SELECTION OF OPTIMAL THRESHOLD AND NEAR-OPTIMAL INTERVAL USING PROFIT FUNCTION AND ROC CURVE: A RISK MANAGEMENT APPLICATION

A Dissertation
Submitted to
the Temple University Graduate Board

---

in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

---

by
Jingru Chen
January, 2011

Examining Committee:

Francis Hsuan, Advisory Chair, Statistics
Damaraju Raghavarao, Statistics
Milton Parnes, Statistics
Jiawei Liu, Statistics, Georgia State University

ABSTRACT

SELECTION OF OPTIMAL THRESHOLD AND NEAR-OPTIMAL INTERVAL USING PROFIT FUNCTION AND ROC CURVE: A RISK MANAGEMENT APPLICATION

Jingru Chen

DOCTOR OF PHILOSOPHY

Temple University, January, 2011

Professor Francis Hsuan, Chair

The ongoing financial crisis has had major adverse impact on the credit market. As the financial crisis progresses, the skyrocketing unemployment rate puts more and more customers in such a position that they cannot pay back their credit debts. The deteriorating economic environment and growing pressures for revenue generation have led creditors to re-assess their existing portfolios. The credit re-assessment is to accurately estimate customers' behavior and distill information for credit decisions that differentiate bad customers from good customers. Lending institutions often need a specific rule for defining an optimal cut-off value to maximize revenue and minimize risk.

In this dissertation research, I consider a problem in the broad area of credit risk management: the selection of critical thresholds, which comprises of the "optimal cut-off point" and an interval containing cut-off points near the optimal cut-off point (a "near-optimal interval"). These critical thresholds can be used in practice to adjust credit lines, to close accounts involuntarily, to re-price, etc. Better credit re-assessment practices are essential for banks to prevent loan loss in the future and restore the flow of credit to entrepreneurs and individuals.

The Profit Function is introduced to estimate the optimal cut-off and the near-optimal interval, which are used to manage the credit risk in the finan-

cial industry. The credit scores of the good population and bad population are assumed from two distributions, with the same or different dispersion parameters. In a homoscedastic Normal-Normal model, a closed-form solution of optimal cut-off and some properties of optimal cut-off are provided for three possible shapes of the Profit Functions. The same methodology can be generalized to other distributions in the exponential family, including the heteroscedastic Normal-Normal Profit Function and the Gamma-Gamma Profit Function. It is shown that a Profit Function is a comprehensive tool in the selection of critical thresholds, and its solution can be found using easily-implemented computing algorithms.

The estimation of near-optimal interval is developed in three possible shapes of the bi-distributional Profit Function. The optimal cut-off has a closed-form formula, and the estimation results of near-optimal intervals can be simplified to this closed-form formula when the tolerance level is zero.

Two nonparametric methods are introduced to estimate critical thresholds if the latent risk score is not from some known distribution. One method uses the Kernel density estimation method to derive a tabulated table, which is used to estimate the values of critical thresholds. A ROC Graphical method is also developed to estimate critical thresholds.

In the theoretical portion of the dissertation, we use Taylor Series and the Delta method to develop the asymptotic distribution of the non-constrained optimal cut-off. We also use the Kernel density estimator to derive the asymptotic variance of the Profit function.

ACKNOWLEDGEMENTS

First and foremost I offer my sincerest gratitude to my advisor, Dr. Francis Hsuan, who has supported me thoughout my thesis with his patience and expertise. His insights into many aspects of risk scoring and logical way of thinking have been of great value for me. Without him, this thesis would not have been completed or written. I am indebted to him more than he knows.

Besides my advisor, I am deeply grateful to the rest of my thesis committee members: Dr. Damaraju Raghavarao, Dr. Milton Parnes and Dr. Jiawei Liu. Their tremendous support and guidance have been of great value in this study and have made this dissertation definitely better. A great deal of thanks goes to all the professors in Statistics Department, Temple University.

My special thanks should go to Dr. Mark Wardell and Dr. James Stewart, who were my advisors at The Pennsylvania State University. Their encouragement and tremendous support helped me transfer to the filed of Statistics.

I would like to express my endless thanks to my family members, especially my father, Zujin Chen, for their unconditional support and steady encouragement, for listening to my complaints and frustration, and for believing me. My mother sadly died before I could tell her I would complete my PhD studies. I also want to express my great appreciation to two friends, Dr. Wei Huang and Dr. Yang Lu, for their sincere help in the past.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background

The ongoing financial crisis has had major adverse impact on the credit market. As the financial crisis progresses, the skyrocketing unemployment rate puts more and more customers in such a position that they cannot pay back their credit debts. The deteriorating economic environment and growing pressures for revenue generation have led creditors to re-assess their existing portfolios. In this dissertation research, I consider a problem in the broad area of credit risk management: the selection of critical thresholds, which comprises of the "optimal cut-off point" and a "near-optimal interval" (the mathematical definitions for the two quoted terms are given later in this section). These critical thresholds can be used in practice to adjust credit lines, to close accounts involuntarily, to re-price, etc. Better credit re-assessment practices are essential for banks to prevent loan loss in the future and maintain the flow of credit to entrepreneurs and individuals.

Most credit re-assessment models used in practice concern the Receiver Operating Characteristic (ROC) Curve, which is a common method for organizing, visualizing, and comparing model classifiers based on the simple default probability and its associated statistics. However, models based on a measurement on the profit maximization is perhaps more attractive to banks

and lending institutions, since creditors need to balance the objectives of maximizing the profit and minimizing the exposure to risks. Lending institutions often need a specific rule for defining an optimal cut-off value to maximize revenues and minimize risks.

Both Profit Functions and Receiver Operating Characteristics (ROC) curves can be used to find the optimal cut-off point as well as a near-optimal interval. The Profit Function is defined as the expected net income, resulting from either a correct or wrong decision. Good/Bad status and Accept/Reject decision form a 2-by-2 table, and classify the applicants into four categories. A Profit Function, rigorously defined in equation (1.2) on page 5, incorporates four classification probabilities and the revenue/cost associated with the correct/wrong decision. It is a comprehensive tool in the selection of critical thresholds. This dissertation presents my investigation on the problem of the selection of the optimal threshold and a near-optimal interval. Chapter 3 and Chapter 4 contain my main theoretical results of a parametric Profit Function. The theoretical results of a nonparametric Profit Function are discussed in the first part of Chapter 5.

An ROC curve is a plot of two conditional probabilities, with the true positive probability as the y-axis and the false positive probability as the x-axis. ROC curves have been extensively studied in the literature; for a cursory review, see Section 2.2. The evaluation of ROC curves usually focuses on several criteria, such as Kolmogorov-Smirnov test (KS), Area Under Curve (AUC), and confidence-bands/confidence intervals. For my dissertation, I propose a descriptive method to find the optimal cut-off point and the near-optimal interval in the second part of Chapter 5.

1.2 A Mathematical Setup

Suppose n applicants apply for credit loans at time 0. The personal and financial characteristics of an individual, such as income, current debt, number of credit inquiries, etc, are denoted by a vector of attributes $\boldsymbol{x} = (\boldsymbol{x_1}, \ldots, \boldsymbol{x_m})$.

Let D be a binary variable to reflect the status of one's observed net cash flow (NCF) during the last 12 months. The value of D is either Bad(B), indicating a non-positive NCF, or Good(G), indicating a positive NCF.

A typical process of credit re-assessment is to develop a scorecard that assigns a score y^* to each existing customer (ECM), and then to select a cut-off point that determines acceptance or rejection. This score can be risk score, promotion score, or collection score, etc, in different business environments. This dissertation focuses on the portfolio score to decide whether further credit approval should be granted. We let y^* be a latent variable derived from generalized linear model to represent the creditworthiness. Once the creditworthiness $y^{*\prime}s$ are generated from the credit-scoring model for all n applicants, we rank the values from low to high and define the range of y^* as $[y^*_{min}, y^*_{max}]$. If the creditworthiness value is greater than a selected threshold, k, the creditor approves the credit application of the applicant. Let Y_k be the binary decision variable based on the creditworthiness score and the threshold k, with the value B = Rejection and G = Acceptance, that is

$$Y_k = \begin{cases} G(Granting\ credit) & \text{if } y^* > k, \\ B(Denying\ credit) & \text{if } y^* \leq k \end{cases} \tag{1.1}$$

The two binary variables Y and D form a 2-by-2 contingency table, as shown in Table 1.1. Because the decision of acceptance or rejection may be either correct or incorrect, there are four conditional probabilities associated with this table. The sensitivity, also named as the True Positive rate (TP), is the conditional probability of making the correct decision when an applicant has good creditworthiness. And the specificity, also named as the True Negative rate (TN), is the conditional probability of making the correct decision when an applicant has bad creditworthiness. Two complementary terms, False Positive rate (FP) and False Negative rate (FN), are denoted as $FP = 1-TN$ and $FN = 1 - TP$. From definition (1.1), it is clear that when k decreases, more applicants will pass the criterion $y^* > k$, and therefore TP (sensitivity) and FP(1-specifity) will increase. The ROC curve is a plot of TP(the y-axis)

Table 1.1: Classification of Decision Results by Population Status

		Y (Decision)		
	Probability	B(Rejection)	G(Acceptance)	
D	Bad(B)	TN P_B=$(1-FP)$ P_B	FP P_B	P_B
(Actual)	Good(G)	FN P_G=$(1-TP)$ P_G	TP P_G	P_G

versus FP(the x-axis), when k varies.

The conditional probabilities in Table 1.1 are associated with costs or revenues. These payoffs have multiple sources, either from a one-time charge or some long-term accumulations. TN and FN are results from the decision of denying loan applications, and are associated with r_B (revenue of true negative) and c_G (cost of false negative) respectively. Some lending processes charge certain application fees, which are the main source of r_B. Lending institutions usually have a second selection process for the groups of true negative and false negative, in order to further filter these rejected applicants, which incur more operational expense. It usually takes more effort to study the false negative group than to study the true negative group. That is the main source of c_G. TP and FP are results from the decision of granting credit and are associated with r_G and c_B, respectively. These revenue and cost variables could construct a payoff matrix. In the analysis, the revenues of correct decisions (r_G, r_B) and the costs of false decisions (c_G, c_B) are usually assumed to be independent of the model (Stein, 2005). In Chapter 6, we present a numerical study that deploys several payoff matrices for two reasons. One is that a major credit institution has large portfolio. Each segment of customers in the whole portfolio may have very different characteristics, which justify the existence of various payoff matrices. The other is that we want to observe the impact of using various factors in the sensitivity study.

The two row marginal probabilities of Table 1.1, P_G and P_B, also play roles in the determination of the final revenue. Some literatures assume that the percentage of bad population (P_B) is 2% (Altman et al. (1977), Stein

(2005)). However, our practical experience suggests that 2% is too low and perhaps covers only the bankruptcy rate. In the current severe economic situation, the overall loss rate, including charge off and bankruptcy, is already between 6.5% and 10.5% in the credit card industry. Our practical experience suggests that the unprofitable population in the credit risk is usually in the range between 25% and 40%. For example, a significant amount of customers often takes advantage of promotion offers from various credit-card companies. These customers often have very good credit risk scores and are unlikely to file bankruptcy, but credit-card companies may not make profits from such balance-transfer behaviors.

We now define a Profit Function that incorporates conditional probabilities, marginal probabilities, as well as payoff factors. Let $R(k)$ be the *Expected Profit* per outcome / decision / customer, associated with the decision rule. The value of the expected profit curve depends on the cut-off point k and the creditworthiness variable y^*. Given the threshold k, for any model the expected profit depicts the tradeoff between revenue and cost of credit decision.

$$R(k) = \underbrace{P_G r_G TP_k + P_B r_B TN_k}_{E[Revenue(k)]} - \underbrace{(P_G c_G FN_k + P_B c_B FP_k)}_{E[Cost(k)]} \tag{1.2}$$

where

P_G = the proportion of applicants having good creditworthiness,

$P_B = 1 - P_G$,

r_G = the revenue associated with a True Positive outcome,

c_B = the cost associated with a False Positive outcome,

r_B = the revenue associated with a True Negative outcome,

c_G = the cost associated with a False Negative outcome.

And, for a given threshold k,

E[Revenue(k)] = the expected revenue of a credit applicant for the given k,

E[Cost(k)] = the expected cost of a credit applicant for the given k.

$TP_k = P(y^* > k|D = G) = P(Y = G|D = G)$,

$FP_k = P(y^* > k|D = B) = P(Y = G|D = B)$,

$TN_k = P(y^* \leq k|D = B) = P(Y = B|D = B)$,
$FN_k = P(y^* \leq k|D = G) = P(Y = B|D = G)$

Note that $TP_k + FN_k = 1$, $FP_k + TN_k = 1$, and that the expected profit is the probability-weighted sum of costs and revenues.

In most occasions, the expected profit per customer is positive and the lending institutions can make profits. In some severe occasions the expected profit may be negative and the lending institutions can lose money. For example, in the first quarter of 2009, the credit loss rates were skyrocketing to more than 8.0% in major credit-card issuers (Earning release). An issuer can make profit if the loss rate is controlled at 7.5% or below, therefore the expected profits of these issuers were negative. As a response, these credit card issuers cut marketing expenses but still mailed out millions of credit-approval letters to attract outstanding balance, with the intention to control the loss rate appearing on the quarterly and annual earnings reports.

Our objective in this dissertation is to find the optimal threshold, denoted by k^*, which maximizes the expected profit. We provide both point and interval estimations of k^*. Researchers usually prefer interval estimation to point estimation. In the business environment, the creditor institutions are more interested in some interval of k^* to implement various marketing strategies and to absorb the impact of population shift in the applicants over time. Here we introduce the idea of "*near-optimality*" and construct an interval $[k_L^*, k_U^*]$ around the optimal cut-off value k^*. Any value of k is considered to be *near-optimal*, if $R(k) \geq R(k^*) - \Delta$ for a specified tolerance level, Δ. The value of Δ can be a fixed positive constant or equal to $\psi|R(k^*)|$, where ψ is a proportion in [0, 1], such as $\psi = 0.10$, or $\psi = 0.20$.

1.3 Research Contribution

The study of the optimal cut-off and its near-optimal interval is a relatively new topic in the diagnostic studies. To our best knowledge, no studies have been published to apply a Profit Function or a ROC curve to analyze near-

optimality in the financial industry. Our work may be the first attempt to provide a comprehensive study on the optimal cut-off and the near-optimal interval, with an easily implemented solution.

The research makes contributions in policy implications. From a policy standpoint, the study provides guidance for the proposed overhaul of banking regulations in risk managements, especially in credit risk. Future extensions of the work presented in this dissertation may be used by policy makers to manage credit risk and the implementation of Basel II in the financial industry. We know that Basel II is the second of the Basel Accords, which are issued by the Basel Committee on Banking Supervision. Recently, the financial industry has started to implement Basel II to set up more rigorous risk and capital management requirements to ensure the stability of the international financial system.

Methodologically, our study has the following features. (1) We use the observed P&L performance of credit customers as the dependent variables. (2) Our study shows that the Profit Function has various types of shapes, as opposed to the only shape of a concave function in the literature. (3) We apply the Profit Function and ROC curve to look for the optimal cut-off point, which determines an acceptance or a rejection of credit applications. (4) We prove that under the assumption of y^* belonging to distributions in the exponential family, there exists at least a closed-form equation for the solution of the optimal cut-off point. Here, as well as for distribution not belonging to the exponential family, numerical algorithms may be needed to search for the optimal cut-off point. (5) We drive the asymptotic distribution of the optimal cut-off point under the normality assumption and find the asymptotic variance of the Profit Function. (6) We employ the idea of near-optimality for credit re-assessment as an alternative to confidence bands or confidence intervals of the optimal threshold. The idea can be implemented in both Profit Function method and ROC method.

A Profit Function offers a comprehensive tool in the selection of optimal threshold and near-optimal interval. The ROC analysis can be used to estimate

optimal thresholds in the nonparametric way. Then it is a convenient tool to practitioners. The numerical study confirms that the proposed methodology provides an excellent forecast on the optimal cut-off point and the near-optimal interval. Further development of the methodology in this framework can be undertaken later.

1.4 Dissertation Structure

The remaining chapters of this dissertation are organized as follows. Chapter 2 provides the literature review on the ROC curve, the Profit Function, generalized linear models, exponential family, and numerical algorithms for solving non-linear equations. Chapter 3 presents a number of new mathematical or statistical results for the problem of finding the optimal cut-off point and near-optimal intervals, under a widely-adopted assumption of normality and using the approach of maximizing the Profit Function. In Chapter 4, the theoretical results in Chapter 3 are generalized to the exponential family. We use the heteroscedastic Normal-Normal model and the homoscedastic Gamma-Gamma model to illustrate these generalized results. Chapter 5 proposes two nonparametric methods as the complementary to the parametric method. A numerical study is presented in Chapter 6. Chapter 7 concludes the dissertation by discussing the contributions of this work and offering suggestions for future research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The literature review chapter consists of six parts. Section 2.2 reviews the basics of the traditional ROC curve, its evaluation criteria, and its application to the diagnostic study. Section 2.3 discusses the application of the Profit Function to the diagnostic study. In section 2.4, we describe generalized linear model, which is a flexible generalization of the ordinary least square regression and is popular for modeling non-normal data. Section 2.5 describes the general form and the canonical form of the exponential family and proposes to use Normal and Gamma distribution in the following study. Section 2.6 discusses two numerical algorithms used in the estimation of critical thresholds. The last section (2.7) provides a summary of the chapter.

2.2 Receiver Operating Characteristic (ROC) Curve

A Receiver Operating Characteristic (ROC) Curve is a method for organizing, visualizing, and comparing model classifiers based on their performance. The ROC curve was first introduced to analyze radar signals during World War

II before it was used in signal detection theory in 1945. The signal detection theory consists of two distinct parts, the decision theory and the distribution theory (Egan 1975). The decision theory considers evidence and risk, and makes rational decisions with the goal of reaching the optimal decision. A decision goal can be maximization of an expected value, maximization of the percentages of correct responses, or maximization of a weighted combination. The maximization of a weighted combination is to maximize the probability of correct acceptance of the hypothesis while minimizing the probability of an incorrect acceptance of the hypothesis. Egan's distribution theory assumes some specific distributions and can be categorized as one of the Profit Function approaches.

Since it was proposed in the 1950s, the ROC curve has been widely used in many fields, psychophysics, medicine, epidemiology, radiology, social sciences, machine learning, etc. The evaluation of the ROC curve focuses on four aspects, Kolmogorov-Smirnov (KS) Curve, Area under Curve (AUC), confidence-bands/confidence intervals, and the threshold to classify the scoring function. The partial area under the curve, pAUC, is a summary index that restricts attention to certain area under the AUC. Indices KS, AUC, and pAUC are widely used to compare models for the ROC analysis. These statistics use either the relative distance or the area to evaluate the model performance in binary tests.

Pepe (2003), Yan, Mozer, and Wolniewicz (2003), Marzban (2004), Pepe and Cai (2004), and Fawcett (2006) discussed the AUC. Reiser and Faraggi (1997), Cortes and Mohri (2005), Macskassy and Provost (2004), Macskassy, Provost, and Rosset (2005) presented some results on the confidence band of the ROC curve and the confidence interval of the diagnostic likelihood ratio (ratio between true-positive rate and false-positive rate). Pepe (2003) proposed three approaches, non-parametric method, parametric method, and semi-parametric method, to estimate ROC curves, its summary indices, and confidence intervals. Pepe (1997), Pepe (1998), Pepe (2000), Pepe and Cai (2004), Leisenring and Pepe (1998), Pepe (2006), Zou and Hall (2000), Song

and Zhou (2006) discussed modeling approaches to study the impact of covariates on the ROC curve.

These researchers provide conceptual ideas on the comparison and selection of model candidates. But none is directly related to the optimal cut-off point and the near-optimal interval. Metz (1978) suggested that a graphical approach could be used to derive the optimal cut-off point. Stein (2005) proposed to use the tangent line to fit to the ROC curve and then manually estimate the optimal cut-off, but did not supply a detailed solution.

2.3 Bi-distribution-Based Profit Function

In contrast to huge volumes of papers on ROC curves, relatively few papers incorporate financial factors into the diagnostic study. Green and Swets (1966) proposed to maximize the expected value in a binary decision situation and provided a closed-form formula of the slope function of an ROC Curve. Egan (1975) discussed several assumptions of bi-distribution, including Normal-Normal distribution, and applied the likelihood-ratio criterion to maximize the correct percentage of a decision function in the signal detection. The likelihood-ratio method compares posterior probability to determine a decision rule. The decision rule in the signal detection is to separate two groups, including signal plus noise and noise alone. Egan's decision function did not include payoff factors, thus the decision rule is a simplified version of our result shown in equation (3.1). Noe (1983) suggested applying the ROC curve to the diagnostic utility in the field of chemistry but his work did not include selection of the optimal point. Greene (1992) proposed to use the expected profit as the objective function in the study of the credit analysis. Chauchat, Rakotomalala, Carloz, and Pelletier (2001) found that although cost-gain hypotheses do not make much difference on scoring results, the expected profit depends on the cost-gain hypothesis. Pepe (2003) used the cost function to compare the relative performance of diagnostic tests. Stein (2005) suggested that the cost function incorporated pricing relationships, and therefore, was a

more complete pricing approach. Drummond and Holte (2006) presented some simulation studies to show the superior performance of cost curves over ROC curves. These studies discussed the maximal value of the expected value for various forms of Profit Function, but did not explore the near-optimal interval.

Practical experiences suggest that the creditworthiness is a continuous latent variable. Since the creditworthiness of good population ($D = G$) and bad population ($D = B$) have very different characteristics, it is reasonable to assume that the creditworthiness of these two populations are from different distributions. Egan (1975) proposed to consider two different distributions, including Normal, Chi-square, and Bernoulli, from the same distribution family in the ROC analysis. Many other distributions are theoretically eligible for the bi-distribution-based model.

In particular, the Normal-Normal model (sometimes called the bi-normal model) plays an important role in the rigorous scientific evaluation of diagnostic tests to identify disease conditions (Pepe, 2003). Green and Swets (1966) pointed out two reasons to assume the underlying normality. One reason is that lots of data are approximately normal. Another reason is the mathematical convenience of using normal distributions. Blume (2008), Faraggi and Reiser (2002), Cai and Moskowitz (2004) provided reviews of the Normal-Normal model in diagnostic studies and concluded that the bi-normal model is a reasonable choice in diagnostic studies. Marzban (2004) compared several bi-distributional models and concluded that the Normal-Normal model offers the most realistic approximation to the underlying distribution of the diagnostic study. Our experiences have shown that, the distributions of the latent variable y^* in the good population and bad population are close to the Normal-Normal model, i.e. the probit model works well.

2.4 Generalized Linear Model

Generalized linear model, as the name implies, is a flexible generalization of the ordinary least square regression. The classical linear regression model

assumes that the dependent variable is continuous and normally distributed with constant variance, and that the dependent variable is a linear function of a set of regression parameters. The classical linear model takes the following form

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{2.1}$$

where

$\mathbf{Y}$ is an $n \times 1$ column vector containing the continuous dependent variable,

$\mathbf{X}$'s are the independent variables and can be categorical, continuous, or a combination of both,

and $\boldsymbol{\epsilon}$ is assumed to be a normal distributed error vector

The dependent variable $\mathbf{Y}$ is decomposed into two components, including a systematic component $\boldsymbol{X\beta}$ and an error component $\boldsymbol{\epsilon}$. The systematic component is the expected value of y, $E(y)$, for the given set of values for the $\mathbf{X}$'s. That is, $E(y)$ is a conditional mean vector that depends on the values of the covariates $\mathbf{X}$'s and may be written as

$$\boldsymbol{E(Y)} = \boldsymbol{\mu} = \boldsymbol{X\beta} \tag{2.2}$$

In many applications, the dependent variable is categorical variable, or a count variable, or a continuous but nonnormal variable. The assumptions of the classical linear regression model do not hold when the distribution of the dependent variable is nonnormal and its variance is a function of its mean, μ. The reason is that the least square estimates are not always equal to maximum likelihood estimates, as they are for the normal distribution (Dunteman and Ho, 2006).

Nelder and Wedderburn (1972) introduced generalized linear models to address these limitations. As a generalization of the classical linear regression model, generalized linear models extend analyses to predict the mean of variables that are not reasonably assumed to be from normal distribution, at the same time, retain all the power of the models on the right-side of equation

(2.2). One extension is that the mean of the response variable can come from any distribution in the exponential family. The other extension is that the link function can be any monotonic differentiable function, $g(\boldsymbol{\mu})$. Equation (2.2) can be rewritten as

$$\boldsymbol{g(\mu) = X\beta} \tag{2.3}$$

Here $g(\mu)$ is often called the link function. The link function can have various forms, including identity link, logit link, log link, reciprocal link, etc, which correspond to Normal, Poisson, Binormal, Gamma, respectively. For example, the logit link function is expressed as a linear function of the independent variables, $g(\boldsymbol{\mu}) = logit(\boldsymbol{\mu}) = \boldsymbol{X\beta}$. We will discuss more about the link function in Chapter 4.

McCullagh and Nelder (1989) provided a very good summary of generalized linear models. The explicit assumptions of generalized linear model consist of three parts, including random component, systematic component, and link function.

(1) Random component: Each component of Y is independently distributed with mean $E(\boldsymbol{Y}) = \boldsymbol{\mu}$ and variance $Var(\boldsymbol{Y}) \propto \sigma^2$.

(2) Systematic component: The p covariates, $\boldsymbol{x = (x_1, \ldots, x_p)}$, are combined to produce a linear predictor $\boldsymbol{\eta}$ by $\boldsymbol{\eta = X\beta}$.

(3) Link Function: The random component and the systematic component are related together via a link function, $\boldsymbol{\eta} = g(\boldsymbol{\mu})$. The link function of $\boldsymbol{\eta = \mu}$ is called the identity link.

2.5 Exponential Family

Generalized linear models are usually formulated within the framework of the exponential family. The exponential family of distributions offers much more flexibility in specifying the variance structure of the response variable.

From the practical perspective, this is a big step towards more realistic modeling of latent variables, while preserving the advantages of ordinary regression theory such as the ability to calculate standard errors of estimated parameters.

The exponential family is a class of probability distributions, including continuous or discrete types, which share a certain form of density functions.

$$\underline{General\ Form:} \qquad f(y_i;\theta) = h(y_i)exp\left(d(\theta)T(y_i) - A(\theta)\right), \qquad (2.4)$$

where $A(.)$, $d(.)$, $T(.)$, and $h(.)$ are all known functions that have the same form for all y_i, $i = 1, 2, \ldots, n$.

$$\underline{Canonical\ Form:} \qquad f(y_i|\eta,\phi) = exp\left(\left(\frac{y_i\eta - b(\eta)}{a(\phi)}\right) + c(y_i,\phi)\right) \qquad (2.5)$$

for some specific function $a(.)$, $b(.)$, and $c(.)$. This is an exponential family model with canonical parameter η. ϕ is the dispersion (scale) parameter. In equation (2.5), both y_i and η can be vector valued.

McCullagh and Nelder (1989) introduced that the general form of the exponential family can be rewritten in the canonical form, for which $\eta = d(\theta)$ and $y_i = T(y_i)$. The function $b(\eta)$ is commonly called the cumulant function of Y because one can generate the cumulants of Y by successively differentiating $b(\eta)$ with respect to η. In particular, the mean $\mu = E(Y)$ and the variance $\sigma^2 = Var(Y)$ obey the rule $\mu = b'(\eta) = \frac{\partial b(\eta)}{\partial \eta}$ and $\sigma^2 = b''(\eta)a(\phi) = (\frac{\partial^2 b(\eta)}{\partial \eta^2})a(\phi)$. The inverse function of $b' : \eta \to \mu$ is called the canonical link function of Y. Note that the canonical link function is a one to one function. In what follows, we will use $g = b'^{-1} : \mu \to \eta$ to denote the canonical link function. The canonical form can also be written in terms of the mean μ, rather than η. Let the transformation between μ and η be $\mu = g^{-1}(\eta)$.

The canonical form includes a dispersion function $a(\phi)$ instead of a simple constant ϕ. This apparent complication provides an important extra degree of flexibility to model the distributions of y_i whose mean and variance are not of the same values but have the same form. That is, y_i are independently, but not identically distributed. A scoring practitioner can use a dispersion function to help improve a model by setting $a(\phi) = \phi/w_i$, where ϕ is constant for

all observations and w_i is a weight variable that may vary by observations. A prior weight, w_i, is selected to make correction for unequal variances among the observations, which would otherwise violate the assumption that ϕ is constant. The value of w_i is usually 1, or can be selected using external information or empirical experiences. In this study, we set the value of w_i to be 1, and assume the constant variance.

As described in Section 2.3, the latent variable, y^*, is a continuous variable. Normal, Gamma, and Inverse Gaussian are three candidates for the assumptions of continuous distribution. Gamma and Inverse Gaussian share the property that both distributions have the scale parameter and shape parameter, and that they have one mode in the interior of the range of possible values. Both distributions are possibly skewed to the right, sometimes with an extremely long tail. The appropriately chosen parameter values can make Gamma and Inverse-Gaussian practically indistinguishable (Johnson, 2006). In the rest of the study, our discussion of the exponential family will focus on two continuous distributions, Normal and Gamma.

In the canonical form, the probability distribution of the normal variable, Y, appears as

$$f(y|\mu,\sigma^2) = exp\left(\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2}\left(\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right)\right) \tag{2.6}$$

The canonical link function is the identity link, that is, $\eta = \mu$, with $b(\eta) = \eta^2/2$, $a(\phi) = \phi = \sigma^2$, and $c(y,\phi) = -(y^2/\phi + \ln(2\pi\phi))/2$. The MLEs of μ and σ^2 have closed-form solution as follows,

$$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad and \quad \hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

The density function of a gamma random variable, denoted as $Y \sim \Gamma(\alpha,\beta)$, contains the shape parameter α and the scale parameter β. Expressing the Gamma probability function in terms of the canonical form

$$f(y|\mu,\phi) = exp\left(\frac{-y/\mu - \ln\mu}{\phi} + \frac{1}{\phi}\ln\left(\frac{y}{\phi}\right) - \ln\Gamma\left(\frac{1}{\phi}\right) - \ln y\right) \tag{2.7}$$

The canonical link is $\eta = g(\mu) = \frac{-1}{\mu}$, and $b(\eta) = -\ln(-\eta)$, $a(\phi) = \phi = \frac{1}{\alpha}$, $c(y,\phi) = \left(\frac{1}{\phi}\ln(\frac{y}{\phi}) - \ln\Gamma(\phi^{-1}) - \ln y\right)$. The MLEs of α and β are solutions to a set of non-linear equations, and do not have closed-form formulas. From the practical point of view, one might use the moment estimator $\hat{\alpha} = (\bar{Y}^2/s^2)$ and $\hat{\beta} = (s^2/\bar{Y})$, which are consistent estimators of α and β.

2.6 Numerical Algorithms

Since the Profit Function and the ROC curve have only one independent variable of interest, the one-dimensional problem is to find the root of the derivative function. Let $T(.)$ refer to such a function. Since the closed-form expression is not always available, we introduce two numerical algorithms, bisection method and Newton-Raphson method, to resolve the one-dimensional problem.

(1) Bisection method

The bisection method is an example of a bracketing method, which bounds a root within an iterative sequence of nested intervals of decreasing length. A root is *bracketed* in the interval (a,b) if $T(a)$ and $T(b)$ have opposite signs. The function, $T(.)$, is known to pass through zero at a point in the interval (a,b). At each iteration, the bisection method uses the midpoint of the interval to replace one boundary limit of the interval. Hence the interval gets halved. The algorithm of Bisection method is as follows.

Algorithm 1: Bisection method

The steps to apply the bisection method to find the root of the equation $T(\upsilon) = 0$ are

1. Choose υ_l and υ_u as two guesses for the root such that $T(\upsilon_l)T(\upsilon_u) < 0$, or in other words, $T(\upsilon)$ changes sign between υ_l and υ_u.

2. Estimate the root, υ_m of the equation $T(\upsilon) = 0$ as the midpoint between

υ_l and υ_u as $\upsilon_m = \frac{\upsilon_l + \upsilon_u}{2}$.

3. Now check the following

a) If $T(\upsilon_l)T(\upsilon_m) < 0$ then the root lies between υ_l and υ_m; so reset $\upsilon_l = \upsilon_l$ and $\upsilon_u = \upsilon_m$.

b) If $T(\upsilon_m)T(\upsilon_u) < 0$ then the root lies between υ_m and υ_u; so reset $\upsilon_l = \upsilon_m$ and $\upsilon_u = \upsilon_u$.

c) If $T(\upsilon_l)T(\upsilon_m) = 0$ then the root is υ_m. Stop the algorithm if this is true.

4. Find the new estimate of the root $\upsilon_m^{new} = \frac{\upsilon_l + \upsilon_u}{2}$ and calculates the absolute convergence criterion or the relative convergence criterion

$$\mid \varepsilon \mid = \mid \upsilon_m^{new} - \upsilon_m^{old} \mid \quad or \quad \mid \varepsilon \mid = \mid \frac{\upsilon_m^{new} - \upsilon_m^{old}}{\upsilon_m^{new}} \mid$$

where υ_m^{new} = estimated root from present iteration

υ_m^{old} = estimated root from previous iteration

5. Compare the absolute error $\mid \varepsilon \mid$ with the pre-specified relative error tolerance. If $\mid \varepsilon \mid >$ pre-specified tolerance, then return to Step 3, else stop the algorithm. Also, check whether the number of iterations has exceeded the maximum number of iterations.

6. $\upsilon^* = \upsilon_m$; Return υ^*.

Advantage:

a) The bisection method is guaranteed to converge to a point in the interval (a, b) since the method is based on reducing the gap between two guesses. If the interval contains more than one root, bisection will find one of roots.

b) The number of iterations required to achieve a given tolerance level in the solution is predictable.

$$n = \log_2 \left(\frac{\mid \upsilon_u - \upsilon_l \mid}{\varepsilon_0} \right)$$

where ε_0= the tolerance level in the absolute convergence criterion

Disadvantage:

a) The convergence of bisection method is slow as it is based on halving the interval at each iteration.

b) If a function $T(\upsilon)$, for example $T(\upsilon) = \upsilon^2$, just touches the x-axis, it is not able to find the lower guess, υ_l, and the upper guess, υ_u.
c) If $T(\upsilon)$ is a singularity function, for example $T(\upsilon) = \upsilon^{-1}$, straddles a singularity and contains no root, it will converge to the singularity.

(2) Newton-Raphson method

The Newton-Raphson method is an open method, that is, the root is not bracketed. One initial guess of the root is needed to start the iterative process to find the root of an equation, $T(\upsilon) = 0$. The Newton-Raphson method is based on the following principle: if the tangent line at the current point, υ_i, is extended to cross zero, then the point υ_{i+1} where the tangent line crosses the abscissa is an improved estimate of the root.

Algebraically, the Newton-Raphson method is derived from the familiar Taylor series expansion. For a general function, $T(\upsilon)$, the Taylor series is

$$T(\upsilon_{i+1}) = T(\upsilon_i) + T'(\upsilon_i)(\upsilon_{i+1} - \upsilon_i) + \frac{1}{2!}T''(\upsilon_i)(\upsilon_{i+1} - \upsilon_i)^2 + \dots$$

For small difference between υ_i and υ_{i+1}, and for well-behaved functions, the terms beyond linear are not important. As an approximation, take only the first two terms on the right hand side,

$$T(\upsilon_{i+1}) = T(\upsilon_i) + T'(\upsilon_i)(\upsilon_{i+1} - \upsilon_i)$$

Hence, $T(\upsilon_{i+1}) = 0$ implies

$$\upsilon_{i+1} = \upsilon_i - \frac{T(\upsilon_i)}{T'(\upsilon_i)} \tag{2.8}$$

Equation (2.8) is critical for the algorithm of the Newton-Raphson method. The detailed algorithm is shown below.

Algorithm 2: Newton-Raphson method

The steps to apply the Newton-Raphson method to find the root of the equation $T(\upsilon) = 0$ are

1. Evaluate $T'(\upsilon)$ symbolically

2. Use an initial guess of the root, υ_i, to estimate the new value of the root υ_{i+1} as

$$\upsilon_{i+1} = \upsilon_i - \frac{T(\upsilon_i)}{T'(\upsilon_i)}$$

3. Calculate the absolute convergence criterion or relative convergence criterion

$$\mid \varepsilon \mid = \mid \upsilon_{i+1} - \upsilon_i \mid \quad or \quad \mid \varepsilon \mid = \mid \frac{\upsilon_{i+1} - \upsilon_i}{\upsilon_{i+1}} \mid$$

4. Compare the absolute error $\mid \varepsilon \mid$ with the pre-specified relative error tolerance. If $\mid \varepsilon \mid >$ pre-specified tolerance, then repeat Steps $2 \sim 4$, else stop the algorithm. Also, check whether the number of iterations has exceeded the maximum number of iterations.

5. $\upsilon^* = \upsilon_{i+1}$; Return υ^*.

Advantage:

a) The Newton-Raphson method is an extremely fast root-finding approach.

b) The method is not restricted to one-dimension and can be easily generalized to multiple dimensions.

Disadvantage:

a) The Newton-Raphson method may diverge at inflection points.

b) If $T'(\upsilon_i)$ is zero or near zero, it will give a large magnitude for the next value, υ_{i+1}.

2.7 Chapter Summary

In this dissertation, the Profit Function and the ROC curve will be applied to select the optimal cut-off point and its near-optimal interval to maximize the expected profit of a customer in the credit re-assessment process. This chapter gives a literature review of several topics in the field, the ROC curve, the bi-distribution-based Profit Function, generalized linear models, exponential family, and the numerical algorithms of finding the solution of an equation. Numerical algorithms are usually needed in the estimation of the

critical thresholds in the non-Normal-Normal models. All topics discussed in this chapter will be illustrated in the numerical study in Chapter 6.

CHAPTER 3

Optimal Cut-off and Near-optimal Interval in the Normal-Normal Model

3.1 Introduction

This chapter will develop some schemes to evaluate the Profit Function under the widely-adopted assumption of normality. The main outputs of this chapter include the estimation of the optimal threshold and the near-optimal interval as well as some related theorems.

The rest of the chapter describes a number of new mathematical and statistical results for the study of optimal thresholds in the Normal-Normal Profit Function. Section 3.2 introduces the selection of optimal thresholds in three types of Profit Functions, concave, decreasing, and increasing. This section also suggests some theorems to identify the approximate location of the optimal threshold, without developing a statistical model. Section 3.3 discusses the variance and asymptotic distribution of the non-constrained optimal threshold. The final section 3.4 gives a summary of the chapter and an introduction to the next chapter.

3.2 Selection of Optimal Cut-off and Near-optimal Interval

The Normal-Normal model assumes two distributions for y^*, $N_1(\mu_G, \sigma_G^2)$ and $N_2(\mu_B, \sigma_B^2)$, which represent distributions of good and bad population, respectively. It is reasonable to assume the expected creditworthiness of the good population is better than that of the bad population. Therefore, μ_G can be reasonably assumed to be greater than μ_B.

The relation between σ_G and σ_B is more complex and has generated many discussions in the literature. Egan (1975), in a textbook about signal detection theory, assumed a bi-normal model in which $\sigma_G = \sigma_B$. Green and Swets (1966) pointed out that unless the ratio of σ_G/σ_B is considerably different from unity, the fact that $\sigma_G \neq \sigma_B$ has a very limited impact on the Normal-Normal model. Irwin, Hautus, and Stillman (1992), and Hautus and Irwin (1995) demonstrated that the equal-variance Normal-Normal model usually provides a good fit to the empirical studies of the discriminiability of food and beverages. Pepe (2003) proposed, since $\sigma_G \neq \sigma_B$ usually occurs over only a very small part of diagnostic studies, most real studies on diagnostic tests can assume the equal-variance Normal-Normal model. Blume (2008) assumed equal variance when estimating the Area Under the ROC Curve (AUC). Faraggi and Reiser (2002) compared simulated samples with normal populations of equal and unequal variances and concluded that the estimation results of AUC differ very little. Zou and Hall (2000) applied several estimation methods to estimate the ratio of σ_G/σ_B in the bi-normal model. The results, using the semi-parametric method, the parametric transformation method, and Metz's LABROC4 method, are 0.993, 0.999, and 1.048, all of which are close to 1. Cai and Moskowitz (2004) compared five estimation methods, including MLE, PMLE, Zou & Hall's method, GLM, and LABROC. Results show that in the diagnostic tests, the estimated ratios of σ_G/σ_B in the bi-normal model are 1.065, 1.020, 1.007, 1.017, and 1.002, again all of which are close to 1.

All these previous studies demonstrate that we can further assume that the standard deviations, σ_G and σ_B, are 1 in the bi-normal model. Credit re-assessment is always based on large population. Therefore, in this paper, we make the following assumption in the Normal-Normal model: $(1)\mu_G > \mu_B$ and $(2)\sigma_G = \sigma_B = 1$.

After all these assumptions, we assume two distributions, $f_G(y^* \mid D = G, \mu_G, k)$ and $f_B(y^* \mid D = B, \mu_B, k)$, which represent creditworthiness of applicants with good credit backgrounds and bad credit backgrounds. Figure 3.1 is an example about the distribution of y^*, along with a selection of a cut-off, k, for good population and bad population. The shaded area in Figure 3.1.a demonstrates FN_k in the good population for the cut-off point k, when y^* fall into $[y^*_{min}, k]$. The shaded area in Figure 3.1.b demonstrates FP_k in the bad population for the cut-off point k, when y^* fall into $(k, y^*_{max}]$. The values of y^*_{min} and y^*_{max} are the lower bound and upper bound of y^*, respectively. Given changing economic environment, lending institution usually adjusts the threshold towards either y^*_{min} or y^*_{max}.

3.2.1 Selection of the Optimal Cut-off

There are many possible ways to find the optimal cut-off point, k^*. Noe (1983) proposed to obtain the root of the first derivative of $R(k)$ over TN_k, i.e. $\frac{\partial R(k)}{\partial TN_k}$. Stein (2005) showed that the slope of the ROC curve at the optimal cut-off k^* satisfies $\frac{\partial TP_k}{\partial FP_k}\Big|_{k=k^*} = C = \frac{P_B(r_B+c_B)}{P_G(r_G+c_G)}$. He suggested using this equation to find k^*, but did not provide a solution. Actually, the standard way to find the optimal cut-off k^* is to take the first derivative of $R(K)$ with respect to k, set it to zero, and solve the equation. Since each k uniquely determines a unique TP_k, TN_k, $R(k)$, and $cost(k)$, it is easy to see that, using the chain-rule, we can find k^* by setting any of $\frac{\partial R(k)}{\partial k}$, $\frac{\partial R(k)}{\partial FP_k}$, $\frac{\partial R(k)}{\partial TN_k}$, $\frac{\partial R(k)}{\partial cost(k)}$, etc., to zero.

The above method is valid only when the maximization of $R(k)$ is subject to no constraints on k. In practice, however, the optimal threshold must be subject to the constraints, $k \in [k_1 = y^*_{min}, k_2 = y^*_{max}]$. Our main results concerning the constrained maximization problem, i.e., to find k^* to maximize

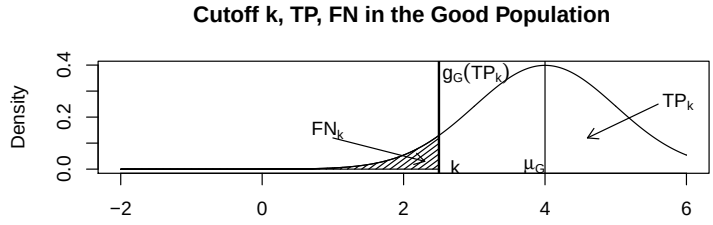

Figure 3.1.a

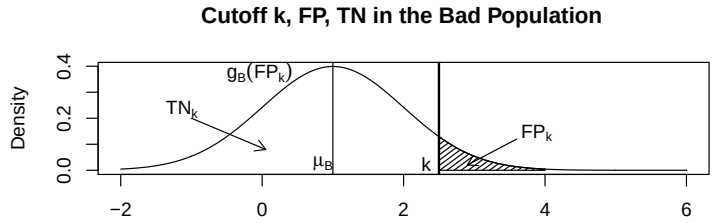

Figure 3.1.b

Figure 3.1: $f(y^*|D = G, \mu_G, k)$ and $f(y^*|D = B, \mu_B, k)$ with $\mu_G > \mu_B$, $\sigma_G = \sigma_B = 1$

$R(k)$ subject to $k \in [k_1 = y^*_{min}, k_2 = y^*_{max}]$, are summarized in the following three theorems. In the following theorems, let k_0 be a critical point satisfying $R'(k_0) = 0$. Under the assumption of Normal-Normal model, k_0 can be shown to be unique and have a closed-form expression determined by P_G, P_B, c_G, c_B, r_G, r_B, μ_G, and μ_B (See Appendix A.1):

$$k_0 = \frac{\ln C}{\mu_G - \mu_B} + \frac{\mu_G + \mu_B}{2} \tag{3.1}$$

$$where \quad \ln C = \ln\left(\frac{P_B(r_B + c_B)}{P_G(r_G + c_G)}\right) \tag{3.2}$$

Theorem 3.1 *Assuming a Normal-Normal model with $N(\mu_G, 1)$ and $N(\mu_B, 1)$ with $\mu_G > \mu_B$. Let $C = \frac{P_B(r_B+c_B)}{P_G(r_G+c_G)}$ and k_0 be defined as in equation (3.1). Then*
(i) $R'(k) \geq 0$ when $k \leq k_0$
(ii) $R'(k) \leq 0$ when $k \geq k_0$
(iii) $R''(k_0) < 0$

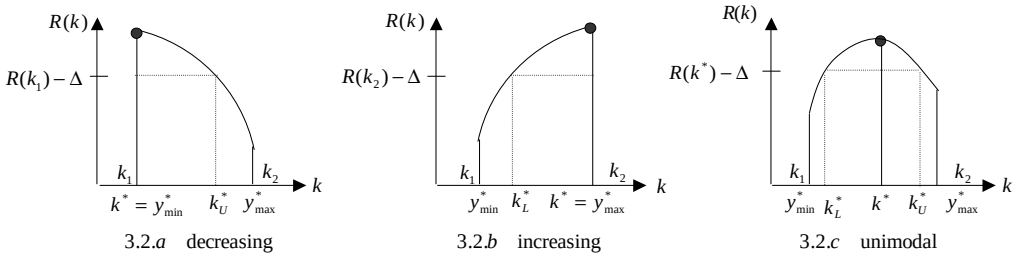

Figure 3.2: Three Types of Profit Functions

Proof is in the Appendix A.1.

Theorem 3.2 below follows immediately from Theorem 3.1.

Theorem 3.2 *Let $R(k)$ be defined in the range $k \in [k_1, k_2]$, and k_0 be defined as in equation (3.1). The maximization of $R(k)$ can be separated in three cases.*
(i) If $k_0 \leq k_1$, then $R(k)$ is maximal when $k^* = k_1$ with $R(k^*) = (P_G r_G - P_B c_B)$.
(ii) If $k_0 \geq k_2$, then $R(k)$ is maximal when $k^* = k_2$ with $R(k^*) = (P_B r_B - P_G c_G)$.
(iii) If $k_1 < k_0 < k_2$, then $R(k)$ is maximal when $k^* = k_0$ with
$R(k^*) = P_B(r_B+c_B)\Phi(\frac{\ln C}{\mu_G-\mu_B}+\frac{\mu_G-\mu_B}{2}) - P_G(r_G+c_G)\Phi(\frac{\ln C}{\mu_G-\mu_B}-\frac{\mu_G-\mu_B}{2}) + R(k_1)$

Figure 3.2 illustrates three cases of Profit Functions in Theorem 3.2. Figure 3.2.a illustrates case (i), in which TP_k and FP_k are equal to one, and all credit applications get approved. Figure 3.2.b illustrates case (ii), in which all credit applications are rejected. Figure 3.2.c illustrates the most popular case, case (iii), where $max(R(k))$ exists somewhere in the interval of (y^*_{min}, y^*_{max}).

Relationship between k^* and $[\mu_B, \mu_G]$ can give the lending company some indication how the distribution of customer portfolios impacts net income. Again there are three possibilities for this relationship: k^* can be less than μ_B, be within the close interval $[\mu_B, \mu_G]$, or be greater than μ_G. These three possibilities are unrelated to three cases in Theorem 3.2; rather, as indicated

in the following Theorem 3.3, they are determined by the relationship between two quantities, $\frac{(\mu_G-\mu_B)^2}{2}$ and $C = \frac{P_B(r_B+c_B)}{P_G(r_G+c_G)}$.

Theorem 3.3 *Let k^* be the optimal cut-off discussed in Theorem 3.2. Then*
(i) If $\ln C \leq -\frac{(\mu_G-\mu_B)^2}{2}$, then $k^* \leq \mu_B$.
(ii) If $\ln C \geq \frac{(\mu_G-\mu_B)^2}{2}$, then $k^* \geq \mu_G$.
(iii) If $-\frac{(\mu_G-\mu_B)^2}{2} < \ln C < \frac{(\mu_G-\mu_B)^2}{2}$, then k^* is located within the interval of (μ_B, μ_G).

Proof is in Appendix A.2.

3.2.2 Selection of the Near-optimal Interval

The previous section discusses the determination of optimal cut-off point, which is critical to the classification of credit scoring results. This section discusses near-optimal interval, which provides the practitioners more flexibility in the lending practices. The concept of near-optimal interval, $[k_L^*, k_U^*]$, is described in Section 1.2. As shown in Figure 3.2, the two end points, k_L^* and k_U^*, are the solutions to equation (3.3).

$$R(k^*) - R(k) = \psi|R(k^*)| \tag{3.3}$$

We investigate the near-optimal interval of $R(k^*)$, corresponding to three cases in Theorem 3.2, concerning the shapes of $R(k)$, whether it is decreasing, increasing, or concave.

If the Profit Function is a decreasing function as in Figure 3.2.*a*, then k_L^* is located at the left bound, i.e., $k_L^* = k_1$. k_U^* is the solution of k in equation (3.3). Provided that a linear approximation is adequate for the Profit Function, k_U^* can be estimated from equation (3.4)

$$k_U^* \approx k_1 - \frac{\psi|R(k_1)|}{R'(k_1)} \tag{3.4}$$

Similarly, if the Profit Function is an increasing function as in Figure 3.2.*b*, k_U^* is known to locate at the right bound, i.e., $k_U^* = k_2$. Provided that a linear

approximation is adequate for the Profit Function, k_L^* can be estimated from equation (3.5).

$$k_L^* \approx k_2 - \frac{\psi|R(k_2)|}{R'(k_2)} \tag{3.5}$$

If the Profit Function is concave-down as shown in Figure 3.2.*c*, k_L^* and k_U^* can be estimated from equation (3.6). Note that $R''(k^*)$ is always negative under Theorem 3.1.

$$k \approx k^* \pm \sqrt{\frac{-2\psi|R(k^*)|}{R''(k^*)}} \tag{3.6}$$

with $k_L^* = k^* - \sqrt{\frac{-2\psi|R(k^*)|}{R''(k^*)}}$ and $k_U^* = k^* + \sqrt{\frac{-2\psi|R(k^*)|}{R''(k^*)}}$.

When ψ is equal to 0, equations (3.4), (3.5), and (3.6) are simplified as equation (3.1), as it should. Proofs of deriving these equations are in Appendix A.3.

3.3 Asymptotic Distribution of Non-constrained Optimal Cut-off

The previous sections have demonstrated how to derive the optimal threshold and near-optimal thresholds under the Normal-Normal assumption. Practical experiences suggest that the characteristics of the population gradually shift over time. A practitioner often wants to know the consistency of estimates of critical thresholds. This section provides some study on the variance of optimal thresholds and their asymptotic distributions. We focus on the case where the optimal threshold k^* is inside the interval of (y_{min}^*, y_{max}^*).

The asymptotic distribution of $\hat{k}^*$ can be derived using the Delta method and Taylor Series. Starting from equation (3.1), note that $\bar{y}_G^*$ and $\bar{y}_B^*$ are estimates of μ_G and μ_B, the estimator of k^* is written as

$$\hat{k}^* = h(\bar{y}_G^*, \bar{y}_B^*) = \frac{\ln C}{\bar{y}_G^* - \bar{y}_B^*} + \frac{\bar{y}_G^* + \bar{y}_B^*}{2} \tag{3.7}$$

where y_G^*'s and y_B^*'s are from two independent groups with sample sizes n_G and n_B

$$\bar{y}_G^* \sim N(\mu_G, \frac{1}{n_G}) \quad and \quad \bar{y}_B^* \sim N(\mu_B, \frac{1}{n_B})$$

A first-order two-variable Taylor series expansion of $\hat{k}^*$ about μ_G and μ_B gives the approximation.

$$\begin{aligned} h(\bar{y}_G^*, \bar{y}_B^*) &\approx h(\mu_G, \mu_B) + (\bar{y}_G^* - \mu_G) \left. \frac{\partial h(\bar{y}_G^*, \bar{y}_B^*)}{\partial \bar{y}_G^*} \right|_{\mu_G, \mu_B} + (\bar{y}_B^* - \mu_B) \left. \frac{\partial h(\bar{y}_G^*, \bar{y}_B^*)}{\partial \bar{y}_B^*} \right|_{\mu_G, \mu_B} \\ &= h(\mu_G, \mu_B) + (\bar{y}_G^* - \mu_G) \left[\frac{1}{2} - \frac{\ln C}{(\mu_G - \mu_B)^2} \right] + (\bar{y}_B^* - \mu_B) \left[\frac{1}{2} + \frac{\ln C}{(\mu_G - \mu_B)^2} \right] \end{aligned}$$

It follows that the variance of $\hat{k}^* = h(\bar{y}_G^*, \bar{y}_B^*)$ can be written as

$$var(\hat{k}^*) \approx \left[\frac{1}{2} - \frac{\ln C}{(\mu_G - \mu_B)^2} \right]^2 var(\bar{y}_G^*) + \left[\frac{1}{2} + \frac{\ln C}{(\mu_G - \mu_B)^2} \right]^2 var(\bar{y}_B^*)$$

The approximate variance of the optimal threshold is

$$var(\hat{k}^*) \approx \left[\frac{1}{2} - \frac{\ln C}{(\mu_G - \mu_B)^2} \right]^2 \frac{1}{n_G} + \left[\frac{1}{2} + \frac{\ln C}{(\mu_G - \mu_B)^2} \right]^2 \frac{1}{n_B} \tag{3.8}$$

Therefore, the asymptotic distribution of $\hat{k}^*$ is

$$\sqrt{n} \left[h(\bar{y}_G^*, \bar{y}_B^*) - h(\mu_G, \mu_B) \right] \quad approximately \ \ follows \ \ N(0, \sigma_0^2) \tag{3.9}$$

where $\frac{\sigma_0^2}{n} = var(\hat{k}^*) = \left[\frac{1}{2} - \frac{\ln C}{(\mu_G - \mu_B)^2} \right]^2 \frac{1}{n_G} + \left[\frac{1}{2} + \frac{\ln C}{(\mu_G - \mu_B)^2} \right]^2 \frac{1}{n_B}$

Proof is straightforward and presented in Appendix A.4.

The variance of $\hat{k}^*$ is jointly determined by P_G, P_B, c_G, c_B, r_G, r_B, μ_G, μ_B, and n. Under the normal assumption, as $n \to \infty$, variance of optimal threshold is close to zero. The estimate of $\hat{k}^*$ is asymptotically consistent, as long as P_G, P_B, c_G, c_B, r_G, and r_B are fixed.

3.4 Chapter Summary

Lending institutions often need a specific rule for defining an optimal cut-off value to accept or reject loan applications. The credit will be granted if the

score is above the specific optimal cut-off. Actually, a near-optimal interval offers more flexibility to lending institutions in adoption of various marketing campaigns and risk strategies. We studied optimal thresholds in three shapes of the Profit Function. The large sample size in practice assures that the estimate of k^* is asymptotically consistent.

The next chapter will extend the selection of critical thresholds to the exponential family. Two models, the heteroscedastic Normal-Normal model and the Gamma-Gamma model, will be discussed.

CHAPTER 4

OPTIMAL CUT-OFF AND NEAR-OPTIMAL INTERVAL IN THE EXPONENTIAL FAMILY

4.1 Introduction

The main results in Chapter 3 were derived using a homoscedastic Normal-Normal model for the latent credit score (i.e. $Y_G \sim N(\mu_G, \sigma_G^2)$ and $Y_B \sim N(\mu_B, \sigma_B^2)$ with $\sigma_G^2 = \sigma_B^2$). In this chapter, we consider situations when the actual data of the credit scores do not completely meet the assumptions of the equal-variance Normal-Normal model, and develop solutions for some alternative models, specifically, a heteroscedastic Normal-Normal model, and a Gamma-Gamma model based on the gamma distribution (which is skewed).

Section 4.2 discusses the selection of optimal cut-off points, in case that Y_G and Y_B are from two distributions in the exponential family (called an E_G-E_B model, and an E-E model when Y_G and Y_B are from the same distribution with a common dispersion parameter). In Section 4.3, we extend Theorems 3.1 and 3.2 to a heteroscedastic Normal-Normal model. The extension to a

Gamma-Gamma model is discussed in Section 4.4. Section 4.5 discusses the selection of near-optimal cut-off points. The last section (4.6) will summarize the chapter and make a brief introduction to Chapter 5.

4.2 Selection of Optimal Cut-off in the Exponential Family

The first derivative of the payoff function in equation (1.2) can be written as follows.

$$R'(k) = -\varsigma f_G(k; \eta_G, \phi_G) + \gamma f_B(k; \eta_B, \phi_B) \tag{4.1}$$

where $\varsigma = P_G r_G + P_G c_G$, $\gamma = P_B r_B + P_B c_B$

We use the notation E_G-E_B to indicate that the distributions of Y_G and Y_B are both from the exponential family (2.5) but two distributions need not be the same, so we will use the notations a_G, a_B, η_G, η_B, ϕ_G, ϕ_B, etc. As shown in Appendix A.5, the sign of $R'(k)$ in this situation matches with that of the following

$$d(k) = \ln C - k\left(\frac{\eta_G}{a_G(\phi_G)} - \frac{\eta_B}{a_B(\phi_B)}\right) + \left(\frac{b_G(\eta_G)}{a_G(\phi_G)} - \frac{b_B(\eta_B)}{a_B(\phi_B)} - c_G(k, \phi_G) + c_B(k, \phi_B)\right) \tag{4.2}$$

The solution to the equation $d(k) = 0$ is denoted as k_0. Since we may not have an explicit solution, some numerical approximation is usually needed to find k_0. As a special case, Y_G and Y_B can be distributed from the same distribution in the exponential family. We will denote a special E_G-E_B model as E-E when (1) f_G and f_B belong to the same distribution in the exponential family, and (2) The dispersion parameters are identical. That is, $a_G(\phi_G) = a_B(\phi_B) = a(\phi)$. In an E-E model, the functions b_G and b_B are the same, denoted by $b_G = b_B = b$, and $c_G(k, \phi_G) = c_B(k, \phi_B)$. Consequently, from (4.2), both Theorems 3.1 and 3.2 still hold, except that the formula for the critical point k_0 is replaced by

$$k_0 = (\eta_G - \eta_B)^{-1}\left[a(\phi)\ln C + b(\eta_G) - b(\eta_B)\right] \tag{4.3}$$

In the Normal-Normal model, the dispersion parameter is equivalent to the variance, thus an E_G-E_B model is often known as the heteroscedastic Normal-Normal model. In a particular case of Normal-Normal model with equal dispersion parameters, an E-E model is called the homoscedastic Normal-Normal model, and (4.3) simplifies to equation (3.1). Derivations of the formula (4.2) and (4.3) are given in Appendix A.5.

4.3 Selection of Optimal Cut-off in the Heteroscedastic Normal-Normal Model

Earlier in Chapter 2 we assumed $\mu_G > \mu_B$ and $\sigma_G^2 = \sigma_B^2$, then, $d(k)$ in (4.2) is a linear function of k with a negative slope, which leads to the main results in the previous chapter: equation (3.1), Theorem 3.1 and Theorem 3.2. This section extends the results in Chapter 3 to a heteroscedastic Normal-Normal model. We may assume either $\sigma_G^2 > \sigma_B^2$ or $\sigma_G^2 < \sigma_B^2$.

For the heteroscedastic Normal-Normal model, (4.1) is rewritten as

$$R'(k) = -\varsigma exp\left[\frac{k\eta_G - b(\eta_G)}{a(\phi_G)} + c(k, \phi_G)\right] + \gamma exp\left[\frac{k\eta_B - b(\eta_B)}{a(\phi_B)} + c(k, \phi_B)\right] \tag{4.4}$$

Again, the sign of $R'(k)$ matches with that of

$$\begin{aligned} d(k) &= \ln C - k\left(\tfrac{\mu_G}{\sigma_G^2} - \tfrac{\mu_B}{\sigma_B^2}\right) \\ &+ \left[\tfrac{\mu_G^2/2}{\sigma_G^2} - \tfrac{\mu_B^2/2}{\sigma_B^2} + \tfrac{1}{2}\left(\tfrac{k^2}{\sigma_G^2} + \ln(2\pi\sigma_G^2)\right) - \tfrac{1}{2}\left(\tfrac{k^2}{\sigma_B^2} + \ln(2\pi\sigma_B^2)\right)\right] \end{aligned}$$

Set $d(k)$ to zero and simplify the equation, we obtain the estimation equation for k_0 in the heteroscedastic Normal-Normal model, as follows.

$$\frac{k^2}{2}\left(\frac{1}{\sigma_G^2} - \frac{1}{\sigma_B^2}\right) - k\left(\frac{\mu_G}{\sigma_G^2} - \frac{\mu_B}{\sigma_B^2}\right) + \left[\ln C + \ln\left(\frac{\sigma_G}{\sigma_B}\right) + \frac{1}{2}\left(\frac{\mu_G^2}{\sigma_G^2} - \frac{\mu_B^2}{\sigma_B^2}\right)\right] = 0 \tag{4.5}$$

Let $D_0 = \left(\frac{\mu_G}{\sigma_G^2} - \frac{\mu_B}{\sigma_B^2}\right)^2 - 4\frac{1}{2}\left(\frac{1}{\sigma_G^2} - \frac{1}{\sigma_B^2}\right)\left[\ln C + \ln\left(\frac{\sigma_G}{\sigma_B}\right) + \frac{1}{2}(\frac{\mu_G^2}{\sigma_G^2} - \frac{\mu_B^2}{\sigma_B^2})\right]$ be the discriminant of the quadratic equation (4.5). In the homoscedastic Normal-Normal model, the quadratic equation (4.5) is simplified to a linear function of k, which leads to the unique solution of k_0 in equation (3.1). In the heteroscedastic Normal-Normal model, the discriminant D_0 can be zero, negative, or positive, which will be discussed in the following.

If the discriminant D_0 is zero, equation (4.5) has one real solution. That is

$$k_0 = \frac{\mu_G\sigma_B^2 - \mu_B\sigma_G^2}{\sigma_B^2 - \sigma_G^2} \tag{4.6}$$

If the discriminant D_0 is negative, then equation (4.5) does not have any real solutions, and the parabola $y = R'(x)$ is entirely on one side of the x-axis, either above the x-axis (when $\sigma_G^2 < \sigma_B^2$) or below the x-axis (when $\sigma_G^2 > \sigma_B^2$). If $\sigma_G^2 < \sigma_B^2$, then $R'(k) > 0$, the function $R(k)$ is strictly increasing, thus k^* is equal to k_2. If $\sigma_G^2 > \sigma_B^2$, then $R'(k) < 0$, the function $R(k)$ is strictly decreasing, thus k^* is equal to k_1. The conditions under which the parabola of $R'(k)$ is entirely on one side of the x-axis include

$$R'(k) > 0 \;\; or \;\; R'(k) < 0 \tag{4.7}$$

If the discriminant D_0 is positive, two distinct real solutions can be derived from equation (4.5). That is

$$k = \frac{(\frac{\mu_G}{\sigma_G^2} - \frac{\mu_B}{\sigma_B^2}) \pm \sqrt{D_0}}{\frac{1}{\sigma_G^2} - \frac{1}{\sigma_B^2}} \tag{4.8}$$

Let the two expressions for k in (4.8) be $k^{(1)}$ and $k^{(2)}$ with $k^{(1)} < k^{(2)}$. If $\sigma_G^2 < \sigma_B^2$, k_0 is equal to $k^{(1)}$, and $R'(k)$ is a parabola opening up. Since the function $R'(k)$ goes from positive to negative around $k^{(1)}$, thus $R(k^{(1)})$ is a local maximum. One needs to compare the values of $R(k)$ at $k^{(1)}$ and two end points (k_1 and k_2) to determine the global maximum. See Figure 4.1.*a*. If $\sigma_G^2 > \sigma_B^2$, then k_0 is equal to $k^{(2)}$, and $R'(k)$ is a parabola opening down. Since the format of $R'(k)$ goes from positive to negative around $k^{(2)}$, thus $R(k^{(2)})$ is

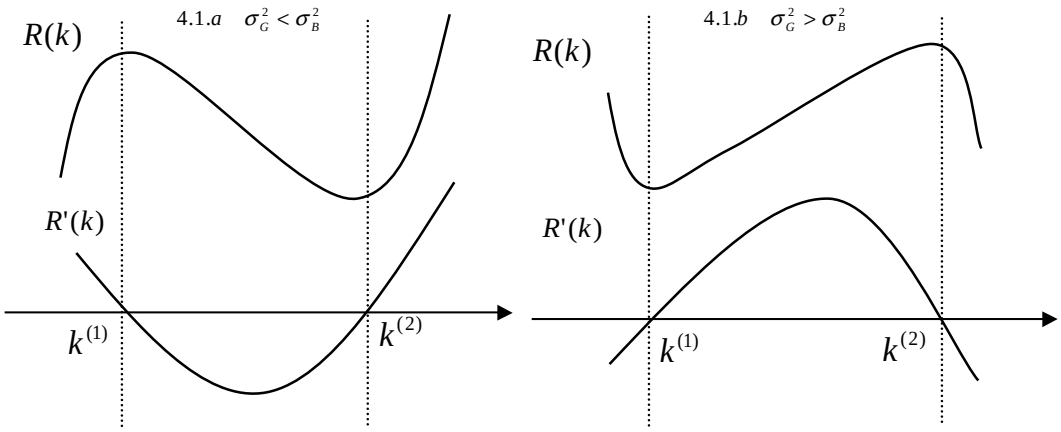

Figure 4.1: Two Types of Profit Function and its First Derivative

a local maximum. Similarly, one needs to compare $R(k^{(2)})$, $R(k_1)$, and $R(k_2)$ to determine the global maximum. See Figure 4.1.*b*.

Our result of the constrained maximization problem is to find k^* to maximize $R(k)$ subject to $k \in [k_1 = y^*_{min}, k_2 = y^*_{max}]$ as summarized in the following theorems.

Theorem 4.1 *Assuming the distribution model of Y_G and Y_B is Normal-Normal, that is, $N(\mu_G, \sigma_G^2)$ and $N(\mu_B, \sigma_B^2)$ with $\mu_G > \mu_B$. Let $C = \frac{P_B(r_B+c_B)}{P_G(r_G+c_G)}$ and $k^{(1)}$ and $k^{(2)}$ be defined as in (4.8) with $k^{(1)} < k^{(2)}$. Then*
when $\sigma_G^2 < \sigma_B^2$,
(i) $R'(k) \geq 0$ when $k \leq k^{(1)}$ or $k \geq k^{(2)}$
(ii) $R'(k) \leq 0$ when $k^{(1)} \leq k \leq k^{(2)}$
(iii) $R''(k^{(1)}) < 0$
when $\sigma_G^2 > \sigma_B^2$,
(iv) $R'(k) \geq 0$ when $k^{(1)} \leq k \leq k^{(2)}$
(v) $R'(k) \leq 0$ when $k \leq k^{(1)}$ or $k \geq k^{(2)}$
(vi) $R''(k^{(2)}) < 0$

Proof is in Appendix A.6.
Theorem 4.2 below follows immediately from Theorem 4.1.

Theorem 4.2 *Let $R(k)$ be a function in the range $k \in [k_1, k_2]$, and let $k^{(1)}$ and $k^{(2)}$ be defined as in (4.8) with $k^{(1)} < k^{(2)}$. The maximization of $R(k)$ can be derived in four cases.*

(i) If $k^{(1)} \leq k_1$, then $R(k)$ is maximal when $k^* = k_1$ with

$$R(k^*) = (P_G r_G - P_B c_B)$$

(ii) If $k^{(2)} \geq k_2$, then $R(k)$ is maximal when $k^* = k_2$ with

$$R(k^*) = (P_B r_B - P_G c_G)$$

(iii) If $k_1 < k^{(1)} < k^{(2)} < k_2$ and $\sigma_G^2 < \sigma_B^2$, then the local maximum of $R(k)$ is located at $k^* = k^{(1)}$ with

$$R(k^*) = P_B(r_B + c_B)\Phi(\tfrac{k^{(1)}-\mu_B}{\sigma_B}) - P_G(r_G + c_G)\Phi(\tfrac{k^{(1)}-\mu_G}{\sigma_G}) + R(k_1)$$

(iv) If $k_1 < k^{(1)} < k^{(2)} < k_2$ and $\sigma_G^2 > \sigma_B^2$, then the local maximum of $R(k)$ is located at $k^* = k^{(2)}$ with

$$R(k^*) = P_B(r_B + c_B)\Phi(\tfrac{k^{(2)}-\mu_B}{\sigma_B}) - P_G(r_G + c_G)\Phi(\tfrac{k^{(2)}-\mu_G}{\sigma_G}) + R(k_1)$$

4.4 Selection of Optimal Cut-off in the Gamma-Gamma Model

A Gamma-Gamma model with equal dispersion parameters makes the following assumptions, including (1) f_G and f_B are the Gamma density functions; and (2) the two dispersion parameters are identical, i.e. $\phi_G = \phi_B = \phi$. The second assumption is equivalent to $\alpha_G = \alpha_B = \alpha$. The assumption of $\mu_G > \mu_B$ in Chapter 3 also holds true in the Gamma-Gamma model, which leads to $\beta_G > \beta_B$. With some algebra, it can be derived that under the assumption of Gamma-Gamma model with equal dispersion parameters, the formula of k_0 in equation (4.3) is rewritten as follows.

$$k_0 = \left(\frac{\mu_G \mu_B}{\mu_G - \mu_B}\right)\left[\phi \ln C + \ln\left(\frac{\mu_G}{\mu_B}\right)\right] \tag{4.9}$$

where $\mu_G = \alpha\beta_G$, $\mu_B = \alpha\beta_B$, $\phi = 1/\alpha$.

Similarly, the following theorems are derived from our results of the constrained maximization problem in the Gamma-Gamma model with equal dis-

persion parameters.

Theorem 4.3 *Assuming a Gamma-Gamma model,* $Y_G \sim \Gamma(\alpha, \beta_G)$ *and* $Y_B \sim \Gamma(\alpha, \beta_B)$ *with* $\beta_G > \beta_B$. *Let* $C = \frac{P_B(r_B+c_B)}{P_G(r_G+c_G)}$ *and* k_0 *be defined as in (4.9). Then*
(i) $R'(k) \geq 0$ *when* $k \leq k_0$
(ii) $R'(k) \leq 0$ *when* $k \geq k_0$
(iii) $R''(k_0) < 0$

Proof is in the Appendix A.7.
Theorem 4.4 below follows immediately from Theorem 4.3.

Theorem 4.4 *Assuming a Gamma-Gamma model,* $Y_G \sim \Gamma(\alpha, \beta_G)$ *and* $Y_B \sim \Gamma(\alpha, \beta_B)$ *with* $\beta_G > \beta_B$. *Let* $R(k)$ *be defined in the range* $k \in [k_1, k_2]$, *and* k_0 *be defined as in (4.9). The maximization of* $R(k)$ *can be derived in three cases.*
(i) If $k_0 \leq k_1$, *then* $R(k)$ *is maximal when* $k^* = k_1$ *with* $R(k^*) = (P_G r_G - P_B c_B)$
(ii) If $k_0 \geq k_2$, *then* $R(k)$ *is maximal when* $k^* = k_2$ *with* $R(k^*) = (P_B r_B - P_G c_G)$
(iii) If $k_1 < k_0 < k_2$, *then the local maximum of* $R(k)$ *is located at* $k^* = k_0$ *with*

$$R(k^*) = P_B(r_B + c_B)F(k_0; \alpha, \beta_B) - P_G(r_G + c_G)F(k_0; \alpha, \beta_G) + R(k_1)$$

where $\mu_G = \alpha\beta_G$, $\mu_B = \alpha\beta_B$, $\sigma_G^2 = \alpha\beta_G^2$, $\sigma_B^2 = \alpha\beta_B^2$. *It is clear that* $\mu_G > \mu_B$ *and* $\sigma_G^2 > \sigma_B^2$.

In a Gamma-Gamma model with unequal dispersion parameters, the selection of the optimal cut-off point, k^*, becomes complex since there is no closed-form solution of k_0. We suggest applying some numerical method, such as bisection method or Newton-Raphson method, to approximate the value of k_0, which satisfies the following estimation equation.

$$\begin{aligned} 0 = d(k) &= k\left(\tfrac{1}{\mu_G\phi_G} - \tfrac{1}{\mu_B\phi_B}\right) - \ln(k)\left(\tfrac{1}{\phi_G} - \tfrac{1}{\phi_B}\right) \\ &+ \left[\ln C + \ln\left(\tfrac{\Gamma(\phi_G^{-1})}{\Gamma(\phi_B^{-1})}\right) + \tfrac{1}{\phi_G}\ln(\mu_G\phi_G) - \tfrac{1}{\phi_B}\ln(\mu_B\phi_B)\right] \end{aligned} \tag{4.10}$$

where $\mu_G = \alpha_G\beta_G$, $\mu_B = \alpha_B\beta_B$, $\sigma_G^2 = \alpha_G\beta_G^2$, $\sigma_B^2 = \alpha_B\beta_B^2$, and $\mu_G > \mu_B$.

4.5 Selection of Near-optimal Interval

This section extends the selection of near-optimal interval to the case of heteroscedastic Normal-Normal model or a Gamma-Gamma model with equal dispersion parameters. Figure 4.1 suggests that the shape of the Profit Function is no longer concave-down. We investigate the near-optimal interval of $R(k^*)$, corresponding to three cases concerning the shapes of $R(k)$ in a heteroscedastic Normal-Normal model or a Gamma-Gamma model with equal dispersion parameters, whether it is decreasing, increasing, or some other shape.

If the Profit Function is a decreasing function as in Figure 3.2.*a*, or an increasing function as in Figure 3.2.*b*, then the values of k_L^* and k_U^* can be derived from equations (3.4) and (3.5) in Section 3.2, respectively. If the Profit Function is similar to one of the two shapes shown in Figure 4.1, we will study the near-optimal interval whose range in k covers the value of k_0. The values of k_L^* and k_U^* can be derived from equation (3.6) in Section 3.2.

4.6 Chapter Summary

This chapter discussed the selection of optimal cut-off point in two typical models in the exponential family. One is the heteroscedastic Normal-Normal model. The conclusions of the heteroscedastic Normal-Normal model are similar to those in Chapter 3, but will vary by different cases of discriminant D_0 and variances (i.e. $\sigma_G^2 < \sigma_B^2$ or $\sigma_G^2 > \sigma_B^2$). The other is the Gamma-Gamma model with equal dispersion parameters.

It should be noted that it is not easy to derive some theorems for the near-optimal interval in these two models. A typical method is usually to identify the shape of $R(k)$ and then use the formulas in Section 3.2 to estimate the values of k_L^* and k_U^*.

The next chapter will provide two nonparametric methods in the estimation of optimal cut-off and near-optimal interval.

CHAPTER 5

ESTIMATION OF OPTIMAL CUT-OFF AND NEAR-OPTIMAL INTERVAL USING NONPARAMETRIC METHODS

5.1 Introduction

My practical experiences suggest that it is not easy to find a right parametric distribution for the data. When data can not be easily modeled using known parametric distributions, it is desirable to develop a *nonparametric* method to estimate the optimal and near-optimal cut-off points. A major advantage of the nonparametric estimation method is that it is free of distribution assumption and hence can avoid any bias caused by a mis-specification of the distribution function in the payoff function.

We propose two nonparametric methods in this chapter. One is a purely numerical approach and uses the kernel density estimation method to estimate

the quantities (k^*, k_L^*, k_U^*). Another is a graphical approach and uses the ROC curve. Since most statistical software can construct the ROC curve directly from data, the latter method has some appeal.

Section 5.2 introduces the kernel density estimator and its measurement error, and also discusses several estimation methods in the determination of optimal bandwidth. This section also introduces Azzalini's method in the selection of critical thresholds. In Section 5.3, we discuss the application of kernel density method to the determination of critical cut-off points. Section 5.4 discusses the selection of critical cut-offs using a graphical method from an existing ROC curve. The last section (5.5) will summarize this chapter and make introduction to the numerical study in Chapter 6.

5.2 Kernel Density Estimator

The kernel density estimator is by now the most widely studied nonparametric technique in the literature. Given a set of independent and identically distributed observations, $y_1, \ldots, y_n$, the kernel estimator of the probability density $f(.)$ is

$$\hat{f}_{w,h}(y) = \left(\sum_{i=1}^{n} w_i\right)^{-1} \sum_{i=1}^{n} \frac{w_i}{h_i} s\left(\frac{y - y_i}{h_i}\right) \tag{5.1}$$

where the notations $s(.)$ and h_i are discussed below. Equation (5.1) is a highly parametrised mixture model. As the special case of unit weight, equation (5.1) simplifies to

$$\hat{f}(y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i} s\left(\frac{y - y_i}{h_i}\right) \tag{5.2}$$

Let $\boldsymbol{w'} = (w_1, \ldots, w_n)$ be the weight parameter vector, and $\boldsymbol{h'} = (h_1, \ldots, h_n)$ be the scale parameter vector, with $s(.) \geq 0$ and $h_i \geq 0$, $i = 1, \ldots, n$. The scale parameter h_i is often called the *bandwidth* and determines the window width. The density $s(.)$ is a bounded density function called the *kernel* which

is a non-negative real-valued integrable function, that satisfies $\int s(t)dt = 1$ and $s(t) = s(-t)$ for all real t (hence. $\int ts(t)dt = 0$). The kernel function s determines the shape of the bumps of the estimated density function. There are numerous choices for the kernel function (Silverman, 1986). One typical option is that the kernel function is assumed to be a standard Gaussian function with mean zero and variance 1 (Storvik, 1999; Jones and Henderson, 2009).

The kernel density estimator has drawn some attention over years. Many efforts have been devoted to single adaptations in the kernel estimator, such as variable location, variable bandwidth, variable weight, high order kernel, etc. Abramson (1982) proposed the variable bandwidth kernel density estimators. Samiuddin and el-Sayyad (1990) discussed the variable location kernel density estimators. Hall and Turlach (1999) proposed the variable weight kernel density estimators. The recent research seems to favor the combination of variations in location, bandwidth, weights, etc. Storvik (1999) proposed the idea of maximum likelihood kernel density estimators that allow simultaneous variations of location, weight, and bandwidth. Jones and Henderson (2009) discussed the application of an enhanced version of Storvik's method, but their simulation study supported the simple location-only maximum likelihood kernel density estimator for practical use. Jones and Signorini (1997) reviewed and compared many of these methods and concluded that no method is significantly better than other methods. A recent study suggested that the variable-bandwidth kernel estimator can achieve desirable local adaptivity and the intuitive interpretability (Salgado-Ugarte and Perez-Hernandez, 2003).

Silverman (1986) and Izenman (1991) had a very good review of characteristics of kernel estimator, including measurement errors and asymptotic properties. Bias, mean squared error, and integrated mean squared error are often used to measure the discrepancy between the kernel density estimator and the true density. The derivation of these statistics starts from the expected

value of the kernel estimator defined by Whittle (1958)

$$E(\hat{f}(y)) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}E\left(s\left(\frac{y-y_i}{h_i}\right)\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}\int s\left(\frac{y-t}{h_i}\right)f(t)dt \quad (5.3)$$

and the bias follows immediately

$$bias_h(y) = E(\hat{f}(y)) - f(y) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}\int s\left(\frac{y-t}{h_i}\right)f(t)dt - f(y) \quad (5.4)$$

The mean squared error (MSE) and integrated mean squared error (IMSE) are related to equations (5.3) and (5.4) with the following forms

$$MSE_h(\hat{f}) = E\left(\hat{f}(y) - f(y)\right)^2 \quad (5.5)$$

$$IMSE(\hat{f}) = E\int\left(\hat{f}(t) - f(t)\right)^2 dt \quad (5.6)$$

In our application, obtaining an estimate of the density function is not enough. We need to find a reliable estimate of the cumulative distribution function (CDF) of Y, which allows the estimation of TP and FP in $R(k)$. Azzalini (1981) discussed the application of kernel estimation to cumulative distribution function and quantiles. His study adopted constant *bandwidth* h. The kernel estimate for the cumulative function of Y, corresponding to equation (5.2), is

$$\hat{F}(y) = \int_{-\infty}^{y}\hat{f}(t)dt = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{h}\int_{-\infty}^{y}s\left(\frac{t-y_i}{h}\right)dt\right) = \frac{1}{n}\sum_{i=1}^{n}\left(S\left(\frac{y-y_i}{h}\right)\right) \quad (5.7)$$

where $S(t) = \int_{-\infty}^{t}s(r)dr$

Nadaraya (1964) proved that under mild conditions, the mean and variance of $\hat{F}$ approach to those of F asymptotically. The mean squared error of $\hat{F}$ is as follows

$$E(\hat{F}(y) - F(y))^2 \approx F(y)(1-F(y))/n - uh/n + vh^4 \quad (5.8)$$

where $u = f(y)\left(h - \int_{-h}^{h}S^2(t)dt\right)$, $v = \left(\frac{1}{2}f'(y)\int_{-h}^{h}t^2s(t)dt\right)^2$

Asymptotic statistical properties of equation (5.2) have been well established. Parzen (1962), and Devroye and Wagner (1980) showed that under some assumptions of kernel and the bandwidth, provided f is continuous at y, the kernel density estimator $\hat{f}(y)$ is a consistent estimator of $f(y)$. Silverman (1986) showed that under suitable regularity conditions, the approximate value of IMSE converges to zero at the rate $n^{-\frac{1}{2}}$. If an optimal h is selected, the IMSE would converges to zero at a faster convergence rate of $n^{-\frac{4}{5}}$. Parzen (1962) gave a rather complex formula of the optimal bandwidth, which is derived from minimizing the integrated mean squared error in equation (5.6). Similarly, the asymptotically optimal bandwidth h can be derived from equation (5.8).

From practical point of view, a difficult and rather subjective part in kernel density estimation is to select a suitable bandwidth. The bandwidth, h, should increase with the density function $f(x)$ to reduce variance and should decrease with $|f''(x)|$ to reduce bias. Variable bandwidth method decreases the window width at high densities and increases it at intervals with high counts. The method is less vulnerable to noise in any low count interval of the distribution and catches distribution details in areas with high density of data. Several procedures in the selection of bandwidth, such as least-squares cross validation, optimal of Silverman, biased cross-validation, Sheather Jones plug-in, etc., appear in the literature (Silverman, 1986; Hardle, 1991; Scott, 1992; Wand and Jones, 1995; Simonoff, 1996; Bowman and Azzalini, 1997; Salgado-Ugarte et al. 1993, 1995b, 1997, 2003). Each method has its own strength and weakness. The bandwidth derived from the least-squared cross validation bandwidth fails to maintain the stability of the bandwidth in a relatively small data set. The procedures of the optimal Silverman bandwidth and biased cross-validation width bandwidth often oversmooth the probability densities in multi-modal densities. Jones et al. (1996) found that the Sheather Jones Plug-in bandwidth has the best performance.

After a kernel estimator is determined, one can integrate the kernel estimator of the density function and estimate the cumulative distribution function,

$\hat{F}$. Azzalini (1981) presented an application of $\hat{F}$ in a numerical example of the mean squared error of $\hat{F}$. $\hat{F}$ is derived at the *pth* quantile ξ_p, defined as $p = F(\xi_p)$ $(0 < p < 1)$. The mean squared error of $\hat{F}$ is derived from the comparison between the kernel estimation $\hat{F}$ and the empirical distribution function $\hat{F}_n$. The numerical results show that $\hat{F}$ is generally better than the empirical distribution function $\hat{F}_n$, with mean squared error $p(1-p)/n$. The optimum choice of *bandwidth* h is proportional to $n^{-\frac{1}{5}}$ in the density estimation, while the asymptotically optimum value of bandwidth is proportional to $n^{-\frac{1}{3}}$ in the mean squared error of $\hat{F}$ in equation (5.8). For detail, refer to Azzalini's paper on the estimation of distribution function (Azzalini, 1981).

We propose an approach to use Azzalini's method in the determination of optimal cut-off point k^*, and the near-optimal cut-off points (k_L^*, k_U^*). TP and FP are two quantities from the cumulative distributions in the good population and the bad population, respectively. Some kernel estimators can be used to estimate TP and FP. The *pth* quantiles, ξ_p, are then used to estimate these cut-off points.

5.3 Kernel Density Method

In this section, we propose a purely numerical approach outlined in Section 5.2. Our objective is to take the data Y_G and Y_B as input, then use equation (5.7) to produce a numerical table for $R(k)$, from which numerical values of k^*, k_L^*, and k_U^* are obtained. In the purely numerical approach, one uses the kernel method to estimate the quantities TP and FP, and then to compile a numerical table of k versus $R(k)$, where the value of k ranges from a given minimum k_1 to a given maximum k_2 with a given step size τ. From such a table it is easy to find (k^*, k_L^*, k_U^*) with a pre-specified precision. The step size τ determines the precision. Many functions can serve as the kernel in density estimates. We simply use the default kernel in R density() function, the Gaussian kernel. A computing algorithm is presented below to illustrate this approach.

Step 1. Find the minimum k_1 and the maximum k_2 from the estimated latent risk score, y^*. The step size τ is used to determine the precision and achieves a total of $\frac{k_2-k_1}{\tau}$ steps.

Step 2. Use equation (5.7) to create a table of k versus TP.

Step 3. Use equation (5.7) to create a table of k versus FP.

Step 4. Use the four payoff factors, plus the quantities P_G, P_B, TP, and FP, to compile a table of k versus $R(k)$.

Step 5. Look up the complied table and find the maximum value of $R(k)$ and its corresponding k^*.

Step 6. Find the value of $\left(\mathrm{R(k^*)} - \psi|R(k^*)|\right)$ and its corresponding k_L^* and k_U^* for the given allowance ratio ψ.

The variance of the payoff function $R(k)$ can be derived from Azzalini's result of the kernel estimate of the cumulative function of Y. Proof is in the Appendix A.8.

$$\begin{aligned} Var(\hat{R}(k)) &= \varsigma^2 \left(\tfrac{\hat{F}_G(k)[1-\hat{F}_G(k)]}{n_G}\right) + \gamma^2 \left(\tfrac{\hat{F}_B(k)[1-\hat{F}_B(k)]}{n_B}\right) \\ &+ \varsigma^2 \left(\tfrac{-u_G h}{n_G} + v_G h^4\right) + \gamma^2 \left(\tfrac{-u_B h}{n_B} + v_B h^4\right) \end{aligned} \tag{5.9}$$

where $\varsigma = P_G r_G + P_G c_G$, $\gamma = P_B r_B + P_B c_B$

$\hat{F}(G)$ and $\hat{F}(B)$ are the kernel estimators of the CDF of the good and bad populations, respectively. u_G and u_B are defined in equation (5.8) and are for the good and bad populations, respectively. v_G and v_B are defined in equation (5.8) and are for the good and bad populations, respectively.

When h approaches to zero in the order $o(n^{-1/4})$, (5.9) reduces to

$$Var(\hat{R}(k)) = \varsigma^2 \left(\frac{\hat{F}_G(k)[1-\hat{F}_G(k)]}{n_G}\right) + \gamma^2 \left(\frac{\hat{F}_B(k)[1-\hat{F}_B(k)]}{n_B}\right) \tag{5.10}$$

5.4 A Graphical Method

In a graphical approach, we draw several lines (defined below) on the ROC curve, and use the y-intercepts of these lines to determine the FP values corresponding to k^*, k_L^*, and k_U^*. The algorithm and its theoretical (geometrical)

foundation are given here, while a computing algorithm is provided in Chapter 6.

An ROC curve is a parametric plot of $y = TP(k)$ against $x = FP(k)$ for all given k in $[k_1, k_2]$. In ROC method, our goal is to use a graphical method to locate FP^*, FP^*_L, FP^*_U, the x-coordinates on ROC corresponding to k^*, k^*_L, and k^*_U. Call these points $A = (FP^*, TP^*)$, $B = (FP^*_L, TP^*_L)$, and $C = (FP^*_U, TP^*_U)$. In Chapter 1 we assume that the *near-optimal* payoff is defined via a parameter $0 < \psi < 1$, that is, $R(k)$ is called near-optimal if $R(k) \geq R(k^*) - \psi|R(k^*)|$. The proposed graphical method (see Figure 5.1) is based on the following two mathematical results. Let $L1$ be the tangent line at point A to the ROC curve, and $L2$ the secant line connecting points B and C. Two results follow (proofs are given in Appendix A.9). The first result has been approved in Stein (2005).

Result 1. Let the constants ς and γ be defined in equation (5.9). Then,

$$\textit{The slope of } L1 = \textit{the slope of } L2 = \gamma/\varsigma \tag{5.11}$$

Result 2. Let y_1 and y_2 be respectively the y-intercepts of $L1$ and $L2$. Then they satisfy the equation

$$y_2 = (1 - \psi)y_1 - \psi(d_0/\varsigma) \tag{5.12}$$

where $d_0 = P_B r_B - P_G c_G$

Based on these results, we propose the following graphical method to find the values of cut-off points, FP^*, FP^*_L, and FP^*_U. (Refer to Figure 5.1)

Step 1. From data, draw a smooth ROC curve.

Step 2. On the ROC plot, draw a line $L0$ passing through the origin and having the slope γ/ς.

Step 3. Draw a line $L1$ parallel to $L0$ and tangent to the ROC curve. Label the tangent point by "A". Its x-coordinate is FP^*. This ensures that the slope of $L1$ satisfies equation (5.11).

Step 4. On the y-axis, label the y-intercept point of $L1$ by "G", and the point $(0, -d_0/\varsigma)$ by "E".

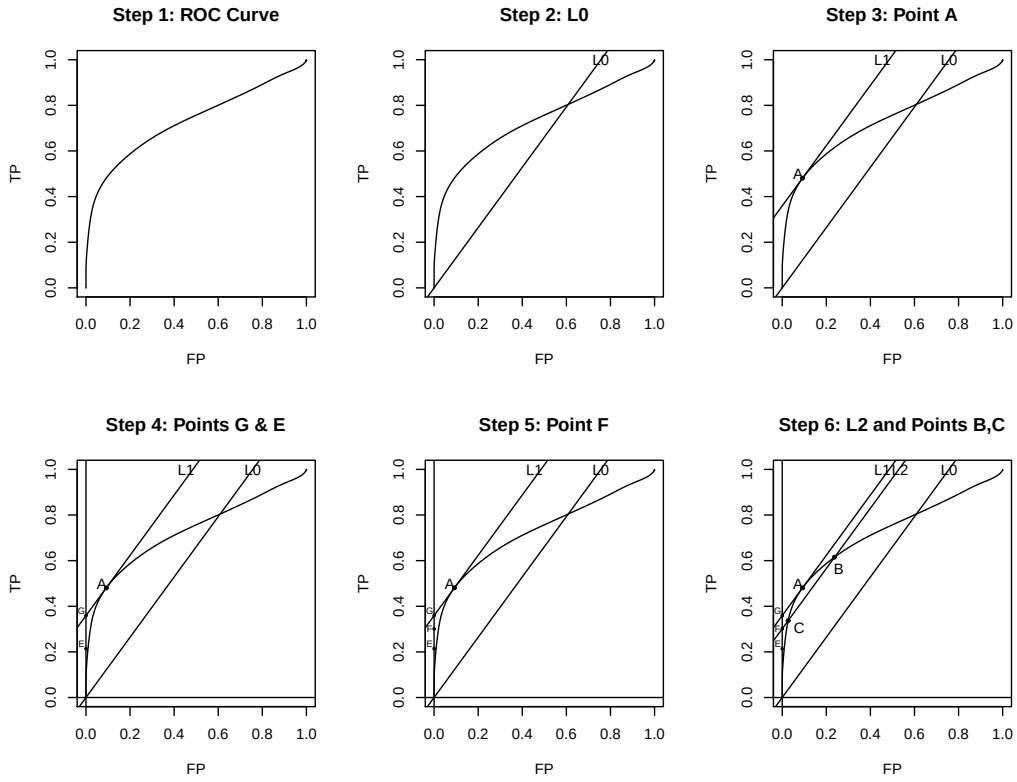

Figure 5.1: Using a ROC Plot to Determine FP^*, FP^*_L, FP^*_U

Step 5. On the y-axis, locate the point "F" on the line connecting G and E, such that the ratio of the two distances $\overline{GF} : \overline{FE}$ is $\psi : (1 - \psi)$. This ensures that the y-coordinates of G and F, y_1 and y_2 respectively, satisfy equation (5.12).

Step 6. Draw a third parallel line $L2$ passing through the point F. This ensures that the slope of $L2$ satisfies equation (5.11). Line $L2$ cuts the ROC curve at two points. Label the upper point by "B", and the lower one by "C". The x-coordinate of B is FP^*_L, and the x-coordinate of C is FP^*_U.

The algorithm above finds the optimal and near-optimal points FP^*, FP^*_L, and FP^*_U. If needed, one can find the optimal cut-off point k^* and the near-optimal cut-off points k^*_L and k^*_U using a tabulated table of k versus $FP(k)$.

5.5 Chapter Summary

This chapter presented two algorithms for estimating the critical thresholds of the Profit Function based on two nonparametric methods. In this first algorithm, the kernel estimator was applied to derive a tabulated table, from which one can find an approximate value for the critical threshold. The second algorithm used two results to estimate FP^*, FP_L^*, and FP_U^* on the ROC curve and find critical threshold on some similar tabulated table.

The next chapter will explain the data source and the steps taken to obtain the values of critical thresholds introduced in Chapter 3. The numerical study also demonstrates the results derived from the two nonparametric methods.

CHAPTER 6

A NUMERICAL STUDY

6.1 Introduction

This chapter will use simulated data to illustrate all the theorems and methodologies in the previous chapters. The next Section 6.2 will describe the data source and explain the simulation method used for the numerical study. Section 6.3 will introduce a numerical study to illustrate all theorems in Chapters 3. The numerical study applies generalized linear model to the simulated data set and estimates the optimal cut-off point and its near-optimal interval for a homoscedastic Normal-Normal model. Other related theorems are also discussed in the numerical study. In Section 6.4, we will illustrate how two nonparametric methods are used to select the optimal cut-off point and its near-optimal interval. The last section (6.5) concludes the numerical study and introduces the next chapter.

6.2 Data Sources

Several data files from the UCI Repository and Greene (1992) were integrated to construct individually based credit history for the purpose of credit re-assessment. Some variables contain information from credit bureau, including bureau risk score, number of recent credit inquiries, ownership of residence,

Table 6.1: Sample Data of Credit Re-assessment

ID	Dur	Credit	Inq	Curr Balance	Bureau Score	Length Resid	Offer Count	Co-App	Past Due	Edu	Sex	Income	Own/ Rent	Age	12Mon -NCF
1	28	$65,550	9	$53,595	657	8.4	6	0	0	1	2	6	1	54	$2,314
2	2	$69,800	1	$4,659	776	12.5	5	0	0	1	1	9	1	48	-$267
3	5	$58,650	3	$3,263	782	6.4	5	0	0	3	2	7	1	72	-$633
4	1	$78,232	2	$29,385	758	10.9	5	0	0	2	1	7	1	38	$178
5	16	$50,076	2	$11,158	773	14.4	6	0	0	1	1	6	1	66	-$861

length of residence, number of extended credit offers and co-applications. Some other variables include individual financial information such as existing credit line, current credit balance, number of late payments, years of established credit history, and 12 months of net cash flow (12Mon-NCF). The others are age, education level, gender, and income. Each observation in the integrated data consists of 15 variables concerning the above economic, financial, and personal information of a credit applicant.

The simulation is based on multivariate-normality and pairwise correlation coefficients are derived from the integrated file. The simulated data has a size of 20,000 observations. A sample of 5 records is listed as in Table 6.1. Our dependent variable D is defined as 1 if 12Mon-NCF $>$ \$0, or 0 if 12Mon-NCF $\leq$ \$0.

Various modeling techniques can be used to derive estimates of latent variable, y^*. In this paper, we apply probit model to the simulated data set using D as the dependent variable, and y^* is derived from the equation $y^* = \Phi^{-1}(P(\hat{D} = 1))$. The resulting y^* in the simulated data of size 20,000 is in the range of $[k_1 = -3.048, k_2 = 5.290]$. Our simulated data is based on the homoscedastic Normal-Normal model with $\sigma_G = \sigma_B = 1$, $\mu_G = 1.1867$, $\mu_B = 0.5628$, $P_G = 0.735$, and $P_B = 0.265$, where, the last four values were the parameter estimates from the probit model to the integrated data. The histograms of y^* in the good group and bad group are close to normality as seen in Figure 6.1.

As stated in Theorem 3.2 of Chapter 3 (Figure 3.2), the values of four payoff factors, r_G, r_B, c_G, and c_B, as wells as P_G and P_B, significantly impact the shape of Profit Function. All payoff factors play roles in the determination

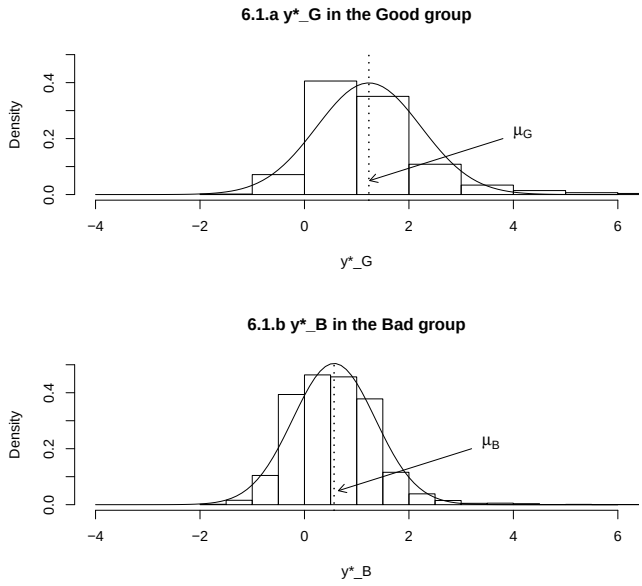

Figure 6.1: Histograms of y^* in The Data

of revenue and loss. r_G and c_B are related to accepted customers and therefore contribute to realized gain and loss, and r_B and c_G are related to potential gain and loss. Our numerical study fix P_G and P_B, but deploys six combinations of costs and revenues. We then examine the impact of C in equation (3.2) to the location of the optimal and near-optimal cut-off points, as well as the value of the optimal profit function. The sensitivity study uses six payoff matrices in Table 6.2. For convenience, all six payoff matrices share same values in r_B, r_G, and c_G, but have various values in c_B.

Each of 6 payoff matrices reflects a different stage in a dynamic economic environment, in which a typical economic circle includes several major phases, such as expansion, contraction, and recession. The third matrix is more popular when the economy is strong. In a booming economy, it is more likely for credit borrowers to maintain stable income and pay back credit debt. Since the expected profit per outcome / decision / customer, $R(k)$, is usually high, lending institutions tend to lower the lending criteria (including the value of

Table 6.2: Six Exemplary Payoff Matrices

		Case 1		Case 2		Case 3		Case 4		Case 5		Case 6	
		Y Decision		Y Decision		Y Decision		Y Decision		Y Decision		Y Decision	
		0	1	0	1	0	1	0	1	0	1	0	1
D	0	\$280	\$42	\$280	\$1680	\$280	\$2800	\$280	\$4900	\$280	\$7000	\$280	\$105000
(Actual)	1	\$560	\$1400	\$560	\$1400	\$560	\$1400	\$560	\$1400	\$560	\$1400	\$560	\$1400
c_B:r_G			0.03		1.2		2		3.5		5		75

k^*) in order to attract more potential customers and generate more revenues. During the economic contraction, the average loss from charge-off and delinquent behaviors starts to grow, at the same time, the capacity of making profit significantly decreases, as we can see from the fourth payoff matrix. Lending institutions tend to tighten the lending criteria (including the value of k^*) to decline credit seekers without solid credit profiles. During the recession, the loss was skyrocketing so that some banks may face huge loss or even fall down. A typical example is the financial crisis in 2008 and 2009. The fifth payoff matrix reflects the situation that most financial companies lost money in 2009. Without loss of generality, we will use the third matrix and the fourth matrix in the numerical study in Section 6.3 and Section 6.4, respectively.

6.3 An Illustration of Parametric Methods

In this section we demonstrate how the Profit Function method is used to determine the optimal threshold and near-optimal interval in the Normal-Normal model. We use a simulated data to demonstrate numerically the theories discussed in Chapter 3.

6.3.1 Selection of Optimal Cut-off Point

First, we illustrate the computational procedure of deriving the optimal and near-optimal cut-off points using the third payoff matrix, which represents a common situation in a booming economy. We compute $\ln C$ according to equation (3.2),

$$\ln C = \ln\left(\frac{0.265(280+2800)}{0.735(560+1400)}\right) = -0.5682$$

and $\hat{k}_0$ according to equation (3.1),

$$\hat{k}_0 = \frac{-0.5682}{1.1867 - 0.5628} + \frac{1.1867 + 0.5628}{2} = -0.0360$$

Since $\hat{k}_0$ is between k_1 and k_2, Theorem 3.2 stipulates that $R(k)$ is maximal at $\hat{k}^* = \hat{k}_0$, with value

$$\begin{aligned} R(\hat{k}^*) &= P_G(r_G + c_G)TP_{\hat{k}_0} - P_B(r_B + c_B)FP_{\hat{k}_0} - P_G c_G + P_B r_B \\ &= 0.735(560 + 1400)\Big(1 - \Phi(-0.0360 - 1.1867)\Big) \\ &\quad -0.265(280 + 2800)\Big(1 - \Phi(-0.0360 - 0.5628)\Big) \\ &\quad -0.735(560) + 0.265(280) \\ &= 351.67 \end{aligned}$$

To compute the near-optimal cut-off points $\hat{k}_L^*$ and $\hat{k}_U^*$, we set ψ to 0.20, and use $\hat{k}^*$ for the value of k_0 in equation (3.6), and equation (A.6) for the value of $R''(k_0)$. The results of C, $\hat{k}_0$, $R(\hat{k}^*)$, $\hat{k}_L^*$ and $\hat{k}_U^*$ are given in Table 6.3, case 3 and Figure 6.2.*c*.

6.3.2 Impact of C on $R(k)$ and $R(k^*)$

We are interested in the impact of the change in the payoff matrices on the Profit Function. As shown in Table 6.3, the change in the payoff matrices influences the value C, which, in turn, influences the shape of the Profit function, as seen in the graphs of Figure 6.2. The net effect is that $R(k^*)$ decreases when C increases (see Table 6.3). When C is small, the Profit function is monotonic decreasing, and decreases precipitously as k increases past k_U^* (Figure 6.2.*a*), suggesting that a more aggressive marketing strategy ought be used. As the value of C increases, the shape of the Profit function becomes uni-modal, as shown in Figures 6.2.*b*, 6.2.*c*, 6.2.*d*, and 6.2.*e*. Furthermore, the $R(k)$ curve is

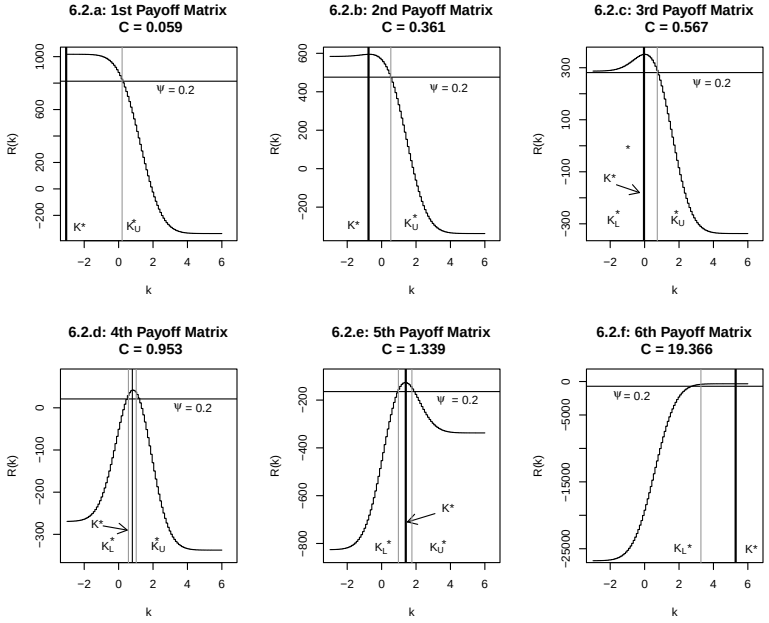

Figure 6.2: Profit Functions with 80% Near-optimal Intervals (with $\psi = 0.20$) and Optimal Thresholds

often asymmetrical with respect to the two sides of the vertical line $k = k^*$. For example, in Figure 6.2.d, $R(k)$ drops very quickly on the left-hand-side of the $k = k^*$ line, therefore the lending institution can employ a more "conservative" marketing strategy and select a relatively large cut-off point. Eventually, as C gets large, the Profit function changes to monotonically decreasing, and the value of $R(k^*)$ is well below zero. The shape of the Profit function increases precipitously as k increases, as shown in Figure 6.2.f. That indicates that the market is too risky to make profit, therefore the marketing program should be highly conservative, or even be suspended.

6.3.3 Impact of C on k^*

Sometimes, practitioners are interested in knowing the approximate location of k^* without building a risk model. Theorem 3.3 in Chapter 3 addresses this practical need. We next discuss how to use C, μ_G and μ_B to identify

Table 6.3: Numerical Results of Critical Cut-offs in the Parametric Methods

Variables	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
C values	0.059	0.361	0.567	0.953	1.339	19.366
k_0 values	-3.655	-0.760	-0.036	0.797	1.343	5.624
k^* values	-3.048	-0.760	-0.036	0.797	1.343	5.290
$R(k^*)$ values	\$1,018	\$595	\$352	\$42	-\$126	-\$337
k_L^* values ($\psi = 0.2$)	-3.048	-3.048	-3.048	0.570	0.975	3.272
k_U^* values ($\psi = 0.2$)	0.203	0.535	0.737	1.022	1.760	5.290

approximate location of k^*. As suggested in Theorem 3.3, the location of k^* relies on the relationship between C, $e_1 = e^{\frac{-(\mu_G-\mu_B)^2}{2}}$, and $e_2 = e^{\frac{(\mu_G-\mu_B)^2}{2}}$. Figure 6.3 lists the distributions of critical thresholds under six payoff matrices. In the first three cases, Figure 6.3.*a* clearly demonstrates that when the value C is less than e_1, the corresponding k_0 value is less than μ_B. It implies that the mean of the bad population has more impact on the value of threshold point. When the value of C is between e_1 and e_2, the corresponding k_0 value is located in the open interval of (μ_B, μ_G), as seen in case 4 in Figure 6.3.*a*. When the value C is larger than e_2, the corresponding k^* values are on the right side to μ_G (case 5 on Figure 6.3.*a* and case 6 in Figure 6.3.*b*). It implies that the mean of the good population has more impact on the value of threshold point.

6.4 An Illustration of Nonparametric Methods

The main objective of this section is to illustrate the application of two nonparametric methods in the selection of critical cut-off points. The first method is based on the kernel density estimation, while the second method uses the ROC curve to estimate critical cut-off points.

6.4.1 Kernel Density Method

We will use the simulated data file to derive kernel estimates of TP_k and FP_k. First, we set up a list of k values using the range $[k_1 = -3.048, k_2 = 5.290]$ and the step size $\tau = 0.005$. This step size will ensure that the resulting k^* is precise up to the second decimal place. We use Gaussian kernel function,

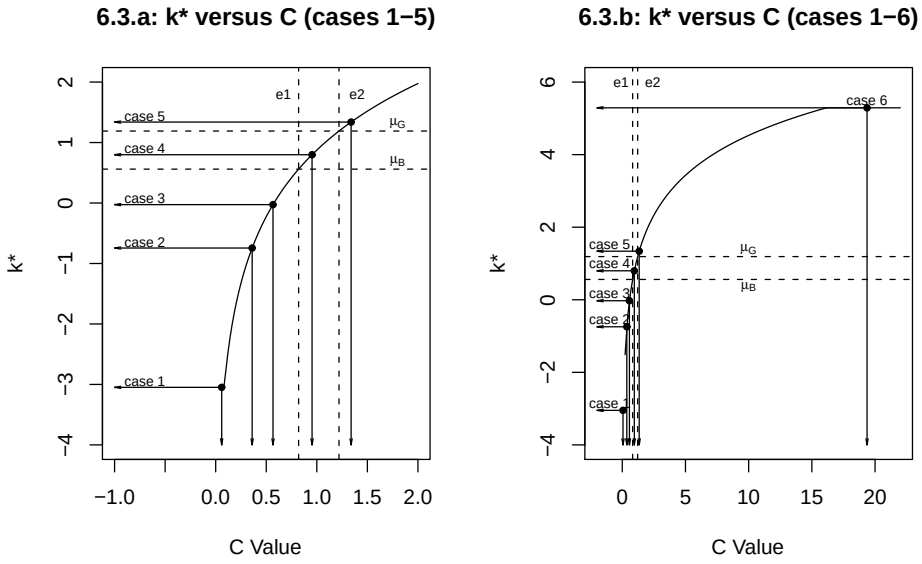

Figure 6.3: Locations of C, μ_B, μ_G, and k^*
Note: $e_1 = e^{\frac{-(\mu_G-\mu_B)^2}{2}}$, and $e_2 = e^{\frac{(\mu_G-\mu_B)^2}{2}}$

with bandwidths 0.20 for the good population and 0.23 for the bad population. Due to the capacity constraint of R and system ram size, a sub-sample of 5,000 observations is further randomly selected from the simulated data file of 20,000 observations.

We implement equation (5.7) in R to compute TP_k and FP_k, and the payoff factors in case 4 of Table 6.2 to compute $R(k)$. The results are graphed and presented in Figure 6.4. A set of numerical values of k, $\hat{T}P_k$, $\hat{F}P_k$, and $\hat{R}(k)$ is presented in Table 6.4. The first three rows are from the lower bound of $[k_1, k_2]$ and the last three rows are from the upper bound of $[k_1, k_2]$. The estimated values of (k_L^*, k^*, k_U^*) and $R(k^*)$, corresponding to the 722th, the 769th and the 814th row in Table 6.4, are reasonably close to those estimated by the parametric model (Case 4 in Table 6.3).

6.4.2 Confidence Interval of $R(k)$

Given the fourth payoff matrix, the value of $R(k^*)$ estimated from the Kernel

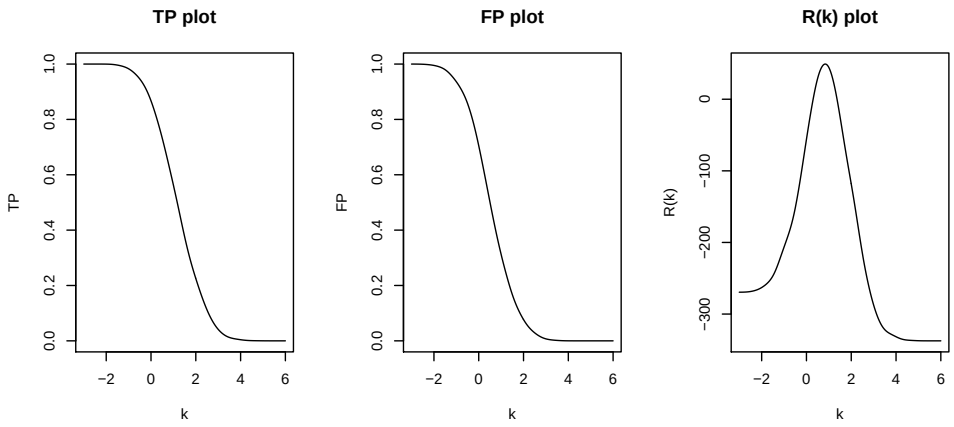

Figure 6.4: Kernel Estimation of TP_k, FP_k, and $R(k)$

Table 6.4: Numerical Results of Critical Cut-offs in the Kernel Density Method

Location	k	$\hat{TP}$	$\hat{FP}$	$\hat{R}(k)$
1	-3	1	0.99999	-$269.49
2	-2.995	1	0.99999	-$269.48
3	-2.99	1	0.99999	-$269.48
722	0.605	0.69891	0.45885	$39.58
769	0.84	0.61898	0.36802	$49.12
814	1.065	0.53914	0.29114	$39.63
1799	5.99	0.00000	0	-$337.40
1800	5.995	0.00000	0	-$337.40
1801	6	0.00000	0	-$337.40

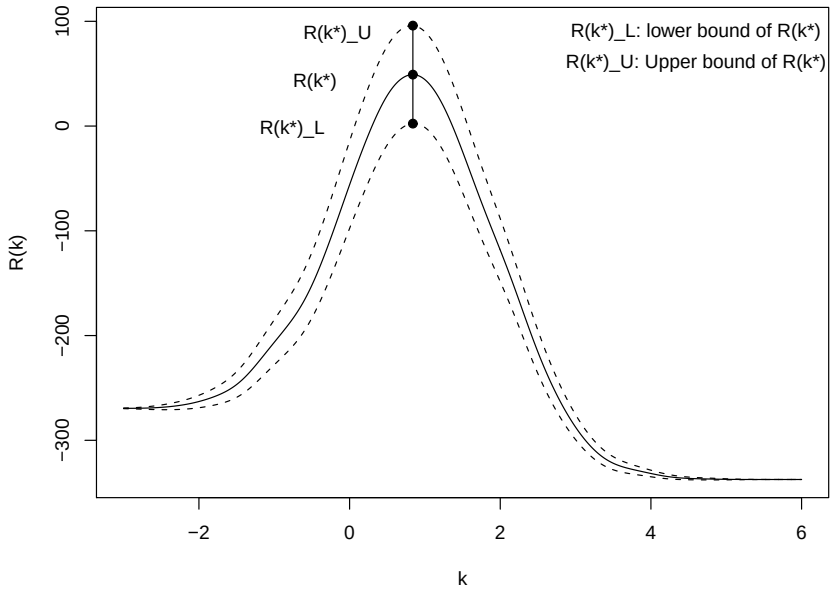

Figure 6.5: 90% Confidence Interval of $R(k)$

density method is \$49.12, as shown in the 769th row of Table 6.4. Equation (5.10) is used to derive the 90% confidence band for $R(k^*)$, with results shown in Figure 6.5. The same equation also leads the 90% confidence interval of $R(k^*)$ as [2.336,95.895]. From Figure 6.5, it is obvious that the variance of $R(k)$ is fairly small at the two tails of the payoff function and reaches the peak around the optimal cut-off point, k^*. This phenomenon is aligned to our initial conjecture that the payoff function $R(k)$ has small variances near the boundary of the range of k, and has more volatility around the optimal cut-off point. Thus this numerical result also explains why we need to propose the idea of the near-optimal interval.

6.4.3 ROC Graphical Method

We illustrate the computational procedure of deriving (k^*, k_L^*, k_U^*) in the ROC graphical method. First, we compute the slope, γ/ς, according to Result 1 in Chapter 5,

$$\frac{\gamma}{\varsigma} = \frac{0.265(280 + 4900)}{0.735(560 + 1400)} = 0.9529$$

and d_0

$$d_0 = P_B r_B - P_G c_G = 0.265(280) - 0.735(560) = -337.4$$

The y-intercept of $L1$ is estimated from the ROC curve on Figure 6.6,

$$y_1 \approx 0.27$$

The y-axis values of points E and F, y_E and y_2 respectively, are derived as

$$y_E = \frac{-d_0}{\gamma} = \frac{-(-337.4)}{0.735(560 + 1400)} = 0.2342$$
$$y_2 = y_1(1 - \psi) + y_E\psi \approx 0.27(1 - 0.2) + 0.2342(0.2) \approx 0.26$$

The x-axis values of points A, B, and C are estimated from the ROC curve in Figure 6.6,

$$\hat{F}P^* \approx 0.37$$

$$\hat{F}P_L^* \approx 0.46$$

$$\hat{F}P_U^* \approx 0.30$$

We also need to use k, y_G^*, and y_B^* to construct a new decision variable, Y. Y is used jointly with D and k to compile a similar table as Table 6.4 The estimated values of (k^*, k_L^*, k_U^*) are found to be as follows.

$$k^* \approx 0.835$$

$$k_L^* \approx 0.600$$

$$k_U^* \approx 1.035$$

These estimated values are reasonably close to the estimated critical cut-off points estimated by the parametric method (Case 4 in Table 6.3) and by the nonparametric kernel estimation method (Table 6.4).

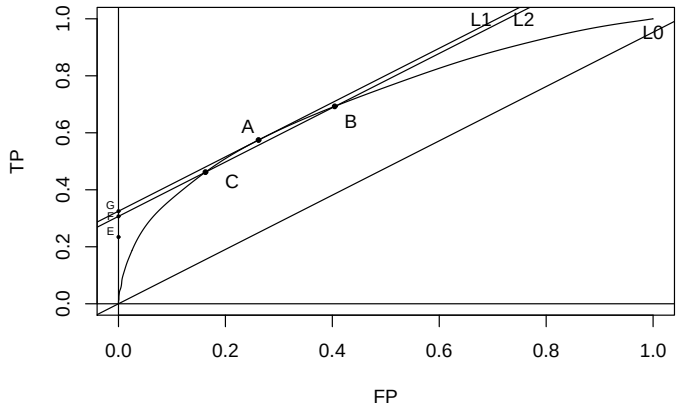

Figure 6.6: Application of ROC Graphical Method

6.5 Chapter Summary

A numerical study was presented in this chapter to illustrate theorems and methodologies in previous chapters. The resulting estimates for the critical thresholds were reasonably close, whether the parametric method or the non-parametric method was used. Parametric method can generate more accurate results if the underlying distribution of the data can be confirmed to be close to some specific distribution. If the underlying distribution of the data is far from any known distribution, nonparametric methods should provide robust results.

The numerical study implemented six payoff matrices. The numerical study suggests that the value of $R(k)$ changes with respect to changes in C in an opposite direction. That is, given the values of payoff factors and population percentages, if we increase the value of C, the value of $R(k)$ will decrease. We found that the change in $R(k^*)$ is nonlinear with respect to the change in C.

CHAPTER 7

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

7.1 Conclusion

The financial industry always faces risk in the lending business, and needs to enhance risk management, as more regulations will become effective. An actionable process is critical to the creditor to manage various loans, including mortgage loan, student loan, etc. with the goal of maximizing net income and assuring the financial safety. A typical loan process starts with collecting applicant's credit history, developing a scorecard and assigning a score to each applicant. The creditor evaluates the credit score of each applicant, selects a cut-off point, and makes a decision to accept or reject the loan application.

Our study assumed that the latent credit scores of good population and bad population were from two different distributions. A total of 10 parameters, including 4 conditional decision probabilities, 2 population probabilities, and 4 payoff factors were used to construct the Profit Function. The methodology of Profit Function provided a flexible tool to maximize revenue and minimize cost, eventually achieved maximization of net income. The derived optimal cut-off

point was used to classify loan applicants into two exclusive groups, acceptance population and rejection population, and achieved the maximization of the Profit Function. The study gave the closed-form solution of the optimal cut-off point for the Normal-Normal distribution, and provided a general form of solution for the E-E model. The analysis showed that when the dispersion parameters of two distributions were different, i.e. in a E_G-E_B model, it became more complex to determine the locations of the critical cut-off points. A numerical method was usually needed.

This dissertation provided three types of interval in the study of the Profit Function. The Delta method and Taylor Series were used to derive the variance of the optimal cut-off point, which in turn was used to construct the confidence interval for the optimal cut-off point and the optimal value of the Profit Function (Section 3.3). We also discussed the typical confidence interval of the Profit Function in Section 5.3. A third type of interval, which we called the near-optimal interval, was perhaps the most attractive solution as it could provide more flexibility in practical risk management.

Besides the parametric method, we introduced two nonparametric methods. The kernel estimator was used to construct a numerical table to estimate critical cut-off points. Also, we derived an ROC-based method to estimate critical thresholds since the ROC curve is widely used in various fields. The results from parametric methods and nonparametric methods were reasonably close.

Our paper contributed to the literature in several areas. The following three were the most valuable. One was that this dissertation assumed a general exponential family for the distribution of the underlying creditworthiness and discussed the selection of the optimal cut-off point for various shapes of the Profit Functions. The other was to introduce the concept of near-optimal interval, which gave practitioners the flexibility of implementing various strategies in marketing and risk management. The third contribution was that this paper provided nonparametric as well as parametric solutions. The study in this dissertation formed a systematic credit re-assessment process for creditors.

7.2 Future Extensions

The proposed methodology in this dissertation can be extended in a variety of directions.

Much of the literature assumes that the decision variable has only two classes, Good or Bad, or Yes or No. With more than two classes, the complexity of the diagnostic study increases exponentially. With n classes, the payoff matrix becomes an $n \times n$ matrix, with n correct classifications along the diagonal line and $n^2 - n$ incorrect classifications off the diagonal line. One needs to study n benefits and $n^2 - n$ errors under the $n \times n$ matrix. Fawcett (2006) proposed an easy and feasible solution, which is to generate n different ROC curves. Provost and Domingos (2001) introduced the weighted AUCs in the study of multi-classes dependent variable. Hand and Till (2001) suggested a non-weighted approach to the multi-class generalization of AUCs. These methods may be more difficulty to be adapted to the selection of optimal cut-off point in the ROC method, over the method of Profit Function.

The second extension would assume that the payoff factors are random. Our study assumes 6 different payoff matrices. A major loan-lending company usually have many loan products and large market shares, and needs loan-specific payoff matrices; for the same loan product, the customer populations in various locations / segments may have different payoff matrices. In this setting, the researcher is interested in predicting the net income in a dynamic environment.

Another extension is to extend the usage of the constant kernel function and kernel bandwidth discussed in Chapter 5 to apply the variable bandwidth kernel density method to enhance the accuracy of kernel estimators. Some other kernel functions, such as Biweight, Triangular, or Rectangular kernels, can be used to compare with the Gaussian kernel, which is implemented in our study. Silverman (1986) proposed to compare the efficiency of various kernels to decide an ideal kernel function.

REFERENCES

Abramson, I.S. (1982). *On bandwidth variation in kernel estimates - a square root law.* Ann. Statist., 9, 168-176.

Altman, E.I., R. Haldeman, and P. Narayanan (1977). *ZETA analysis, a new model for Bankruptcy classification.* Journal of Banking and Finance.

Azzalini, A (1981). *A note on the estimation of a distribution function and quantiles by a kernel method.* Biometrika, 68, 1, 326-328.

Bairagi and Suchindran (1982). *An Estimator of the cut-off point maximizing sum of sensitivity and specificity.*

Blakesley, Vicky (2005). *Current Methodology to Assess Bioequivalence of Levothyroxine Sodium Products Is Inadequate.* The AAPS Journal 2005, 7 (1) Article 5.

Blume, Jeffrey D. (2008). *Bounding Sample Size Projections for the Area Under a ROC Curve.* Journal of Statistical Planning and Inference.

Bowman, A.W. and A. Azzalini (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*, Volum 18, Oxford University Press.

Cai, Tianxi and Chaya S. Moskowitz (2004). *Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test.* Biostatistics, 5, 4, 573-586

Chauchat, J-H., R. Rakotomalala, M. Carloz, and C. Pelletier (2001). *Targeting Customer Groups using Gain and Cost Matrix: a Marketing Application.* Proceedings of Data Mining for Marketing.

Cortes, Corinna and Mehryar Mohri (2004). *Confidence Intervals for the Area under the ROC Curve.*

Cortes, Corinna and Mehryar Mohri (2005). *AUC Optimization vs. Error Rate Minimization.*

Devroye, L.P. and T. J. Wagner (1980). *Distribution-Free Consistency Results in Nonparametric Discrimination and Regression Function Estimation.* The Annals of Statistics, Vol. 8, No. 2, 231-239

Di, Lascio and Ewens (2007). *Comparison between K-Means and Hierarchical Clustering of Dependent and Independent Data Generated from Multivariate Gaussian Copula Function.* NZSA 2007 Conference.

Drasgow, Fritz (1982). *Polychoric and Polyserial Correlations.* Encyclopedia of Statistical Sciences, Volume 7, 68-74.

Drummond, Chris and Robert C. Holte (2006). *Cost Curves: An improved method for visualizing classifier performance.* Machine Learning, 65, 95-130.

Dunteman, George H. and Moon-Ho R. Ho (2006). *An Introduction to Generalized Linear Models.* Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-145. Thousand Oaks, CA: Sage.

Egan, James P. (1975), *Academic Press*. Signal Detection Theory and ROC Analysis.

Emrich, L.J. and M.R. Piedmonte (1991). *A Method for generating high-dimensional multivariate binary variables*. The American Statistician, Vol. 45, No. 4, 302-304.

Faraggi, David and Benjamin Reiser (2002). *Estimation of the area under the ROC curve*. Statistics in Medicine, 21, 3093-3106.

Fawcett, Tom (2006). *An Introduction to ROC analysis*. Pattern Recognition Letters, 27, 861-874.

FDA (2008). *Publication of Guidances for Industry Describing Product-Specific Bioequivalence Recommendations*. http://www.fda.gov/OHRMS/DOCKETS/98fr/FDA-2007-D-0369-NAD.pdf

Gange, Stephen J. (1995). *Generating Multivariate Categorical Variates Using the Iterative Proportional Fitting Algorithm*. The American Statistician, Vol. 49, No. 2, 134-138.

Gill, Jeff (2000). *Generalized Linear Models - A Unified Approach*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-134. Thousand Oaks, CA: Sage.

Green, M. David and A. John Swets (1966). *Signal Detection Theory and Psychophysics*. Peninsula Publishing.

Greene, William H. (1992). *A Statistical Model for Credit Scoring*. Working Papers 92-29, New York University, Leonard N. Stern School of Business, Department of Economics.

Greene, William H. (2003). *Econometric Analysis*. Pearson Education.

Hall, P. and B.A. Turlach (1999). *Reducing bias in curve estimation by use of weights*. Computational Statistics & Data Analysis, 30(1), 67-86.

Hand, D.J., and R.J. Till (2001). *A simple generalization of the area under the ROC curve to multiple class classification problem*. Mach. Learning, 45 (2), 171-186.

Hamdan, M.A. (1970). *The Equivalence of Tetrachoric and Maximum Likelihood Estimates of ρ in 2 $\times$ 2 Tables*. Biometrika, Vol. 57, 212-215.

Hardin, James and Joseph Hilbe (2001). *Generalized Linear Models*. Stata Press.

Hardle, W. (1991). *Smoothing techniques: With implementation in S*. New York: Springer-Verlag.

Harshinder Singh, Shengqiao Li, and Michael Andew (2007). *A Quantitative Assessment of Diagnostic Cut-off Point Selection Methods*.

Hautus, Michael J., and R. John Irwin (1995). *Two Models for Estimating The Discriminability of Foods and Beverages*. Journal of Sensory Studies, Vol 10. Issue 2, 203-215.

Irwin, R. John, Michael J. Hautus, and Jennifer A. Stillman (1992). *Use of The Receiver Operating Characteristic in The Study of Taste Perception*. Journal of Sensory Studies, Vol 7. Issue 4, 291-314.

Izenman, A.J. and L.T. Tran(1990). *Kernel Estimation of the Survival Function and Hazard Rate Under Weak Dependence*. Journal of Statistical Planning and Inference, 24, 233-247.

Izenman, A. J. (1991). *Recent Developments in Nonparametric Density Estimation*. Journal of the American Statistical Association, 86, 205-224.

Johnson, Paul E. (2006). Inverse Gaussian Distribution. http://pj.freefaculty.org/stat/Distributions/InverseGaussian-01.pdf

Jones and Henderson (2009). *Maximum likelihood kernel density estimation: On the potential of convolution sieves*. Computational Statistics and Data Analysis, Volume 53 , Issue 10.

Jones, M.C. and D.F. Signorini (1997). *A comparison of higher-order bias kernel density estimators*. Journal of the American Statistical Association, 92, 1063-1073.

Jones, M.C., J.S. Marron, and S.J. Sheather (1996). *Progress in data-based bandwidth selection for kernel density estimation*. Computational Statistics, 11, 337-381.

Juras, Josip, and Pasaric Zoran (2006). *Application of tetrahoric and polychoric correlation coefficients to forecast verication*. GEOFIZIKA, Vol. 23, No.1.

Lee, A.J. (1993). *Generating Random Binary Deviates Having Fixed Marginal Distributions and Specified Degrees of Association*. The American Statistician, Vol. 47, No. 3, 209-215.

Leisenring, Wendy, and Margeret Sullivan Pepe (1998). *Regression Modelling of Diagnostic Likelihood Ratios for the Evaluation of Medical Diagnostic Tests*. Biometrics, Vol. 54, 444-452.

Lunn, A.D. and S.J. Davies (1998). *A note on Generating Correlated Binary Variables*. Biometrika, Vol. 85. No. 2. 487-490.

Macskassy, Sofus, and Foster Provost (2004). *Confidence Bands for ROC Curves*. CeDER Working Paper 02-04, Stern School of Business, New York University, NY, NY 10002.

Macskassy, Sofus, Foster Provost, and Saharon Rosset (2005). *ROC Confidence Bands: An Empirical Evaluation*. Proceedings of the 22nd International Conference on Machine Learning.

Marzban, Caren (2004). *A Comment on the ROC Curve and the Area Under it as Performance Measures*.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models*. New York: Chapman & Hall.

Metz, Charles (1978). *Basic Principles of ROC Analysis*. Seminars in Nuclear Medicine, Vol. VIII, No.4 (October).

Murrell, Paul (2006). *R Graphics*. Chapman Hall / CRC.

Nadaraya, E.A. (1964). *Some new estimates for distribution functions*. Theory Prob. Applic, 15, 497-500.

Nelder, J.A. and R.W.M Wedderburn (1972). *Generalized Linear Models*. Journal of the Royal Statistical Society, Series A, 135, 370-385

Noe, D.A. (1983). *Selecting a diagnostic study's cut-off value by using its receiver operating characteristic curve*. Clinical Chemistry, 29, 571, Vol 29, No.3.

Park, Chul Gyu, Taesung Park, Dong Wan Shin (1996). *The American Statistician*, Vol. 50, No. 4, 306-310.

Parzen, Emanuel (1962). *On estimation of a probability density function and mode*. Annals of Mathematical Statistics, Vol. 33, 1065-1076.

Pearson, K. (1900). *Mathematical Contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable*. Philos. Tr. R. Soc., S.-A, Vol. 195, 1-47.

Pepe, Margeret Sullivan (1997). *A Regression Modelling Framework for Receiver Characteristic Curves in Medical Diagnostic Testing*. Biometrika, Vol. 84, No.3, 595-608.

Pepe, Margeret Sullivan (1998). *Three Approaches to Regression Analysis of Receiver Operating Characteristic Curves for Continuous Test* . Biometrics, Vol. 54, No.1, 124-135.

Pepe, Margeret Sullivan (2000). *An Interpretation for the ROC Curve and Inference Using GLM Procedures*. Biometrics, Vol. 56, 352-359.

Pepe, Margeret Sullivan (2006). *Adjusting for Covariate Effects on Classification Accuracy Using the Covariate-Adjusted ROC Curve*. The Berkeley Electronic Press.

Pepe, Margeret Sullivan and Tianxi Cai (2004). *Combining Predictors for Classification Using the Area Under the ROC Curve*. The Berkeley Electronic Press.

Pepe, Margeret Sullivan (2000). *Receiver Operating Characteristic Methodology*. Journal of the American Statistical Association, Vol. 95, No. 449, 308-311.

Pepe, Margeret Sullivan (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press.

Prentice, R.L. (1986). *Binary regression Using an Extended Beta-Binomial Distribution, with Discussion of Correlation Induced by Covariate Measurement Errors*. Journal of the American Statistical Association, Vol. 81, No. 394, 321-327.

Prentice, R.L. (1988). *Correlated Binary Regression with Covariates Specific to Each Binary Observation*. Biometrics, Vol. 44, No. 4, 1033-1048.

Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery (1992). *Numerical Recipes in C*. Cambridge University Press.

Provost, F., and T. Rawcett (1977). *Analysis and visualization of classifier performance: Comparison under imprecise class and cost distribution*. Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97).

Provost, F. and P. Domingos (2001). *Well-trained PETs: Improving Probability Estimation Trees*. CeDER Working Paper IS-00-04, Stern School of Business, New York University, NY, NY 10012.

Reiser, Benjamin and David Faraggi (1997). *Confidence Intervals for the Generalized ROC Criterion*. Biometrics, Vol. 53, No.2, 644-652.

Salgado-Ugarte, I. H., M. Shimizu, and T. Taniuchi (1993). snp6: *Exploring the shape of univariate data using kernel density estimators*. Stata Technical Bulletin 16: 8 C 19. In Stata Technical Bulletin Reprints, vol. 3, 155 C 173. College Station, TX: Stata Press.

Salgado-Ugarte, I. H., M. Shimizu, and T. Taniuchi (1995b). snp6.2: *Practical rules for bandwidth selection in univariate density estimation* . Stata Technical Bulletin 27: 5 C 19. In Stata Technical Bulletin Reprints, vol. 5, 172 C 190. College Station, TX: Stata Press.

Salgado-Ugarte, I. H., M. Shimizu, and T. Taniuchi (1997). snp13: *Nonparametric assessment of multimodality for univariate data*.

Stata Technical Bulletin 38: 27 C 35. In Stata Technical Bulletin Reprints, vol. 7, 232 C 243. College Station, TX: Stata Press.

Salgado-Ugarte and Perez-Hernandez (2003). *Exploring the use of variable bandwidth kernel density estimators*. The Stata Journal, 3, Number 2, 133-147.

Samiuddin, M. and G.M. el-Sayyad (1990). *On nonparametric kernel density estimates*. Biometrika, 77, 865-874.

Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: John Wiley & Sons.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Simonoff, J.S. (1996). *Smoothing Methods in Statistics*. New York: Springer.

Song, Xiao and Xiao-hua Zhou (2006). *Covariate Specific ROC Curve With Survival Outcome*. The Berkeley Electronic Press.

Stein, Roger M. (2005). *The Relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing*. Journal of Banking and Finance, 29, 1213-1236.

Storvik, B.E. (1999). *Contributions to Nonparametric Curve Estimation*. Dr.Scient. thesis. University of Oslo.

Wand, M.P. and M.C. Jones (1995). *Kernell Smoothing*. London: Chapman and Hall.

Wherry, R. J., SR. (1984). *Contributions to correlation analysis*. New York: Academic Press.

Whittle, P. (1958). *On the smoothing of probability density functions*. J. Roy. Statist. Soc. Ser. B. 20 334-343.

Yan, Lian, Michael C. Mozer, and Richard Wolniewicz (2003). *Optimizing Classifier Performance via an Approximation to the Wilcoxon-Mann-Whitney Statistic*. Proceeds of the Twentieth International Conference on Machine Learning (ICML-2003). Washington DC.

Zou, Keely H. and W.J. Hall (2000). *Two transformation models for estimating an ROC curve derived from continuous data*. Journal of Applied Statistics, Vol. 27, No. 5, 621-631.

Zou, Kelly H. (2006). *Receiver Operating Characteristic (ROC) Literature Research*. http://www.spl.harvard.edu/archive/spl-pre2007/pages/ppl/zou/roc.html

APPENDIX A

Mathematical Proofs

A.1 Proof for Theorem 3.1

The first derivative of $R(k) = P_G(r_G+c_G)TP_k - P_B(r_B+c_B)FP_k + (P_B r_B - P_G c_G)$ is as follows:

$$\begin{aligned} R'(k) &= P_B(r_B + c_B)\phi(\mu_B - k) - P_G(r_G + c_G)\phi(\mu_G - k) \\ &= P_G(r_G + c_G)\phi(\mu_B - k)\left(C - \tfrac{\phi(\mu_G - k)}{\phi(\mu_B - k)}\right) \end{aligned}$$

where $C = \frac{P_B}{P_G}\frac{(r_B+c_B)}{(r_G+c_G)}$ and ϕ is the density function of the standard normal distribution.

Then $R'(k) \geq 0$ is true if and only if

$$P_G(r_G + c_G)\phi(\mu_B - k)\left(C - \frac{\phi(\mu_G - k)}{\phi(\mu_B - k)}\right) \geq 0 \tag{A.1}$$

Since P_G, r_G, c_G, and $\phi(\mu_B - k)$ are positive, (A.1) is equivalent to the following condition.

$$C \geq \frac{\phi(\mu_G - k)}{\phi(\mu_B - k)} \tag{A.2}$$

Taking logarithmic transformation on both sides, and under the N-N model, (A.2) can be rewritten as

$$\ln C \geq (\mu_G - \mu_B)\left(k - \frac{\mu_G + \mu_B}{2}\right) \tag{A.3}$$

which is equivalent to the following condition under the fist assumption in Section 3.2, $\mu_G > \mu_B$,

$$k \leq \frac{\ln C}{\mu_G - \mu_B} + \frac{\mu_G + \mu_B}{2} = k_0 \tag{A.4}$$

Similarly, $R'(k) \leq 0$ is true if and only if the following condition holds,

$$k \geq k_0 \tag{A.5}$$

The proof of (iii) in Theorem 3.1 is shown in three scenarios.

$$\frac{\partial^2 R(k)}{\partial k^2} = \frac{-(\mu_G - k)}{\sqrt{2\pi}} e^{\frac{-(k-\mu_G)^2}{2}} P_G(r_G + c_G) + \frac{(\mu_B - k)}{\sqrt{2\pi}} e^{\frac{-(k-\mu_B)^2}{2}} P_B(r_B + c_B) \tag{A.6}$$

Substituting $k = k_0 = \frac{\ln C}{\mu_G - \mu_B} + \frac{\mu_G + \mu_B}{2}$ into (A.6), where $C = \frac{P_B}{P_G} \frac{(r_B + c_B)}{(r_G + c_G)}$, after some math, we get $R''(k_0) < 0$ if and only if

$$(\mu_B - k_0) < (\mu_G - k_0) \tag{A.7}$$

(I) If $k_0 < \mu_B$, then $k_0 < \mu_B < \mu_G$ (by assumptions in Section 3.2). That is, $\mu_G - k_0 > \mu_B - k_0$, (A.7) holds.
(II) If $\mu_B \leq k_0 \leq \mu_G$, then $\mu_B - k_0 \leq 0 \leq \mu_G - k_0$. And for the special case that $\mu_B - k_0 = 0$, $\mu_G - k_0 = \mu_G - \mu_B > 0$, then (A.7) holds again.
(III) If $\mu_G < k_0$, then $\mu_B < \mu_G < k_0$. So $(\mu_B - k_0) < (\mu_G - k_0)$ and (A.7) is true.
Therefore, $R''(k_0) < 0$ is always satisfied.

A.2 Proof for Theorem 3.3

(I) When $\ln C \leq \frac{-(\mu_G - \mu_B)^2}{2}$, we get $(\frac{\ln C}{\mu_G - \mu_B} + \frac{\mu_G + \mu_B}{2}) - \mu_B \leq 0$. Given $k^* = \frac{\ln C}{\mu_G - \mu_B} + \frac{\mu_G + \mu_B}{2}$, then $k^* \leq \mu_B$ is satisfied.
(II) When $\ln C \geq \frac{(\mu_G - \mu_B)^2}{2}$, we get $(\frac{\ln C}{\mu_G - \mu_B} + \frac{\mu_G + \mu_B}{2}) - \mu_G \geq 0$. Given $k^* = \frac{\ln C}{\mu_G - \mu_B} + \frac{\mu_G + \mu_B}{2}$, then $k^* \geq \mu_G$ is satisfied.
(III) When $\frac{-(\mu_G - \mu_B)^2}{2} < \ln C < \frac{(\mu_G - \mu_B)^2}{2}$, we then get $(\frac{\ln C}{\mu_G - \mu_B} + \frac{\mu_G + \mu_B}{2}) - \mu_B > 0$ from the left inequality and $(\frac{\ln C}{\mu_G - \mu_B} + \frac{\mu_G + \mu_B}{2}) - \mu_G < 0$ from the right inequality. Therefore, when $k^* = (\frac{\ln C}{\mu_G - \mu_B} + \frac{\mu_G + \mu_B}{2})$, then $\mu_B < k^* < \mu_G$.

A.3 Derivation of the near-optimal interval in the Normal-Normal model

(I) If the Profit function is a decreasing function, k_L^* is located at $k^* = k_1$. The Taylor series of $R(k)$ at k^* can be written as the following:

$$R(k) = R(k^*) + (k - k^*)R'(k^*) + \frac{1}{2}(k - k^*)^2 R''(k^*) + \dots$$

Provided that a linear approximation is adequate for the profit function, k_U^* can be estimated from the following function.

$$R(k) \approx R(k^*) + (k - k^*)R'(k^*)$$

$$R(k^*) - R(k) = -(k - k^*)R'(k^*) \tag{A.8}$$

Following equation (3.3), (A.8) is rewritten as

$$\psi|R(k^*)| = -(k - k^*)R'(k^*)$$

In this case, the creditworthiness interval between $[k_L^*, k_U^*]$ is $[k_1, k_U^*]$. The value of k_U^* is derived from equation (A.9).

$$k_U^* \approx k_1 - \frac{\psi|R(k_1)|}{R'(k_1)} \tag{A.9}$$

(II) Similarly, if the Profit function is an increasing function, k_U^* is located at $k^* = k_2$. The creditworthiness interval between $[k_L^*, k_U^*]$ is $[k_L^*, k_2]$. Provided that a linear approximation is adequate for the profit function, the value of k_L^* is derived from equation (A.10).

$$k_L^* \approx k_2 - \frac{\psi|R(k_2)|}{R'(k_2)} \tag{A.10}$$

(III) If the Profit function is a concave function, $k^* = \frac{\ln C}{\mu_G - \mu_B} + \frac{\mu_G + \mu_B}{2}$ and $R'(k^*) = 0$. $R(k)$ can be written as Taylor series at $k = k^*$.

$$\begin{aligned} R(k) &= R(k^*) + (k - k^*)R'(k^*) + \tfrac{1}{2}(k - k^*)^2 R''(k^*) + \dots \\ &\approx R(k^*) + \tfrac{1}{2}(k - k^*)^2 R''(k^*) \end{aligned} \tag{A.11}$$

Equation (A.11) is rewritten as

$$-\psi|R(k^*)| = \frac{1}{2}(k-k^*)^2 R''(k^*)$$

k_L^* and k_U^* can be estimated from equation (A.12).

$$k \approx k^* \pm \sqrt{\frac{-2\psi|R(k^*)|}{R''(k^*)}} \tag{A.12}$$

with $k_L^* = k^* - \sqrt{\frac{-2\psi|R(k^*)|}{R''(k^*)}}$ and $k_U^* = k^* + \sqrt{\frac{-2\psi|R(k^*)|}{R''(k^*)}}$

When ψ is equal to zero, equations (A.9), (A.10), and (A.12) are simplified to equation (3.1).

A.4 Use the Delta method to derive the asymptotic distribution of optimal cut-off in the Normal-Normal model

$$\hat{k}^* = h(\bar{y}_G^*, \bar{y}_B^*) = \frac{\ln C}{\bar{y}_G^* - \bar{y}_B^*} + \frac{\bar{y}_G^* + \bar{y}_B^*}{2}$$

Let n_G and n_B be the respective sample size from good and bad populations, and $n = n_G + n_B$. Since y_G^*'s and y_B^*'s are from independent groups, we get $cov(\bar{y}_G^*, \bar{y}_B^*) = 0$. By central limit theorem, we have the asymptotic distributions of $(\bar{y}_G^*, \bar{y}_B^*)$, for large n,

$$\sqrt{n}\left[\begin{pmatrix}\bar{y}_G^* \\ \bar{y}_B^*\end{pmatrix} - \begin{pmatrix}\mu_G \\ \mu_B\end{pmatrix}\right] \textit{ is approximately distributed as } N_2\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\frac{n}{n_G} & 0 \\ 0 & \frac{n}{n_B}\end{pmatrix}\right]$$

Next, we will derive the asymptotic distribution of $h(\bar{y}_G^*, \bar{y}_B^*)$. Using the Delta method,

$$\sqrt{n}\left[h(\bar{y}_G^*, \bar{y}_B^*) - h(\mu_G, \mu_B)\right] \textit{ is approximately distributed as } N(0, \sigma_0^2) \tag{A.13}$$

Note that $h(\mu_G, \mu_B) = \frac{\ln C}{\mu_G - \mu_B} + \frac{\mu_G + \mu_B}{2}$. The approximate variance σ_0^2 is

$$\begin{aligned}
\sigma_0^2 &= \begin{pmatrix} \frac{\partial h(\mu_G,\mu_B)}{\partial \mu_G} \\ \frac{\partial h(\mu_G,\mu_B)}{\partial \mu_B} \end{pmatrix}^T \begin{pmatrix} n/n_G & 0 \\ 0 & n/n_B \end{pmatrix} \begin{pmatrix} \frac{\partial h(\mu_G,\mu_B)}{\partial \mu_G} \\ \frac{\partial h(\mu_G,\mu_B)}{\partial \mu_B} \end{pmatrix} \\
&= \left[\tfrac{1}{2} - \tfrac{\ln C}{(\mu_G - \mu_B)^2}\right]^2 \tfrac{n}{n_G} + \left[\tfrac{1}{2} + \tfrac{\ln C}{(\mu_G - \mu_B)^2}\right]^2 \tfrac{n}{n_B}
\end{aligned}$$

The approximate distribution of $\hat{k}^* = h(\bar{y}_G^*, \bar{y}_B^*)$ is a normal distribution with mean $h(\mu_G, \mu_B)$ and variance σ_0^2/n.

A.5 The sign of $R'(k)$ is the same as the sign of the $d(k)$ function in (4.2)

$$\begin{aligned}
d(k) &= \ln C - k\left(\tfrac{\eta_G}{a_G(\phi_G)} - \tfrac{\eta_B}{a_B(\phi_B)}\right) \\
&+ \left(\tfrac{b_G(\eta_G)}{a_G(\phi_G)} - \tfrac{b_B(\eta_B)}{a_B(\phi_B)} - c_G(k,\phi_G) + c_B(k,\phi_B)\right) \qquad (4.2)
\end{aligned}$$

The first derivative of the payoff function is rewritten as follows.

$$R'(k) = -\varsigma f_G(k;\eta_G,\phi_G) + \gamma f_B(k;\eta_B,\phi_B) \qquad (A.14)$$

where $\varsigma = P_G r_G + P_G c_G, \gamma = P_B r_B + P_B c_B$

$$R'(k) = \varsigma f_B(k;\eta_B,\phi_B)\left(\frac{\gamma}{\varsigma} - \frac{f_G(k;\eta_G,\phi_G)}{f_B(k;\eta_B,\phi_B)}\right) = A\left(C - \frac{f_G(k;\eta_G,\phi_G)}{f_B(k;\eta_B,\phi_B)}\right)$$

where $A = \varsigma f_B(k;\eta_B,\phi_B)$, and $C = \frac{P_B(r_B+c_B)}{P_G(r_G+c_G)}$

Note that $A > 0$, the sign of $R'(k)$ depends on the sign of $\left(C - \frac{f_G(k;\eta_G,\phi_G)}{f_B(k;\eta_B,\phi_B)}\right)$. It is equivalent to check whether the sign of $\left(\ln C - \ln\left(\frac{f_G(k;\eta_G,\phi_G)}{f_B(k;\eta_B,\phi_B)}\right)\right)$ is positive or negative, and hence to check the sign of the below expression for an E_G-E_B model,

$$\ln C - k\left(\frac{\eta_G}{a_G(\phi_G)} - \frac{\eta_B}{a_B(\phi_B)}\right) + \left(\frac{b_G(\eta_G)}{a_G(\phi_G)} - \frac{b_B(\eta_B)}{a_B(\phi_B)} - c_G(k,\phi_G) + c_B(k,\phi_B)\right) \tag{A.15}$$

For an E-E model with $\phi_G = \phi_B$, the optimal cut-off point k_0, that satisfies $R'(k_0) = 0$, can be easily derived from (A.15).

$$k_0 = (\eta_G - \eta_B)^{-1}[a(\phi)\ln C + b(\eta_G) - b(\eta_B)] \tag{A.16}$$

A.6 Proof of Theorem 4.1

(i) $R'(k) \geq 0$ is equivalent to $C \geq \frac{f_G(k;\eta_G,\phi_G)}{f_B(k;\eta_B,\phi_B)}$. Taking logarithm on both sides, we get the following inquality,

$$\frac{k^2}{2}\left(\frac{1}{\sigma_G^2} - \frac{1}{\sigma_B^2}\right) - k\left(\frac{\mu_G}{\sigma_G^2} - \frac{\mu_B}{\sigma_B^2}\right) + \left[\ln C + \ln\left(\frac{\sigma_G}{\sigma_B}\right) + \frac{1}{2}\left(\frac{\mu_G^2}{\sigma_G^2} - \frac{\mu_B^2}{\sigma_B^2}\right)\right] \geq 0$$

Under the assumption that $\sigma_G^2 < \sigma_B^2$, the conditions that $R'(k) \geq 0$ is equal to the following conditions

$$k \leq k^{(1)} \quad or \quad k \geq k^{(2)}$$

(ii) Similarly, under the assumption that $\sigma_G^2 < \sigma_B^2$, the conditions that $R'(k) \leq 0$ is equal to the following conditions

$$k^{(1)} \leq k \leq k^{(2)}$$

(iii) From (4.8), $R'(k^{(1)}) = 0$. Under the assumption that $\sigma_G^2 < \sigma_B^2$, results (i) and (ii) suggest that $R'(k)$ changes from positive to negative around $k = k^{(1)}$, that is, $R'(k) > 0$ for $k < k^{(1)}$, and $R'(k) < 0$ for $k > k^{(1)}$. Therefore $R''(k^{(1)}) < 0$ and $R(k)$ obtains a local maximum at $k^{(1)}$.

Similarly, (iv), (v) and (vi) can be proved.

A.7 Proof of Theorem 4.3

(i) $R'(k) \geq 0$ is equivalent to $C \geq \frac{f_G(k;\eta_G,\phi_G)}{f_B(k;\eta_B,\phi_B)}$. Taking logarithm on both sides, we get

$$\ln C \geq \ln f_G(k;\eta_G,\phi_G) - \ln f_B(k;\eta_B,\phi_B)$$

For a Gamma-Gamma model with equal dispersion parameters, the above expression can be rewritten as

$$\ln C \geq k\left(\frac{\eta_G - \eta_B}{\phi}\right) - \left(\frac{b(\eta_G) - b(\eta_B)}{\phi}\right)$$

so, $$k \leq (\tfrac{\phi \ln C}{\eta_G - \eta_B}) + (\tfrac{b(\eta_G)-b(\eta_B)}{\eta_G-\eta_B}) = k_0$$

Similarly, (ii) can be proved.

(iii) It is easy to show that the derivative of function $d(k)$ in (4.2) is negative, which suggests that $d(k)$ is a strictly decreasing function of k for a Gamma-Gamma model with equal dispersion parameters. The formula (4.9) suggests $R'(k_0) = 0$. $R'(k)$ changes from positive to negative at k_0, then $R(k)$ has a local maximum at k_0.

A.8 Proof of equation (5.9)

$$R(k) = \varsigma T P_k - \gamma F P_k + d_0 \tag{A.17}$$

Since $\hat{f}(y) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i} s(\frac{y-l_i}{h_i})$ and $\hat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} S(\frac{y-l_i}{h})$, where $S(t) = \int_{-\infty}^{t} s(r)dr$. We get

$$\hat{T}P_k = \hat{P}(y_G^* > k | D = G) = 1 - \hat{F}_G(k)$$

$$\hat{F}P_k = \hat{P}(y_G^* > k | D = B) = 1 - \hat{F}_B(k)$$

Thus,

$$\hat{R}(k) = \varsigma \hat{T} P_k - \gamma \hat{F} P_k + d_0 = \varsigma(1 - \hat{F}_G(k)) - \gamma(1 - \hat{F}_B(k)) + d_0$$

$$= (\varsigma - \gamma + d_0) - \varsigma \hat{F}_G(k) + \gamma \hat{F}_B(k)$$

Because samples from the good and bad populations are independent, $Cov(\hat{F}_G(k), \hat{F}_B(k)) = 0$. So,

$$\begin{aligned} Var(\hat{R}(k)) &= \varsigma^2 Var(\hat{F}_G(k)) + \gamma^2 Var(\hat{F}_B(k)) \\ &= \varsigma^2 E(\hat{F}_G(k) - F_G(k))^2 + \gamma^2 E(\hat{F}_B(k) - F_B(k))^2 \end{aligned}$$

Note that in Azzalini's paper,

$$E(\hat{F}(x) - F(x))^2 \approx \frac{F(x)[1 - F(x)] - uh}{n} + vh^4$$

where $u = f(x)\left[h - \int_{-h}^{h} S^2(t)dt\right]$ and $v = \left[\frac{1}{2}f'(x)\int_{-h}^{h} t^2 s^2(t)dt\right]^2$

So,

$$E(\hat{F}_G(k) - F_G(k))^2 \approx \frac{\hat{F}_G(k)[1 - \hat{F}_G(k)] - u_G h}{n_G} + v_G h^4$$

where $u_G = \hat{f}_G(k)\left[h - \int_{-h}^{h} S^2(t)dt\right]$ and $v_G = \left[\frac{1}{2}\hat{f}'_G(k)\int_{-h}^{h} t^2 s^2(t)dt\right]^2$

Similarly,

$$E(\hat{F}_B(k) - F_B(k))^2 \approx \frac{\hat{F}_B(k)[1 - \hat{F}_B(k)] - u_B h}{n_B} + v_B h^4$$

where $u_B = \hat{f}_B(k)\left[h - \int_{-h}^{h} S^2(t)dt\right]$ and $v_B = \frac{1}{2}\left[\hat{f}'_B(k)\int_{-h}^{h} t^2 s^2(t)dt\right]^2$

Therefore, the variance of $\hat{R}(k)$ can be approximated by

$$\begin{aligned} Var(\hat{R}(k)) &= \varsigma^2\left[\frac{\hat{F}_G(k)[1-\hat{F}_G(k)]-u_G h}{n_G} + v_G h^4\right] + \gamma^2\left[\frac{\hat{F}_B(k)[1-\hat{F}_B(k)]-u_B h}{n_B} + v_B h^4\right] \\ &= \varsigma^2\left(\frac{\hat{F}_G(k)[1-\hat{F}_G(k)]}{n_G}\right) + \gamma^2\left[\frac{\hat{F}_B(k)[1-\hat{F}_B(k)]}{n_B}\right] \\ &+ \varsigma^2\left[\frac{-u_G h}{n_G} + v_G h^4\right] + \gamma^2\left[\frac{-u_B h}{n_B} + v_B h^4\right] \end{aligned} \tag{A.18}$$

A.9 Proof of Equations (5.11 and 5.12)

Proof of Equation (5.11)

By setting the first derivative of equation (4.1) be zero, $R'(k) = \varsigma\frac{\partial TP_k}{\partial k} -$

$\gamma \frac{\partial FP_k}{\partial k} = 0$, we get the slope of the ROC curve as follows,

$$\left.\frac{\partial TP_k}{\partial FP_k}\right|_{k=k^*} = \frac{\gamma}{\varsigma}$$

Because $\varsigma = P_G r_G + P_G c_G \geq 0$ and $\gamma = P_B r_B + P_B c_B \geq 0$, the range of slope is $[0, +\infty)$. When $k = k_1$, then $TP_k = FP_k = 1$ and $P_B = 0$, the slope is 0. When $k = k_2$, then $TP_k = FP_k = 0$ and $P_G = 0$, the slope is $+\infty$.

Proof of Equation (5.12)

The payoff function is rewritten as

$$TP_k = \frac{\gamma}{\varsigma} FP_k + \frac{R(k) - d_0}{\varsigma}$$

When FP_k is 0, the y-intercept of $L1$ is $y_1 = \frac{R(k)-d_0}{\varsigma}$;
and the y-intercept of $L2$ is $y_2 = \frac{(1-\psi)R(k)-d_0}{\varsigma} = (1-\psi)\left(\frac{R(k)-d_0}{\varsigma}\right) - \psi\frac{d_0}{\varsigma}$.
Hence we get result 2,

$$y_2 = (1-\psi)y_1 - \psi\frac{d_0}{\varsigma} \tag{A.19}$$

APPENDIX B

Notations

D - True Status of creditworthiness $= \begin{cases} G & Good\ creditworthiness \\ B & Bad\ creditworthiness \end{cases}$

Y - Decision variable of credit application $= \begin{cases} G & Approve\ credit\ application \\ B & Decline\ credit\ application \end{cases}$

G and B - Subscripts for good and bad creditworthiness

y^* - A continuous score to indicate creditworthiness, such as FICO, or a score derived from a non-binary credit-diagnostic test

$[y^*_{min}, y^*_{max}]$ - The range of creditworthiness y^*

y^*_G - Creditworthiness from the derived Good population

y^*_B - Creditworthiness from the derived Bad population

$\bar{y}^*_G$ - Average of creditworthiness from the derived Good population

$\bar{y}^*_B$ - Average of creditworthiness from the derived Bad population

k - Numerical threshold value of cut-off point

k_1 - Minimum value of numerical threshold in the range of $[y^*_{min}, y^*_{max}]$, i.e. $k_1 = y^*_{min}$

k_2 - Maximum value of numerical threshold in the range of $[y^*_{min}, y^*_{max}]$, i.e. $k_2 = y^*_{max}$

k_0 - Root to the first derivative of Profit Function (i.e. $R'(k_0) = 0$)

k^* - Numerical threshold value of optimal cut-off point

k^*_L - Low bound of near-optimal interval

k_U^* - High bound of near-optimal interval

D_0 - Discriminant of the estimation equation for k_0

TP_k - True Positive rate = $P(y^* > k|D = G) = P(Y = G|D = G)$. Given k, conditional probability of making the correct decision when an applicant has good creditworthiness

FP_k - False Positive rate = $P(y^* > k|D = B) = P(Y = G|D = B)$. Given k, conditional probability of incorrectly saying that an applicant has good creditworthiness

TN_k - True Negative rate = $P(y^* \leq k|D = B) = P(Y = B|D = B)$. Given k, conditional probability of making the correct decision when an applicant has bad creditworthiness

FN_k - False Negative rate = $P(y^* \leq k|D = G) = P(Y = B|D = G)$. Given k, conditional probability of incorrectly saying that an applicant has bad creditworthiness

t - false positive fraction variable, $t = FP_k \in [0, 1]$

$ROC(t)$ - Receiver operating characteristic curve at $t = FP_k$

c_G - Cost associated with a False Negative outcome

c_B - Cost associated with a False Positive outcome

r_G - Revenue associated with a True Positive outcome

r_B - Revenue associated with a True Negative outcome

P_G - Proportion of applicants having good creditworthiness

P_B - Proportion of applicants having bad creditworthiness

n - Total size of applicants

n_G - Size of applicants having good creditworthiness

n_B - Size of applicants having bad creditworthiness

$R(k) = P_G r_G TP_k + P_B r_B TN_k - (P_G c_G FN_k + P_B c_B FP_k)$ - Expected profit per outcome / decision / customer

$R(k^*)$ - Maximum value of $R(k)$ at the optimal cut-off point k^*

ψ - Allowance Ratio of Creditworthiness-equivalence

$C = \frac{P_B}{P_G} \frac{(r_B + c_B)}{r_G + c_G}$ - optimal ROC slope

Φ - Cumulative normal distribution function

ϕ - Standard normal density function

$N(\mu_G, \sigma_G^2)$ - Normal distribution of Good Population with mean μ_G and variance σ_G^2

$N(\mu_B, \sigma_B^2)$ - Normal distribution of Bad Population with mean μ_B and variance σ_B^2

$g(.)$ - Link function in Generalized Linear Model

ς - The constant $(P_G r_G + P_G c_G)$

γ - The constant $(P_B r_B + P_B c_B)$

$d(k)$ - A mathematically-convenient function of k used to determine the sign of $R'(k)$

$k^{(1)}$ - The lower root to a quadratic equation of k in the heteroscedastic Normal-Normal model

$k^{(2)}$ - The upper root to a quadratic equation of k in the heteroscedastic Normal-Normal model