sohn (2008) found that discount rate, unemployment rate, and GDP growth have high correlation in the fraction downgrading to upgrading policy. Additionally they realized that, when unemployment and GDP growth rate increase credit rating is desired to upgrade whereas the discount rate decrease.

In the following Chapter 3, we start to discuss data collection and filtration step by step. In other words, we show the data mining processes for classification of country and country risk rating.

Chapter 3

DATA COLLECTION AND VARIABLE SELECTION

In this chapter data collection and the procedure of the selection of variables are discussed. The data in this study was collected from two main databases, which are World Bank Institute (WB) and International Monetary Fund (IMF).

The World Bank is an international financial organization that provides variety of financial services to many countries. The organization involves five agencies which are International Bank for Reconstruction and Development (IBRD), International Development Association (IDA), International Finance Corporation (IFC), Multilateral Investment Guarantee Agency (MIGA), and International Centre for Settlement of Investment Disputes (ICSID). Each of these agencies provides different services to developing or developed countries. The International Monetary Fund (IMF) is an organization of 188 countries, working to promote global monetary cooperation. The organization objectives are secure financial stability, facilitate international trade, promote high employment and sustainable economic growth, and reduce poverty around the world.

3.1 Data Collection

As mentioned before, this thesis has two main objectives: classification of countries and countries risk rating. Consequently, the data in this study is collected based on the necessities in those two problems. For both sections, training set and test set data are collected and analyzed.

3.1.1 Collected Data for Classification of Countries

The data used in the classification of countries ie taken from the World Bank database website. According to the World Bank, the countries under consideration are categorized into four classes by their income levels:

I. High-income economies (class C_4) are mostly European ones and the United States, Canada, Australia, New Zealand, Japan, and Hong Kong.

II. Upper-middle income economies (class C_3) are countries from Europe (e.g., Poland and Hungary), South and Eastern Asia, and Latin America.

III. Lower-middle income economies (class C_2) are Eastern Europe, Asia, Africa, and Latin America.

IV. Low-income economies (class C_1) are mostly in Africa and Asia.

Criteria are selected according to their importance and effect on the countries' economic and political situations. Countries are chosen by considering the data availability of selected criteria for the alternatives (countries), meaning that if an alternative does not have enough information about one or more criteria in the data set, the alternative is eliminated automatically from the data sample. When there are no data related to some criteria, it is not possible to compare alternatives and classify them into the organized classes. The number of classes depends on predefined ranges and can vary, but according to the World Bank, countries are classified into four groups according to their income levels. The filtration steps are as follows.

We selected all the countries and criteria available in the World Bank database as raw information. In total, 241 countries and 43 criteria (political and economic) are available in the database. For each factor, the average of the data from 1990 to 2008 was considered for the analysis. By considering the combination of countries, criteria and specified model, the most important criteria were recognized. In each filtration step, some countries or some criteria were eliminated because of lack of available data. In this study, 44 alternatives (countries) were selected for the analysis, and each country falls in a specified set (***a*** or ***b***) shown in Table 3.1.

Table 3.1. Countries with Their Income Classes and Sets

No.	Country Name	Set /Class	No.	Country Name	Set/Class
1	Argentina	b/C_3	23	Japan	a/C_4
2	Australia	a/C_4	24	Korea, Rep.	a/C_4
3	Austria	a/C_4	25	Luxembourg	a/C_4
4	Bolivia	b/C_2	26	Mexico	b/C_3
5	Belgium	a/C_4	27	Netherlands	a/C_4
6	Brazil	b/C_3	28	New Zealand	a/C_4
7	Bulgaria	b/C_3	29	Norway	a/C_4
8	Canada	a/C_4	30	Oman	a/C_4
9	China	b/C_3	31	Paraguay	b/C_2
10	Colombia	b/C_3	32	Poland	b/C_3
11	Czech Republic	a/C_4	33	Portugal	a/C_4
12	Dominican Republic	b/C_3	34	Russian Federation	b/C_3
13	Ecuador	b/C_2	35	Slovak Republic	a/C_4
14	Denmark	a/C_4	36	Spain	a/C_4
15	Finland	a/C_4	37	Sweden	a/C_4
16	France	a/C_4	38	South Africa	b/C_3
17	Germany	a/C_4	39	Switzerland	a/C_4
18	Hungary	a/C_4	40	Turkey	b/C_3
19	Iceland	a/C_4	41	Uruguay	b/C_3
20	India	b/C_2	42	Venezuela, RB	b/C_3
21	Indonesia	b/C_2	43	United Kingdom	a/C_4
22	Italy	a/C_4	44	United States	a/C_4

This classification constitutes the basis for developing the appropriate country risk assessment model. The classes were divided into two sets. The C_1, C_2 and C_3 classes belong to set ***b***, and C_4 belongs to set ***a***. Using the data available in the World Bank database, we considered 19 criterion (criteria), including both economic and political criteria. Each criterion has three levels in its general form, which are low, medium and high; however, some specified criteria were divided to six levels. The 19 criteria with their levels are shown in Table 3.2.

Table 3.2. Criteria with Their Levels

Evaluation Criteria	Criteria	Levels
g_1	Electric power consumption (kWh per capita)	6
g_2	Energy use (kg of oil equivalent per capita)	3
g_3	Exports of goods and services (% of GDP)	3
g_4	Fertility rate, total (births per woman)	3
g_5	GDP (current US$)	3
g_6	GDP growth (annual %)	6
g_7	GNI per capita, Atlas method (current US$)	3
g_8	GNI per capita, PPP (current international $)	3
g_9	GNI, Atlas method (current US$)	3
g_{10}	GNI, PPP (current international $)	3
g_{11}	Gross capital formation (% of GDP)	6
g_{12}	Imports of goods and services (% of GDP)	3
g_{13}	Inflation, GDP deflator (annual %)	6
g_{14}	Military expenditure (% of GDP)	3
g_{15}	Mobile cellular subscriptions (per 100 people)	6
g_{16}	Net migration	3
g_{17}	Population growth (annual %)	6
g_{18}	Population, total	3
g_{19}	Surface area (sq. km)	6

GDP: Gross Domestic Product **GNI**: Gross National Income **PPP**: Purchase Power Parity

After analyzing the training set by MHDIS method which is discussed in detail in Chapter 5, test set is collected for further study. The main reason of collecting test set is the verification and validation of the developed model in classification of countries. Table 3.3 shows the new 39 countries (test set) that is used for verification of the model. All of the countries, with three exceptions of those having a star symbol, are classified correctly. All of the data which are used in the classification of countries are gathered from WB database. As stated before, the number of countries and variables that are utilized in this study is dictated by data availability. In other words, there are limited numbers of countries in the WB databases. As a result, we have to select those which have adequate information about criteria. As described before, limitation in data availability forced us to select a restricted number of countries and criteria in the data set and training set for both the classification of countries and countries risk rating models.

Table 3.3. Countries of Test Set

No.	Country Name	No.	Country Name	No.	Country Name
1	Bahrain	14	Bulgaria	27	Pakistan
2	Cyprus	15	Cameroon	28	Panama
3	Greece	16	Chile *	29	Philippines *
4	Ireland	17	Costa Rica	30	Romania
5	Israel	18	Côte d'Ivoire	31	South Africa
6	Kuwait	19	Cuba	32	Sudan
7	Saudi Arabia	20	Egypt, Arab Rep.	33	Tajikistan
8	Singapore	21	Georgia	34	Turkmenistan
9	United Arab Emirates	22	Ghana	35	Vietnam
10	Albania	23	Iran, Islamic Rep.	36	Belarus
11	Armenia	24	Jamaica *	37	Bosnia and Herzegovina
12	Azerbaijan	25	Kazakhstan	38	Libya
13	Bangladesh	26	Lebanon	39	Malaysia *

3.1.2 Collected Data for Countries Risk Rating

Similar datasets (training set and test set) are collected for analyzing country risk model. There is only one difference in the sources. The data in training set and test set are collected from two different databases, which are IMF and WB. All of the data collected are based on information of the year 2010.

In the training set, 71 countries and 30 economic factors are used to elaborate the approximate rating method. Countries are chosen according to their availability of information in WB and IMF database, and the factors which have been selected are designed to measure country's performance in economic and other sectors. Besides, we have considered factors such that the theoretical literature has stressed their importance in risk rating of countries (Hammer *et, al.* 1996). The following table presents a list of the factors or indicators collected from WB and IMF databases to develop countries risk rating.

Table 3.4. Economical, Environmental, Educational, and Infrastructure Factors

Indicator and (Unit)	**Index**	**Indicator and (Unit)**	**Index**
General government gross debt (% of GDP)	G2	International migrant stock (% of population)	G28
General government net lending/borrowing (% of GDP)	G3	Land area (sq. km)	G29
General government total expenditure (% of GDP)	G5	Mobile cellular subscriptions (per 100 people)	G30
Gross domestic product based on purchasing-power-parity (PPP) per capita GDP (Current international dollar USD)	G6	Net income (BOP, current US$)	G31
Gross domestic product based on purchasing-power-parity (PPP) valuation of country GDP (Current international dollar USD)	G8	Net migration	G32
Gross domestic product, constant prices (Percentage change)	G10	Population ages 0-14 (% of total)	G33
Unemployment rate (Percent of total labor force)	G15	Population ages 65 and above (% of total)	G34
Burden of customs procedure, WEF (1=extremely inefficient to 7=extremely efficient)	G16	Population density (people per sq. km of land area)	G35
Business extent of disclosure index (0=less disclosure to 10=more disclosure)	G17	Population growth (Annual %)	G36
Cost to import (USD per container)	G20	Population, total	G37
Current account balance (% of GDP)	G21	Secure Internet servers (per 1 million people)	G39
Domestic credit to private sector (% of GDP)	G22	Time required to register property (days)	G41
Ease of doing business index (1=most business-friendly regulations)	G23	Time required to start a business (days)	G42
Export value index (2000 = 100)	G24	Time to export (days)	G43
GDP, PPP (current international $)	G25	Urban population (% of total)	G45

The first 7 indicators (indices: G2, G3, G5, G6, G8, G10, and G15) which are mostly economical factors were collected from IMF database. The rest of indictors (23 indicators) which are mostly environmental, educational, and infrastructure factors were collected from WB database. The names of the countries that are used in the training set for countries risk rating are listed in the Table 3.5.

Table 3.5. Selected Countries for Training Set

No.	Country	No.	Country	No.	Country	No.	Country
1	Austria	19	Chile	37	Mauritius	55	Portugal
2	Canada	20	China	38	Bulgaria	56	Turkey
3	Denmark	21	Saudi Arabia	39	Colombia	57	Egypt, Arab Rep.
4	Finland	22	Czech Republic	40	Croatia	58	Georgia
5	France	23	Estonia	41	Hungary	59	Albania
6	Germany	24	Israel	42	Iceland	60	Dominican Republic
7	Luxembourg	25	Slovenia	43	Latvia	61	Venezuela, RB
8	Netherlands	26	Cyprus	44	Panama	62	Vietnam
9	Norway	27	Poland	45	Peru	63	Bosnia and Herzegovina
10	Singapore	28	Malaysia	46	Azerbaijan	64	Ukraine
11	Sweden	29	South Africa	47	Indonesia	65	Argentina
12	United Kingdom	30	Lithuania	48	Ireland	66	Jamaica
13	United States	31	Mexico	49	Morocco	67	Moldova
14	Belgium	32	Russian Federation	50	Uruguay	68	Nicaragua
15	Hong Kong SAR, China	33	Thailand	51	Armenia	69	Pakistan
16	Italy	34	Tunisia	52	El Salvador	70	Greece
17	Japan	35	Brazil	53	Jordan	71	Ecuador
18	Spain	36	Kazakhstan	54	Philippines		

These 71 countries are rated by Moody's agencies in 2010. The selected countries are analyzed by country risk rating model which is discussed in detail in Chapter 5. Afterward, an additional of 34 countries were used to verify the approximation method. The name of the 34 countries that constitute the test set is listed below.

Table 3.6. Selected Countries for Test Set

No.	Country	No	Country
1	Angola	18	Lebanon
2	Bahrain	19	Malta
3	Bangladesh	20	Mongolia
4	Barbados	21	Montenegro
5	Belarus	22	Namibia
6	Belize	23	New Zealand
7	Bolivia	24	Nicaragua
8	Botswana	25	Oman
9	Cambodia	26	Papua New Guinea
10	Costa Rica	27	Paraguay
11	Cyprus	28	Qatar
12	Fiji Islands	29	Romania
13	Guatemala	30	Slovenia
14	Hong Kong	31	Sri Lanka
15	India	32	Switzerland
16	Korea	33	Trinidad and Tobago
17	Kuwait	34	United Arab Emirates

In the next two chapters methodologies and techniques that are utilized for classification of countries and countries risk rating are discussed in detail.

Chapter 4

LOGICAL ANALYSIS OF DATA

Logical Analysis of Data (LAD) is a classification method (Boros et al., 1997) that can be used when there are only two classes, and the objects are described by the same attribute set. LAD was applied both to separate out the high-income countries and to reconstruct Moody's rating system. One of the key purposes of LAD is to classify new data or observation in a way consistent with past categorizations. The available information consists of an archive of previous observations.

One of the main characteristic of LAD is to create a Boolean function to distinguish observation from one class to another one (Boros et al., 2009).

4.1 Boolean Variables and Function

A Boolean variable is a variable with its only possible values being 0 and 1. A Boolean function is a mapping from a Boolean vector to a Boolean variable. The following are the Boolean function definition by Crama and Hammer (2011):

"*A Boolean function of n variables is a function on B^n into B, where B is the set {0,1}, n is a positive integer, and B^n denoted the n-fold Cartesian product of the set B with itself*."

There are three basic operations related to Boolean function (Crama and Hammer 2011):

- The binary operation Disjunction, with symbol $\vee$ (Boolean OR),

- The binary operation Conjuction, with symbol $\wedge$ (Boolean AND),
- The unary operation Complementation, Negation with symbol $\bar{.}$ (Boolean NOT)

For instance, in equations 4.1 and 4.2 C and D are expression the form of an elementary conjunction and disjunction:

$$C = \bigwedge_{i \in A} x_i \bigwedge_{j \in B} \bar{x}_j \quad , \qquad \text{where } A \cap B = \emptyset \tag{4.1}$$

$$D = \bigvee_{i \in A} x_i \vee \bigvee_{j \in B} \bar{x}_j \quad , \quad \text{where } A \cap B = \emptyset \tag{4.2}$$

In the equation 4.1 and 4.2 x_i and $\bar{x}_j$ stand for finite collection of Boolean variables that belong to disjoint subsets A and B respectively.

4.1.1 Disjunctive Normal Form (DNF)

In the original form of LAD it creates a Boolean function of the Boolean variables which are also created by LAD on the attributes of the objects. It is well-known (Crama and Hammer 2011, Theorem 1.4.) that any Boolean function can be given in the Disjunctive Normal Form (DNF) which is used by LAD, as well in the following equation 4.3:

$$\bigwedge_{k=1}^{m} C_k = \bigvee_{k=1}^{m} \left(\bigwedge_{i \in A_k} x_i \bigwedge_{j \in B_k} \bar{x}_j \right) \tag{4.3}$$

4.2 Real life Examples and Application of LAD

The following examples of LAD in this chapter are country risky rating result of decision tree. Find more information in Chapter 5.

LAD has many applications, for examples a bank wants to have some methods to recognize the customers who may potentially have problems in paying, before granting the loan. Based on historical data LAD can create rules for the bank. There are plenty of other applications. Distinguishing successful and unsuccessful locations of oil drilling can save millions of dollars. Separation of patients who suffer and don't suffer in a certain disease may help to cure them.

LAD assumes that all objects in the two classes are characterized by the same parameter set. In the case of rating the sovereign debts, all parameters of the countries are numerical values like the GDP per capita and not categorical as man/woman. LAD is able to use both types of data; however categorical data are not discussed further on. The way how LAD works is shown on the example of sovereign debt in Chapter 5.

In the following we present some examples which belong to Chapter 6 country risk rating. If the objects have numerical attributes then LAD claims that the objects must have high/low value in a certain attribute. This is the first step to distinguish the element of the two classes. For example if the Aaa countries are to be distinguished from the other countries then

the increase of the GDP is *at least* 0.57 percent

is an example for claiming the value of a parameter to be high. An example for claiming a parameter value to be low is:

general government gross debt is *at most* 99.54 percent of the GDP.

Notice that an object either satisfies the constraint or violates it. These two options mean that the claim is either **true** or **false**, i.e. the sets of objects are mapped into a Boolean variable. Generally, such a constraint is satisfied by the elements of both the positive and negative classes, although the majority belongs to the positive class.

Therefore, LAD collects constraints into groups in ***the second step*** such that all constraints of the group are satisfied by the elements of the positive class only.

For example, the constraints

GDP per capita based on purchasing-power-parity (PPP) is *at least* \$29411

and

the change of GDP measured in constant prices (Percentage) is *at least* 0.31

are satisfied by only countries having Aa3 or better rating. The name of this type of groups of constraints is a *pattern*. If all elements of the positive class satisfy all constraints of a group then it is a *perfect pattern*. If the positive class consists of the countries having rating Aaa or Aa1 then

GDP per capita based on purchasing-power-parity (PPP) is *at least* \$29411

and

the change of GDP measured in constant prices (Percentage) is *at least* 0.31

and

the percentage of the population of age 0 to 14 is *at most* 20.13

is a perfect pattern, that is all countries of rating Aaa and Aa1 satisfy it, however none of the countries having Aa2 or lower rating does it. Notice that a pattern is the conjunction of the Boolean variables which are equivalent to the constraints of the pattern. No perfect pattern exists in general. Therefore LAD collects patterns in ***the third step*** such that the elements of the positive class satisfy at least one of the patterns. Countries having Aa3 or better ratings satisfy at least one of the following three patterns; however none of the countries having A1 or lower rating satisfy any of the three patterns.

Pattern 1.

GDP per capita based on purchasing-power-parity (PPP) is *at least* \$29411

and

the change of GDP measured in constant prices (Percentage) is *at least* 0.31

Pattern 2.

GDP per capita based on purchasing-power-parity (PPP) is *__at least__* $29411

and

the percentage of the population of age 0 to 14 is *__at most__* 20.13

Pattern 3.

Burden of customs procedure, WEF is at least 4.475

and

Gross domestic product based on purchasing-power-parity (PPP) valuation of country GDP is at least 602.3.

A theory of LAD is a DNF formed from the Boolean variables which are contained in the patterns. In LAD methodology it is allowed that a pattern is satisfied the elements of both classes. Then the quantity of a pattern is measured mainly by two quantities. **Prevalence** is the percentage of the positive class which satisfies the pattern. **Homogeneity** is the percentage of the elements of the positive class among all objects which satisfy the pattern. In the case of perfect pattern both prevalence and homogeneity are 100 percent.

In the following chapters 5 and 6 there are real case application of LAD which helps us to classify or rate the countries by considering specific criteria.

Chapter 5

CLASSIFICATION OF COUNTRIES

The chapter objective is to analyze whether the classification of countries provided by the World Bank (WB) can be reconstructed with a linear and/or integer-programming model known as Multi-Group Hierarchical Discrimination, using only data published by the WB.

The WB has a public database containing countries' economic-financial and political criteria. The model's parameters were determined for a collection of 44 countries, and the model was verified using another 39 countries. Only four out of 39 countries were misclassified, which shows the elaborated model's power. Logical Analysis of Data (LAD) also analyzed the problem. The attempt to reconstruct the classification uses 19 criteria.

5.1 Introduction to Classification of Countries

Since the 1990s, financial risk management has become an important subject for operation researchers because it provides important information for the field of financial engineering (John et al., 1997). Operations researchers, financial investigators, statisticians, and econometricians have proposed many practical approaches to measure and assess financial risks.

Financial risk management assists decision makers and financial managers in making effective financial decisions. It allows investors and financial managers to decide where, when, and how to invest their funds. Moreover, global companies are able to make decisions about where to locate their new branches or invest their capital.

5.2 The World Bank Classification Procedure and Criteria

In this section, countries were classified according to the World Bank categorization. For operational and analytical purposes, the World Bank's main criterion for classifying countries is Gross National Income (GNI) per capita. In the past, the World Bank used the Gross National Product (GNP) instead of GNI to classify countries. Based on GNI per capita, each country is categorized into one of four economic classes:

- low income ($995 or less),
- middle income (which subdivided into two classes, lower middle $996-$3,945 and upper middle $3,946-$12,195), and
- high income ($12,196 or more).

In addition to the GNI per capita criterion, two other criteria are utilized to classify countries:

Geographic region: Classifications reported for geographic regions are for low-income and middle-income countries only. Low-income and middle-income countries are sometimes labeled as developing countries. The use of the term is inconvenient, as it does not mean that all the countries in that class are experiencing similar development or that other countries have reached a preferred or final phase of development. It is important to know that classification by income does not necessarily reflect development status.

Lending category: International Development Association (IDA) countries are those that had a per capita income in 2009 of less than $1,165 and lack the financial ability to borrow from the International Bank for Reconstruction and Development (IBRD). IDA loans are deeply concessional or interest-free loans and grants for economic growth to improve programs aimed at boosting living conditions. The World Bank publishes income classifications every year on the 1st of July. These official classifications are fixed during the World Bank's fiscal year, which ends in June. Countries remain in the predefined categories into which they are classified, regardless of any revisions to their income data.

In this study, we use two different methods to classify countries. The first method is the Multi-Group Hierarchical Discrimination (MHDIS) method, which is based on the Multi-Criteria Decision Aid (MCDA). Zopounidis and Doumpos suggested the MHDIS in 2000. The Multi-Group Hierarchical Discrimination (MHDIS) method classifies a set of alternatives to the specified class. A set of additive utility functions is developed by linear and/or mixed integer programming. The alternatives are classified when the value of the utility functions is above or below a certain threshold. For a good summary, see (Zopounnidis and Doumpos, 2002).

The second method is a novel technique in risk management called Logical Analysis of Data (LAD). Logical Analysis of Data is a qualitative method which is able to distinguish two sets of alternatives. LAD generates sets of Boolean constraints such that the alternative belonging to one class must completely satisfy the constraints, and the alternatives of the opposite class must not satisfy them completely. The basic description of LAD is given in (Boros et al., 1997). Its application in a similar area is shown in (Hammer et al., 2007).

5.3 Models of Multi Hierarchical Discrimination

The MHDIS method has been used to develop this study's model as a non-parametric approach (Doumpos and Zopounidis, 2001). The problem involves two or more ordered groups of alternatives for comparison, and this model is also based on regression analysis. The notation and formulas are as follows:

f is the objective function of the basic model. It is the total error of the utility function in country misclassification, and it should be minimized. Variable S is on the right side of the constraints and is either a non-negative constant number defining the gap of separation of the two classes or is the objective function if perfect separation is possible. The results of the model are sensitive to its value. In the following section, we will discuss the value of S.

Initially, a reference set, A, consisting of n alternatives, $a_1, a_2, ..., a_n$, classified into q ordered classes, C_1 C_2, ..., C_q (C_q is preferred to C_{q-1}, C_{q-1} is preferred to C_{q-2}, etc.), is used for model development (i.e., a training sample). The alternatives are described (evaluated) along a set of m evaluation criteria, $g = \{g_1, g_2, ..., g_m\}$. The evaluation of an alternative, a, on criterion g_i is denoted as $g_i(a)$, which is the level of a at alternative i. The criteria set may include both criteria of increasing and decreasing preference. For example, high GDP is preferred to low GDP, but as an alternative in the case of the inflation rate, a low rate is preferred to a high rate.

A criterion, g_i, is assumed to have p_i different levels, which are rank-ordered from the lower $g^1{}_i$ (the least preferred value) to the higher $g^{pi}{}_i$ (the most preferred value). The number of criterion levels is specified according to the evaluations of the alternatives integrated in the training sample. In this model, r_{ai} denotes the position of the evaluation of the alternative, $\boldsymbol{a}$, on criterion $\boldsymbol{g_i}$ within the rank ordering of the criterion levels from the lower $\boldsymbol{g^1_i}$ to the higher $\boldsymbol{g^{pi}_i}$. In general, $\boldsymbol{W_{ij}}$ is the award function for the existing criteria levels, and $\boldsymbol{M_{ij}}$ is the penalty function for the missing criteria levels. Index $\boldsymbol{k}$ indicates the class of alternative; index $\boldsymbol{i}$ indicates criteria; and index $\boldsymbol{j}$ indicates the criteria level.

The basic model below is a linear programming model used in our classification study (Doumpos and Zopounidis, 2001). The quantity of $\boldsymbol{S}$ can be a fixed positive value, or it can be considered a variable. In the objective function below $\boldsymbol{e(a)}$ and $\boldsymbol{e(b)}$ indicates error. And the developed model is used for two classes which are class **α** and class **β**.

$$\text{Min } f = \sum e(a) + \sum e(b) \tag{5.1}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{r_{ai}-1} W_{ij} - \sum_{i=1}^{m} \sum_{j=r_{ai}}^{p_i-1} M_{ij} - e(a) \geq S \quad \forall a \in \alpha \tag{5.2}$$

$$\sum_{i=1}^{m} \sum_{j=r_{bi}}^{p_i-1} M_{ij} - \sum_{i=1}^{m} \sum_{j=1}^{r_{bi}-1} W_{ij} - e(b) \geq S \quad \forall b \in \beta \tag{5.3}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{p_i-1} M_{ij} = 1, \qquad \sum_{i=1}^{m} \sum_{j=1}^{p_i-1} W_{ij} = 1 \tag{5.4}$$

$$S, W_{ij}, M_{ij}, e(a), e(b) \geq 0 \tag{5.5}$$

Note that in the original paper (Doumpos and Zopounidis, 2001), the following clarification is not explained.

Lemma: If $\boldsymbol{S}$ is a variable in problems (5.1)-(5.5), there is an optimal solution with $\boldsymbol{S}$=0.

Proof: Assume that $\boldsymbol{S}$ is positive if all e(a) and e(b) are zero. The solution remains feasible if $\boldsymbol{S}$ is decreased to zero, and the value of the objective function is not changed, i.e., the solution is still optimal. If $\boldsymbol{S}$ is positive and some errors (e's) are also positive, then let:

$\varepsilon = \min\{S, \min\{e(a) \mid e(a) > 0, a \in \alpha\}, \min\{e(b) \mid e(b) > 0, b \in \beta\}\} > 0.$

Then, all positive e's and **S** can be decreased by ε. The constraints are still satisfied, and the objective function is decreased by:

$\varepsilon \left(|\{a \mid e(a) > 0, a \in \alpha\}| + |\{b \mid e(b) > 0, b \in \beta\}| \right) \geq \varepsilon > 0.$

Thus, the previous solution was not optimal. □

If the optimal value of problems (5.1)-(5.5) is zero, then the maximal separation gap for perfect classification can be obtained by the following model:

Max **S** (5.6)

$$\sum_{i=1}^{m} \sum_{j=1}^{r_{ai}-1} W_{ij} - \sum_{i=1}^{m} \sum_{j=r_{ai}}^{p_i-1} M_{ij} \geq S \quad \forall a \in \alpha \tag{5.7}$$

$$\sum_{i=1}^{m} \sum_{j=r_{bi}}^{p_i-1} M_{ij} - \sum_{i=1}^{m} \sum_{j=1}^{r_{bi}-1} W_{ij} \geq S \quad \forall b \in \beta \tag{5.8}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{p_i-1} M_{ij} = 1, \qquad \sum_{i=1}^{m} \sum_{j=1}^{p_i-1} W_{ij} = 1 \tag{5.9}$$

$$S, W_{ij}, M_{ij} \geq 0 \tag{5.10}$$

An issue with models (5.1)-(5.5) and (5.6)-(5.10) is that although there are several classes in the underlying problem, they separate the countries (alternatives) into only two classes. It is possible to classify them into the required classes using the following simple model. The model is formalized for four classes, as the classification problem in question has four classes; however, the generalization is straightforward for more classes.

$$\text{Min} \quad f = \sum_{a\in C_1,C_2,C_3} e_{U(a)} + \sum_{a\in C_2,C_3,C_4} e_{L(a)} \tag{5.11}$$

$$L(a) = \sum_{i=1}^{m} \sum_{j=1}^{r_{ai}-1} W_{ij} - \sum_{i=1}^{m} \sum_{j=r_{ai}}^{p_i-1} M_{ij} \qquad \forall a \in \alpha \tag{5.12}$$

For all $\in C_1$: $U_1 \geq L(a) - e_U(a)$,

For all $\in C_2$: $L_2 \leq L(a) + e_L(a)$, $U_2 \geq L(a) - e_U(a)$,

For all $\in C_3$: $L_3 \leq L(a) + e_L(a)$, $U_3 \geq L(a) - e_U(a)$,

For all $a \in C_4$: $\quad L_4 \leq L(a) + e_L(a)$, (5.13)

$$\sum_{i=1}^{m} \sum_{j=1}^{p_i-1} M_{ij} = 1, \qquad \sum_{i=1}^{m} \sum_{j=1}^{p_i-1} W_{ij} = 1 \tag{5.14}$$

$$U_1 + S \leq L_2 \qquad U_2 + S \leq L_3 \qquad U_3 + S \leq L_4 \tag{5.15}$$

For all ***a***, S, $e_{U(a)}$, $e_{L(a)}$, ≥ 0 , U_1, U_2, U_3, L_2, L_3 are unrestricted (5.16)

5.4 Application to the Training Set

In the first step, models (5.1)-(5.5) were solved for the training set. The model contained 44 countries, and ***S*** was fixed to zero. Nineteen criteria with three levels each were used in this model. All models in this study were solved using LINGO12.0 package software. As result of the first computation, countries 1 (Argentina) and 39 (Switzerland) were misclassified, meaning that country 1 belonged to the upper class or set ***a,*** and country 39 belonged to the lower class or set ***b***. Important criteria were 6, 8, 11, 13, 15, and 17, which are GDP growth, GNI per capita, Gross capital formation, Inflation (GDP deflator), Mobile cellular subscriptions, and Population growth, respectively. The total error is 0.666.

In the next analysis, ***S*** was raised from 0 to 0.01, and the number of criteria and their levels remained the same. As result, the Lingo solution demonstrates that the total error increased to 0.7333, and countries 1, 4, 13, and 39, which are Argentina, Bolivia, Ecuador, and Switzerland, were misclassified. In addition, important criteria were 1, 6, 11, 13, 15, and 17, which are electric power consumption, GDP growth, gross capital formation, inflation (GDP deflator), Mobile cellular subscriptions, and population growth, respectively. The result of the analysis depends on ***S*****;** however, the number of important criteria does not differ much when the value of ***S*** is increased or decreased.

By refining the criteria levels' system, the equality of the classifications improves. In the second analysis, certain criteria used six levels instead of three (Table 5.2), and criterion 19 (Surface area) was employed as one of the important criteria. Dividing criteria into more levels eliminates the gap that may occur between countries and illustrates the criteria's influence in country classification.

We ran the model after those changes (S=0), and the new result showed that there was no misclassification and that only criteria 13 and 6 were important. There was no misclassification when we ran the model with S=0.01 again, but the number of important criteria increased. The results show that the important criteria were 4, 7, 11, 13, 15 and 19, which are fertility rate, GNI per capita, gross capital formation, inflation, mobile cellular subscriptions, and surface area, respectively. The penalty and reward values are:

M_{42}=0.02 W_{151}=0.51

M_{71}=0.02 W_{191}=0.49

M_{112}=0.48 M_{132}=0.48 (5.17)

The above result will help verify the test set's mathematical model. The results in the previous steps prove that there is no misclassification, allowing us to estimate the possible gap between classes. The following formula discovers the largest gap among classes by maximizing **S** as an objective function with respect to predefined constraints.

The result shows that the value of $\boldsymbol{S}$ is 0.25 (the gap between two classes). The important criteria are 6, 7, 13, 14, 15, and 17,which are GDP growth, GNI per capita, Inflation, Military expenditure, Mobile cellular subscriptions, and Population growth.

The numerical solution for models (5.11)-(5.16) shows separation without error, but the optimal solution was degenerated, i.e., $\boldsymbol{S}=\boldsymbol{U_1}=\boldsymbol{U_2}=\boldsymbol{U_3}=\boldsymbol{L_2}=\boldsymbol{L_3}=\boldsymbol{L_4}=0$. A degenerated solution always exists in this model, e.g., $\boldsymbol{S=0}$, $\boldsymbol{W_{131}}=\boldsymbol{M_{191}}=1$, and the upper and lower bounds are 1 Important criteria are 19 and 13. In the next level, we decided to find the maximal possible gap between the model's upper and lower bounds and developed the following mathematical model:

$$\text{Max } \boldsymbol{S} \tag{5.18}$$

$$L(a) = \sum_{i=1}^{m} \sum_{j=1}^{r_{ai}-1} W_{ij} - \sum_{i=1}^{m} \sum_{j=r_{ai}}^{p_i-1} M_{ij} \qquad \forall a \in C_k \tag{5.19}$$

For all $\in C_1$: $U_1 \geq L(a)$,

For all $\in C_2$: $L_2 \leq L(a)$, $U_2 \geq L(a)$,

For all $\in C_3$: $L_3 \leq L(a)$, $U_3 \geq L(a)$,

For all $\in C_4$: $L_4 \leq L(a)$. (5.20)

$$\sum_{i=1}^{m} \sum_{j=1}^{p_i-1} M_{ij} = 1, \qquad \sum_{i=1}^{m} \sum_{j=1}^{p_i-1} W_{ij} = 1 \tag{5.21}$$

$$U_1 + S \leq L_2, \quad U_2 + S \leq L_3, \quad U_3 + S \leq L_4 \quad L_2 \leq U_2, \quad L_3 \leq U_3 \tag{5.22}$$

For all $\boldsymbol{a}$, $S \geq 0$, U_1, U_2, U_3, L_2, L_3 are unrestricted.

Then, the results are changed as follows:

$S=0.2307692$

$M_{32}=0.1923077$ $M_{133}=0.2307692$ $M_{43}=0.3461538$ $M_{65}=0.2307692$

$W_{62}=0.2307692$ $W_{141}=0.03846154$ $W_{152}=0.2307692$ $W_{153}=0.03846154$

$W_{193}=0.03846154$ $W1_{42}=0.2307692$ $W_{196}=0.1923077$

$U_3=0$ $U_2=-0.5$ $U_1=-0.7307692$

$L_4=0.2307692$ $L_3=-0.2692308$ $L_2=-0.5$ (5.23)

5.5 Verification of the Mathematical Model and Validation for Test Set

For validation and verification, we chose a new data set, defined as the test set, and applied the classification method developed in the previous section. The list of 39 new countries (alternatives) selected to verify our model is shown in Table 3.3 with their levels and classifications. The classifying criteria were the same as in the previous classification. With the new classification, only four of 39 countries were misclassified (have negative value and shown by asterisk (*) in Table 3.3), which are Chile, Jamaica, Malaysia, and the Philippines, providing excellent evidence for our model's acceptance.

Table 5.1. Test Set Countries with Classification and Number of Levels

No.	Country Name	Criteria						S Value
		g_4	g_7	g_{11}	g_{13}	g_{15}	g_{19}	
45	Bahrain	2	1	2	6	6	1	0.48
46	Cyprus	1	1	2	6	6	1	0.48
47	Greece	1	1	2	6	6	1	0.48
48	Ireland	1	2	2	6	6	1	0.5
49	Israel	2	2	2	6	6	1	0.5
50	Kuwait	2	2	1	6	6	1	0.5
51	Saudi Arabia	3	1	2	6	2	1	0.5
52	Singapore	1	2	4	6	6	1	0.98
53	United Arab Emirates	2	2	3	6	6	1	0.98
54	Albania	2	1	2	6	2	1	0.01
55	Armenia	1	1	3	1	1	1	0.52
56	Azerbaijan	1	1	4	1	1	1	0.52
57	Bangladesh	2	1	2	6	1	1	0.52
58	Belarus	1	1	4	1	1	1	0.52
59	Bosnia and Herzegovina	1	1	3	6	1	1	0.04
60	Bulgaria	1	1	2	3	3	1	0.49
61	Cameroon	3	1	1	6	1	1	0.5
62	Chile *	1	1	3	6	3	1	**-0.47**
63	Costa Rica	2	1	2	6	1	1	0.52
64	Côte d'Ivoire	3	1	1	6	1	1	0.5
65	Cuba	1	1	1	6	1	1	0.52
66	Egypt, Arab Rep.	3	1	2	6	1	1	0.5
67	Georgia	1	1	3	1	1	1	0.52
68	Ghana	3	1	2	6	1	1	0.5
69	Iran, Islamic Rep.	2	1	5	6	1	1	0.04
70	Jamaica *	2	1	4	6	3	1	**-0.47**
71	Kazakhstan	1	1	3	1	1	1	0.52
72	Lebanon	2	1	3	6	1	1	0.04
73	Libya	3	1	1	6	1	1	0.5
74	Malaysia *	2	1	4	6	3	1	**-0.47**
75	Pakistan	3	1	1	6	1	1	0.5
76	Panama	2	1	2	6	2	1	0.01
77	Philippines *	3	1	2	6	2	1	**-0.01**
78	Romania	1	1	3	4	2	1	0.01
79	South Africa	2	1	1	6	3	1	0.01
80	Sudan	3	1	2	5	1	1	0.5
81	Tajikistan	3	1	2	1	1	1	0.98
82	Turkmenistan	2	1	5	1	1	1	0.52
83	Vietnam	2	1	4	6	1	1	0.04

5.6 Application of LAD in Classification of Countries

The role of the two classes to be distinguished is not symmetric in the LAD method. LAD wants to separate one of the two sets from the other one. The first class is called *positive* class and the other one is the *negative* class. ***The first step*** in LAD is to select constraints which can be typical for the members of the class to be separated from the other one. Each constraint claims that the value of a parameter must be either high or low. The following set of countries is a pattern that is created by LAD to classifying countries:

$$g_{15} \geq 2,\ g_{13} \geq 3,\ g_{17} \geq 2. \tag{5.24}$$

The first constraint $g_{15} \geq 2$ means that the level of criterion 15 must greater or equal to 2, $g_{13} \geq 3$ means that the level of criterion 13 must greater or equal to 3 and so on. Therefore, the right hand side of each constraint illustrates the level of criterion, and it does not demonstrate the value of it. All countries in the positive class satisfy (5.24), making its prevalence 100%. In (5.24), all positive class elements satisfy it, i.e., 26 in the training set and an additional 8 out of 18 in the negative class, so its homogeneity is 26 out of 34 or 76.47%.

Generally one pattern is not enough for a perfect separation of the two classes, so a subset of patterns must be selected such that each positive object satisfies at least one pattern, i.e., satisfies all the pattern's conditions. Here, perfect patterns separating the upper-middle, medium- and low-income countries from the high-income countries exist. Perfect pattern has 100% prevalence and homogeneity as it is mentioned in Chapter 4. Equation 5.24 describes the perfectly separating patterns.

Table 5.2. Separation Patterns

No.	**Constraints**	**Test Set**	
		High-income countries	**Non-high-income countries**
1	$g_{15}\leq 3, g_4\leq 2, g_1\leq 2$	None	61, 64, 66, 68, 73, 75, 77, 80 81
2	$g_{15}\leq 3, g_4\leq 2, g_2=1$	None	61, 64, 66, 68, 73, 75, 77, 80 81
3	$g_{15}\leq 3, g_7=1, g_4\leq 2$	None	61, 64, 66, 68, 73, 75, 77, 80 81
4	$g_{15}\leq 3, g_8=1, g_4\leq 2$	None	61, 64, 66, 68, 73, 75, 77, 80 81
5	$g_{15}\leq 3, g_{14}\leq 2, g_1\leq 2$	None	All
6	$g_{15}\leq 3, g_{14}\leq 2, g_2=1$	None	All
7	$g_{15}\leq 3, g_{14}\leq 2, g_7=1$	None	All
8	$g_{15}\leq 3, g_{14}\leq 2, g_8=1$	None	All

Notice that the 8 patterns in Table 5.2 combine only 7 constraints. The constraints are given in the order found by LAD. The order reflects the constraints' importance. Thus, the most important constraint is $g_{15}\leq 3$, i.e., the value of criterion 15 (mobile cellular subscription) is below average. The constraints $g_{14}\leq 2$ and $g_4\leq 2$ still have high importance, whereas the other constraints are only supplementary. Table 5.6 analyzes patterns $\mathscr{P}_1$ = (g_{15}≤3, g_4≤2) and $\mathscr{P}_2$ = (g_{15}≤3, g_{14}≤2). Pattern $\mathscr{P}_2$ and patterns containing its two constraints are more robust than pattern $\mathscr{P}_1$ and the children of $\mathscr{P}_1$.

Table 5.3. The Behavior of the Patterns P1 =(g15≤3, g4≤2) and P2 =(g15≤3, g14≤2)

No.	Constraints	Training Set		Test Set	
		High-income countries	**Non-high-income countries**	**High-income countries**	**Non-high-income Countries**
1	g_{15}≤3, g_4≤2	8 (Canada)	all countries	None	54-60, 62, 63, 65, 67, 69-72, 74, 76, 78, 79, 82, 83
2	g_{15}≤3, g_{14}≤2	30 (Oman)	all countries	None	all countries

By using four constraints an additional 31 perfect patterns, are obtained which contain the 7 constraints above and three more: g_9=1, g_{12}≤2, and g_5=1. All 39 perfect patterns contain g_{15}≤3 and exactly one of g_{14}≤2 and g_4≤2. These three constraints are most important in separating upper-middle, medium-and low-income countries from high-income countries.

If the roles of the two sets of countries are interchanged, the results differ. It is a property of LAD that the roles of the two sets are asymmetric. Perfect pattern, i.e., a pattern having 100 % prevalence and homogeneity, does not exist. Tables 5.2 and 5.3 respectively describe the patterns with high prevalence and homogeneity.

Table 5.4. Patterns Separating the High-Income Countries from the Other Ones with Prevalence 1

No.	Constraints	Homogeneity	Training set
			Non-high income countries
1	$g_{15}\geq 2$, $g_{13}\geq 3$	74.86%	7, 10, 26, 31, 32, 38, 40, 41, 42
2	$g_{15}\geq 2$, $g_{17}\geq 2$	72.22%	1, 6, 10, 26, 31, 32, 38, 40, 41
3	$g_{15}\geq 2$, $g_{13}\geq 3$, $g_{17}\geq 2$	76.47%	10, 26, 31, 32, 38, 40, 41, 42
4	$g_{15}\geq 2$, $g_{13}\geq 3$, $g_{18}\leq 2$	74.29%	7, 10, 26, 31, 32, 38, 40, 41, 42

Table 5.5. Patterns Separating the High-Income Countries from the Other Ones with High Homogeneity

No.	Constraints	Prevalence	Homogeneity	Training set	
				High-income countries	Non-high-income countries
1	$g_{15}\geq 4$	92.31%	100%	8, 30	None
2	$g_{15}\geq 3$, $g_{17}\geq 2$, $g_4=1$	96.15%	96.15%	30	32
3	$g_{15}\geq 3$, $4\geq g_{17}\geq 2$	92.30%	96.00%	25, 30	32
4	$g_{15}\geq 3$, $5\geq g_{17}\geq 2$	96.15%	96.15%	30	32

Countries 10, 26, 31, 32, 38, 40, and 41 are non-high income countries, satisfying the patterns with a prevalence of 100%, meaning they come close to being high-income countries. With the exception of country 31(Paraguay), the countries belong to the upper-middle income class. Paraguay is a lower-middle income class country. The majority of these countries could possibly enter the high-income category. Without $g_{15}\geq 4$, it is impossible to achieve 100% homogeneity, which emphasizes the importance of g_{15}.

As previously mentioned, the important criteria (more robust) are 15, 14 and 4, which are mobile cellular subscription (per 100 people), military expenditure (% of GDP), and fertility rate, total (births per woman).

Chapter 6

COUNTRIES RISK RATING

This chapter examines the relative importance of economical, environmental, educational, and infrastructure factors in determining country risk rating. In the first step of the analysis, the economic determinants of the country were collected from World Bank (WB) and International Monetary Fund (IMF) databases. Although the country rating is subjective from time to time and it depends on many different factors (economical, environmental, educational, and infrastructure), changes in economic fundamentals are the main aspects which affect in country risk rating.

In addition the study does some empirical analysis of importance of economic and political factors and it does not stand for exact solution for classification of countries, rather than the study tries to analyze the relative importance of factors in the determination of country risk. It also provides an approximate rating method, which uses only data available in World Bank and International Monetary Fund and everybody, can evaluate easily. Also all data collected are information of year 2010.

6.1 Introduction to Countries Risk Rating

On a daily basis, financial engineering involves developing, planning, and optimization of financial instruments and processes, and formulation of creative models for solving the problems in business and economics (Finnerty, 1988). Until now many scientists in different fields such as operation researchers, statisticians, financial researchers, econometricians, and mathematicians have proposed variety of methodologies for country risk rating (Dahl et al., 1993). Most of the methodologies that are used in country risk rating are based on probabilistic and stochastic models. A typical work is (Hu et al., 2002), which tries to construct rating transition matrices for countries as an input of rating-based credit portfolio model. Another example, (Mulvey et al., 1997) build up strategic financial risk management model using a multi-stage stochastic program for coordinating the asset and liability decision. The study was the continuation of the multi-stage stochastic model that brings together all major financial-related results in a single and unique structure (Mulvey 1996).

Prior to developing rating risk model for countries, bank, or any other financial assets, input data (indicators) have crucial role in the classification. Factors or indicators are important key inputs of country risk rating models that have been developed by many researcher during different periods. Many studies such as (Haque et al., 1996), (Haque et al., 1996), and (Hammer et al., 2007) try to examine the relative importance of economical and political factors that have major effect in country risk rating. However, some of them cannot estimate or evaluate the weight of political factors in country risk rating. A model that is proposed in 2001 was based on the multicriteria decision aid (MCDA) and Multi-group Hierarchical Discrimination (MHDIS), which use different criteria (indicators) to classify number of alternatives (countries) in to specified classes (Doumpos and Zopunidis, 2001). Later on, the proposed model is modified and improved by (Mirzaei and Vizvari, 2011) and utilized to reconstruct the World Bank classification. In all of these models, the economical and political factors have main effect on the result.

(Hammer et al., 2007) has accomplished one of the remarkable works that is applied reverse engineering by utilizing of Logical Analysis of Data (for LAD see Chapter4) in the case of financial risk rating. They developed a high-quality model to identify the financial factors which are most important for bank rating. Furthermore, the model can be used in various stages in credit granting and/or risk rating for different decision processes (Hammer et al., 2007).

They try to develop a reliable model to measure creditworthiness of countries which is independent from rating agencies such as S&P, Moody's and Fitch. Such institution agencies (S&P, Moody's) publish some rating related to country creditworthiness annually or semiannually. As an example Moody's ratings provides investors with a simple system of gradation by which relative creditworthiness of securities are characterized. This system of rating affects countries in many different ways, when the Moody's agency announces downgrading a country, investors may charge higher interest rates or would decline or take out their investment form that country, and local currency value of the country will depreciate. It is a terrible disaster for the countries that are downgraded. Nowadays the ratings provided by Moody's Corporation have great influence on financial markets.

As it is well-known, an economic world crisis started in 2008. Since that time many countries have been downgraded by several scales in one step. For instance during economy crises, outlook of a number of countries in euro zone was darken. As an example, rating of government bound (local currency) of Cyprus in January-2008 was **Aa3** and then it was downgraded after on to **Ba3** in June-2012 (during 4 years the country is downgraded by 6 scales), rating of government bound of Greece was **Ba1** in December-2010 and it was changed to **C** in June-2012, Hungary local currency rating was **Baa1** in July-2009 and then it changed to **Ba1** in December-2011, Iceland local currency rating was **A1** in October-2008 and it was modified to

Baa3 in July-2010, Ireland's outlook was **Aaa** in January-2009, then it changed to **Ba1** in July-2010, Italy's outlook was **Aa2** in June-2011 and then it was changed to **Baa2** in July-2012, Latvia's outlook was **A3** in November-2008 and then it was changed to **Baa3** in June-2011, Portugal's outlook was **Aa2** in October-2009 and then in February-20012 it was **Ba3**, Slovenia's outlook was **Aa2** in March-2009 and then in August-2012 it was **Baa2**, Spain's outlook was **Aaa** in June-2010 and then it was changed to **Baa3** in June 20012 (Moody's agency).

In addition, there are some examples in opposite direction, it means that there are a number of countries that had low rating scales in the past, and then they have improved their economic and political conditions and as result they increased their rating scales. For examples, Brazil's outlook was Ba2 in may-2007 and it was changed to Baa2 in June-2011 (increased by 3 grades), India's outlook was Ba2 in December-2009 and then it was changed to Baa3 in December-2011. The change of the ratings by more than one grade in single step has always some subjective elements. Therefore, downgrading or upgrading outlook of a sovereign is affecting economic and political feature of that country in different ways. In addition, it is crucial for foreign investors and bankers to follow credit risk rating of countries in order to make a decision for investing and/or lending their money. The credit rating agencies make their money on the secrecy of their method. If the method of rating were public information then everybody could calculate the ratings and no income could be made by the agencies. The presence of subjective elements threatens with misleading information. They also exclude the existence of a perfectly defined method. However, approximations may exist and can have value for the financial markets. This study provides one approximation based on the method of Logical Analysis of Data (LAD).

6.2 Moody's Rating Scale and Process

Moody's rating represents a rank-ordering of creditworthiness, or expected loss. Expected loss is a function of the probability of default and the expected severity of loss given a default. Ratings are forward looking in that the rank-ordering is designed to hold across multiple horizons. The moody rating scale, running from high of Aaa to a low of C, comprises 21 notches. It is divided into two sections, investment grade and speculative grade as follow:

1. Investment grade
   - i. **Aaa**: high rating, representing minimum credit risk,
   - ii. **Aa1, Aa2, Aa3**: high grade,
   - iii. **A1, A2, A3**: upper middle grade,
   - iv. **Baa1, Baa2, Baa3**: medium grade,
2. Speculative grade
   - i. **Ba1, Ba2, Ba3**: speculative elements,
   - ii. **B1, B2, B3**: subject high credit risk,
   - iii. **Caa1, Caa2, Ca**a3: bonds of poor standing,
   - iv. **Ca**: high speculative, or near to default,
   - v. **C**: lowest rating, bond typically in default, little prospect for recovery of principal or interest.

Investors use these ratings as limits on their investment parameters and as means for expanding their investment horizons to markets or security types they do not cover by their own analysis. The process of rating is as follows:

- Gathering information sufficient to evaluate risk rating of security, company, industry, countries, or etc,

- Developing a conclusion in committee on the appropriate rating,
- Monitoring on an ongoing basis to determine whether the rating should be changed, and
- Informing the marketplace and market participants of Moody's action.

6.3 Scales and Indicators

In this study, we try to classify the countries into 14 rating scales similar to those used by Moody's agency. Moody's rates countries according to their risk of investment in 21 different ratings scale. As mentioned before, this rating system is based on year 2010 and it includes one year database. There are only few countries in scale rating **B1** and lower (B2, B3, Caa1, Caa2, Caa3, Ca, and C) so we have decided to consider them as one scale. That is why instead of 21 rating system of Moody's we have 14 rating classes.

In this study, 71 countries and 30 economic factors are used to elaborate the approximate rating method. Countries are chosen according to their availability of information in WB and IMF database, and the factors which have been selected are designed to measure country's performance in economic and other sectors. Besides, we have considered factors such that the theoretical literature has stressed their importance in risk rating of countries. The following table demonstrated the list of factors or indicators which collected from WB and IMF to classify countries risk rating:

Table 6.1. Economical, Environmental, Educational, and Infrastructure Factors

Indicator and Unit	Index	Indicator and Unit	Index
General government gross debt (% of GDP)	G2	International migrant stock (% of population)	G28
General government net lending/borrowing (% of GDP)	G3	Land area (sq. km)	G29
General government total expenditure (% of GDP)	G5	Mobile cellular subscriptions (per 100 people)	G30
Gross domestic product based on purchasing-power-parity (PPP) per capita GDP (Current international dollar USD)	G6	Net income (BOP, current US$)	G31
Gross domestic product based on purchasing-power-parity (PPP) valuation of country GDP (Current international dollar USD)	G8	Net migration	G32
Gross domestic product, constant prices (Percentage change)	G10	Population ages 0-14 (% of total)	G33
Unemployment rate (Percent of total labor force)	G15	Population ages 65 and above (% of total)	G34
Burden of customs procedure, WEF (1=extremely inefficient to 7=extremely efficient)	G16	Population density (people per sq. km of land area)	G35
Business extent of disclosure index (0=less disclosure to 10=more disclosure)	G17	Population growth (annual %)	G36
Cost to import (USD per container)	G20	Population, total	G37
Current account balance (% of GDP)	G21	Secure Internet servers (per 1 million people)	G39
Domestic credit to private sector (% of GDP)	G22	Time required to register property (days)	G41
Ease of doing business index (1=most business-friendly regulations)	G23	Time required to start a business (days)	G42
Export value index (2000 = 100)	G24	Time to export (days)	G43
GDP, PPP (current international $)	G25	Urban population (% of total)	G45

The first 7 indicators (indices: G2, G3, G5, G6, G8, G10, and G15) which are mostly economical factors were collected from IMF database. The rest of indictors (23 indicators) which are mostly environmental, educational, and infrastructure factors were collected from WB database.

6.4 Countries and Their Properties

The number of countries, which are used in creating the approximate solution, is limited to 71 for two main reasons. The first reason is that in the WB and IMF databases, several countries have no data available for the indicators in Table 6.1.

The second reason is even if the data are available for a country, it is not rated by Moody's. As result it is not possible to validate or certify of our model for rating that country. The following table shows the name of the countries with their Mood y's rating in 2010. Further 34 countries were used for the verification of the approximate method.

Table 6.2. Selected Countries with Ratings

No.	Country	scale	No.	Country	scale	No.	Country	scale	No.	Country	scale
1	Austria	*Aaa*	19	Chile	*Aa3*	37	Mauritius	*Baa2*	55	Portugal	*Ba2*
2	Canada	*Aaa*	20	China	*Aa3*	38	Bulgaria	*Baa3*	56	Turkey	*Ba2*
3	Denmark	*Aaa*	21	Saudi Arabia	*Aa3*	39	Colombia	*Baa3*	57	Egypt, Arab Rep.	*Ba3*
4	Finland	*Aaa*	22	Czech Republic	*A1*	40	Croatia	*Baa3*	58	Georgia	*Ba3*
5	France	*Aaa*	23	Estonia	*A1*	41	Hungary	*Baa3*	59	Albania	*B1*
6	Germany	*Aaa*	24	Israel	*A1*	42	Iceland	*Baa3*	60	Dominican Republic	*B1*
7	Luxembourg	*Aaa*	25	Slovenia	*A1*	43	Latvia	*Baa3*	61	Venezuela, RB	*B1*
8	Netherlands	*Aaa*	26	Cyprus	*A2*	44	Panama	*Baa3*	62	Vietnam	*B1*
9	Norway	*Aaa*	27	Poland	*A2*	45	Peru	*Baa3*	63	Bosnia and Herzegovina	*B2*
10	Singapore	*Aaa*	28	Malaysia	*A3*	46	Azerbaijan	*Ba1*	64	Ukraine	*B2*
11	Sweden	*Aaa*	29	South Africa	*A3*	47	Indonesia	*Ba1*	65	Argentina	*B3*
12	United Kingdom	*Aaa*	30	Lithuania	*Baa1*	48	Ireland	*Ba1*	66	Jamaica	*B3*
13	United States	*Aaa*	31	Mexico	*Baa1*	49	Morocco	*Ba1*	67	Moldova	*B3*
14	Belgium	*Aa1*	32	Russian Federation	*Baa1*	50	Uruguay	*Ba1*	68	Nicaragua	*B3*
15	Hong Kong SAR, China	*Aa1*	33	Thailand	*Baa1*	51	Armenia	*Ba2*	69	Pakistan	*B3*
16	Italy	*Aa2*	34	Tunisia	*Baa1*	52	El Salvador	*Ba2*	70	Greece	*Caa1*
17	Japan	*Aa2*	35	Brazil	*Baa2*	53	Jordan	*Ba2*	71	Ecuador	*Caa2*
18	Spain	*Aa2*	36	Kazakhstan	*Baa2*	54	Philippines	*Ba2*			

6.5 Decision Tree and Important Indicators

As it is indicated in the decision tree chart, each country rates according to the values of their indicators. For example if a country belongs to class **Aaa** it must have the following attributes:

- General government gross debt is less than or equal to 99. 54 (% of GDP),
- General government total expenditure is greater than or equal to 19.3 (% of GDP),
- GDP based on PPP per capita GDP is greater than or equal to 33960 (current international $),
- GDP, a constant price is greater than or equal to 0.57 (% of change).

Otherwise, it does not belong to class **Aaa**, therefore we should check if it belong to class **Aa1** or lower one. In each step we check the value of some indicators of a country if it satisfies the condition we say that the country belongs to that class, otherwise we go one step further and test the indicators for the lower class and so on. The importance of indicator is achieved by considering Decision Tree chart. In the following section the list of main indicators by order of importance (most important to low important) is provided:

- Gross domestic product based on purchasing-power-parity (PPP) per capita GDP (Current international dollar USD)
- Gross domestic product, constant prices (Percentage change)
- General government gross debt (% of GDP)
- Ease of doing business index (1=most business-friendly regulations)
- Burden of customs procedure, WEF (1=extremely inefficient to 7=extremely efficient)
- Net migration
- Current account balance (% of GDP)
- Domestic credit to private sector (% of GDP)
- Gross domestic product based on purchasing-power-parity (PPP) valuation of country GDP (Current international dollar USD)

- General government net lending/borrowing (% of GDP)
- Urban population (% of total)
- General government total expenditure (% of GDP)
- Unemployment rate (Percent of total labor force)
- Net income (BOP, current US$)
- Population, total
- Secure Internet servers (per 1 million people)
- Time required to start a business (days)
- Time to export (days)

As it listed above totally 18 different indicators are used to construct the Decision Tree chart.

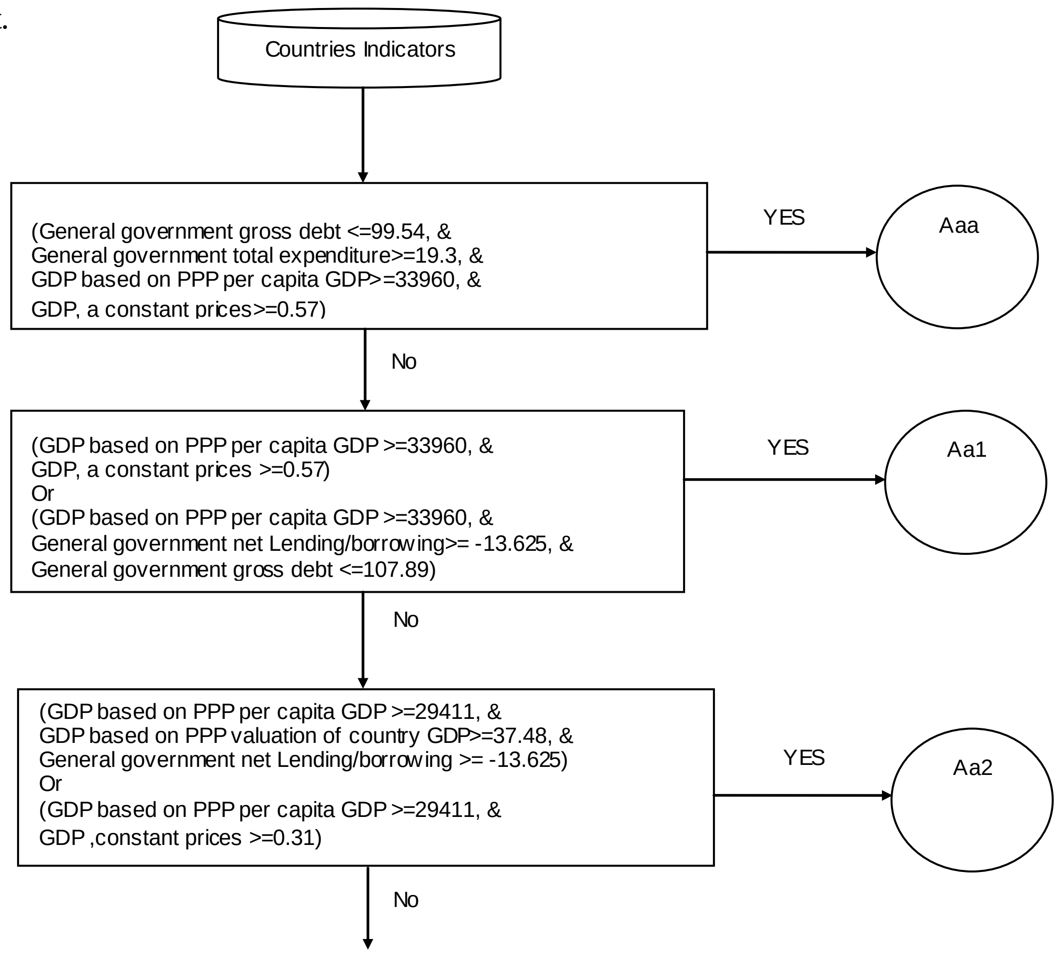

Figure 6.1. Decision Tree Chart for Classification Countries in to Moody's Scale Based on LAD (continued)

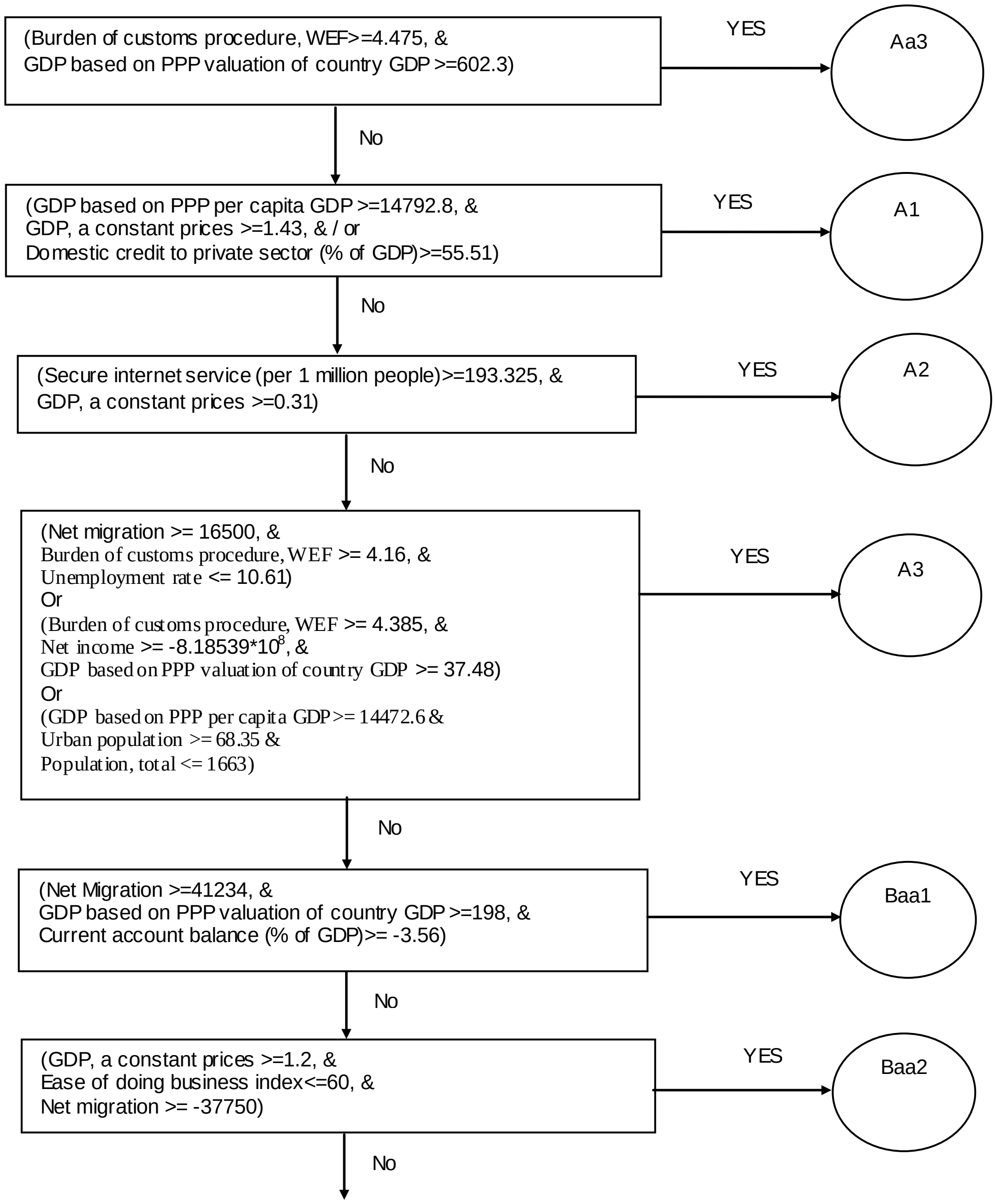

Figure 6.1. Decision Tree Chart for Classification Countries in to Moody's Scale Based on LAD (continued)

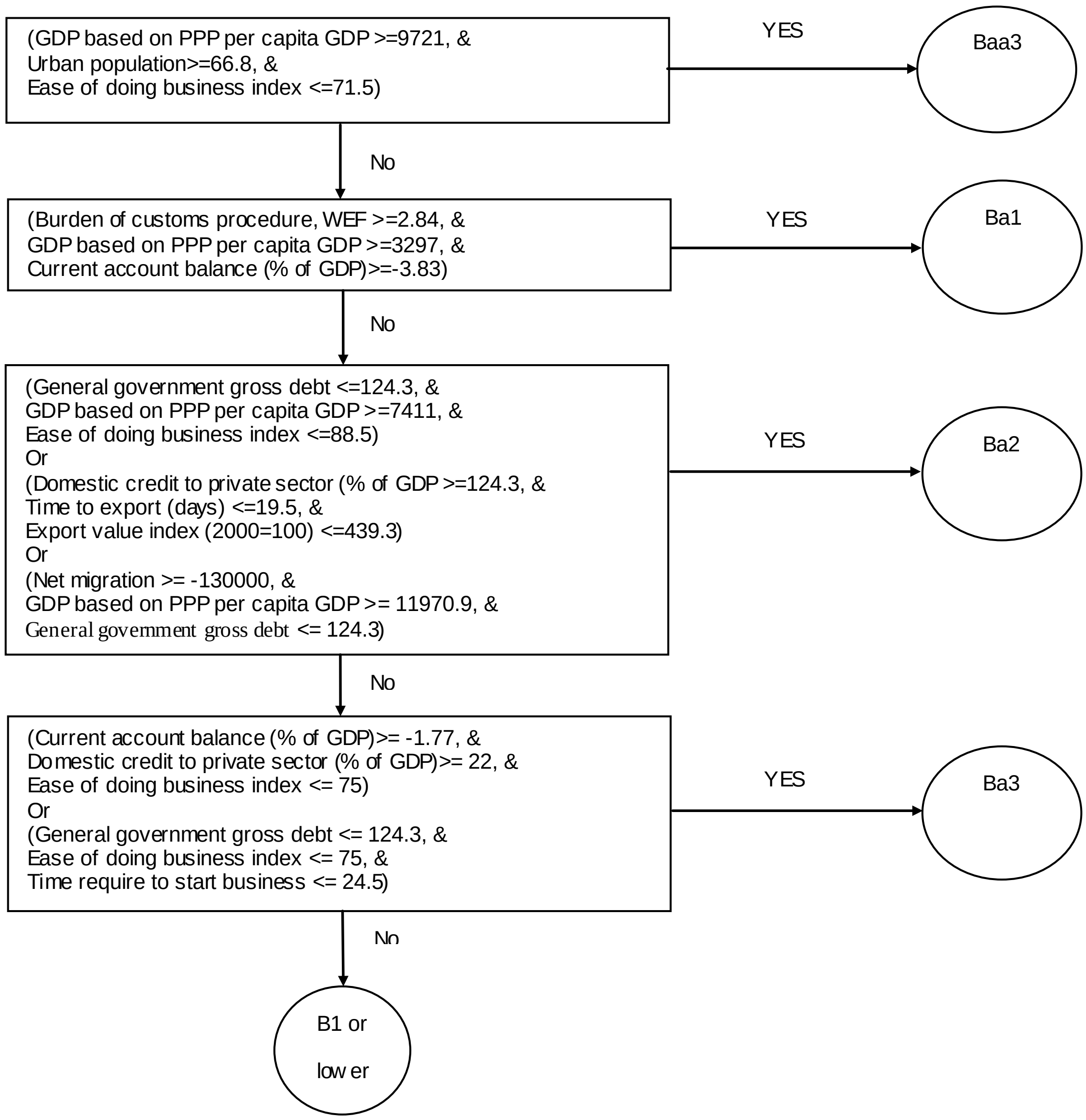

Figure 6.1. Decision Tree Chart for Classification Countries in to Moody's Scale Based on LAD

6.6 Verification of Model Based on Test Set

In order to verify the accuracy of our model, which is build by LAD, a new dataset is collected which is disjoint from the set of 71 countries. The dataset includes 34 countries from different ratings scaled. In Table 6.3 the information related to Moody's scale and the scale that is suggested by our model (Decision Tree Chart) is compared. Those countries were selected which have all the 18 indicators in the World Bank and International Monetary Fund databases which are used in the Decision Tree Charts . There are 10 countries, which have same grading scale with

Moody's rating. All countries having **B1** or lower ratings at Moody's are classified into the aggregate **B1** or lower class of LAD. The difference is one in the case of 5 countries and is two at other 5 countries. The only country where the difference is significant is Lebanon. Moody's rating can be influenced by political factor in this case.

Table 6.3. Test Dataset

Country	Moody's Scale	LAD	Country	Moody's Scale	LAD
Angola	B1	B1 or Low	Lebanon	B1	A1
Bahrain	A3	A3	Malta	A1	A2
Bangladesh	Ba3	Ba3	Mongolia	B1	B1 or Low
Barbados	Baa3	Baa3	Montenegro	Ba3	Ba2
Belarus	B2	B1 or Low	Namibia	Baa3	Ba1
Belize	B3	B1 or Low	New Zealand	Aaa	Aaa
Bolivia	B1	B1 or Low	Nicaragua	B3	B1 or Low
Botswana	A2	A3	Oman	A1	A1
Cambodia	B2	B1 or Low	Papua New Guinea	B1	B1 or Low
Costa Rica	Baa3	Baa3	Paraguay	B1	B1 or Low
Cyprus	A2	A3	Qatar	Aa2	Aaa
Fiji Islands	B1	B1 or Low	Romania	Baa3	Baa3
Guatemala	Ba1	Ba1	Slovenia	Aa2	A1
Hong Kong	Aa1	Aa1	Sri Lanka	B1	B1 or Low
India	Ba1	Ba1	Switzerland	Aaa	Aaa
Korea	A1	Aa2	Trinidad and Tobago	Baa1	Baa1
Kuwait	Aa2	Aaa	United Arab Emirates	Aa2	Aaa

Chapter 7

INEFFECTIVE METHODS USED FOR CLASSIFICATION OR RATING COUNTRIES

In this chapter we attempt to provide information about some mathematical and statistical methods that are unable to classify and/or rate the countries. This means that, the method which is going to be mentioned in this chapter is ineffective for classification and rating of countries base on data that used in the previous chapters.

7.1 Linear Regression Method

Linear regression method is utilized in this research for rating the countries. The rating values of Moody's are considered as dependent variable and numerical value of criteria are independent variables. In this part we used SPSS 17.0 Package software to solve the model. Three methods are applied to find the regression coefficients for criteria, and they are Enter, Forward, and Backward. The study involves totally 72 countries and 30 criteria. Table 7.1 shows the results of linear regression analysis based on three different methods.

First column of the table lists countries name, second column indicated the numerical rating by Mood y's agency (based on data on 2010), and the other column show the rating value which is based on linear regression analysis (Enter, Forward, and Backward).

Table 7.1. Linear Regression Analysis Which Is Based on Enter, Forward, and Backward Methods

Country	Moody's rating	Method: Enter	Method: Backward	Method: Forward	Country	Moody's rating	Method: Enter	Method: Backward	Method: Forward
Albania	5	4	4	6	Jordan	7	4	7	7
Argentina	3	8	8	8	Kazakhstan	10	9	9	9
Armenia	7	4	3	7	Latvia	9	10	11	12
Austria	18	16	15	15	Lithuania	11	10	11	12
Azerbaijan	8	9	8	11	Luxembourg	18	18	18	19
Belgium	17	16	16	16	Malaysia	12	14	15	13
Bosnia and Herzegovina	4	3	4	4	Mauritius	10	9	9	9
Brazil	10	9	7	8	Mexico	11	9	9	11
Bulgaria	9	10	8	10	Moldova	3	3	4	4
Canada	18	18	17	16	Morocco	8	7	6	5
Chile	15	15	15	12	Netherlands	18	16	15	16
China	15	16	16	16	Nicaragua	3	4	3	6
Colombia	9	10	9	10	Norway	18	19	18	19
Croatia	9	10	10	8	Pakistan	3	4	4	6
Cyprus	13	12	12	11	Panama	9	9	9	9
Czech Republic	14	11	11	10	Peru	9	8	8	10
Denmark	18	18	17	17	Philippines	7	4	4	5
Dominican Republic	5	9	7	6	Poland	13	11	11	10
Ecuador	1	4	4	5	Portugal	7	10	11	11
Egypt, Arab Rep.	6	6	6	5	Russian Federation	11	10	9	8
El Salvador	7	5	5	5	Saudi Arabia	15	16	17	16
Estonia	14	14	14	13	Singapore	18	18	19	22
Finland	18	17	16	14	Slovenia	14	15	14	11
France	18	14	15	14	South Africa	12	10	10	10
Georgia	6	10	10	9	Spain	16	14	14	13
Germany	18	16	16	16	Sweden	18	19	19	17
Greece	2	6	8	8	Thailand	11	13	11	10
Hong Kong	17	18	17	19	Tunisia	11	10	11	10
Hungary	9	9	11	11	Turkey	7	6	7	9
Iceland	9	11	11	15	Ukraine	4	6	7	4
Indonesia	8	7	6	6	United Kingdom	18	18	16	16
Ireland	8	12	14	15	United States	18	20	20	19
Israel	14	14	16	14	Uruguay	8	10	9	8
Italy	16	13	13	10	Venezuela, RB	5	5	5	5
Jamaica	3	3	4	6	Vietnam	5	5	5	5
Japan	16	15	15	15					

According to the results, it is impractical to rate counties make use of linear regression method. As it shows in the Table 5.1 there is a gap between Moody's rating and linear regression rating methods. For example Argentina rating score is 3 (by Moody's) but regression analysis results show higher score which is 8, it happened for some other countries such as Ecuador, Greece, Iceland, Ireland, Nicaragua, Latvia, Pakistan, Portugal, Ukraine, and so on. Some of those countries were faced economic crisis in recent years. Also there are some examples for reverse part; some of countries have received high rating score by Moody's agency but the regression analysis results are lower, those countries are Armenia, Brazil, Czech Republic, Netherlands, Poland, Russian Federation, and Spain. Most of these countries are member of European Union. Although, linear regression method is able to rate some countries equivalent to Moody's rating, but it could not find the similar rating with Moody's for most of the countries. Correlation between criteria and nonlinearity relation are two main reasons that cause this dilemma.

An additional to the mentioned results, we can make one more conclusion based on the regression outcomes. Enter and Backward regression methods give more reasonable rating score rather than Forward regression method.

7.1.1 Correlation between Criteria

One of the reason that linear regression model cannot rate countries properly is correlation between criteria. For example, several criteria which are utilized in the analysis and they have high coefficient are listed below:

a. General government total expenditure
b. Gross domestic product (GDP)
c. Unemployment rate
d. Current account balance (% of GDP)
e. Population growth (annual %)
f. Population ages 0-14 (% of total)

As it clear gross domestic products directly depend on general government total expenditure, or current account balance is affected by gross domestic product and general government total expenditure. Furthermore, we know that unemployment rate is indirectly is affected by population growth and GDP. Based on the correlation analysis, 42.07% of criteria have absolute correlation greater than 0.2 with other criteria and 7.36% of criteria have absolute correlation with others more than 0.5. Accordingly, we can say that at least there are two or more criteria that have high correlation between each other.

7.1.2 Nonlinearity Relation between Regressors and Response

In this study we made an assumption and we assumed that criteria (regressors) and rating score (response) have linear relation. Unfortunately, this assumption is rejected after some regression analysis, and it seems that there is nonlinear relation between regressors and response. However, the type of nonlinear relationship function does not identify by any research up to now.

7.2 Clustering Method for Classification Countries

7.2.1 K-mean Clustering

One of the well known techniques in clustering is K-mean method that is utilized for country classification. K-mean clustering is a classification method that aims to separate N objects into K clusters in which each object belongs to the class with the nearest mean. Generally countries are divided to four main classes, as explained in Chapter 4. Therefore, we decide to use K-mean clustering to classify countries in four classes. The following table shows the clustering classification and WB classification for countries. Results in Table 7.2 confirm that there are many misclassifications within K-mean analysis.

For instance, some countries are categorized in higher classes by WB classification compare to K-mean classification, such as Botswana, New Zealand, Oman, Slovenia, Spain, and Switzerland. Those countries have close relation with USA or they are member of European Union. Besides, there are countries such as Croatia, Ecuador, Mauritius, Moldova, Mongolia, Hungary, Iceland, Russia, Thailand, Ukraine, and Uruguay that are classified by WB in lower classes compare to K-mean classification.

Table 7.2. K-mean Analysis for Classification Countries

Country	WB	K-mean	Country	WB	K-mean	Country	WB	K-mean	Country	WB	K-mean
Angola	C2	C2	Malta	C3	C3	El Salvador	C2	C2	Russia	C2	C3
Australia	C4	C4	Mauritius	C2	C3	Estonia	C3	C3	Senegal	C2	C2
Austria	C4	C4	Mexico	C2	C2	Finland	C4	C4	Slovakia	C3	C3
Belgium	C4	C4	Moldova	C2	C3	France	C4	C4	Slovenia	C4	C3
Bolivia	C2	C2	Mongolia	C2	C3	Germany	C4	C4	Spain	C4	C3
Botswana	C3	C2	Morocco	C2	C2	Guatemala	C2	C2	Sri Lanka	C2	C3
Brazil	C2	C2	Netherlands	C4	C4	Honduras	C2	C2	Sweden	C4	C4
Cambodia	C2	C2	New Zealand	C4	C3	Hungary	C2	C3	Switzerland	C4	C3
Canada	C4	C4	Nicaragua	C2	C2	Iceland	C2	C4	Thailand	C2	C3
China	C4	C3	Norway	C4	C4	India	C2	C2	Tunisia	C2	C2
Colombia	C2	C2	Oman	C3	C2	Indonesia	C2	C2	Turkey	C2	C2
Croatia	C2	C3	Papua New Guinea	C2	C2	Italy	C4	C4	Ukraine	C2	C3
Denmark	C4	C4	Peru	C2	C2	Japan	C4	C4	United Kingdom	C4	C4
Dominican Republic	C2	C2	Poland	C3	C3	Jordan	C2	C2	United States of America	C4	C4
Ecuador	C1	C2	Portugal	C2	C3	Latvia	C2	C3	Uruguay	C2	C3
						Lithuania	C2	C3	Venezuela	C2	C2

7.2.2 Support Vector Machine (SVM)

In a few words, Support Vector Machine (SVM) is a technique that use support vector to divide data in two separate classes. Generally, the method applied for classification of data.

Cristianini and Shawe-Taylor in 2000 published a book which explained introduction about SVM, later in 2004 they published another book which have same content and clarified the application area for SVM. Nevertheless, the result of SVM for classification of countries was unsuccessful. It does not mean that the method cannot be applied for classification of countries. However considering available data the method is unable to classify data appropriately. As a case in point, some countries such as USA, Switzerland, Japan, France, and Germany that always are classified in high income countries (C4) categorized in low class (C1) by SVM classification.

Table 7.3. Support Vector Machine Classification Results Compare to WB Classification

Country	WB	SVM	Country	WB	SVM
Angola	C2	C2	Lithuania	C2	C2
Australia	C4	C1	Mauritius	C2	C2
Austria	C4	C1	Moldova	C2	C2
Bolivia	C2	C2	Netherlands	C4	C1
Cambodia	C2	C2	New Zealand	C4	C1
Canada	C4	C1	Nicaragua	C2	C2
China	C4	C1	Norway	C4	C1
Colombia	C2	C2	Oman	C3	C2
Croatia	C2	C2	Papua New Guinea	C2	C2
Denmark	C4	C1	Peru	C2	C2
Dominican Republic	C2	C2	Poland	C3	C2
Estonia	C3	C2	Portugal	C2	C2
Finland	C4	C1	Russia	C2	C2
France	C4	C1	Senegal	C2	C2
Germany	C4	C1	Slovakia	C3	C2
Guatemala	C2	C2	Slovenia	C4	C1
Honduras	C2	C2	Spain	C4	C1
Iceland	C2	C2	Switzerland	C4	C1
India	C2	C2	Turkey	C2	C2
Indonesia	C2	C2	Ukraine	C2	C2
Italy	C4	C1	United States of America	C4	C1
Japan	C4	C1	Venezuela	C2	C2

Chapter 8

CONCLUDING REMARKS

In the first part of this thesis, we attempt to find the most important criteria affecting the World Bank's country classification. To reach this goal, the MHDIS method was used to develop mathematical models. The training set was gathered and filtered, and the mathematical models were run on LINGO 12.0 and Xpress software package to discover significant criteria and to identify misclassified countries. In the last step, a test set from the database was developed to verify and validate the models, verify the results and complete the study.

The results of the MHDIS method, discussed in Chapter 5, and it is shown that based on the analysis the most important criteria are GDP growth (annual %), GNI per capita (PPP, current international $), gross capital formation (% of GDP), inflation (GDP deflator, annual %), mobile cellular subscriptions (per 100 people), and population growth (annual %). Subsequently, we determined the maximal gap (***S***) between classes and generalized the model to four classes instead of two. In the last step of the MHDIS method, the numerical methods determined by the training set were applied to a test set of countries, which verified and validated our model, illustrating that the method is powerful, as only four out of 39 countries were misclassified. In addition, LAD is applied for classification of countries. LAD determines a set of constraints or pattern to separate the two classes. A pattern is perfect if it satisfies all objects on one side and if no object of the other side satisfies its constraints. Several perfect patterns separated upper-middle, medium and low-income countries from the high-income countries in the training set. A perfect pattern does not exist in the opposite direction. There were some good-quality patterns. The significant constraints with a crucial effect on separation are based on criteria 15 (mobile cellular subscription), 14 (military expenditure), and 4 (fertility rate). The use of criteria 4 and 14 has a mutually exclusive character, and all patterns use criterion 15. The patterns based on 15 and 14, i.e., mobile cellular subscription and military expenditure, are the most robust, and provide a solid basis to determine whether a country belongs to the high-income category. Another valuable result from the LAD analysis was the number of non-high income countries (10, 26, 31, 32, 38, 40, and 41) with the potential to become high-income countries, meaning that these countries could belong to the high-income countries because they satisfy the patterns of the high-income countries with a prevalence of 100%.

Although the World Bank classification is based on GNI per capita, GNI per capita has only a marginal role in the classification. The more important criteria are GDP growth (annual %), GNI per capita (PPP, current international US dollar), gross capital formation (% of GDP), inflation (GDP deflator, annual %), mobile cellular subscriptions (per 100 people), population growth (annual %), military expenditure (annual %), and fertility rate (total birth per woman). As a result, we are able to classify countries with two different methods.

In Chapter 6 of thesis we found that Moody's rating system of sovereign debts is highly non-linear system containing subjective elements. The same is true for any other rating system of sovereign debts. Further difficulty is that their method is secret. Therefore the agents of the financial markets cannot justify of refute if it is necessary of the announced changes in the rating.

This Chapter provides an approximate grading system which is based on public information of World Bank and International Monetary Fund databases. The method is elaborated by Logical Analysis of Data methodology. It is summarized in a Decision Tree Chart. It does not need to make any calculation, however, just the comparison of the data of countries by certain threshold values. Thus, it can be used easily by everybod y.

Finally in Chapter 7 we discussed those methodologies that are used for classification or rate countries in this study but they are impotent to give accurate results in those fields.

REFERENCES

Alesina, A., Broeck, M. D., Prati, A., & Tabellini, G. (1992). Default Risk. *Economic Policy*. 15, 427-463.

Alfonso, A. (2003). Understanding the Determinant of Sovereign Debt Ratings: Evidence for the two leading Agencies. *Journal of Economic and Finance*. 27, 56-74.

Altman, E. I., & Rijken, H. A. (2004). How Rating Agencies Achieve Rating Stability. *Journal of Banking and Finance*. 28, 2679-2714.

Altman, E. I., Saunders A. (1997). Credit Risk Measurement: Developments over the Last 20 Years. *Journal of Banking and Finance*. 21 (11-12), 1721-1742.

Boros, E., Hammer, P. L., Ibaraki, T., & Kogan, A. (1997). Logical Analysis of Numerical Data. *Mathematical Programming*. 79, 163-190.

Boros, E., Hammer, P.L., Ibaraki, T., Kogan, A., Crama, Y., Ibaraki, T., & Makino K. (2009). Logical Analysis of Data: Classification with Justification. RUTCOR Research Report (RRR5), Operation Research Rutgers University, New Jersey, USA.

Boros, E., Hammer, P.L., Ibaraki, T., Kogan, A., Mayoraz, E., & Machnick, I. (2000). An Implementation of Logical Analysis of Data. *IEEE transaction on knowledge and data mining*. 12, 292-306.

Brewer, T. L., & Rivoli, P. (1990). Politics and Perceived Country Creditworthiness in International Banking. *Journal of Money, Credit and Banking*. 22, 357-369.

Cantor, R., & Packer, F. (1996). Determinants and Impact of Sovereign Credit Ratings. *Federal Reserve Bank of New York Economic Policy Review*. 2, 37-53.

Cantor, R., & Packer, F. (1994). The Credit Rating Industry. *Federal Reserve Bank of New York Quarterly Review*. 1, 1-26.

Christensen, J. H. E., Hansen, E., & Lando, D. (2004). Confidence sets of continuous-time rating transaction probabilities. *Journal of Banking and Finance*. 28, 2575–2602.

Cosset, JC., & Roy, J. (1991). The Determinants of Country Risk Ratings. *Journal of International Business Studies*. 22, 135-142.

Cosset, JC., Siskos, Y., & Zopounidis, C. (1992).Evaluating Country Risk: a Decision Support Approach. *Global Finance Journal*. 3, 79-95.

Crama, Y., & Hammer, P. L. (2011). Boolean Functions: Theory, Algorithms, and Applications. Cambridge University Press.

Cristianini, N., & Shawe, T. J. (2000). An introduction to support vector machines: and other kernel-based learning methods. Cambridge University Press, New York, USA.

Dahl, H., Meeraus, A., & Zenios, S. A. (1933). Some Financial optimization models: I. Risk management. Financial optimization. Cambridge: Cambridge University Press.

Doumpos, M., & Zopounidis, C. (2001). Assessing financial risks using a multicriteria sorting procedure: the case of country risk assessment. *International Journal of Management Science*. 29, 97-109.

Edwards, S. (1984). Foreign Borrowing and Default Risk: An Empirical Investigation, 1976- 80. *American Economic Review*. 74, 726-734.

Fayyad, M. U., Piatesky, S. G., Smuth, P., & Uthurusamy, R. (1996). Advances in Knowledge Discovery and Data Mining. AAAI Press.

Ferri, G., Liu, L. G., & Stiglitz, J. (1999). The Procyclical Role of Rating Agencies: Evidence from the East Asian Crisis. *Economic Notes*. 3, 335-355.

Finnerty, D. J. (1988). Financial engineering in corporate finance: an overview. *Financial Management*. 17, 14-33.

Guha, S., Rastogi, R., & Shim, K. (1998). CURE: An Efficient Clustering Algorithm for Large Databases. In Proceedings of the ACM SIGMOD Conference.

Hammer, A. B., Hammer, P. L., & Muchnock, I. (1996).Logical Analysis of Chinese Productivity Pattern. RUTCOR Report RR-1-96. Rutgers University.

Hammer, P.L., Kogan, A., & Lejeune, M. A. (2007). Reverse-Engineering Country Risk Rating: A Combinatorial Non-recursive Model. The State University of New Jersey and George Washington University, March.

Haque, N., Mark, N., & Mathieson, D. (1998). The Relative Importance of Political and Economic Variables in Creditworthiness Ratings. IMF Working Paper 98/46, April.

Haque, N., Kumar, M., Mark, N., & Mathieson, D. (1996). The Economic Content of Indicators of Developing Country Creditworthiness.IMF Staff Papers 43: 688-724.

Haque, N., Mark, N., & Mathieson, D. (1997). Rating the Raters of Country Creditworthiness. *Finance & Development*. 34, 10-13.

Hlakidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems*. 17, 107-145.

Hu, Y. T., Kiesel, R., & Perraudin, W. (2002). The Estimation of Transition Matrices for Sovereign Credit Ratings. *Journal of Banking and Finance*. 26, 1383-1406.

International Monetary Fund (IMF). World Economic Outlook Databases: http://www.imf.org/external/ns/cs.aspx?id=28

Jain, A. K., Murty, M. N., & Flyn, P. J. (1999). Data Clustering: A Review. ACM Computing Surveys. 31(3), 264–323.

Kim, Y., & Sohn, S. Y. (2008). Random Effect Model for Credit Rating Transitions. *European Journal of Operation Research*. 184, 561-573.

Lando, D., & Skodeberg, T. M. (2002). Analyzing Ratings Transitions and Rating Drift with Continuous Observations. *Journal of Banking and Finance*. 26, 423–444.

Lee, S. (1993). Are the Credit Ratings Assigned by Bankers Based on the Willingness of LDC Borrowers to epay?. *Journal of Development Economics*. 40, 349-359.

Lingo 12.0 software package.

Mirzaei, N., & Vizvari, B. (2011). Reconstruction of World Bank’s Classification of Countries: *African Journal of Business Management*. 5(32), 12577-12585.

Moody’s Credit Outlook and Analytical Reports. http://www.moodys.com/researchandratings

Moon, C., & Stotsky, J. (1993). Testing the Differences between the Determinants of Moody's and Standard and Poor's Ratings. *Journal of Applied Econometrics*. 8, 51-69.

Mulvey, J. M., Rosenbaum, D. P., & Shetty, B. (1997). Strategic Financial Risk Management and Operation Research. *European Journal of Operation Research*. 97, 1-16.

Mulvey, J. M. (1996). Financial Planning Via Multi-Stage Stochastic Optimization. *Journal Computers & Operations Research*. 31(1), 1-20.

Oral, M., Kettani, O., Cosset, J. C., & Daouas, M. (1992). An Estimation Model for Country Risk Rating. *International Journal of Forecasting*. 8, 583-93.

Sohn, S. Y., & Choi, H. (2006). Random Effects Logistic Regression Model for Data Envelopment Analysis with Correlated Decision Making Units. *Journal of the Operational Research Society*. 57, 552–560.

SPSS 12.0 software package.

Taylor, J., & Cristianini, N. (2004). Kernel Method for Pattern Analysis. Cambridge University Press, New York, USA.

The World Bank (WB) Database: http://data.worldbank.org/

Zopounidis, C., Doumpos, M. (2000). Multicriteria Sorting Methods. Encyclopedia of optimization. Academic Publishers, 2000.

Zopounnidis, C., Doumpos, M. (2002). Multicriteria classification and sorting method: A literature review. *European Journal of Operational Resaerch*. 138, 229-246.

Xpress Optimiozation Suit: http://www.fico.com/en/Products/DMTools/Pages/FICO-Xpress-Optimization-Suite.aspx