COUNTRY RISK CLASSIFICATION

AND

MULTICRITERIA DECISION-AID

COUNTRY RISK CLASSIFICATION

AND

MULTICRITERIA DECISION-AID

By

XIJUN WANG, M.SC.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2004)   McMaster University

COMPUTING AND SOFTWARE   Hamilton, Ontario

TITLE: Country Risk Classification and Multicriteria Decision-Aid

AUTHOR: Xijun Wang, B.SC., M.SC.

SUPERVISOR: Dr. Jiming Peng

NUMBER OF PAGERS: xvi, 89

Abstract

Country risk is an important concern in international business. Country risk classification refers to determining the risk level at which a country will not repay its international debt. Traditionally, country risk classification resorts to statistics methods such as discriminant analysis. In the past two decades, the so-called multicriteria decision aid (MCDA) methods have been proved to enjoy better performance than the standard statistics methods. Nevertheless, the performance of the MCDA methods is still far away from satisfactory and can be improved significantly.

The better performance of several MCDA methods, such as UTADIS (UTilités Additives DIScriminantes) and MHDIS (Multigroup Hierarchical DIScrimination), is achieved by exploiting the rater's background knowledge. In the standard MCDA model, we assume that the criterion function for every factor is monotone and all the factors are independent. Then, we approximate the impact of every factor and use the sum of the corresponding criterion functions to determine the risk level of a country. By discretizing the feasible domain of the factor, the MCDA method solves a linear program to find a classifier for country risk classification.

This thesis tries to enhance the capability of MCDA methods by allowing a class of non-monotone criteria: the unimodal ones. For this purpose,

we developed an integer quadratic (non-convex) program for general unimodal criteria. Further, if we restrict ourselves to convex or concave unimodal criteria, then we can still use a linear program to find a classifier. For the case where all the factors are correlated, a simple quadratic form of aggregation is proposed to deal with it. Compared with the original UTADIS model, our generalized model is more flexible and can deal with more complex scenarios.

Finally, our generalized model is tested based on cross-validation and our experiment is carried out under the AMPL+sovers environment. Promising numeric results indicate that except for its theoretical advantages, our generalized model exhibits practical efficiency and robustness as well.

Acknowledgements

Finally and foremost, I thank my wife Lili Wang and my son Suyu Wang for the love and happiness they gave me.

Contents

List of Figures

List of Tables

List of Notations

Chapter 1

Introduction

In this chapter, we first give a brief introduction to country risk and country risk analysis. Then, we describe the motivation that sparks this thesis and outline the organization of the thesis.

1.1 Country Risk

Country risk refers to uncertainties or potential risk related to investing in or loaning to different countries in international business. There are various reasons that might change the capability and willingness of a country to repay its international debt. For example, higher taxes or tariffs, limited currency conversion, inflation and currency depreciation, economic recession, workers striking, war etc. These potential risks make investment in the country less

1

profitable even losing money, as investors may not get their repayments on time.

Many institutions have been providing country risk rating services for a long time. *Euromoney, Institutional Investor* and *International Country Risk Guide*, for example, started publishing their country risk ratings periodically as early as 1980's. Among those rating providers, *Standard & Poor's* and *Moody's* are the most prominent ones and both of them have a history over 100 years.

The profusion of country risk providers is a direct consequence of the trend of globalization and internationalization. As more and more countries involve in international business, more and more country risks arise. For example, the total amount of cross-border interbank lending increased from 174.4 billion US dollars in 1971 [IMF, 1971] to 6262.7 billion US dollars in 1994 [IMF, 1995], which shows a dramatic increase of international loan. The total Net Foreign Direct Investment around the world increased from 6 billion US dollars in 1970 to 400 billion in 2000 [The World Bank, 2002]. The number of transnational corporations across the globe has increased from 15,000 in 1980 to 64,000 in 2003, with their total foreign affiliates increased from 35,000 to 850,000 during the same period [UNCTAD, 2002]. These corporations have to take care of their transnational assets, depend on country risk ratings to

make their management decisions. Thus, a vast demand of reliable country risk ratings exists among bankers and investors who are usually unfamiliar with foreign countries and have to look for help from intermediate agencies.

The ability to manage country risk may be different for different banks, which makes no external rating perfectly suitable for every bank. Thus, many banks have to build an internal country risk evaluation system by themselves, not simply referring to an external risk rating provider. This diversity of requirement of different users is another strong motivation for country risk research besides the vast amount of users.

Country risk ratings impact both sides of the interested bankers (investors) and the rated countries. Investors and bankers make their decisions on whether to invest in or loan to a country according to its risk rating. For the rated countries, their ratings influence the interest rates that investors and bankers expect from them, i.e., a country with lower risk will usually be charged a lower interest rate.

1.2   Country Risk Evaluation

Usually, the country risk of one country is represented by a single index, which shows the degree of the overall risk to invest in or loan to this country. There are two types of indices used to represent the degree of country risk, discrete

and continuous. For the discrete type, several risk levels are predefined and every country is in one level. The number of risk levels may vary from two, simply high or low, to twenty, from 1 to 20. Table 1.1 lists the levels used in *Standard & Poor's* and *Moody's*. Table 1.2 lists the ratings provided by *Standard & Poor's* [Alexe et al., 2003] in December 1998. For the continuous type, a number between 0 and 1 is used to indicate the magnitude of risk to loan to or invest in a country. The rating provided by *International Country Risk Guide* is an example of using continuous type of representation for country risk.

The single index representing the degree of country risk is a composition of various factors about the country. The main interested factors are political and economic-financial ones, and the total number of factors used may vary from less than ten to more than twenty. Some providers publish separate ratings on the risks associated with political, economic and financial aspects of each country. Political risk is often regarded as unwillingness to pay while economic and financial risk are thought of as inability to pay. Table 1.2 lists some possible factors and values of some countries on these factors. It should be noticed that there is no agreement about whether political factors have significant impact on country risk evaluation [Brewer and Rivoli, 1990; Haque et al., 1998].

Table 1.1: Country Risk Levels Used by *Standard & Poor's* and *Moody's*

Risk Level Group	S&P's	Moody's	Description
Investment Levels	AAA	Aaa	Capability to meet financial commitments ranges from extremely strong to strong.
	AA+	Aa1	
	AA	Aa2	
	AA-	Aa3	
	A+	A1	
	A	A2	
	A-	A3	
	BBB+	Baa1	Default is unlikely due to adequate protection parameters, while it lack of certain protective elements.
	BBB	Baa2	
	BBB-	Baa3	
Speculative Levels	BB+	Ba1	Payment capability is a little vulnerable due to more uncertainty and less protection.
	BB	Ba2	
	BB-	Ba3	
	B+	B1	Payment capability is more vulnerable, but it's currently still able to meet commitments.
	B	B2	
	B-	B3	
Non-Investment Levels	CCC+	Caa1	Payment vulnerability is currently evident, and default becomes likely.
	CCC	Caa2	
	CCC-	Caa3	
	CC	Ca	
	C	C	
	SD		Default already occurred, either partially or entirely.
	D		

Table 1.2: An Example of Country Risk Rating and Related Factors (1998)

Level	Country	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
BB	Argentina	0.605	7.37	202.7	0.4	4.0	0.844	8.0	52.2	3.9	0.9	23.3	57.7	5.7
AA+	Canada	0.467	3.18	637.8	1.7	5.6	0.940	5.9	33.1	3.3	1.0	82.3	0.0	1.0
BBB-	Egypt, Arab Rep.	0.676	4.74	64.6	2.0	4.7	0.642	10.2	34.4	5.6	4.2	42.6	9.8	10.0
AAA	France	0.526	5.07	583.9	2.2	5.9	0.928	6.0	32.7	3.4	0.7	49.6	0.0	2.2
AAA	Germany	0.458	5.47	567.1	2.3	4.6	0.925	4.8	28.3	2.1	0.9	56.2	0.0	1.9
BB	India	0.645	4.37	21.9	0.6	3.2	0.577	15.9	32.5	6.0	13.2	25.6	20.6	5.7
BB+	Korea, Rep.	0.399	4.00	432.6	2.7	4.1	0.882	12.8	31.6	-6.7	7.5	86.0	12.9	5.1
BB	Mexico	0.627	5.18	103.5	0.4	4.2	0.796	3.8	54.6	5.0	15.9	63.5	20.0	2.4
BBB-	Poland	0.479	3.19	227.6	0.7	5.4	0.833	5.7	31.6	4.8	11.7	61.6	9.8	6.1
CCC-	Russian Federation	0.461	1.75	198.5	1.0	3.5	0.781	17.9	45.6	-4.9	27.7	57.3	12.4	1.7
BB+	South Africa	0.617	3.34	120.4	0.6	6.1	0.695	5.2	59.3	0.7	6.9	50.6	12.2	1.8
AAA	United Kingdom	0.535	5.47	554.1	1.8	4.7	0.928	7.2	36.0	2.6	3.4	53.9	0.0	0.9
AAA	United States	0.518	3.82	665.3	2.5	5.0	0.939	15.8	40.8	4.4	1.6	23.9	0.0	1.3

(1) Age dependency ratio (dependents to working-age population); (2) GDP per unit of energy use (PPP $ per kg of oil equivalent); (3) Telephone mainlines (per 1,000 people); (4) Research and development expenditure (% of GNI); (5) Public spending on education, total (% of GDP); (6) Human Development Index; (7) Military expenditure (% of central government expenditure); (8) GINI index (%); (9) GDP growth (annual %); (10) Inflation, consumer prices (annual %); (11) Trade (% of GDP); (12) Total debt service (% of exports of goods and services); (13) Gross international reserves in months of imports.

There are two typical ways to composite a single index representing the country risk. The first is based on a rating survey of country risk experts, purely qualitative. For example, a group of experts are asked to evaluate every factor of a country on a scale from 0 to 100, then the scores on every factor are averaged. The weighted sum of these averaged scores is consequently used to construct the country risk index. In this way, the weight on each factor is predetermined and will be used for a long time. The second way is based on classification techniques using a mixture of quantitative and qualitative factors. The rater chooses a classification model, compiles data on various factors and on risk levels of each country, then uses the data to determine the parameters in the model and possibly to point out the most relevant factors. The data on qualitative factors can be collected in the same manner as in the first way. Once the parameters of the model is determined, the rater can compute the risk index for each country. Table 1.2 is an example of quantitative factors.

Country risk classification in the literature refers to the problem of determining in which risk level a given country will not repay its international debt. For example, the risk levels of countries listed in Table 1.3 have to be determined according to available information on the listed factors.

Table 1.3: An Example of Country Risk Classification

level	country	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
?	Brazil	0.538	6.58	120.5	0.8	4.6	0.757	5.2	58.5	0.2	3.2	17.3	74.8	5.3
?	China	0.479	3.81	69.6	0.1	2.2	0.726	25.1	44.7	7.8	-0.8	39.2	8.6	9.9
?	Indonesia	0.566	4.34	27.0	0.1	1.4	0.684	4.8	34.3	-13.1	57.6	96.2	31.7	5.2
?	Israel	0.609	5.84	471.0	3.4	7.7	0.896	19.7	35.5	2.6	5.4	72.3	0.0	6.4
?	Japan	0.455	6.15	534.0	2.8	3.5	0.933	4.1	24.9	-1.1	0.7	19.5	0.0	5.2
?	Malaysia	0.634	3.90	201.5	0.4	5.2	0.782	7.4	49.2	-7.4	5.3	209.5	7.2	4.3
?	Netherlands	0.467	4.89	592.4	2.0	4.9	0.935	6.3	32.6	3.7	2.0	116.3	0.0	1.4
?	New Zealand	0.523	3.96	492.5	1.2	7.2	0.917	3.7	36.2	0.0	1.3	60.7	0.0	2.6
?	Philippines	0.692	6.87	34.1	0.2	3.2	0.754	7.9	46.1	-0.6	9.7	110.9	10.8	3.1
?	Singapore	0.409	3.49	459.9	1.1	3.1	0.885	20.0	42.5	0.1	-0.3	293.8	0.0	7.4
?	Spain	0.460	6.00	413.7	0.8	4.5	0.913	6.1	32.5	4.3	1.8	54.6	0.0	4.0
?	Turkey	0.538	5.87	265.2	0.5	3.0	0.742	13.4	40.0	3.1	84.6	52.2	24.0	4.1
?	Venezuela, RB	0.643	2.47	111.5	0.4	5.0	0.770	6.3	49.1	0.2	35.8	41.0	28.2	7.2

The name of the factors are the same as that in Table 1.2.

1.3   Country Risk Analysis

The research on country risk analysis dates back to as early as 1950's. After the second oil price shock of 1979-1980, country risk has caught more and more attention. Various models have been proposed to predict the potential risk related to countries that investors and bankers are interested in.

Before 1980's, most researchers considered only two risk levels, and multivariate statistical techniques were the main tools. For example, discriminant analysis [Anderson and Bahadur, 1962; Frank and Cline, 1971; Somerville and Taffler, 1995], logit analysis [Feder and Just, 1977], regression analysis [Alexe et al., 2003] have all been applied in country risk classification. The early research provided us many important factors for country risk classification. In Chapter 2, we will give a brief review on classical discriminant analysis techniques. A comprehensive survey of the statistical techniques used in country risk classification can be found in Saini and Bates [1984].

However, since most statistical techniques use explicitly or implicitly assumptions on the distribution of the factor values, which is typically unknown in practice, no bank exclusively rely on statistical methods to determine the country risk levels. Another limitation of statistical techniques is that, qualitative factors are difficult to use, e.g., in discriminant analysis.

9

Since 1980's, *multicriteria decision aid* methods (MCDA[1]) for country

risk classification have been proposed and proved to be more reliable than sta-

tistical methods [Doumpos and Zopounidis, 2001; Doumpos et al., 2001b]. For

example, UTADIS (UTilités Additives DIScriminantes) methods [Doumpos et

al., 2001b; Spathis et al., 2002; Zopounidis and Doumpos, 1999] and MHDIS

(Multigroup Hierarchical DIScrimination) method [Doumpos and Zopounidis,

2002; Doumpos et al., 2001a,b] are typical MCDA methods. Comparing with

statistical methods, they are not restricted by statistical assumptions and can

easily deal with qualitative factors. These methods incorporate the back-

ground knowledge of the rater by using criterion functions that characterize

the relation between values on a single factor and the country risk. Chapter 2

will also give the details in UTADIS and MHDIS classification methods.

For the country risk classification methods mentioned above, it is nec-

essary to predefine the possible risk levels. This is usually done by experience.

Alternatively, Smet and Gilbart [2001] suggested an approach to help bankers

decide how many risk levels are suitable to them.

---

[1]The MCDA model can also provide solutions for the other two types of multicriteria
decision problems: choice and ranking problems. See Vincke [1992] and Zopounidis and
Doumpos [2002] for more about general MCDA methods.

1.4   Motivation and Organization

Many aspects of the existing country risk rating systems still have to be improved, as suggested in Alexe et al. [2003]. One of them is the comprehensibility of the rating results. A well-known example is described in Alexe et al. [2003]: Japan, as the second largest economic body in the world, was rated by some "big rating agencies" in the same risk level in 2001 as an African country (Botswana) to which Japan was providing assistance. Such uncomprehensible ratings make it clear that more attention should be paid on the reliability of country risk rating system.

We note that, MCDA methods [Doumpos and Zopounidis, 2002; Doumpos et al., 2001b; Spathis et al., 2002; Vincke, 1992; Zopounidis and Doumpos, 1999] require each criterion function to be monotone. Take for example, the criterion function of financial reserve: the more financial reserve a country holds, the larger the criterion function value is. This definitely restricts the application of the MCDA models as the criteria on some factors related to country risk might not be monotone. The main purpose of this thesis is to extend MCDA methods so that the extended MCDA can deal with non-monotone criteria. In particular, we discuss how the MCDA methods can be extended to deal with unimodal criteria, a typical kind of non-monotone criteria.

The rest of this thesis is organized as follows. In Chapter 2, a review

of statistical methods and MCDA methods for country risk classification are provided. Then, the MCDA methods are extended to deal with unimodal criteria in Chapter 3. Chapter 4 extends one class of MCDA methods by considering the dependency among various factors. Chapter 5 describes the testing data and reports out experiments. Chapter 6 gives conclusions and some suggestions for further work.

Chapter 2

From Statistical to MCDA

Methods

In this chapter, we review the typical statistical methods for country risk classification and introduce the multicriteria decision-aid methods UTADIS and MHDIS.

2.1 Multivariate Statistical Methods

In country risk literatures, most statistical methods consider only two-level classification. In this section, we first give a brief introduction of statistical methods for country risk classification based on discriminant analysis and logit analysis.

Let π_1 be the population of countries reaching the limit of their debt servicing capacity, and π_2 be the contrary. Let x be an observation over n indicators of the country risk of one country. $Z = f(x) \in \mathbb{R}$ stands for a composite index of these n indicators, and $Z^* \in \mathbb{R}$ represents a critical value. For the given observation x, if $Z \geq Z^*$ the country will be classified as coming from π_1, while if $Z < Z^*$ the country will be thought of as coming from π_2. To build such a classification model, our task is to find an appropriate function $f(x)$ and a suitable threshold Z^* such that the model will have a good performance.

2.1.1 Bayesian Discriminant Analysis

Define *type I error* of classification as misclassifying a country from π_1 as from π_2, and *type II error* as misclassifying a country from π_2 as from π_1. For a given discriminant function $f(x)$ and threshold Z^*, denote $p(\mathrm{I})$ and $p(\mathrm{II})$ the probability of making type I and type II error respectively, i.e.

$$p(\mathrm{I}) = \Pr\{f(x) < Z^* \mid x \in \pi_1\} \tag{2.1a}$$

$$p(\mathrm{II}) = \Pr\{f(x) \geq Z^* \mid x \in \pi_2\}. \tag{2.1b}$$

Roughly speaking, Bayesian discriminant analysis method tries to minimize probabilities of both types of errors. Some specific objectives that we might want to minimize are

14

(1) the probability of one error when the other is given;

(2) the maximum probability of the two type of errors;

(3) the total probability of misclassification when a priori probability of each population is given.

Definition 2.1 *One classifier $(f_1(x), Z_1^*)$ is* better *than another $(f_2(x), Z_2^*)$ if the corresponding classification errors satisfy the following condition*

$$p_1(I) \le p_2(I) \ and \ p_1(II) \le p_2(II),$$

and at least one of these inequalities is strict, where $p_i(I)$ and $p_i(II)$ are the two types of errors for classifier $i = 1, 2$. A classifier is admissible *if no other classifier is better than it.*

It can be easily seen that the solution to minimizing each of the objective mentioned must be admissible. Because of this, we first discuss the properties of the admissible classifiers.

Admissible Classifiers

Suppose x is normally distributed on both populations. Denote the mean vectors and covariance matrices of π_1 and π_2 as (μ_1, Σ_1) and (μ_2, Σ_2) respectively, where $\mu_1 \ne \mu_2$ and both Σ_1 and Σ_2 are nonsingular, i.e., positive definite.

15

Consider linear discrimination function $f(x) = b^T x$ and threshold c. When x is coming from π_i, $b^T x$ will be a normal distributed univariate random variable with mean $b^T \mu_i$ and variance

$$E[(b^T x - b^T \mu_i)^2] = E[b^T (x - \mu_i)(x - \mu_i)^T b] = b^T \Sigma_i b.$$

$p(\mathrm{I})$ and $p(\mathrm{II})$ becomes

$$
\begin{aligned}
p(\mathrm{I}) &= \Pr\{b^T x < c \mid x \in \pi_1\} \\
&= \Pr\left\{ \frac{b^T x - b^T \mu_1}{(b^T \Sigma_1 b)^{1/2}} < \frac{c - b^T \mu_1}{(b^T \Sigma_1 b)^{1/2}} \mid x \in \pi_1 \right\} \\
&= \Phi\left(\frac{c - b^T \mu_1}{(b^T \Sigma_1 b)^{1/2}} \right) \\
&= 1 - \Phi\left(\frac{b^T \mu_1 - c}{(b^T \Sigma_1 b)^{1/2}} \right)
\end{aligned}
\tag{2.2a}
$$

$$
\begin{aligned}
p(\mathrm{II}) &= \Pr\{b^T x \geq c \mid x \in \pi_2\} \\
&= \Pr\left\{ \frac{b^T x - b^T \mu_2}{(b^T \Sigma_2 b)^{1/2}} \geq \frac{c - b^T \mu_2}{(b^T \Sigma_2 b)^{1/2}} \mid x \in \pi_2 \right\} \\
&= 1 - \Phi\left(\frac{c - b^T \mu_2}{(b^T \Sigma_2 b)^{1/2}} \right),
\end{aligned}
\tag{2.2b}
$$

where $\Phi(\cdot)$ is the cumulative distribution function of standardized normal distribution

$$\Phi(z) = (2\pi)^{-1/2} \int_{-\infty}^{z} e^{-t^2/2} \, dt.$$

Let us define

$$y(\mathrm{I}) = \frac{b^T \mu_1 - c}{(b^T \Sigma_1 b)^{1/2}} \tag{2.3a}$$

$$y(\mathrm{II}) = \frac{c - b^T \mu_2}{(b^T \Sigma_2 b)^{1/2}}. \tag{2.3b}$$

Since $\Phi(z)$ is an increasing function of z, smaller errors $p(\mathrm{I})$ and $p(\mathrm{II})$ correspond to larger arguments of $\Phi(\cdot)$ in (2.2a,b), i.e., larger value of $y(\mathrm{I})$ and $y(\mathrm{II})$. Solving c from (2.3b) and substituting it into (2.3a), we get

$$y(\mathrm{I}) = \frac{b^T \delta - y(\mathrm{II})(b^T \Sigma_2 b)^{1/2}}{(b^T \Sigma_1 b)^{1/2}}, \tag{2.4}$$

where $\delta = \mu_1 - \mu_2$. The right hand side of (2.4) is homogeneous in b with degree 0, thus we can restrict b to be on an ellipse. This proves that $y(I)$ has a finite maximum value with respect to b for any given $y(\mathrm{II})$.

Since $y(\mathrm{I})$ is a decreasing function of $y(\mathrm{II})$ for any given b,

$$y^*(\mathrm{I}) = \max_b y(\mathrm{I}) \tag{2.5}$$

is also decreasing in $y(\mathrm{II})$. It can be easily verified that the classifier (b^*, c^*) corresponding to $(y^*(\mathrm{I}), y(\mathrm{II}))$ is admissible. It's also obvious that the set of classifiers corresponding to $\{(y^*(\mathrm{I}), y(\mathrm{II})) \mid \forall y(\mathrm{II})\}$ contains all possible admissible classifiers.

We should also notice that $y^*(\mathrm{I})$ is a convex function of $y(\mathrm{II})$ while the corresponding $p^*(\mathrm{I})$ is not convex in $p(\mathrm{II})$. This is the reason to work with $y(\mathrm{I})$ and $y(\mathrm{II})$ rather than $p(\mathrm{I})$ and $p(\mathrm{II})$.

The following lemma characterizes the admissible classifiers in general case. The lemma and its proof are attributed to Anderson and Bahadur [1962].

17

Lemma 2.1 *If (b^*, c^*) is an admissible classifier corresponding to $(y^*(I), y(II))$ with $y^*(I) > 0, y(II) > 0$, then there exist $t_1 > 0, t_2 > 0$ such that*

$$b^* = (t_1 \Sigma_1 + t_2 \Sigma_2)^{-1} \delta, \tag{2.6a}$$

$$c^* = b^{*T} \mu_1 - t_1 b^{*T} \Sigma_1 b^* = b^{*T} \mu_2 + t_2 b^{*T} \Sigma_2 b^*. \tag{2.6b}$$

Proof $(y^*(I), y(II))$ is on the following line in y_1-y_2 plane defined by b^*

$$y_1 = \frac{b^{*T} \mu_1 - c}{(b^{*T} \Sigma_1 b^*)^{1/2}} \text{ and } y_2 = \frac{c - b^{*T} \mu_2}{(b^{*T} \Sigma_2 b^*)^{1/2}}, \tag{2.7}$$

where c is parameter. It can be easily seen that there exists an ellipse

$$\frac{y_1^2}{t_1} + \frac{y_2^2}{t_2} = k, \tag{2.8}$$

where t_1, t_2, k are all positive, such that it is tangent to the line defined above at point $(y^*(I), y(II))$.

Now consider the line defined by an arbitrary b. This line is tangent to some ellipse with the same t_1 and t_2 as (2.8) but a different k. Note that the slope of the tangent line of ellipse (2.8) at point (y_1, y_2) is $-\frac{y_1 t_2}{y_2 t_1}$. By equating it to the slope of the line defined by b, we get the tangent point

$$y_1 = \frac{t_1 (b^T \Sigma_1 b)^{1/2} b^T \delta}{t_1 b^T \Sigma_1 b + t_2 b^T \Sigma_2 b}, \ y_2 = \frac{t_2 (b^T \Sigma_2 b)^{1/2} b^T \delta}{t_1 b^T \Sigma_1 b + t_2 b^T \Sigma_2 b}, \tag{2.9}$$

and the corresponding c is

$$c = \frac{t_1 b^T \Sigma_1 b b^T \mu_2 + b^T \Sigma_2 b b^T \mu_1}{t_1 b^T \Sigma_1 b + t_2 b^T \Sigma_2 b}. \tag{2.10}$$

18

The ellipse tangent to the line defined by b is then

$$\frac{y_1^2}{t_1} + \frac{y_2^2}{t_2} = \frac{(b^T \delta)^2}{b^T(t_1 \Sigma_1 + t_2 \Sigma_2)b}. \tag{2.11}$$

The right hand side of (2.11) is homogeneous in b of degree 0, and its maximum in b is achieved when b is equal to the right hand side of (2.6a).

However, k in (2.8) must be equal to this maximum; otherwise, the line defined by $\bar{b} = (t_1 \Sigma_1 + t_2 \Sigma_2)^{-1} \delta$ will be on the upright of $(y^*(\mathrm{I}), y(\mathrm{II}))$ and there will be a point on this line which corresponds to a better classifier. Hence, (2.6a) has to hold. Substituting b^* into (2.10), we get the corresponding c^* as in (2.6b). ∎

Substituting b^* into (2.9), we get

$$y^*(\mathrm{I}) = t_1(b^{*T} \Sigma_1 b^*)^{1/2}, \tag{2.12a}$$

$$y(\mathrm{II}) = t_2(b^{*T} \Sigma_2 b^*)^{1/2}. \tag{2.12b}$$

We see that $(y^*(\mathrm{I}), y(\mathrm{II}))$ is homogeneous in (t_1, t_2) of degree 0. We can then normalize it by restricting

$$\begin{cases} t_1 + t_2 = 1 & \text{for } t_1 > 0, t_2 > 0; \\ t_1 - t_2 = 1 & \text{for } t_1 > 0, t_2 < 0; \\ t_2 - t_1 = 1 & \text{for } t_1 < 0, t_2 > 0. \end{cases} \tag{2.13}$$

For the first case, it can also be verified that $y^*(\mathrm{I})$ is increasing in t_1 (for $0 \leq t_1 \leq 1$) while $y(\mathrm{II})$ is decreasing in t_1.

The following theorem [Anderson and Bahadur, 1962] is the converse of Lemma 1.

Theorem 2.2 *For any t_1 and t_2 such that $t_1\Sigma_1 + t_2\Sigma_2$ is positive definite, the classifier (b^*, c^*) defined by (2.6) is admissible.*

Interested readers are referred to Anderson and Bahadur [1962] for the proof.

This theorem establishes a one-to-one correspondence between the admissible classifiers (b^*, c^*) and the normalized (t_1, t_2) pairs. Thus, the task of searching for an optimal classifier can be accomplished by searching the normalized (t_1, t_2) pairs.

We next describe several different measurements used in the statistics methods for country risk analysis.

(1) Minimizing the probability of one error when the other is given.

Suppose $p(II)$ or equivalently $y(II)$ is given. If $y(II) \geq 0, y^*(I) \geq 0$, then the corresponding (t_1, t_2) can be easily determined since $y(II)$ is decreasing in t_1 and $t_1 + t_2 = 1$. If $y(II) \geq 0, y^*(I) < 0$, $y(II)$ is a decreasing function of t_2 so the corresponding (t_1, t_2) can be determined.[1] The case of $y(II) < 0, y^*(I) \geq 0$ can be similarly handled.

(2) Minimizing the maximum probability of the two types of errors.

Because $y^*(I)$ is strictly decreasing with respect to $y(II)$, it has to be

---

[1]See Anderson and Bahadur [1962] for an exceptional case.

20

$y^*(\text{I}) = y(\text{II}) > 0$ in this case, or equivalently,

$$y^*(\text{I})^2 - y(\text{II})^2 = 0 \Leftrightarrow b^{*T}(t_1^2 \Sigma_1 - (1 - t_1)^2 \Sigma_2)b^* = 0. \tag{2.14}$$

Since $y^*(\text{I})$ is increasing in t_1 and $y(\text{II})$ is decreasing in t_1, a solution to (2.14) can be easily found, from which b^* and c^* can be computed.

(3) Minimizing the probability of misclassification given a priori probability of each population. Let q_1 and q_2 are the a priori probability of population π_1 and π_2, respectively. The probability of misclassification is then

$$q_1 p^*(\text{I}) + q_2 p(\text{II}) = 1 - (q_1 \Phi(y^*(\text{I})) + q_2 \Phi(y(\text{II}))).$$

Minimizing this probability, as a function of t_1, is a non-convex optimization problem.

2.1.2 Logit Analysis

Logit analysis probably is the most widely used statistical method for two-class classification problems. Unlike discriminant analysis approaches which usually need normal distribution assumption on risk indicators, logit analysis assumes that for a given observation x the probability of coming from population π_1 has the form

$$P(x) = \frac{e^{\beta^T x}}{1 + e^{\beta^T x}},$$

21

where β is a vector of parameters to be determined. If we use $y = 1$ to indicate that a country comes from π_1 and $y = 0$ from π_2, the previous probability assumption can also be described as

$$\Pr(y = 0|x) = \frac{1}{1 + e^{\beta^T x}},$$

$$\Pr(y = 1|x) = \frac{e^{\beta^T x}}{1 + e^{\beta^T x}}.$$

Thus, for a set of observations $(x_1, y_1), \cdots, (x_m, y_m)$, the likelihood function L that the observations happened is

$$L = \frac{\prod_{i=1}^m e^{y\beta^T x_i}}{\prod_{i=1}^m (1 + e^{\beta^T x_i})}.$$

The parameters β can then be determined by maximizing this likelihood function.

Once β is determined, the probability function will be used as a discriminant function during the classification procedure, the same as in discriminant analysis. The user can classify a country into the two classes by choosing a critical value of $P(x)$ according to his/her preference.

While logit analysis use different assumptions from discriminant analysis, experimental results showed that there is no significant difference between these two approaches [Somerville and Taffler, 1995].

2.2 Multicriteria Decision Aid Methods

2.2.1 Risk Levels as Preferences

Multicriteria decision-aid methods for country risk classification were first introduced by researchers in decision science [Doumpos et al., 2001b]. From the point view of decision making, the risk level of a country can be regarded as the loan preference of a banker.

Suppose that every country has to be in some risk level and the banker is indifferent between countries in the same level. Then the risk level defines an ordering relation on countries.

Definition 2.2 *For any pair of countries a and b, $a \succ b$ if and only if the country risk of a is lower than the country risk of b; $a = b$ if and only if they are in the same risk level; $a \succeq b$ if and only if $a \succ b$ or $a = b$.*

The risk levels are denoted by $C_1, C_2, \cdots, C_q$ where q is the number of levels and a smaller subscript indicates a lower risk.

2.2.2 Preference on A Factor

Similar to the global preference on countries, the rater may prefer one value to another for a given factor. This preference often comes from the background

knowledge of a rater that might be different for different raters. It will be helpful to build a classification model which is more consistent with the rater's preference on every factor.

The preference of the rater on a factor can be mathematically represented by three binary relations between any two values of the factor. We use $x\mathrm{P}y$ to represent that the rater prefers x to y, and use $x\mathrm{I}y$ to represent that the rater is indifferent between x and y. Sometimes, two values of a factor might not be comparable in term of preference, which we denote it by $x\mathrm{J}y$.

The traditional method to model the preference of a rater is utilizing a function g defined on the value set A of the factor such that

$$\forall x, y \in A : \begin{cases} x\mathrm{P}y & \Leftrightarrow g(x) > g(y) \\[2mm] x\mathrm{I}y & \Leftrightarrow g(x) = g(y). \end{cases} \qquad (2.15)$$

While more complex methods were introduced by researchers [Vincke, 1992], we only use the above conventional method to model the rater's preference in this thesis. A function g which satisfies condition 2.15 is called a *criterion* which corresponds to *true criterion* in Vincke [1992]. By using this preference modelling method, we actually made the following implicit assumptions on the binary relations,

(1) P is transitive, i.e., $\forall x, y, z \in A$, $x\mathrm{P}y$ and $y\mathrm{P}z \Rightarrow x\mathrm{P}z$;

24

(2) I is transitive, i.e., $\forall x, y, z \in A$, $x\mathrm{I}y$ and $y\mathrm{I}z \Rightarrow x\mathrm{I}z$;

(3) J is empty,

which are simple deductions of relation among numbers. The first two assumptions ensure that the preference of the rater on all values of the factor are not contradictive, and the last assumption ensures that any two values of the factor are comparable. For the country risk case, these are mild assumptions that allow us to analyze the country risk level based on both quantitative and qualitative factors of a country.

In the current literature, most researchers assume that each criterion is monotone with respect to its corresponding numerical factors. In other words, the rater prefers a larger or smaller value for each numerical factor. For many factors, this is true. Since $g(x)$ is decreasing if and only if $g(-x)$ is increasing, we can assume without loss of generality that the monotone factors used are increasing.

When we have multiple factors to be considered and model each factor using the above method, we get n criteria $g_1, g_2, \cdots, g_n$ for a country. This is where the term *multicriteria* comes. For factor value x_i, the ith criterion value is denoted by $g_i(x_i)$. For a country c with value x_i^c on factor i, the ith criterion value can be abbreviated as $g_i(c)$.

In MCDA methods, we also assume the independence between criteria.

Suppose $F = \{g_1, \cdots, g_n\}$ is a family of criteria, G and $\overline{G}$ are a pair of complementary nonempty subsets of F. The following definition of the independence of criteria G was first introduced in Vincke [1992].

Definition 2.3 G *is* preferentially independent *in F if, for any four countries* a, b, c, d *such that*

$$g_i(a) = g_i(b), \forall i \in \overline{G} \tag{2.16a}$$

$$g_i(c) = g_i(d), \forall i \in \overline{G} \tag{2.16b}$$

$$g_i(a) = g_i(c), \forall i \in G \tag{2.16c}$$

$$g_i(b) = g_i(d), \forall i \in G, \tag{2.16d}$$

we will also have

$$a \succ b \Leftrightarrow c \succ d. \tag{2.16e}$$

2.2.3   UTADIS Method

Utility Function

The term utility function first appeared in the utility theory [Fishburn, 1970; Keeney and Raiffa, 1993] of decision science on which one class of MCDA methods including UTADIS is based. One fundamental axiom of the utility theory is that, when making the best choice from a list of choices (countries to

26

be loaned in our case), any decision-maker (rater in our case) unconsciously uses some function

$$U(c) = U(g_1(c), g_2(c), \cdots, g_n(c)),$$

which aggregates his preferences in different aspects (criteria of different factors in our case) of an choice, to rank all of these optional choices. This function is called *utility function*. Using such a utility function, one can model the global preference of a rater on countries.

To use utility function for classification, UTADIS introduces a sequence of strictly decreasing critical values $\{\mu_k\}_{k=1}^{q-1}$ and classifies the risk level of a country c by the following rule

$$\begin{cases} \text{if } U(c) \geq \mu_1, \text{ then } c \in C_1 \\ \\ \text{if } \mu_k \leq U(c) < \mu_{k-1} \text{ for some } 2 \leq k \leq q-1, \text{ then } c \in C_k \qquad (2.17) \\ \\ \text{if } U(c) < \mu_{q-1}, \text{ then } c \in C_q, \end{cases}$$

where the utility function is usually confined to be *normalized* without loss of generality, i.e., with minimal value 0 and maximal value 1. This classification model is demonstrated in Figure 2.1.

In other words, we use the utility function and a set of critical values to model the global preference relation $\succeq$. This is different from the way we model the preference on a single factor. The main task of UTADIS classification

27

model is to estimate the utility function and its critical values for different risk levels.

In order for the global preference on countries to be consistent with the preference on every factor, an additional constraint on the utility function is imposed, i.e., for any two countries a and b and for any factor i,

$$\left.\begin{array}{c} g_i(a) > g_i(b) \\ g_j(a) = g_j(b), \forall j \neq i \end{array}\right\} \Rightarrow a \succ b. \tag{2.18}$$

The Additive Form of Utility Function

To estimate the utility function, we have to define a function space where the approximation utility function will be searched for. The simplest and commonly used utility function takes the following additive form

$$U(g_1, \cdots, g_n) = \sum_{i=1}^{n} U_i(g_i),$$

where each $U_i(g_i)$ is a strictly increasing function.

By using this additive form of utility function, we impose another implicit assumption, i.e., any subset of the criteria used is preferentially independent. This is due to the fact that, for any subset G of the whole criteria set F and for any four countries a, b, c, d which satisfy the left hand side of (2.16),

$$U(a) - U(b) = \sum_{j \in G} [U_j(g_j(a)) - U_j(g_j(b))] \tag{2.19a}$$

28

$$= \sum_{j \in G}[U_j(g_j(c)) - U_j(g_j(d))] \qquad (2.19b)$$

$$= U(c) - U(d) \qquad (2.19c)$$

implies the right hand side of (2.16).

For a factor i, the composition function $U_i(g_i(x_i))$ is called a *marginal utility function*, abbreviated as $U_i(x_i)$ or $U_i(c)$ for country c. It's obvious that $U_i(x_i)$ is an increasing function. When the utility function is normalized, the maximal value of each marginal utility function is a percentage that indicates the relative importance of that factor in the country risk classification model. We call these maximal values of marginal utility functions *weights* of the factors.

Estimating the Utility Function

With this additive form, UTADIS method estimates the utility function by estimating every marginal utility function using a piecewise linear function. Note that the criterion function on each factor will not be estimated. We will directly estimate the composition function $U_i(g_i(x_i))$, i.e., the marginal utility functions.

Denote x_{i*} and x_i^* the smallest and largest possible value of quantitative factor i, and choose a set of increasing values $x_i^j, j = 1, \cdots, r_i$ between $x_{i*} = x_i^1$

29

and $x_i^* = x_i^{r_i}$, where r_i is the number of gridding points. The marginal utility function values at these points are denoted by U_i^j respectively. The piecewise linear function defined by the sequence $U_i^1, \cdots, U_i^{r_i}$ is the estimation of the ith marginal utility function. If the factor value x_i^c of a country c falls in interval $[x_i^j, x_i^{j+1})$, then its ith marginal utility will be calculated by

$$U_i(c) = \frac{x_i^{j+1} - x_i^c}{x_i^{j+1} - x_i^j} U_i^j + \frac{x_i^c - x_i^j}{x_i^{j+1} - x_i^j} U_i^{j+1}.$$

Estimation of utility function is now to determine the value of U_i^j and μ_i for $i = 1, \cdots, n, j = 1, \cdots, r_i$.

For given $\{U_i^j\}$ and $\{\mu_i\}$, the two types of classification error are illustrated in Figure 2.1. If country c actually comes from level C_k while $U(c) < \mu_k$ (or $U(c) \geq \mu_{k-1}$), we say the classification error is $\sigma^+(c) = \mu_k - U(c)$ (or $\sigma^-(c) = U(c) - \mu_{k-1}$). To find the best utility function that represents the rater's preference, we solve the following linear program, which aims to minimizing the total classification errors:

$$\min_{\sigma, \mu, U} \sum_{c \in C_k : k < q} \sigma^+(c) + \sum_{c \in C_k : k > 1} \sigma^-(c) \tag{2.20a}$$

$$\text{subject to } \sum_{i=1}^n U_i(c) - \mu_k + \sigma^+(c) \geq 0, \quad \forall 1 \leq k \leq q-1, \forall c \in C_k, \tag{2.20b}$$

$$\sum_{i=1}^n U_i(c) - \mu_{k-1} - \sigma^-(c) \leq -\delta, \quad \forall 2 \leq k \leq q, \forall c \in C_k, \tag{2.20c}$$

$$\mu_{k-1} - \mu_k \geq s, \quad \forall k = 2, \cdots, q-1, \tag{2.20d}$$
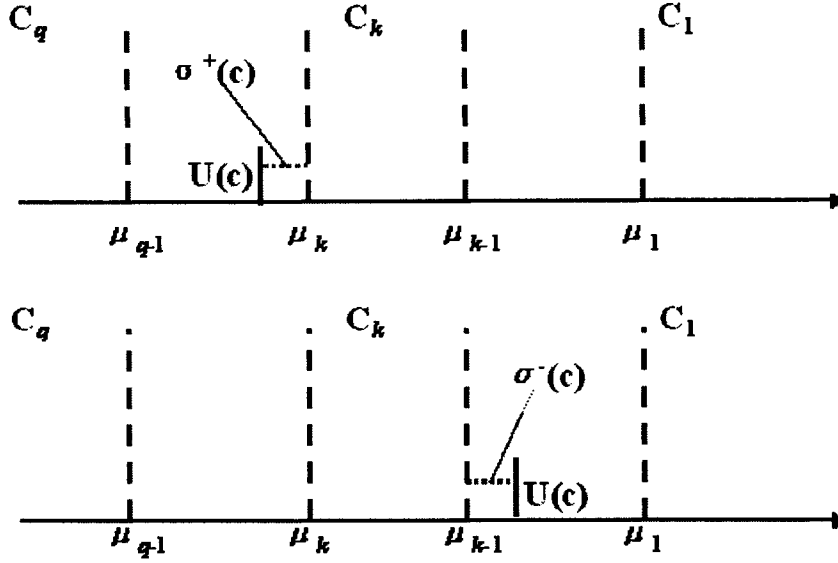
Figure 2.1: Illustration of Classification Errors on Country c

$$U_i^{j+1} - U_i^j \geq 0, \quad \forall i = 1, \cdots, n, \forall j = 1, \cdots, r_i - 1, \qquad (2.20e)$$

$$U_i^1 = 0, \quad \forall i = 1, \cdots, n, \qquad (2.20f)$$

$$\sum_{i=1}^{n} U_i^{r_i} = 1, \qquad (2.20g)$$

$$\sigma^+(c) \geq 0, \sigma^-(c) \geq 0, \quad \forall c. \qquad (2.20h)$$

In the above LP model, (2.20b,c,h) define the classification errors for each training country, (2.20e) is the monotonically increasing assumption, (2.20f,g) normalize the utility function and (2.20d) is used to keep the increasing behavior of the utility thresholds. s and δ are small positive constants used to ensure the strict inequality. These two parameters are usually chosen such

31

that $\frac{1}{q} > s > \delta > 0$.

Postoptimality Analysis

The solution to (2.20) may be sensitive to the data, i.e., a little change on the data may cause a big change on the solution. Spathis et al. [2002]; Zopounidis and Doumpos [1999] proposed a procedure to produce a classifier which is hopefully less sensitive to the data. This procedure is called *postoptimality analysis* in MCDA literature.

Suppose E^* is the optimal total classification error given by the solution to (2.20). In the postoptimality analysis, we allow the objective function to be near-optimal and change the objective function to the following constraint

$$\sum_{c \in C_k : k < q} \sigma^+(c) + \sum_{c \in C_k : k > 1} \sigma^-(c) \leq (1 + z)E^*, \qquad (2.21)$$

where $z > 0$ is a parameter indicating how much disturbance is allowed. To obtain a less sensitive classifier, Zopounidis and Doumpos [1999] proposed to solve the following sequence of linear programs

$$\max_{U, \mu, \sigma} U_i^{r_i} + \sum_{k=1}^{q-1} \mu_k, \quad \forall i = 1, \cdots, n \qquad (2.22a)$$

$$\min_{U, \mu, \sigma} U_i^{r_i} + \sum_{k=1}^{q-1} \mu_k, \quad \forall i = 1, \cdots, n \qquad (2.22b)$$

subject to the same constraints as in (2.20) with the additional one (2.21). Then the average of these solutions is proposed to be used as the final solution

which is hopefully less sensitive to the data than the one obtained by (2.20).

Note that, because of the increasingness of each marginal utility function, the average of the marginal utility functions on the same factor is also an increasing function and the average utility function of those obtained from the above linear programs, together with the solution to (2.20), is also normalized. This means that the average of solutions to (2.20) and (2.22) corresponds to a normalized classifier.

2.2.4   MHDIS Method

Another useful approach for country risk classification is MHDIS. While UTADIS determines the risk level of a country in one stage according to its utility as defined in the last section, MHDIS uses a sequential and hierarchical process to decide it. We still use $C_1, C_2, \cdots, C_q$ to denote the q risk levels, from the lowest to the highest. Given a country c, MHDIS decides in the fist stage whether it's in level C_1. If so, c is classified as in C_1, and the classification process is done. If not, MHDIS continues to decide whether it's in level C_2, and so on. The maximum number of these repeated stages is $q - 1$. If a country is decided not in level C_{q-1}, it will be classified as in level C_q, and the classification process is done.

To determine whether a country c is in level C_k or $C_{\neg k} = C_{k+1} \bigcup \cdots \bigcup C_q$,

MHDIS uses a *similarity* function $U_k(c) = U_k(g_1(c), \cdots, g_n(c))$ and a *dissimilarity* function $U_{\neg k}(c) = U_{\neg k}(g_1(c), \cdots, g_n(c))$: if $U_k(c) > U_{\neg k(c)}$, then c is in C_k; otherwise c is in $C_{\neg k}$. As indicated by the names, function U_k represents the degree of similarity between a given country and the countries at risk level C_k, while function $U_{\neg k}$ represents the degree of similarity between a given country and the countries at risk level C_{k+1} or higher. The estimation of these two functions can be done via using the additive form as in UTADIS, e.g.,

$$U_k(g_1, \cdots, g_n) = \sum_{i=1}^{m} U_{ki}(g_i),$$

where $g_i, i = 1 \cdots, n$ are criterion values on n factors. Each composition of $U_{ki}(g_i)/U_{\neg ki}(g_i)$ and $g_i(x_i)$, abbreviated as $U_{ki}(x_i)/U_{\neg ki}(x_i)$ or $U_{ki}(c)/U_{\neg ki}(c)$ for country c, is called a *marginal similarity/dissimilarity function*. Each marginal similarity/dissimilarity function indicates the similarity aforementioned from the point of view of the ith factor. MHDIS assumes that $U_{ki}(x_i)$ is monotonically increasing with respect to the ith factor value, and contrarily $U_{\neg ki}(x_i)$ is monotonically decreasing. Piecewise linear functions are again used to estimate marginal similarity/dissimilarity functions.

To understand better MHDIS, let's first look at an alternative model to estimate $U_k(\cdot)$ and $U_{\neg k}(\cdot)$. In this model (2.23), the number of misclassified

training countries is to be minimized:

$$\min_{I, U_k, U_{\neg k}} \sum_{c \in C_k \bigcup C_{\neg k}} I(c) \tag{2.23a}$$

$$\text{subject to } \sum_{i=1}^{n} U_{ki}(c) - \sum_{i=1}^{n} U_{\neg ki}(c) + I(c) \geq s, \quad \forall c \in C_k, \tag{2.23b}$$

$$\sum_{i=1}^{n} U_{\neg ki}(c) - \sum_{i=1}^{n} U_{ki}(c) + I(c) \geq s, \quad \forall c \in C_{\neg k}, \tag{2.23c}$$

$$U_{ki}^{j}, j = 1, \cdots, r_i \text{ is a increasing sequence}, \quad \forall k, \forall i \tag{2.23d}$$

$$U_{\neg ki}^{j}, j = 1, \cdots, r_i \text{ is a decreasing sequence}, \quad \forall k, \forall i \tag{2.23e}$$

$$U_k, U_{\neg k} \text{ are normalized}, \tag{2.23f}$$

$$I(c) \in \{0, 1\}, \quad \forall c, \tag{2.23g}$$

where s is a small positive constant to ensure the strict inequalities and $I(c)$ indicates whether c is misclassified or not. Minimizing the number of misclassified countries actually has been used in UTADIS(II) method [Doumpos et al., 2001b]. However, the number of training countries might be large . Because we use one binary variable $I(c)$ for every training country c, the large number of training countries results in a large number of integer variables, thus a big MIP has to be solved.

MHDIS also tries to minimize the number of misclassifications. To reduce the computation effort and improve the performance, MHDIS first uses an LP to recognize those countries that can be easily classified. Usually, many

training countries belong to this category. Then MHDIS requires these countries to be classified correctly and tries to decrease the number of misclassifications in the rest countries. In this way, the number of integer variables are decreased significantly.

We now describe the details in MHDIS. As mentioned above, MHDIS first recognizes those countries which can be easily classified by minimizing the sum of classification error $e(c)$ as shown in linear program (2.24):

$$\min_{e, U_k, U_{\neg k}} \sum_{c \in C_k \bigcup C_{\neg k}} e(c) \tag{2.24a}$$

$$\text{subject to } \sum_{i=1}^{n} U_{ki}(c) - \sum_{i=1}^{n} U_{\neg ki}(c) + e(c) \geq s, \quad \forall c \in C_k, \tag{2.24b}$$

$$\sum_{i=1}^{n} U_{\neg ki}(c) - \sum_{i=1}^{n} U_{ki}(c) + e(c) \geq s, \quad \forall c \in C_{\neg k}, \tag{2.24c}$$

$$U_{ki}^{j}, j = 1, \cdots, r_i \text{ is a increasing sequence,} \quad \forall k, \forall i \tag{2.24d}$$

$$U_{\neg ki}^{j}, j = 1, \cdots, r_i \text{ is a decreasing sequence,} \quad \forall k, \forall i \tag{2.24e}$$

$$U_k, U_{\neg k} \text{ are normalized,} \tag{2.24f}$$

$$e(c) \geq 0, \quad \forall c, \tag{2.24g}$$

where a small positive constant s is again used to ensure the strict inequalities and (2.24b,c) defines $e(c)$. The set of countries classified correctly after (2.24) is denoted by COR, and the set of countries misclassified is denoted by MIS.

To decrease the number of misclassified countries in MIS, MHDIS

directly minimizes it as shown in mixed integer program (2.25):

$$\min_{I, U_k, U_{\neg k}} \sum_{c \in MIS} I(c) \tag{2.25a}$$

$$\text{subject to } \sum_{i=1}^{n} U_{ki}(c) - \sum_{i=1}^{n} U_{\neg ki}(c) \geq s, \quad \forall c \in C_k \cap COR, \tag{2.25b}$$

$$\sum_{i=1}^{n} U_{\neg ki}(c) - \sum_{i=1}^{n} U_{ki}(c) \geq s, \quad \forall c \in C_{\neg k} \cap COR, \tag{2.25c}$$

$$\sum_{i=1}^{n} U_{ki}(c) - \sum_{i=1}^{n} U_{\neg ki}(c) + I(c) \geq s, \quad \forall c \in C_k \cap MIS, \tag{2.25d}$$

$$\sum_{i=1}^{n} U_{\neg ki}(c) - \sum_{i=1}^{n} U_{ki}(c) + I(c) \geq s, \quad \forall c \in C_{\neg k} \cap MIS, \tag{2.25e}$$

$$U_{ki}^{j}, j = 1, \cdots, r_i \text{ is a increasing sequence}, \quad \forall k, \forall i \tag{2.25f}$$

$$U_{\neg ki}^{j}, j = 1, \cdots, r_i \text{ is a decreasing sequence}, \quad \forall k, \forall i \tag{2.25g}$$

$$U_k, U_{\neg k} \text{ are normalized}, \tag{2.25h}$$

$$I(c) \in \{0, 1\}, \quad \forall c \in MIS, \tag{2.25i}$$

where s is the same constant as in (2.24) and $I(c)$ indicates whether c is misclassified or not. If we denote COR' the set of countries classified correctly and MIS' misclassified after (2.25), then MIS' surely might not contain the minimum number of misclassified countries as obtained in (2.23), while this two-step method significantly reduces computation effort.

MHDIS does not stop here. There might exist many different U_k and $U_{\neg k}$ pairs that can result in the same number of misclassifications. However, these different pairs might enjoy different generalization ability. Remember

that $U_k(c)$ and $U_{\neg k}(c)$ represent the similarity of country c to countries in C_k and $C_{\neg k}$ respectively. Similarity to $C_{\neg k}$ actually measures dissimilarity to C_k, and vice versa. Thus, we might expect that a similarity/dissimilarity function pair which has larger differences for those correctly classified countries can achieve better generalization performance[2]. This is what MHDIS does in its last step, as shown in the following linear program:

$$\max_{d, U_k, U_{\neg k}} d \tag{2.26a}$$

$$\text{subject to} \quad \sum_{i=1}^{n} U_{ki}(c) - \sum_{i=1}^{n} U_{\neg ki}(c) - d \geq s, \quad \forall c \in C_k \cap COR', \tag{2.26b}$$

$$\sum_{i=1}^{n} U_{\neg ki}(c) - \sum_{i=1}^{n} U_{ki}(c) - d \geq s, \quad \forall c \in C_{\neg k} \cap COR', \tag{2.26c}$$

$$\sum_{i=1}^{n} U_{ki}(c) - \sum_{i=1}^{n} U_{\neg ki}(c) \leq 0, \quad \forall c \in C_k \cap MIS', \tag{2.26d}$$

$$\sum_{i=1}^{n} U_{\neg ki}(c) - \sum_{i=1}^{n} U_{ki}(c) \leq 0, \quad \forall c \in C_{\neg k} \cap MIS', \tag{2.26e}$$

$$U_{ki}^{j}, j = 1, \cdots, r_i \text{ is a increasing sequence,} \quad \forall k, \forall i \tag{2.26f}$$

$$U_{\neg ki}^{j}, j = 1, \cdots, r_i \text{ is a decreasing sequence,} \quad \forall k, \forall i \tag{2.26g}$$

$$U_k, U_{\neg k} \text{ are normalized,} \tag{2.26h}$$

$$d \geq 0, \tag{2.26i}$$

where s is the same small constant as in (2.24) and d is the minimum difference between the similarity to C_k and $C_{\neg k}$ of countries in COR'. The countries in

---

[2]Such an idea was also used in methods UTADIS(I) and UTADIS(III) [Doumpos et al., 2001b].

MIS remains misclassified in program (2.26).

Using these three steps (2.24,2.25,2.26), MHDIS gets a classifier to classify C_k and $C_{\neg k}$. In total, there are $q-1$ two-class classifiers that constitute one multiple-class country risk classifier. Each step in MHDIS has actually considered in UTADIS family methods. However, MHDIS takes a more efficient way to derive a classifier.

We point out here that it is reasonable to expect that a pair of similarity and dissimilarity functions that has less differences for those misclassified countries can achieve better generalization performance. This can be done by introducing another variable d' similar to d in (2.26) to measure the maximum difference of similarities to C_k and $C_{\neg k}$ of countries in COR'. By minimizing a weighted sum of d and d', a classifier with better generalization performance can be obtained. Such an idea is used in UTADIS (III) [Doumpos et al., 2001b], but not in MHDIS.

'

Chapter 3

Dealing with Complex Criteria

We extend the UTADIS method to GUTADIS that allows more complex criteria to be incorporated into the model by using integer programming and nonlinear programming technique.

3.1 Unimodal Criteria

As we described in Section 2.2.2, most MCDA methods assume that each criterion function is monotone with respect to the corresponding factor and decision makers are able to choose factors that satisfy this assumption. However, criteria on some important factors might fail to satisfy this assumption. Take the ratio of trade to GDP for example. A very high trade ratio makes a country more vulnerable to international demand changes, hence a higher country risk, while a low ratio provides a country little capability to pay foreign debts. Another example is value added in industry as a percentage of

41

GDP, on which most countries at low risk level take a value in the middle of the two extremes. These two cases are confirmed by Figure 3.1 and Figure 3.2, where the countries at lower risk level have values more adjacent to the av-

Figure 3.1: A Non-Monotone Criterion on Trade (% of GDP)
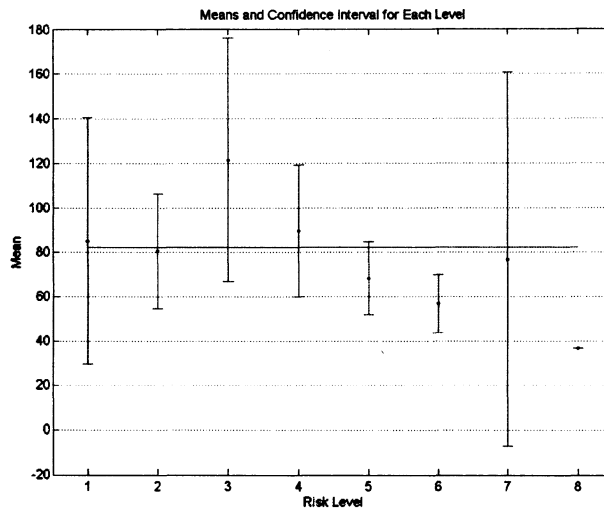
erage values. The above figures show that these factors are non-monotone, or in other words, the monotonicity assumption is unrealistic for these factors.

In order to use non-monotone criteria for country risk classification, we need to relax the monotone constraints on the criterion functions such that more complex and practical criteria can be used in MCDA models. It's not a good idea to simply drop the monotonicity assumptions, because nobody in practice have a preference that corresponds to a highly vibrating criterion function. For example, perhaps nobody would have a complex preference $g(x)$

42

Figure 3.2: A Non-Monotone Criterion on Industry (Value Added, % of GDP)
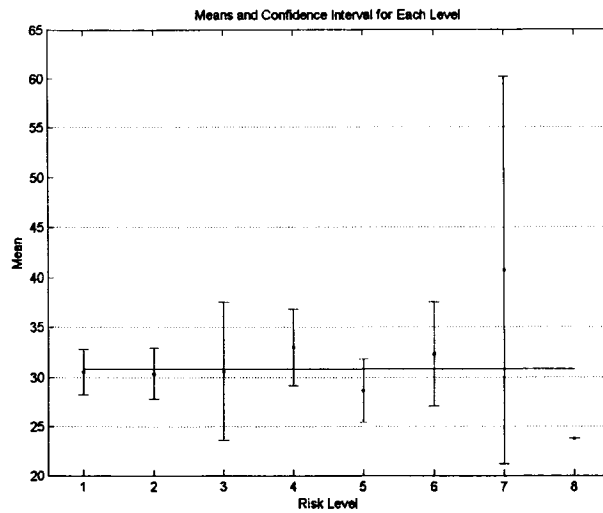
on the trade ratio like this

$$g(0.46) > g(0.49) > g(0.42) > g(0.44) > g(0.48).$$

Besides monotone functions, another class of functions representing the trend of a banker's preference on a single factor is the class of unimodal functions.

Definition 3.1 *(See Stout [2000]) A* unimodal *function is a univariate function which is monotonically increasing before a point reaching the maximal function value and monotonically decreasing after that point. A point where the unimodal function reaches its maximal function value is called a* mode *of the unimodal function.*

43

Note that the set of unimodal functions is a superset of monotone functions and include monotone functions as special cases. When a criterion is thought of as a unimodal one, the solution to a MCDA model that can deal with unimodal criteria may claim that it's simply increasing or decreasing. This may be used by decision-makers to verify their monotone preferences on factors.

3.2   The Generalized UTADIS Method

In this section, we generalize the UTADIS method so that it can use unimodal criteria. This generalized UTADIS method is called GUTADIS in this thesis.

Denote the set of factors with monotonically increasing criteria and unimodal criteria by I_m and I_u respectively. GUTADIS model can then be formulated as follows:

$$\min_{\sigma, \mu, U} \sum_{c \in C_k : k < q} \sigma^+(c) + \sum_{c \in C_k : k > 1} \sigma^-(c) \tag{3.1a}$$

$$\text{subject to } \sum_{i=1}^{n} U_i(c) - \mu_k + \sigma^+(c) \geq 0, \quad \forall 1 \leq k \leq q-1, \forall c \in C_k, \tag{3.1b}$$

$$\sum_{i=1}^{n} U_i(c) - \mu_{k-1} - \sigma^-(c) \leq -\delta, \quad \forall 2 \leq k \leq q, \forall c \in C_k, \tag{3.1c}$$

$$\mu_{k-1} - \mu_k \geq s, \quad \forall k = 2, \cdots, q-1, \tag{3.1d}$$

$$U_i^j, j = 1, \cdots, r_i \text{ is a increasing sequence}, \quad \forall i \in I_m \tag{3.1e}$$

44

$$U_i^j, j = 1, \cdots, r_i \text{ is a unimodal sequence,} \quad \forall i \in I_u \qquad (3.1f)$$

$$U \text{ is normalized,} \qquad (3.1g)$$

$$\sigma^+(c) \geq 0, \sigma^-(c) \geq 0, \quad \forall c. \qquad (3.1h)$$

It is unclear whether the new formulation is an LP program. In this formulation, that $U_i^j, j = 1, \cdots, r_i$ is a unimodal sequence means, there exists a l between 1 and r_i such that

$$U_i^1 \leq U_i^2 \leq \cdots \leq U_i^l \geq U_i^{l+1} \geq \cdots \geq U_i^{r_i}. \qquad (3.2)$$

For general unimodal functions, the model (3.1) is much harder to solve than (2.20).

3.2.1 The Normalization Constraints

The most difficult constraints in (3.1) are (3.1f) and (3.1g). Let's first consider the normalization constraints (3.1g). Using strict inequalities, the model can be rewritten as follows.

$$\min_{\sigma, \mu, U} \sum_{c \in C_k : k < q} \sigma^+(c) + \sum_{c \in C_k : k > 1} \sigma^-(c) \qquad (3.3a)$$

$$\text{subject to } \sum_{i=1}^{n} U_i(c) - \mu_k + \sigma^+(c) \geq 0, \quad \forall 1 \leq k \leq q-1, \forall c \in C_k, \qquad (3.3b)$$

$$\sum_{i=1}^{n} U_i(c) - \mu_{k-1} - \sigma^-(c) < 0, \quad \forall 2 \leq k \leq q, \forall c \in C_k, \qquad (3.3c)$$

45

$$\mu_{k-1} - \mu_k > 0, \quad \forall k = 2, \cdots, q-1, \tag{3.3d}$$

$$U_i^j, j = 1, \cdots, r_i \text{ is a increasing sequence}, \quad \forall i \in I_m, \tag{3.3e}$$

$$U_i^j, j = 1, \cdots, r_i \text{ is a unimodal sequence}, \quad \forall i \in I_u, \tag{3.3f}$$

$$\left. \begin{array}{l} \min_j U_i^j = 0, \quad \forall i \\[2mm] \sum_i \max_j U_i^j = 1, \end{array} \right\} \text{(Normalization)} \tag{3.3g}$$

$$\sigma^+(c) \geq 0, \sigma^-(c) \geq 0, \quad \forall c, \tag{3.3h}$$

where I_m is the set of monotonically increasing criteria, and I_u is the set of unimodal criteria.

When all criteria are monotone, we know where the minimum and maximum marginal utility function values are achieved, and the normalization constraints can be easily formulated as linear constraints (2.20f,g). However, when unimodal criteria are used, we don't know both the minimum and the maximum of the corresponding marginal utility functions. This makes normalization constraints in (3.3g) difficult to be realized.

Consider relaxing the normalization constraints as in (3.4g).

$$\min_{\sigma,\mu,U} \sum_{c \in C_k : k < q} \sigma^+(c) + \sum_{c \in C_k : k > 1} \sigma^-(c) \tag{3.4a}$$

$$\text{subject to } \sum_{i=1}^n U_i(c) - \mu_k + \sigma^+(c) \geq 0, \quad \forall 1 \leq k \leq q-1, \forall c \in C_k, \tag{3.4b}$$

$$\sum_{i=1}^{n} U_i(c) - \mu_{k-1} - \sigma^-(c) < 0, \quad \forall 2 \leq k \leq q, \forall c \in C_k, \qquad (3.4c)$$

$$\mu_{k-1} - \mu_k > 0, \quad \forall k = 2, \cdots, q-1, \qquad (3.4d)$$

$$U_i^j, j = 1, \cdots, r_i \text{ is a increasing sequence}, \quad \forall i \in I_m, \qquad (3.4e)$$

$$U_i^j, j = 1, \cdots, r_i \text{ is a unimodal sequence}, \quad \forall i \in I_u, \qquad (3.4f)$$

$$0 \leq U_i^j \leq 1, \quad \forall i = 1, \cdots, n, \forall j = 1, \cdots, r_i, \qquad (3.4g)$$

$$\sigma^+(c) \geq 0, \sigma^-(c) \geq 0, \quad \forall c. \qquad (3.4h)$$

While the optimal solution of (3.3) might not be the optimal solution of (3.4), the optimal solution of (3.3) can be recovered from the optimal solution of (3.4), as shown in our next discussion.

Suppose (U, μ) is feasible for (3.4), and σ is its corresponding classification error vector. Let

$$\alpha_i = \min_j U_i^j, \quad \forall i = 1, \cdots, n, \qquad (3.5a)$$

$$U_i^{j'} = U_i^j - \alpha_i, \quad \forall i = 1, \cdots, n, \forall j = 1, \cdots, r_i, \qquad (3.5b)$$

$$\mu_k' = \mu_k - \sum_{i=1}^{n} \alpha_i, \quad \forall k = 1, \cdots, q-1. \qquad (3.5c)$$

Note that $\min_j U_i^{j'} = 0$ for all i. In this thesis, we call a utility function *semi-normalized* if its minimum value is 0. System of equations (3.5) provides a procedure to transform a general utility function into a semi-normalized one. It can be easily verified that (U', μ') is also feasible for (3.4) with the same

classification error vector $\sigma' = \sigma$. Thus, we can confine the optimal utility function of (3.4) to be semi-normalized.

We next introduce the following normalization procedure

$$\beta_i = \max_j U_i^j, \quad \forall i = 1, \cdots, n, \tag{3.6a}$$

$$U_i^{j'} = \frac{U_i^j}{\sum_{l=1}^n \beta_l}, \quad \forall i = 1, \cdots, n, \forall j = 1, \cdots, r_i, \tag{3.6b}$$

$$\mu_k' = \frac{\mu_k}{\sum_{i=1}^n \beta_i}, \quad \forall k = 1, \cdots, q-1. \tag{3.6c}$$

Theorem 3.1 *Suppose (U, μ) is non-constant and semi-normalized. If (U, μ) is optimal for (3.4), then (U', μ'), provided by the procedure (3.6), is optimal for (3.3).*

Proof Since (U, μ) is non-constant and feasible for (3.4), we claim that $\sum \beta_i > 0$. Consider the following program (3.7) which is different from (3.3) only in the normalization constraints:

$$\min_{\sigma, \mu, U} \sum_{c \in C_k : k < q} \sigma^+(c) + \sum_{c \in C_k : k > 1} \sigma^-(c) \tag{3.7a}$$

$$\text{subject to } \sum_{i=1}^n U_i(c) - \mu_k + \sigma^+(c) \geq 0, \quad \forall 1 \leq k \leq q-1, \forall c \in C_k, \tag{3.7b}$$

$$\sum_{i=1}^n U_i(c) - \mu_{k-1} - \sigma^-(c) < 0, \quad \forall 2 \leq k \leq q, \forall c \in C_k, \tag{3.7c}$$

$$\mu_{k-1} - \mu_k > 0, \quad \forall k = 2, \cdots, q-1, \tag{3.7d}$$

$$U_i^j, j = 1, \cdots, r_i \text{ is a increasing sequence}, \quad \forall i \in I_m, \tag{3.7e}$$

48

$$U_i^j, j = 1, \cdots, r_i \text{ is a unimodal sequence,} \quad \forall i \in I_u, \qquad (3.7\text{f})$$

$$\left. \begin{array}{l} \min_j U_i^j = 0, \quad \forall i \\[2em] \sum_i \max_j U_i^j = \gamma, \end{array} \right\} \text{(normalization)} \qquad (3.7\text{g})$$

$$\sigma^+(c) \geq 0, \sigma^-(c) \geq 0, \quad \forall c, \qquad (3.7\text{h})$$

where γ is a positive constant. Obviously, if $(U^{\gamma_1}, \mu^{\gamma_1})$ is the optimal solution for (3.7) with $\gamma = \gamma_1$, then $\frac{\gamma_2}{\gamma_1}(U^{\gamma_1}, \mu^{\gamma_1})$ is the optimal solution for (3.7) with $\gamma = \gamma_2$, since their classification error vectors are proportional.

Let $\gamma_3 = \sum_{i=1}^n \beta_i$ where β_i is defined in (3.6a). If (U, μ) is optimal for (3.4), then (U, μ) is also the optimal solution of (3.7) with $\gamma = \gamma_3$. Thus, as discussed above, $\frac{1}{\gamma_3}(U, \mu)$ is the optimal solution for (3.7) with $\gamma = 1$, which provides a solution to (3.3). ∎

According to Theorem 3.1, we can first solve (3.4), then use procedure (3.5) and (3.6) to transform the solution into an equivalent but normalized one. In this way, we avoid dealing with the nonlinear constraint (3.3g).

In actual computation, we again introduce parameter δ and s to strengthen the strict inequalities to non-strict ones as shown in the following

$$\min_{\sigma, \mu, U} \sum_{c \in C_k : k < q} \sigma^+(c) + \sum_{c \in C_k : k > 1} \sigma^-(c) \qquad (3.8\text{a})$$

$$\text{subject to } \sum_{i=1}^n U_i(c) - \mu_k + \sigma^+(c) \geq 0, \quad \forall 1 \leq k \leq q - 1, \forall c \in C_k, \qquad (3.8\text{b})$$

$$\sum_{i=1}^{n} U_i(c) - \mu_{k-1} - \sigma^-(c) \leq -\delta, \quad \forall 2 \leq k \leq q, \forall c \in C_k, \quad (3.8c)$$

$$\mu_{k-1} - \mu_k \geq s, \quad \forall k = 2, \cdots, q-1, \quad (3.8d)$$

$$U_i^j, j = 1, \cdots, r_i \text{ is a increasing sequence}, \quad \forall i \in I_m, \quad (3.8e)$$

$$U_i^j, j = 1, \cdots, r_i \text{ is a unimodal sequence}, \quad \forall i \in I_u, \quad (3.8f)$$

$$0 \leq U_i^j \leq 1, \quad \forall i = 1, \cdots, n, \forall j = 1, \cdots, r_i, \quad (3.8g)$$

$$\sigma^+(c) \geq 0, \sigma^-(c) \geq 0, \quad \forall c. \quad (3.8h)$$

The relation between (3.8) and (3.1) is characterized by the following theorem, which is similar to Theorem 3.1.

Theorem 3.2 *Suppose (U, μ) is non-constant and semi-normalized. If (U, μ) is optimal for (3.8), then (U', μ') provided by procedure (3.6) is optimal for (3.1) with parameter setting*

$$\delta' = \frac{\delta}{\sum_{i=1}^{n} \beta_i}, s' = \frac{\delta}{\sum_{i=1}^{n} \beta_i}. \quad (3.9)$$

The theorem can be proved by introducing an auxiliary model the same as (3.7) except with strict inequalities changed. Note that procedure (3.5) won't change the optimality of (U, μ) for both (3.1) and (3.8). Thus, we are still guaranteed to get an optimal solution that is also semi-normalized.

However, if the scale of the solution to (3.8) is small, we will get a small $\sum_{i=1}^{n} \beta_i$ that makes equation (3.9) correspond to large tolerances. Thus, to

obtain the solution to (3.1) under a certain tolerance setting, we need to know the scale of the solution to (3.8) if we want to use procedure (3.6) to construct the solution to (3.1). This is impractical because knowing the scale of the solution is a problem the same as our ultimate problem (3.1).

Our strategy to attack this difficulty is using two stages. Given the tolerance setting, we first solve (3.8) under this setting and get its scale. Second, we use the obtained scale to adjust the tolerance parameters and solve (3.8) again. The normalized solution of the second stage is used as an approximate solution to (3.1) with given tolerance parameters.

3.2.2 Concave Marginal Utility Functions

Starting from this section, we will focus on solving model (3.8).

Concave marginal utility functions are among the unimodal ones that can be relatively easily handled. If we assume that $U_i(x)$ is concave, the corresponding constraint (3.8f) can be relaxed to the following linear inequalities

$$\frac{U_i^j - U_i^{j-1}}{x_i^j - x_i^{j-1}} \geq \frac{U_i^{j+1} - U_i^j}{x_i^{j+1} - x_i^j}, \quad \forall j = 2, \cdots, r_i - 1. \tag{3.10}$$

It can be verified that (3.10) is a sufficient and necessary condition for a piece-wise linear function with nodes $(x_i^j, U_i^j), j = 1, \cdots, r_i$ to be a concave function. Thus, concave marginal utility functions do not bring extra difficulty

51

into model (3.8) and it's again a linear program. Convex marginal utility functions can be handled in a similar manner as in (3.10).

Convex/concave constraints may also be used to produce smooth marginal utility functions of monotone criteria. For example, a marginal utility function obtained by (2.20) might alike the one shown in Figure 3.3. One might

Figure 3.3: An Example of Steep Marginal Utility Function
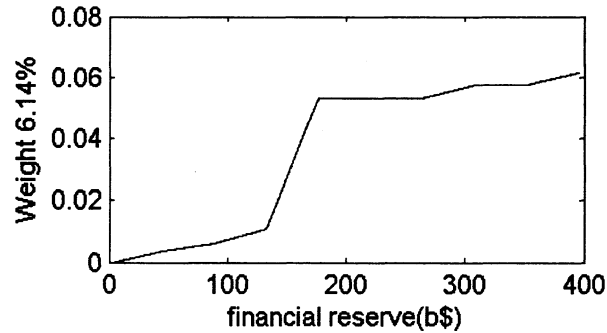
believe that the impact of financial reserve changes should not be so unevenly, or in other words, the corresponding marginal utility function should not be so steep. A classical methods to deal with this situation is using the following constrains for increasing marginal utility function

$$0 \leq U_i^{j+1} - U_i^j \leq \tau U_i^{r_i}, \quad \forall j = 1, \cdots, r_i - 1, \qquad (3.11)$$

where $\tau > 0$ is a parameter.

However, if we confine the marginal utility functions to be convex or concave, then the constraint (3.10) will ensure that the resulting utility func-

tion will not be very steep. We shall provide a numerical test of using concave or convex constraints smoothing the marginal utility function in Chapter 5, which verify our conclusion.

3.2.3   General Unimodal Criteria

If the mode of each criterion function is known in prior, the unimodality can be represented by two monotone segments separated by the mode, one increasing and the other decreasing, and consequently, the unimodal constraints can be expressed by linear inequality constraints. This situation was briefly discussed in Zopounidis and Doumpos [1999].

In case that the modes are unknown, one way to solve (3.8) is enumerating all possible mode combinations of the unimodal criteria to find the optimal solution. It involves solving $\prod_{i \in I_u} r_i$ number of linear programs (3.8) where (3.8f) is alternatively changed to linear constraints (3.2) with different modes. However, even if only 3 criteria are unimodal, this procedure, programmed using MATLAB 6.5, takes more than an hour on an IBM RS6000 workstation, which makes the generalized model impractical.

Our approach to deal with the unimodal constraints is reformulating (3.8) into an integer program and exploiting the power of existing integer-program solvers. We will show that the number of integer variables introduced

to tackle the unimodal constraints is about the same as the number of grid points of unimodal criteria.

We introduce an integer variable $Y_{ij} = 1$ or -1 to indicate whether

$$U_i^j \geq U_i^{j-1} \tag{3.12}$$

or

$$U_i^j \leq U_i^{j-1} \tag{3.13}$$

holds for every $i \in I_u$ and for every j between 2 to r_i. In other words, $Y_{ij} \in \{-1, 1\}$ is confined to qualify the following inequality constraint

$$Y_{ij}(U_i^j - U_i^{j-1}) \geq 0. \tag{3.14}$$

Using variables Y_{ij}, the unimodality of criterion g_i for $i \in I_u$ can be expressed as

$$Y_{ij} = -1 \Rightarrow Y_{i,j+1} = -1, \quad \forall j = 2, \cdots, r_i - 1, \tag{3.15}$$

which can be further expressed as inequality

$$Y_{ij} \geq Y_{i,j+1}, \forall j = 2, \cdots, r_i - 1. \tag{3.16}$$

Variables Y_{ij} together with constraints (3.14) and (3.16) exactly ensure the $\{U_i^j, j = 1, \cdots, r_i\}$ to be a unimodal sequence for each $i \in I_u$. Therefore

54

we can reformulate (3.8) as the following

$$\min_{\sigma,\mu,U,Y} \sum_{c \in C_k : k < q} \sigma^+(c) + \sum_{c \in C_k : k > 1} \sigma^-(c) \tag{3.17a}$$

$$\text{subject to} \sum_{i=1}^{n} U_i(c) - \mu_k + \sigma^+(c) \geq 0, \quad \forall 1 \leq k \leq q-1, \forall c \in C_k, \tag{3.17b}$$

$$\sum_{i=1}^{n} U_i(c) - \mu_{k-1} - \sigma^-(c) \leq -\delta, \quad \forall 2 \leq k \leq q, \forall c \in C_k, \tag{3.17c}$$

$$\mu_{k-1} - \mu_k \geq s, \quad \forall k = 2, \cdots, q-1, \tag{3.17d}$$

$$U_i^j - U_i^{j-1} \geq 0, \quad \forall i \in I_m, \forall j = 2, \cdots, r_i, \tag{3.17e}$$

$$Y_{ij}(U_i^j - U_i^{j-1}) \geq 0, \quad \forall i \in I_u, \forall j = 2, \cdots, r_i, \tag{3.17f}$$

$$Y_{ij} \geq Y_{i,j+1}, \quad \forall i \in I_u, j = 2, \cdots, r_i - 1, \tag{3.17g}$$

$$0 \leq U_i^j \leq 1, \quad \forall i = 1, \cdots, n, \forall j = 1, \cdots, r_i, \tag{3.17h}$$

$$Y_{ij} \in \{-1, 1\}, \quad \forall i \in I_u, j = 2, \cdots, r_i, \tag{3.17i}$$

$$\sigma^+(c) \geq 0, \sigma^-(c) \geq 0, \quad \forall c, \tag{3.17j}$$

which is an integer nonlinear program.

Like linearly constrained integer programs, the solution to nonlinearly constrained program (3.17) mainly resorts to branch and bound or cutting plane algorithms.

3.2.4 Postoptimality Analysis

As mentioned in Section 2.2.3 (page 32), a postoptimality analysis can be done

to produce a hopefully less sensitive solution when all criteria are monotone.

This traditional postoptimality analysis is done by the following way.

First we relax the problem by allowing the classification error to be in an

interval $[E^* - \epsilon, E^*]$, where E^* is obtained by solving the model (2.20). Then

we compute the extreme points of the polyhedron $\mathcal{P}$ defined by (2.20) and

(2.21) and use the average of these extreme points as the final classifier.

Note that the average of these extreme points can be viewed as a sort of

center of the polyhedron $\mathcal{P}$ and there are various methods for computing such

a center in the optimization community, particularly in the area of interior-

point methods [Ye, 1997]. For example, suppose the linear inequalities defined

by (2.20) and (2.21) are rewritten into the following form

$$Ax \geq b, \tag{3.18}$$

where x is the vector of variables. Then one of its analytic center can be

obtained by solving the following optimization problem

$$\max \sum_i \ln(a_i^T x - b_i) \tag{3.19a}$$

$$\text{s.t. } Ax \geq b, \tag{3.19b}$$

56

where a_i^T is the ith row of matrix A and b_i is the ith component of vector b. It has been shown that the above optimization problem can be solved in polynomial time.

So we can perform the postoptimality analysis by solving one convex nonlinear program (3.19) instead of tens of linear programs.

It will be helpful to explore how to apply these powerful optimization techniques in the postoptimality analysis. Due to the time limitation, we leave this as one topics of the future works.

Chapter 4

Using Quadratic Utility

Function

4.1 Quadratic Utility Functions

When we choose to use additive form of utility functions in MCDA methods such as UTADIS and MHDIS, we are assuming that the preferences of raters on countries can be disaggregated into independent preferences on various factors. Since many practically used factors are closely correlated, this assumption is definitely not true. Table 4.1 shows the correlation coefficient matrix of the factors listed in Table 1.2. From this table, we can see that many factors have strong correlation with each other.

59

Table 4.1: The Correlation Coefficients Matrix of Factors in Table 1.2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1)	1.00	0.88	0.60	0.48	0.75	-0.02	-0.22	-0.11	0.15	0.11	-0.33	-0.53	0.10
(2)	0.88	1.00	0.74	0.54	0.87	0.00	-0.27	-0.23	0.05	-0.08	-0.37	-0.66	0.24
(3)	0.60	0.74	1.00	0.54	0.56	0.30	-0.28	-0.06	-0.13	-0.02	-0.26	-0.50	0.31
(4)	0.48	0.54	0.54	1.00	0.60	0.10	-0.35	-0.12	-0.05	-0.07	-0.08	-0.37	-0.18
(5)	0.75	0.87	0.56	0.60	1.00	-0.11	-0.41	-0.31	0.09	-0.04	-0.30	-0.52	0.19
(6)	-0.02	0.00	0.30	0.10	-0.11	1.00	-0.05	-0.21	0.18	0.31	0.14	0.09	-0.14
(7)	-0.22	-0.27	-0.28	-0.35	-0.41	-0.05	1.00	0.14	-0.10	0.10	0.17	0.32	-0.22
(8)	-0.11	-0.23	-0.06	-0.12	-0.31	-0.21	0.14	1.00	-0.18	0.24	-0.20	0.04	0.08
(9)	0.15	0.05	-0.13	-0.05	0.09	0.18	-0.10	-0.18	1.00	-0.02	-0.37	0.02	-0.11
(10)	0.11	-0.08	-0.02	-0.07	-0.04	0.31	0.10	0.24	-0.02	1.00	0.04	0.21	-0.04
(11)	-0.33	-0.37	-0.26	-0.08	-0.30	0.14	0.17	-0.20	-0.37	0.04	1.00	0.60	-0.30
(12)	-0.53	-0.66	-0.50	-0.37	-0.52	0.09	0.32	0.04	0.02	0.21	0.60	1.00	-0.36
(13)	0.10	0.24	0.31	-0.18	0.19	-0.14	-0.22	0.08	-0.11	-0.04	-0.30	-0.36	1.00

The name of the factors are the same as in Table 1.2.

It is more appropriate to take the correlation into consideration in the model when many factors are densely correlated. One possible scheme is to introduce new terms into the additive form of utility function to reflect the correlation among various factors. This kind of more complex utility function may enhance the capability of UTADIS methods.

The simplest extension of the additive form of utility function is the quadratic one as shown in the following general form

$$U(c) = \sum_{i=1}^{n} U_i(g_i(c)) + \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{ij} U_i(g_i(c)) U_j(g_j(c)), \qquad (4.1)$$

where U_i is strictly increasing with respect to $g_i(c)$ for every $i = 1, \cdots, n$ and coefficients $\alpha_{ij}, i, j = 1, \cdots, n$ are parameters. The composition function $U_i(g_i(c))$ are also called *marginal utility functions* for the ith factor. In this form, the effect of the ith factor consists of two parts

$$U_i(g_i(c)) + U_i(g_i(c)) \sum_{j=1}^{n} \alpha_{ij} U_j(g_j(c)),$$

where the first part is contributed by the ith factor itself and the second part is contributed by the correlation between the ith factor and the other factors. Strictly speaking, α_{ii} should be zero, while a positive α_{ii} can be regarded as a quadratic term of the first part.

We assume as in UTADIS methods that every criterion, and hence every marginal utility function, is monotonically increasing to its corresponding

factor value. We can see that, if $\alpha_{ij} \geq 0$ for all $i, j = 1, \cdots, n$, the global preference defined by the quadratic utility function (4.1) and a set of critical values will be guaranteed to be consistent to the individual criterion on each factor, i.e., (2.18) will still hold.

Similarly to (2.20), we can now formulate the training process as the following nonlinear program

$$\min_{\sigma, \mu, U} \sum_{c \in C_k : k < q} \sigma^+(c) + \sum_{c \in C_k : k > 1} \sigma^-(c) \tag{4.2a}$$

$$\text{subject to } U(c) - \mu_k + \sigma^+(c) \geq 0, \quad \forall 1 \leq k \leq q - 1, \forall c \in C_k, \tag{4.2b}$$

$$U(c) - \mu_{k-1} - \sigma^-(c) \leq -\delta, \quad \forall 2 \leq k \leq q, \forall c \in C_k, \tag{4.2c}$$

$$\mu_{k-1} - \mu_k \geq s, \quad \forall k = 2, \cdots, q - 1, \tag{4.2d}$$

$$U_i^{j+1} - U_i^j \geq 0, \quad \forall i = 1, \cdots, n, \forall j = 1, \cdots, r_i - 1, \tag{4.2e}$$

$$U_i^1 = 0, \quad \forall i = 1, \cdots, n, \tag{4.2f}$$

$$\sum_{i=1}^n U_i^{r_i} + \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} U_i^{r_i} U_j^{r_j} = 1, \tag{4.2g}$$

$$\sigma^+(c) \geq 0, \sigma^-(c) \geq 0, \quad \forall c, \tag{4.2h}$$

where δ and s are small positive constants to ensure the strict inequalities, $U(c)$ is defined as in (4.1) and $\alpha_{ij} \geq 0$ are parameters. This formulation is again reduced to a nonlinear quadratic constrained linear program.

We call this model with quadratic utility functions QUDIS.

4.2 A Special Quadratic Utility Function

The quadratic utility function (4.1) is generally neither convex nor concave if we simply have nonnegative coefficients α_{ij}. If there exists a series of $\alpha_j, j = 1, \cdots, n$ such that

$$\alpha_{ij} = \alpha_i \alpha_j, \quad \forall i, j = 1, \cdots, n, \tag{4.3}$$

the utility function is then convex. Actually, the left hand sides of constraints (4.2b,c,d) will be all parabolic: the constraints with $\sigma^-(c)$ are convex, and the ones with $\sigma^+(c)$ concave.

Let's first check a special case of α_{ij}s satisfying condition (4.3):

$$\alpha_{ij} = 1, \text{ for all } i, j = 1, \cdots, n. \tag{4.4}$$

In this case, the utility function becomes

$$\begin{aligned}
U(c) &= \sum_{i=1}^{n} U_i(g_i(c)) + \sum_{i=1}^{n} \sum_{j=1}^{n} U_i(g_i(c)) U_j(g_j(c)) \\
&= \left(\sum_{i=1}^{n} U_i(g_i(c)) + \frac{1}{2} \right)^2 - \frac{1}{4},
\end{aligned} \tag{4.5}$$

which is an ellipsoidal function.

For general α_{ij} which satisfies condition (4.3), we can assume that $\alpha_k \neq 0$ for some k without loss of generality. The utility function then becomes

$$U(c) = \sum_{i=1}^{n} U_i(g_i(c)) + \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j U_i(g_i(c)) U_j(g_j(c))$$

$$= \sum_{i=1}^{n} U_i(g_i(c)) + \left(\sum_{i=1}^{n} \alpha_i U_i(g_i(c)) \right)^2$$

$$= \sum_{i=1}^{n} U_i(g_i(c)) + \alpha_k^2 \left(\sum_{i=1}^{n} \frac{\alpha_i}{\alpha_k} U_i(g_i(c)) \right)^2$$

$$= \sum_{i \neq k} \left(1 - \frac{\alpha_i}{\alpha_k} \right) U_i(g_i(c)) + \alpha_k^2 \left(\sum_{i=1}^{n} \frac{\alpha_i}{\alpha_k} U_i(g_i(c)) + \frac{1}{2\alpha_k^2} \right)^2 - \frac{1}{4\alpha_k^2}, \quad (4.6)$$

which is a parabolic function if not all α_is are equal, or an ellipsoidal function otherwise.

4.3   The Normalization Constraint

Under condition (4.3), the quadratic constraint (4.2h) can be transformed into a similar form to (4.6):

$$1 = \sum_{i \neq k} \left(1 - \frac{\alpha_i}{\alpha_k} \right) U_i^{r_i} + \alpha_k^2 \left(\sum_{i=1}^{n} \frac{\alpha_i}{\alpha_k} U_i^{r_i} + \frac{1}{2\alpha_k^2} \right)^2 - \frac{1}{4\alpha_k^2}, \quad (4.7)$$

where k is chosen such that $\alpha_k \neq 0$. When $\alpha_i = 1$ for all $i = 1, \cdots, n$, it becomes

$$\left(\sum_{i=1}^{n} U_i^{r_i} + \frac{1}{2} \right)^2 = \frac{5}{4}.$$

Since every $U_i(g_i^{r_i})$ is nonnegative, this constraint can be further simplified as a linear one

$$\sum_{i=1}^{n} U_i^{r_i} = \frac{\sqrt{5} - 1}{2}. \quad (4.8)$$

64

4.4 Optimizing α_{ij}

In the previous sections, we use α_{ij} as predetermined parameters to determine the optimal utility function, while different α would give different solution. By forcing these α_{ij} to be also non-negative decision variables, the degree of dependence between factors can also be optimized by (4.2). We should also add upper bound constraints and symmetry constraints on α_{ij}s:

$$0 \leq \alpha_{ij} \leq 1 \text{ and } \alpha_{ij} = \alpha_{ji} \text{ for all } i, j = 1, \cdots, n.$$

When α_{ij} are also variables, (4.2b,c,d,h) become cubic constraints and the problem is hard to be solved. The inclusion of the model optimizing α_{ij} is only intended to show that, by forcing α_{ij} to the special case $\alpha_{ij} = 1$, we will not lose much performance (See Chapter 5 for the performance comparison).

Chapter 5

Data and Result

5.1 Data and Sources

5.1.1 Rating Sources

The country risk ratings used by numerical experiments in this thesis are provided by Standard & Poor's in 1998, 1999 and 2001 [Afonso, 2003; Alexe et al., 2003; Chambers and Beers, 1998]. Standard & Poor's produces its ratings based on information provided by debtor countries themselves and other sources considered to be reliable [Alexe et al., 2003].

As shown in Table 1.1, the total number of risk levels is as large as 23. However, the number of countries rated by Standard & Poor's was only 70 around by December 1998, and only few countries fell into levels with higher

Table 5.1: Distribution of Sample Country-Years

Data Set	AAA	AA	A	BBB	BB	B	CCC	CC – D
No.1	20	22	17	28	33	12	3	3
No.2	11	11	13	15	14	16		

risk. The reason is that most of developing and poor countries obviously have no payment capability and no commercial banks are interested in them.

In this thesis, we use two set of ratings as training samples. One consists of 69 countries in year 1998 and 1999, totally 138 country-years. Another one consists of 80 countries in 2001. Also, we note that countries at positive levels (such as AA+, A+, etc.) and negative levels (such as AA-, A-, etc.) have a little stronger or weaker repayment ability than the corresponding neutral ones. So we combine the positive and negative levels into the corresponding neutral ones. Level CC, C, SD and D, if present, are also combined as one due to the limited number of sample for them. After these combinations, the number of classes of these two data sets are reduced to 8 and 6 respectively.

Finally, the distribution of the sample country-years over the classes is listed in Table 5.1. According to Hu et al. [2002], the number of countries rated as CCC and below shown in this table is all the cases during 1981 and 1999.

5.1.2   Sources of Related Factors

Most of the data on related factors are retrieved from the World Development Indicators database created by The World Bank [2002]. This commercial database recorded information of 207 countries on 575 factors. The online version of this database provides data for years after 2000 and it is another important data sources of this thesis.

Human Development Report[1] created by United Nations Development Programme provides a bunch of statistical information on different aspects of human development. Many of these are overlayed with those provided by The World Bank [2002]. The only one used in this thesis is the factor Human Development Index that is explained in the appendix.

The Polity IV Project polity series database developed by University of Maryland recorded the government status of about 200 countries in the last 200 years. Several factors indicates the property of the political system running in a country. They mainly indicate the degree of autocracy or democracy of a country. The one used in this thesis is named POLITY which is coded as an integer ranging from -10 to 10.

---

[1]See http://hdr.undp.org/statistics/data/.

5.1.3   Selected Factors

The selected factors are supposed to cover the environment and infrastructure, health and education, political and social status, economy and finance aspects of countries. These factors can be organized into five groups as listed in the appendix. Except Human Development Index provided by United Nations and POLITY provided by Polity IV Project, all other factors are extracted from the World Development Indicators database or calculated according to data extracted from that database. A detailed description of the factors is given in the appendix. Six of the forty factors takes a unimodal criterion in this thesis.

Because of the availability of factors in the two main databases, two factors used in the two data sets are slightly different. 'Research and development expenditure (% of GNI)' in the first data set corresponds to 'Research and development expenditure (% of GDP)' in the second. 'Military expenditure (% of GNI)' in the first data set corresponds to 'Military expenditure (% of GDP)' in the second.

5.2   Computation Environment

5.2.1   Solvers

The linear programs (2.20), (2.22), (3.8) with concave constraints (3.10) are solved using solver PCx [Czyzyk et al., 1997] which is an interior-point predictor-

corrector linear programming package. It's developed at the Optimization Technology Center collaborated by Argonne National Laboratory and North-western University.

Mix-integer nonlinear program (3.17) is solved by MINLP [Leyffer, 1999] which implements a branch-and-bound algorithm for nonlinearly con-strained mix-integer programs. The continuous relaxation subproblems are solved using filterSQP [Fletcher and Leyffer, 1998], a Sequential Quadratic Programming solver which is suitable for solving large nonlinearly constrained problems. MINLP was developed by Roger Fletcher and Sven Leyffer at the University of Dundee.

5.2.2   AMPL and NEOS

We design our numerical experiments using the modelling language AMPL [Fourer et al., 1993] and run them on NEOS [Czyzyk et al., 1998; Dolan, 2001; Gropp and Moré, 1997]. Using AMPL developed at Bell Laboratories, we can im-plement the numerical experiments with familiar mathematical notations and concepts. NEOS provides us accesses to powerful computation hardware and accesses to the solvers mentioned in the last section using AMPL as their interfaces, which facilitates the solution of our models.

5.3 Experiments and Numerical Results

5.3.1 k-fold Cross Validation

k-fold cross validation [Kohavi, 1995] is a widely used method to evaluate the prediction ability of the obtained classification model. In this validation method, the data set is randomly partitioned into k subsets, with the samples at each risk level evenly distributed in the k sets. The training procedure is then repeated k times. Each time, one of the k subsets is left out as a validation set and the others are used as a training set. After running the training procedure on the training set, the obtained classifier is validated on the validation set and the validation error rates are recorded. When all the k training-validation procedures are finished, the average validation error rates is used as the approximation of the predication error rates of the classification model.

We can see that each training-validation procedure takes $\frac{k-1}{k}$ of the data set as a training set. On the one hand, the smaller k is chosen, the less the proportion of the training set is, hence most probably a underestimated classifier with poorer generalization ability are produced. On the other hand, a larger k means running the training-validation procedure more times and consumes more computational power. In this thesis, we take k as 10, a widely used trade-off, which means 90 percent of sample is used as training set and

10 percent is used as validation set each time.

5.3.2   Performance of Unimodal Criteria

The first experiment is desired to check whether the obtained classifiers will
enjoy a better performance if we relax some criteria to be unimodal. Table 5.2
and Table 5.3 show the validation accuracy of each level on the two data sets

Table 5.2: 10-fold Cross-validation Accuracy of UTADIS and GUTADIS(1)

S&P's	UTADIS (40C, No Perturbation, 57.0%)							
	AAA	AA	A	BBB	BB	B	CCC	CC–D
AAA	80.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0
AA	13.6	63.6	18.2	4.5	0.0	0.0	0.0	0.0
A	0.0	29.4	52.9	17.6	0.0	0.0	0.0	0.0
BBB	0.0	0.0	10.7	50.0	39.3	0.0	0.0	0.0
BB	0.0	0.0	0.0	27.3	51.5	12.1	6.1	3.0
B	0.0	0.0	0.0	0.0	50.0	41.7	8.3	0.0
CCC	0.0	0.0	0.0	0.0	33.3	0.0	33.3	33.3
CC–D	0.0	0.0	0.0	0.0	33.3	0.0	0.0	66.7
	GUTADIS (40/6C, No Perturbation, 80.6%)							
AAA	80.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0
AA	0.0	95.5	4.5	0.0	0.0	0.0	0.0	0.0
A	0.0	5.9	76.5	11.8	5.9	0.0	0.0	0.0
BBB	0.0	0.0	0.0	71.4	28.6	0.0	0.0	0.0
BB	0.0	0.0	0.0	12.1	72.7	15.2	0.0	0.0
B	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
CCC	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
CC–D	0.0	0.0	0.0	0.0	0.0	0.0	33.3	66.7

respectively. Take for an example, the numbers at row AAA in UTADIS part
means that 80 percent of country-years at AAA level are correctly classified
by UTADIS model, while 20 percent are classified as AA level, and so on. 40C

Table 5.3: 10-fold Cross-validation Accuracy of UTADIS and GUTADIS(2)

S&P's	UTADIS (40C, No Perturbation, 51.8%)					
	AAA	AA	A	BBB	BB	B
AAA	81.8	9.1	9.1	0.0	0.0	0.0
AA	18.2	45.5	36.4	0.0	0.0	0.0
A	0.0	15.4	46.2	30.8	7.7	0.0
BBB	0.0	13.3	20.0	40.0	20.0	6.7
BB	0.0	0.0	0.0	35.7	28.6	35.7
B	0.0	0.0	0.0	6.3	18.8	75.0
	GUTADIS (40/6C, No Perturbation, 86.9%)					
AAA	81.8	18.2	0.0	0.0	0.0	0.0
AA	9.1	81.8	9.1	0.0	0.0	0.0
A	0.0	0.0	84.6	15.4	0.0	0.0
BBB	0.0	0.0	0.0	93.3	6.7	0.0
BB	0.0	0.0	0.0	0.0	100.0	0.0
B	0.0	0.0	0.0	6.3	12.5	81.3

or 40/6C in the table means how many criteria are used and how many are unimodal. 'No Perturbation' indicates that no postoptimality analysis is done.

We can see from Table 5.2 and Table 5.3 that the predication accuracy of GUTADIS is much better than that of UTADIS especially for those at higher risk levels. The overall validation accuracy of these two methods are at the level of 80% and 50% respectively. The accuracy of UTADIS is also tested using other linear solvers in NEOS and there is no much difference with PCx, all of which are roughly the same as the result shown in Doumpos et al. [2001b], i.e., with an overall accuracy 57.7%.

We also present in Table 5.4 an result where the risk levels of sample are relabelled as Investment(1), Speculative(2) or Non-Investment(3) according to

Table 1.1. The accuracy is much better due to the increasing number of sample in each risk level.

Table 5.4: 10-fold Cross-Validation Accuracy of UTADIS and GUTADIS(3)

S&P's	UTADIS			GUTADIS		
	1	2	3	1	2	3
1	94.9	5.1	0.0	96.6	3.4	0.0
2	3.3	88.5	8.2	0.0	91.8	8.2
3	0.0	33.3	66.7	0.0	5.6	94.4
Overall		88.3			94.2	
1	90.0	10.0		88.0	12.0	
2	23.3	76.7		0.0	100.0	
Overall		85.0			92.5	

5.3.3 Performance of Concave Utility Functions

The next experiment shows that, GUTADIS achieves a better performance than UTADIS even if we restrict some of the unimodal and monotone criteria to be concave. Ten of the marginal functions are confined to be concave or convex in this experiment. Note that, by imposing a convexity constraint we reduced the feasible regions of (3.8), hence possibly a larger training and validation error rate.

The validation accuracy of this experiment is shown in Table 5.5 and Table 5.6. The overall validation accuracy is at the level of 70%, lower than GUTADIS but still much better than UTADIS as shown in Table 5.2.

Table 5.5: 10-fold Cross-Validation Accuracy of GUTADIS With Concave and Convex Constraints(1)

S&P's	GUTADIS (40/6/10C, 73.8%)							
	AAA	AA	A	BBB	BB	B	CCC	CC–D
AAA	70.0	25.0	5.0	0.0	0.0	0.0	0.0	0.0
AA	9.1	86.4	4.5	0.0	0.0	0.0	0.0	0.0
A	0.0	11.8	76.5	11.8	0.0	0.0	0.0	0.0
BBB	0.0	0.0	3.6	71.4	25.0	0.0	0.0	0.0
BB	0.0	0.0	0.0	12.1	69.7	18.2	0.0	0.0
B	0.0	0.0	0.0	0.0	16.7	75.0	8.3	0.0
CCC	0.0	0.0	0.0	0.0	0.0	0.0	66.7	33.3
CC–D	0.0	0.0	0.0	0.0	0.0	0.0	66.7	33.3

Table 5.6: 10-fold Cross-Validation Accuracy of GUTADIS With Concave and Convex Constraints(2)

S&P's	GUTADIS (40/6/10C,75.4%)					
	AAA	AA	A	BBB	BB	B
AAA	72.7	27.3	0.0	0.0	0.0	0.0
AA	9.1	81.8	9.1	0.0	0.0	0.0
A	0.0	7.7	61.5	30.8	0.0	0.0
BBB	0.0	0.0	0.0	73.3	26.7	0.0
BB	0.0	0.0	0.0	0.0	92.9	7.1
B	0.0	0.0	0.0	0.0	25.0	75.0

5.3.4   Performance of QUDIS

The performance of QUDIS is tested using nonlinear solver filter [Fletcher and Leyffer, 1998]. The cross-validation accuracy is shown in Table 5.7 and Table 5.8. From these tables, we can see that the overall accuracy of QUDIS is at the level of 70% in the special case of $\alpha_{ij} = 1$.

Comparing to the case that α_{ij} are variables, QUDIS doesn't lose much performance by restricting all α_{ij} to be equal to 1. The computation time

for the first two cases are all very small, less than 5 minutes for the 10-fold cross-validation process, while the third takes about ten minutes. This shows that QUDIS($\alpha_{ij} = 1$) enjoys a better performance than UTADIS without a significant increase in computational cost.

Table 5.7: 10-fold Cross-Validation Accuracy of QUDIS(1)

S&P's	QUDIS($\alpha_{ij} = 1$, 78.6%)							
	AAA	AA	A	BBB	BB	B	CCC	CC–D
AAA	95.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0
AA	0.0	86.4	13.6	0.0	0.0	0.0	0.0	0.0
A	0.0	5.9	76.5	11.8	5.9	0.0	0.0	0.0
BBB	0.0	0.0	3.6	71.4	21.4	3.6	0.0	0.0
BB	0.0	0.0	0.0	9.1	78.8	12.1	0.0	0.0
B	0.0	0.0	0.0	0.0	25.0	66.7	0.0	8.3
CCC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
CC–D	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
	QUDIS($\alpha_{ij} = \alpha_i \alpha_j$ are variables, 80.5%)							
	AAA	AA	A	BBB	BB	B	CCC	CC–D
AAA	90.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0
AA	4.5	86.4	9.1	0.0	0.0	0.0	0.0	0.0
A	0.0	5.9	82.4	5.9	5.9	0.0	0.0	0.0
BBB	0.0	0.0	0.0	67.9	32.1	0.0	0.0	0.0
BB	0.0	0.0	0.0	12.1	81.8	6.1	0.0	0.0
B	0.0	0.0	0.0	0.0	8.3	75.0	16.7	0.0
CCC	0.0	0.0	0.0	0.0	0.0	33.3	66.7	0.0
CC–D	0.0	0.0	0.0	0.0	0.0	0.0	33.3	66.7
	QUDIS(α_{ij} are variables, 78.4%)							
	AAA	AA	A	BBB	BB	B	CCC	CC–D
AAA	90.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0
AA	4.5	77.3	18.2	0.0	0.0	0.0	0.0	0.0
A	0.0	5.9	82.4	5.9	5.9	0.0	0.0	0.0
BBB	0.0	0.0	0.0	64.3	35.7	0.0	0.0	0.0
BB	0.0	0.0	0.0	9.1	78.8	12.1	0.0	0.0
B	0.0	0.0	0.0	0.0	16.7	75.0	8.3	0.0
CCC	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
CC–D	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0

Table 5.8: 10-fold Cross-Validation Accuracy of QUDIS(2)

S&P's	QUDIS($\alpha_{ij} = 1$, 81.4%)					
	AAA	AA	A	BBB	BB	B
AAA	72.7	27.3	0.0	0.0	0.0	0.0
AA	0.0	90.9	9.1	0.0	0.0	0.0
A	0.0	7.7	69.2	23.1	0.0	0.0
BBB	0.0	0.0	0.0	80.0	20.0	0.0
BB	0.0	0.0	0.0	0.0	85.7	14.3
B	0.0	0.0	0.0	6.3	6.3	87.5
S&P's	QUDIS($\alpha_{ij} = \alpha_i \alpha_j$ are variables, 83.3%)					
	AAA	AA	A	BBB	BB	B
AAA	81.8	18.2	0.0	0.0	0.0	0.0
AA	9.1	81.8	9.1	0.0	0.0	0.0
A	0.0	7.7	69.2	23.1	0.0	0.0
BBB	0.0	0.0	0.0	80.0	20.0	0.0
BB	0.0	0.0	0.0	0.0	92.9	7.1
B	0.0	0.0	0.0	0.0	6.3	93.8
S&P's	QUDIS(α_{ij} are variables, 83.3%)					
	AAA	AA	A	BBB	BB	B
AAA	81.8	18.2	0.0	0.0	0.0	0.0
AA	9.1	81.8	9.1	0.0	0.0	0.0
A	0.0	7.7	69.2	23.1	0.0	0.0
BBB	0.0	0.0	0.0	80.0	20.0	0.0
BB	0.0	0.0	0.0	0.0	92.9	7.1
B	0.0	0.0	0.0	0.0	6.3	93.8

Chapter 6

Conclusion and Future Works

In this thesis, we have generalized the UTADIS method for country risk classification in two different ways. First, by utilizing the mix-integer nonlinear optimization technique, we extend the UTADIS model such that it can deal with unimodal criteria. In particular, we show that in case that the unimodal criteria is concave, we can still use a linear program to find a classifier. Secondly, by introducing a quadratic utility function we relaxed the assumption in UTADIS model that the factors used are preferentially independent. This again produces a nonconvex quadratic constrained program.

Numerical results shows that the introduction of unimodality on criteria can improve significantly the performance of UTADIS. Imposing the concavity and unimodality on the marginal utility function can make the resulting classifier more consistent with a rater's background knowledge. Compared with the original UTADIS model, as shown by our numerical result, our new LP model

can achieve higher accuracy without sacrificing the computational efficiency.

There are a few issues that need further study for the UTADIS and GUTADIS models. First, the traditional postoptimality analysis for UTADIS model should be improved as we already mentioned in Section 3.2.4 (page 56).

Secondly, we point out that the traditional postoptimality analysis is not the correct way to find a robust classifier, as it did not address the issue of how to find a robust classifier when the input data contains noise. Regarding to country risk classification, the input data such as the values of numeric factors can hardly be accurate since the collection of data involves huge amount of humane work.

We note that in the optimization community, there are a lot of works dealing with optimization problems for which the input data is inaccurate. This is typically referred to robust optimization. In principle, finding a robust classifier at the presence of noisy data can be modelled as a robust optimization problem. However, it remains an issue how to select a suitable robust optimization model for our task and test the performance of the selected robust optimization model.

Finally, our model can also be applied to other multi-class classification problems. It will be interesting to test and compare our model with other classification models in various scenarios.

Appendix

1) *Population, energy, environment and infrastructure.* These group of factors indicate the general development status of a country which is supposed to have the ability to classify the rated countries in some degree.

   (i) *Birth rate, crude (per 1,000 people)* indicates the number of live births per 1000 population occurring during the year.

   (ii) *Population growth (annual %)* indicates the annual population growth rate. For some countries, there is a big difference between population growth rate and birth rate. We include both of them to find the one more related.

   (iii) *Age dependency ratio (dependents to working-age population)* is the ratio of dependents–people younger than 15 and older than 64– to the working-age population–those ages 15-64. For example, 0.7 means there are 7 dependents for every 10 working-age people.

   (iv) *Employment in agriculture (% of total employment)* is the propor-

tion of total employment recorded as working in the agricultural sector.

(v) *GDP per unit of energy use (PPP $ per kg of oil equivalent)* is the PPP GDP per kilogram of oil equivalent of commercial energy use. PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as a U.S. dollar has in the United States.

(vi) *Commercial energy use (kg of oil equivalent per capita)* refers to apparent consumption, which is equal to indigenous production plus imports and stock changes, minus exports and fuels supplied to ships and aircraft engaged in international transport.

(vii) *Energy imports, net (% of commercial energy use)* are calculated as energy use less production, both measured in oil equivalents. A negative value indicates that the country is a net exporter.

(viii) *CO2 emissions (kg per 1995 US$ of GDP)* are those stemming from the burning of fossil fuels and the manufacture of cement. They include contributions to the carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.

(ix) *Telephone mainlines (per 1,000 people)* are telephone lines connect-

ing a customer's equipment to the public switched telephone network. Data are presented per 1,000 people for the entire country.

(x) *Mobile phones (per 1,000 people)* refers to users of portable telephones subscribing to an automatic public mobile telephone service using cellular technology that provides access to the public switched telephone network, per 1,000 people.

(xi) *Internet users (per 1,000 people)* are people with access to the worldwide network.

2) *Science, education and health.* This group of factors are believed to be important for a country to develop further, thus important for determining country risk.

(i) *Research and development expenditure (% of GNI)* are current and capital expenditures (including overhead) on creative, systematic activity intended to increase the stock of knowledge. Included are fundamental and applied research and experimental development work leading to new devices, products, or processes.

(ii) *High-technology exports (% of manufactured exports)* are exports of products with high R&D intensity. They include high-technology products such as in aerospace, computers, pharmaceuticals, scien-

tific instruments, and electrical machinery.

(iii) *Patent applications, total* are the number of applications filed with a national patent office for exclusive rights for an invention, including both residents'and nonresidents'. A patent provides protection for the invention to the owner of the patent for a limited period, generally 20 years.

(iv) *Mortality rate, infant (per 1,000 live births)* is the number of infants dying before reaching one year of age, per 1,000 live births in a given year.

(v) *Public spending on education, total (% of GDP)* consists of public spending on public education plus subsidies to private education at the primary, secondary, and tertiary levels.

(vi) *Human Development Index* is a summary measure of human development. It measures the average achievements in a country in three basic dimensions of human development: life expectancy at birth, adult literacy rate and GDP per capita (PPP US$).

3) *Political and social status.* These factors characterize the stability, safety and equality in a country.

(i) *POLITY* is a composition index indicating the autocracy and democ-

racy of a country, taking value from -10 to 10.

(ii) *Military expenditure (% of central government expenditure)* is the ratio of military expenses over total government expense.

(iii) *Military expenditure (% of GNI)* is the ratio of military expenses over gross national income.

(iv) *GINI index (%)* measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution. A Gini index of zero represents perfect equality, while an index of 100 implies perfect inequality.

(v) *Unemployment, total (% of total labor force)* refers to the share of the labor force that is without work but available for and seeking employment. Definitions of labor force and unemployment differ by country.

4) *Economy.* These factors are among the most important ones contributing to country risk classification.

(i) *goverment surplus (% of GDP)* is the difference between current revenue, including all revenue to the central government from taxes and nonrepayable receipts, and total expenditure of central govern-

ment measured as a share of GDP.

(ii) *Final consumption expenditure, etc. (% of GDP)* is the sum of household final consumption expenditure (private consumption) and general government final consumption expenditure (general government consumption).

(iii) *GDP (constant 1995 US$)* is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products.

(iv) *GDP growth (annual %)* is annual percentage growth rate of GDP at market prices based on constant local currency.

(v) *Inflation, consumer prices (annual %)* measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a fixed basket of goods and services that may be fixed or changed at specified intervals, such as yearly.

(vi) *Current account balance (% of GDP)* is the sum of net exports of goods, services, net income, and net current transfers.

(vii) *Agriculture, value added (% of GDP)* is the net output of agriculture sector after adding up all outputs and subtracting intermediate inputs.

(viii) *Industry, value added (% of GDP)* comprises value added in mining, manufacturing (also reported as a separate subgroup), construction, electricity, water, and gas.

(ix) *Services, etc., value added (% of GDP)* includes value added in wholesale and retail trade (including hotels and restaurants), transport, and government, financial, professional, and personal services such as education, health care, and real estate services. Also included are imputed bank service charges, import duties, and any statistical discrepancies noted by national compilers as well as discrepancies arising from rescaling.

(x) *Trade (% of GDP)* Trade is the sum of exports and imports of goods and services measured as a share of gross domestic product.

5) Finance.

(i) *Total external debt (% of exports of goods and services)* is the ratio of debt owed to nonresidents repayable in foreign currency, goods, or services to exports. Total external debt is the sum of public, publicly guaranteed, and private nonguaranteed long-term debt, use of IMF credit, and short-term debt.

(ii) *Total external debt (% of GNI)* is the ratio of total external debt to

GNI. GNI (formerly GNP) is the sum of value added by all resident producers plus any product taxes (less subsidies) not included in the valuation of output plus net receipts of primary income (compensation of employees and property income) from abroad.

(iii) *Total debt service (% of exports of goods and services)* is the sum of principal repayments and interest actually paid in foreign currency, goods, or services on long-term debt, interest paid on short-term debt, and repayments (repurchases and charges) to the IMF.

(iv) *Gross international reserves in months of imports* comprise holdings of monetary gold, special drawing rights (SDRs), the reserve position of members in the International Monetary Fund (IMF), and holdings of foreign exchange under the control of monetary authorities. The gold component of these reserves is valued at year-end (December 31) London prices.

(v) *Short-term debt (% of total external debt)* Short-term debt includes all debt having an original maturity of one year or less and interest in arrears on long-term debt.

(vi) *Gross international reserves (includes gold, current US$)* comprise holdings of monetary gold, special drawing rights, reserves of IMF members held by the IMF, and holdings of foreign exchange under

the control of monetary authorities.

(**vii**) *Short-term debt (% of Gross international reserves including gold)* includes all debt having an original maturity of one year or less and interest in arrears on long-term debt.

(**viii**) *Gross national savings, including NCTR (% of GNI)* is equal to gross domestic savings plus net income and net current transfers from abroad.

Bibliography

Afonso, A. (2003). Understanding the determinants of sovereign debt ratings: Evidence for the two leading agencies, *Journal of Economics and Finance* **27**(1): 56–74.

Alexe, S., Hammer, P. L., Kogan, A. and Lejeune, M. A. (2003). A non-recursive regression model for country risk rating, *Technical Report RRR 9-2003*, Rutgers University Center for Operations Research, Piscataway, New Jersey.

Anderson, T. W. and Bahadur, R. R. (1962). Classification into two multivariate normal distributions with defferent covariance matrices, *The Annals of Mathematical Statistics* **33**(2): 420–431.

Brewer, T. L. and Rivoli, P. (1990). Politics and perceived country creditworthiness in international banking, *Journal of Money, Credit and Banking* **22**(3): 357–369.

Chambers, J. and Beers, D. T. (1998). Sovereign ratings display stability over two decades, *Ratings Performance: Stability and Transition* pp. 55–59.

Czyzyk, J., Mehrotra, S., Wagner, M. and Wright, S. J. (1997). PCx user guide, *Technical Report OTC 96/01*, Optimization Technology Center.

Czyzyk, J., Mesnier, M. and Moré, J. (1998). The NEOS server, *IEEE Journal on Computational Science and Engineering* **5**: 68–75.

Dolan, E. (2001). The NEOS server 4.0 administrative guide, *Technical Memorandum ANL/MCS-TM-250*, Mathematics and Computer Science Division, Argonne National Laboratory.

Doumpos, M. and Zopounidis, C. (2001). Assessing financial risks using a multicriteria sorting procedure: the case of country risk assessment, *Omega* **29**: 97–109.

Doumpos, M. and Zopounidis, C. (2002). *Multicriteria Decision Aid Classification Methods*, Vol. 73 of *Applied Optimization*, Kluwer Academic Publishers, Norwell, Massachusetts.

Doumpos, M., Pentaraki, K. and Zopounidis, C. (2001a). Assessing country risk using a multi-group discrimination method: a comparative analysis, *Managerial Finance* **27**(8): 16–34.

Doumpos, M., Zanakis, S. H. and Zopounidis, C. (2001b). Multicriteria preference disaggregation for classification problems with an application to global investing risk, *Decision Sciences* **32**(2): 333–385.

Feder, G. and Just, R. E. (1977). A study of debt servicing capacity applying logit analysis, *Journal of Development Economics* **4**: 25–38.

Fishburn, P. C. (1970). *Utility Theory for Decision Making*, John Wiley & Sons, Inc., Chichester, U.K.

Fletcher, R. and Leyffer, S. (1998). User manual for filterSQP, *Numerical Analysis Report NA/181*, University of Dundee, Dundee, U.K.

Fourer, R., Gay, D. M. and Kernighan, B. W. (1993). *AMPL A Modeling Language for Mathematical Programming*, The Scientific Press.

Frank, Jr, C. R. and Cline, W. R. (1971). Measurement of debt servicing capacity: An application of discriminant analysis, *Journal of International Economics* **1**: 327–344.

Gropp, W. and Moré, J. (1997). Optimization environments and the NEOS server, *in* M. D. Buhmann and A. Iserles (eds), *Approximation Theory and Optimization*, Cambridge University Press, pp. 167–182.

Haque, N. U., Kumar, M. S., Mark, N. and Mathieson, D. (1998). The relative

importance of political and economic variables in creditworthiness ratings, *International Monetary Fund Working Paper* **46**: 1–13.

Hu, Y.-T., Kiesel, R. and Perraudin, W. (2002). The estimation of transition matrices for sovereign credit ratings, *Journal of Banking and Finance* **26**: 1383–1406.

IMF (1971). *International Financial Statistics Yearbook*, English edn, International Monetary Fund, Washington, D.C.

IMF (1995). *International Financial Statistics Yearbook*, English edn, International Monetary Fund, Washington, D.C.

Keeney, R. L. and Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value trade-offs*, Cambridge University Press, Cambridge, U.K.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection, *IJCAI*, pp. 1137–1145.

Leyffer, S. (1999). *User manual for MINLP_BB*, University of Dundee, Dundee, U.K.

Saini, K. G. and Bates, P. S. (1984). A survey of the quantitative approaches to country risk analysis, *Journa of Banking and Finance* **8**: 341–356.

Smet, Y. D. and Gilbart, F. (2001). A class definition method for country risk problems, Retrieved November 2003, from Université Libre de Bruxelles, Service de Mathématiques de la Gestion Web site: http://smg.ulb.ac.be/~ydesmet/download/YDESMET2001_1.doc.

Somerville, R. A. and Taffler, R. J. (1995). Banker judgement versus formal forecasting models: The case of country risk assessment, *Journal of Banking and Finance* **19**: 281–297.

Spathis, C., Doumpos, M. and Zopounidis, C. (2002). Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques, *The European Accounting Review* **11**(3): 509–535.

Stout, Q. F. (2000). Optimal algorithms for unimodal regression, *in* E. J. Wegman and Y. M. Martinez (eds), *Computing Science and Statistics*, Vol. 32, Los Alamos National Laboratories, RAND and Neptune and Company, Inc., Interface Foundation Of North America, Inc., New Orleans, Louisiana, pp. 348–355.

The World Bank (2002). *World Development Indicators on CD-ROM*, IBRD, World Bank, Washington, D.C.

UNCTAD (2002). World investment report 2002 - overview, transnational

corporations and export competitiveness, *Technical report*, United Nations Conference on Trade and Development, New York and Geneva.

Vincke, P. (1992). *Multicriteria Decision-Aid*, John Wiley & Sons Ltd, Chichester, U.K.

Ye, Y. (1997). *Interior Point Algorithms, Theory and Analysis*, John Wiley & Sons, Inc., Chichester, U.K.

Zopounidis, C. and Doumpos, M. (1999). A multicriteria decision aid methodology for sorting decision problems: The case of financial distress, *Computational Economics* **14**: 197–218.

Zopounidis, C. and Doumpos, M. (2002). Multi-criteria decision aid in financial decision making: methodologies and literature review, *Journal of Multi-Criteria Decision Analysis* **11**: 167–186.

6797 20