DETERMINATION OF CUSTOMER SWITCHING USING LOGISTIC REGRESSION: CASE STUDYOF BARCLAYS BANK, GHANA

By

Vincent Agyei Addai (BSc) Agric Engineering

A Thesis submitted to the Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, in partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE
Institute of Distance Learning (IDL)

July 2011

Declaration

I hereby declare that this submission is my own work towards the MSc and that, to the best of my knowledge, it contains no material previously published by another person nor material which has been accepted for the award of any other degree of the university, accept where due acknowledgement has been made in the text.

VINCENT AGYEI ADDAI (20103285)	……………………	………………..
Student Name & ID	Signature	Date

Certified by:

MR KWAKU DARKWAH	**………………….**	**………………..**
Supervisor(s) Name	Signature	Date

Certified by:

MR KWAKU DARKWAH	**………………….**	**…………………**
Head of Dept. Name	Signature	Date

Prof. I.K. Dontwi	………………….	………………….
Dean, IDL	Signature	Date

ABSTRACT

The research is to construct a statistical model of customer switching intentions. It uses a binary logistic regression to model the impact of customer satisfaction, customer demographics on customer switching in Barclays Bank Ghana, and to predict and classify customer data. The model predicted that 18.97% of customers are likely to switch from the bank.

logit(P)(switching) = $0.140 + 0.300AET - 0.183SDS + 0.977EoCS + 0.814P$ - 0.10AGE

Results suggest that the most significant customer satisfaction constructs were Quality of staff, Efficiency of customer service, Electronic transactions and pricing. There was also evidence that customers' age groups and level of education contributed to explaining respondents' propensity to stay with their current banks.

TABLE OF CONTENTS

DECLARATION ii

ABSTRACT iii

LIST OF TABLE viii

LIST OF FIGURES ix

LIST OF MAJOR ACRONYMS x

ACKNOWLEDGEMENTS xi

DEDICATION xii

TABLE OF CONTENT

CHAPTER 1: INTRODUCTION

1.0 BACKGROUND 1

1.1 Profile of Barclays Bank – Ghana 4

1.2 Problem Statement 6

1.3 Research Objectives 6

1.4 Methodology 6

1.5 Justifications 7

1.6 Thesis Outline 7

CHAPTER 2: LITERATURE REVIEW

2.0 INTRODUCTION 8

2.1 The Evolution of Marketing 8

2.2 The Measures Defining Customer Relationship 12

2.2.1 The Customer Satisfaction-Retention-Loyalty Chain (SRLC) 12

2.2.2 The Behavioural and Financial Consequense of Service Quality 13

2.2.3 Customer Satisfaction (CS) 15

2.2.4 Customer Retention (CR) 17

2.2.5 Customer Loyalty (CL) 19

2.3 Marketing or Business Intelligence 21

2.4 The Link between CRM and Database Marketing 22

2.5 Customer as Decision Markers 23

2.6 Customer Value 24

2.7 Customer Segmentation 25

2.8 Customer Activity Measurement 27

2.9 Competitive Advantage 28

2.10 Corporate Image 29

2.11 Switching Barriers 30

CHAPTER 3: METHODOLOGY

3.1 Binary Logistic Regression Analysis 31

3.2 logistic regression 31

3.2.1 The Logistic function 31

3.2.2 The Logistic model 33

3.2.3 Logit Transformation 34

3.3 Measurement Requirement 35

3.4 Assumptions 35

3.5 Sample Size Considerations 36

3.6 Methods for Including Variables 36

3.7 Method For Including Variables 37

3.7.1 Maximum Likelihood Estimation 37

3.7.1.1 Likelihood Equations 40

3.7.1.2 Newton – Raphson Iteration Method 41

3.8 Statistical Inference Using Maximum Likelihood Techniques 43

3.9 The Likelihood Ratio Test 44

3.10 How The LR Test Works 45

3.11 Wald Test 46

3.12 Overall Model Fit 47

3.13 Interpreting Parameters in Logistic Regression 47

3.13.1 Odds 47

3.13.2 Odds Ratio 48

3.13.2.1 Definition 48

3.13.2.2 Derivation of OR Formula 49

3.13.2.3 Properties of the Odds Ratio 50

3.13.3 Regression Co-efficient ($\boldsymbol{\beta}_i$) 51

3.13.4 The Intercept (α) 52

3.13.5 The P – Value 53

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Data Collection 54

4.2 Summary of Data 54

4.2.1 Customer's Demographic and Background Information 54

4.3 Computation 57

4.4 Results 57

4.4.1 The Impact of Customer Satisfaction on Customer Switching Intentions 57

4.4.2 Prediction of Customer satisfaction, Customer's demographic on Customer switching intentions 64

4.5 Discussion 65

4.5.1 Effect of B-values in relation to Customer Switching Intention to stay 65

4.5.2 Effect of Exp B – Odds Ratios (OR) to Customer's Intentions to stay 66

CHAPTER 5: CONCLUSION AND RECOMMENDATION

5.1 Conclusions 68

5.2 Recommendations 68

LIST OF REFERENCES 70

APPENDIX A. Questionnaire Agenda 79

APPENDIX B. SPSS output for customer satisfaction 80

APPENDIX C. SPSS output for customer demographics 81

APPENDIX D. Results of Questionnaire Agenda 82

LIST OF TABLES

Table 2.1 Transaction approach and relationship approach 17

Table 4.1 Demographics of Respondents 55

Table 4.2 Results of the Questionnaire Agenda 56

Table 4.3 logistic regression estimates of the impact of customer attributes on customer switching behaviours 58

Table 4.4 Classification table 59

Table 4.5 Omnibus Test of Model Coefficients 59

Table 4.6 Hosmer and Lemeshow Test 60

Table 4.7 Model Summary 60

Table 4.8 Classification Table for model 61

Table 4.9 Values for Predicted Variables 64

LIST OF FIGURES

Figure 1.1 A typical customer behaviour model 4

Figure 3.1 A graph depicting Nonlinear nature of the logistic regression function 32

Figure 3.2 A graph depicting z-index of combined risk factor of a logistic function 33

Figure 4.1 Predicted probabilities of a customer switching 63

LIST OF MAJOR ACRONYMS

Notation	Explanation
B2B	Business-to-Business
CEM	Customer experience management
CFA	Confirmatory Factor Analysis
CIT	Critical Incident Technique
CL	Customer Loyalty
CLV	Customer Lifetime Value
CR	Customer Retention
CRM	Customer Relationship Management
CS	Customer Satisfaction
IPA	Importance-Performance Analysis
KPIs	Key Performance Indicators
LNP	Local Number Probability
LTV	Lifetime Value
MI	Marketing Intelligence
ML	Maximum Likelihood
SPSS	Statistical Package for the Social Sciences
NPV	Net Present Value

ACKNOWLEDGEMENTS

To the king immortal and only wise God, do I give thanks and praise for, this and many blessings bestowed upon me during the pursuance of this programme, to realize this work.

I would like to thank my supervisor Mr. Kwaku Darkwah for his enthusiastic support and advice, patience and constant energy for idea sharing throughout the research effort. His influence is inherited in both the theoretical and practical aspect of this work.

In terms of exchange of ideas, support, criticisms and intellectual stimulation, thanks are offered to Dr. Samuel Amponsah and my colleaque Mr. Daniel William Oteng .

Finally, I owe special thanks to my family because of their constant and invaluable support all along the two years MSc program. I could not have done it without you all.

DEDICATION

To my lovely wife, Mrs Bernice Agyei Addai

CHAPTER 1

INTRODUCTION

1.0 BACKGROUND

The banking industry is highly competitive, with banks not only competing among each other; but also with non-banks and other financial institutions (Kaynak and Kucukemiroglu, 1992; Hull, 2002). Most bank product developments are easy to duplicate and when banks provide nearly identical services, they can only distinguish themselves on the basis of price and quality. Therefore, customer retention is potentially an effective tool that banks can use to gain a strategic advantage and survive in today's ever-increasing banking competitive environment.

The majority of banks in Ghana has non-domestic owners, and is not very diversified in terms of the products and services they offer (Ghana Banking Survey, 2009). This suggests that the Ghana banking industry has reached the maturity phase of the product lifecycle and has become commoditized, since banks offer nearly identical products. This carries the danger of creating a downward spiral of perpetual price discounting -- fighting for customer share (Mendzela 1999). One strategic focus that banks can implement to remain competitive would be to retain as many customers as possible.

The argument for customer retention is relatively straightforward. It is more economical to keep customers than to acquire new ones. The costs of acquiring customers to "replace" those who have been lost are high. This is because the expense of acquiring customers is incurred only in the beginning stages of the commercial relationship (Reichheld and Kenny, 1990). In addition, longer-term customers buy more and, if satisfied, may generate positive word-of-mouth promotion for the company. Additionally, long-term customers also take less of the company's time and are less sensitive to price changes (Healy, 1999). These findings highlight the opportunity for management to acquire referral business, as it is often of superior quality and inexpensive to obtain. Thus, it is believed that reducing customer defections by as little as five percent can double the profits (Healy, 1999). The key factors influencing customers' selection of a bank include the range of services, rates, fees and prices charged (Abratt and Russell, 1999). It is apparent that superior service, alone, is not sufficient to satisfy customers. Prices are essential, if not more important than service and relationship quality. Furthermore, service excellence, meeting client needs, and providing innovative products are essential to succeed in the banking industry. Most private banks claim that creating and maintaining customer relationships are important to them and they are aware of the positive values that relationships provide (Colgate et al., 1996).

While there have been several studies emphasising the significance of customer retention in the banking industry (see Dawkins and Reichheld, 1990; Fisher, 2001; Marple and Zimmerman, 1999; Page, Pitt, and Berthon, 1996; Reichheld and Kenny, 1990), there has been little empirical research examining the constructs that could lead to customer retention. This paper examines the constructs that impact consumers' decision to stay with or leave their current banks in Ghana. In addition, the paper explores whether there is any association between consumers' demographic characteristics (i.e. age, gender, educational level and income) and loyalty decisions.

Reichheld and Sasser (1990) propose the concept of service profit chain (SPC) which links service quality, customer behaviours and profitability. The SPC concept argues that customer satisfaction is influenced by the value of service quality, which in turn influences customer retention (repurchase and cross-selling) and customer loyalty (word-of-mouth or referral). Consequently, profitability is stimulated by customer retention and loyalty. The concept of service quality would be well established in the marketing literature and several frameworks have been developed (Parasuraman *et al.*, 1988).

Previous research found that there is a strong and positive relationship between service quality attributes and customer satisfaction (Rust and Oliver, 1994; Fornel *et al.* 1996). However, there is also little consensus among experts to explain the relationship between service quality attributes and customer satisfaction.

Finding the critical service attributes that determine customer satisfaction and customer dissatisfaction can lead firms to seek comprehensive strategies for achieving lasting competitive advantage (Matzler *et al.*, 2004). Moreover, customer satisfaction plays as mediating attitude between service quality attributes and customer behaviours (retention and loyalty). A typical customer behaviour model is shown in Figure 1.1. Customer satisfaction may increase the retention of customers through repeated and increased purchase (long-term relationship). Customer satisfaction may also positively affect customer loyalty (word-of-mouth). The combination of improved customer retention and loyalty may in turn increase profitability (Manrodt and Davis, 1993; Emerson and Grimm, 1998).

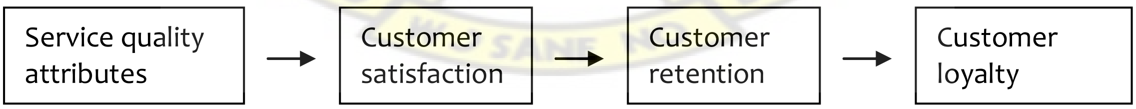


Figure 1.1: A typical customer behaviour model

The marketing literature on customer relationship or behaviour outlines potential benefits available to customers and suppliers for their strategic management and business performance. The literature calls for establishing relationships in order to build trust and loyalty, develop long-term strategies, and to be pro-active to customer needs (Fornell and Lehman, 1994; Anderson *et al.*, 1999). Some of the existing empirical studies seem to lack the necessary theoretical and analytical rigour, and this is seen as a pressing requirement for future customer behaviour analysis (Matzler and Sauerwein, 2002).

1.1 PROFILE OF BARCLAYS BANK - GHANA

Barclays has operated in Ghana for ninety four years. It is wholly owned subsidiary of Barclays Bank PLC. Its vision is to become the best bank in Ghana, making lives much easier for its customers, employees and other stakeholders.

Barclays Bank of Ghana Limited has an extensive retail and commercial banking network in the country, comprising 74 branches, 7 agencies, 10 premier life centres, 2 premier suites and 135 ATMs in all regional capitals and major towns. Barclays offers a wide range of Commercial, Retail and Treasury products and services targeted particularly at business and corporate clients, while extending personal products and services to retail customers. Barclays Ghana also offers local business banking products and services to SMEs and indigenous businesses.

The Bank's premier Banking offers tailor made solutions and one-on-one banking to its high net worth customers. The premier and premier life proposition amongst others offer dedicated banking suites, financial planning, lifestyle alliances, global access to premier lounges (airports) and many other benefits to premier and premier life clients.

The Barclays international Banking unit, the first in Ghana and in Africa south of the Sahara, continues to offer world class banking service to non-resident private clients and corporates.

In June 2010, Barclays Ghana launched its SMS Alerts service which notifies retail customers of any transactions on their accounts such as withdrawals, deposits, etc to enable customers have up-to-the minute awareness of transactions on their account.

At Barclays Ghana, we refer to a three way win when it comes to our community investment. This is in line with the community sustainability programme of the Barclays group. The three areas are Banking on the Brighter Futures, Charity Begins at Work and Looking after local communities. The sustainability agenda is supported with 1% of the Bank's pre-tax profits and has more than 80% of staff involved in voluntary programmes across the country on a yearly basis.

Barclays Bank of Ghana Limited is part of the Barclays Global Retail Banking (GRB) group under Barclays PLC. Barclays is a major global financial services provider engaged in retail banking, credit cards, corporate banking, investment banking, wealth management and investment management services, with an extensive international presence in Europe, the Americas, Africa and Asia. With over 300 years of history and expertise in banking, Barclays operates in over 50 countries and employs approximately 144,000 people. Barclays moves, lends, invests and protects money for over 48 million customers and clients worldwide.

1.2 PROBLEM STATEMENT

Clearly, there are compelling arguments for bank management to carefully consider the factors that might increase customer retention rates. Several studies have emphasised the significance of customer retention in the banking industry (see Dawkins and Reichheld, 1990; Marple and Zimmerman, 1999; Page et al., 1996; Fisher, 2001). However, there has been little effort to investigate factors that might lead to customer retention. Most of the published research has focused on the impact of individual constructs, without attempting to link them in a model to further explore or explain retention. If retention criteria are not well managed, customers might still leave their banks, no matter how hard bankers try to retain them.

1.3 OBJECTIVE

The objectives are;

1. To use a binary logistic regression to model customer switching intensions.
2. To use the logistic regression model to predict and classify customer data.

1.4 METHODOLOGY

The research is to construct a statistical model of customer switching intentions. Data are collected by face to face interview and analysed by standard statistical technique to establish relationships between variables. The research survey instrument was a self administered questionnaire. This was achieved by extracting the data about key service attributes from combinations of surveys, and questionnaire. Respondent data was analysed using statistical package for social sciences (SPSS) and a model developed with logistic regression, logistic regression with dummy variables to test variables and constructs. Libraries and Internet were used to get vital information on related research and publications.

1.5 JUSTIFICATIONS

Understanding the impact of service attributes on customer satisfaction and demographics can help decision makers in the banking industry within resource allocation process which when effectively carried out will reduce cost in acquiring new customers and increase profitability. Profitability may result in new businesses (expansions) contributing to the growth of the economy, society and creating job opportunities for the youth. This research contributes to knowledge in the academia.

1.6 THESIS OUTLINE

Chapter 1 presents the background of the study, profile of Barclays, problems statement, objective of the research, methodology, and justifications.

In Chapter 2 reviewed literature of analytical customer relationship management.

Chapter 3 discusses the research approach and methods undertaken in this thesis.

Chapter 4 presents data analysis and results.

Chapter 5 presents conclusions and recommendations.

CHAPTER 2

LITERATURE REVIEW

2.0 INTRODUCTION

This chapter is a review and appraisal of the literature supporting the research objectives. It examines the search dedicated to service quality and customer behaviours as a major factor in the corporate decision making and strategic planning processes. The material in this chapter focuses on relationship marketing and management science.

2.1. The Evolution of Marketing

During the industrialisation era of the 1920s, the marketing theory pointed particularly to mass marketing because of the nature of mass manufacturing and inception of mass marketing use (radio). The concept continued to expand through the 40s and 50s. It gave corporations an opportunity to approach a wide customer with different needs into buying the same product. Mass manufacturing created a gap between firms and customers. From the firm's perspective, customisation was not economically viable and did not promise greater profits. In addition, individual customer data was not available and there was often very little to almost no interaction between the customer and the firm. Moreover, firms were not open to customer-feedback. Therefore, there was a lack of understanding about the customer service or their needs from the product apart from functionality and durability.

Services marketing pioneers proposed the concept of relationship marketing as means to narrow the gap between companies and their customers.

Leonard Berry (1983) was the first scholar in services marketing who coined the phrase "relationship marketing". However, the concept had been oriented towards how to acquire customers (Storbacka *et al.*, 1994). As a result, such relationships are not necessarily long term relationships where profitability is the main goal of the relationship. The phrase became popular in the late 1980s and early 1990s due to the shift of focus from customer acquisition to customer retention (Morgan and Hunt, 1994; Sheth and Kellstadt, 2002). By comparing relationship marketing (RM) with the traditional transaction marketing, the following can be derived:

1. In relationship marketing the focus is not on service encounters or transactions.

2. Relationship Marketing is focused on retaining customers and enhancing the relationship with the customers.

Reichheld and Sesser, (1990); Shani and Chalasni; (1992) said there are also other accounts for the emergence of relationship marketing, such as the economics of customer retention, the ineffectiveness of the mass media, and higher expectations from customers. Furthermore, Sheth and Kellstadt (2002) categorise the main reasons for the emergence of relationship marketing:

1. The energy crises of the 1970s and economic inflation,

2. Emerging of service marketing, and

3. Supplier partnering.

Later, they also mentioned two other factors that influenced the course and definition of relationship marketing as:

1. Impact of internet and information technology (IT)

2. Selective and targeted relationship (customer segmentation and customisation)

In the past thirty years, there has been a significant number of research and practices in the marketing that have focused on the importance of relationships, networks and interactions. As a result, theories have emerged that contribute to the traditional marketing management. Service marketing and the network approach to business-to-business (B2B) had relatively more than impact on marketing development rather other theories. There were also influences from non-marketing areas such as total quality management (TQM), lean production, customer value chain, balanced scorecard, intellectual capital and organisation theory that further enriched relationship marketing.

Ford, (1980); Christopher *et al.*, (1991); Gummesson, (1991); Lindgreen *et al.*, (2004) in their research said, the concept of the relationship marketing (RM) emerged within the fields of services marketing and industrial marketing. The concept emphasises on customer satisfaction and customer retention as the long-term value for the firm (defensive marketing) rather than customer transactions (offensive marketing) (Kotler, 1991; Varva, 1992). In other words, defensive marketing focuses on reducing customer defection (churning) and increase customer loyalty, whereas offensive marketing focuses on obtaining new customers and increase customers purchase frequency (Fornell and Wernerfelt, 1987). Nowadays, relationship marketing (RM) is considered as a strategy (Berry, 1983; Gummesson, 1993) in which it aims to enhance customer relationship
and profitability (Grönroos, 1994; Storbacka *et al.*, 1994; Rap and Collins, 1990; Blomqvist *et al.*, 1993).

Saren (2007) defines customer relationship (CR) as ‘the creation, maintenance and reproduction of tastes, dreams, aspirations, needs, identities, desires, morality and hedonism'. The concept of RM received considerable criticism, at the beginning, but it is acknowledged that it has made a shift in marketing. According to Gruen (1997):

" the introduction of the relation marketing concept focused business on seeing customers as the centre of the universe and the organisation around them … RM reorients the positions of suppliers and customers through a business strategy of bringing them together in co-operative, trusting and mutually beneficial relationships**."**

According to Bose, (2002); Ahn *et al.*, (2003) said, companies were expecting to gain more market share by shifting to customer orientation from the traditional practices. More importantly, emergence of the One-to-One and the Customer Relationship Management (CRM) concept highlighted the difference between customers; hence attention needs to be paid to how they perceive added value service attributes (Weitz *et al.*, 1995). RM relies upon the acquisition of customer needs and desires with particular relevance to customer satisfaction which, in turn, leads to long-term relationship. According to Gummeson (2008) "Relationship Marketing is the overriding concept for a new marketing type of marketing and Customer Relationship Management as techniques to handle customer relationships in practice." Moreover, He defines Customer Relationship Management as:

"Customer Relationship Management is the values and strategies of Relationship Marketing – with special emphasis on the relationship between a customer and a supplier – turned into practical application and dependent on both human action and information technology."

Despite the advantages that Relationship Marketing offers, practitioners and academics have yet to propose a roadmap to create sustainability and competitive advantages that Relationship Marketing promises to offer (Ganesan, 1994; Morgan and Hunt, 1994). Therefore, it is important to recognise how the competitive advantages can be built through relationship marketing

2.2 The Measures Defining Customer Relationship

In businesses where the underlying products have become commodity-like, quality of service depends heavily on the quality of its personnel. This is well documented in a study by Leeds (1992), who documented that approximately 40 percent of customers switched banks because of what they considered to be poor service. Leeds further argued that nearly three-quarters of the banking customers mentioned teller courtesy as a prime consideration in choosing a bank. The study also showed that increased use of service quality/sales and professional behaviours (such as formal greetings) improved customer satisfaction and reduced customer attrition.

2.2.1 The Customer Satisfaction-Retention-Loyalty Chain (SRLC)

According to Heskett *et al.,* (1994), satisfaction-retention-loyalty-chain (SRLC) is a key concept that needs to be understood due to its link to customer relationship management (CRM) and, in turn, profitability. The concept has been popular since the early 1990s, when measuring and managing customer satisfaction became important to companies. The key point is that improving the performance of service attributes will generate satisfaction (Mousavi *et al.*, 2001). Increased customer satisfaction levels will lead to greater customer retention rate, which is a key determinant for customer loyalty, which may increase the expected profit (Rust and Zahorik, 1993; Anderson and Mittal, 2000). Despite the self-evident nature of these positive links, the empirical evidence of research shows only mixed support (Zeithmal, 2000). There is a lack of research investigating the relationship between perception measures (service attribute quality, customer satisfaction) and action measures (word-of-mouth behaviour, purchase loyalty and long term customer relationship profitability)

2.2.2 The Behavioural and Financial Consequences of Service Quality

According to Reichheld and Sasser, (1990); Parasuraman *et al.*, 1985; Dawkins and Reichheld, (1990), provision of a good quality of service is considered as a key to success in today's competitive business environment. During the 1980s, the primary emphasis of organisations was focused on improving service quality towards customer expectations (Parasuraman *et al.*, 1985). As a result, several methodologies and management framework were proposed (Zeithaml *et al.*, 1996) such as: total quality management (TQM); quality function deployment (QFD); failure modes and effects analysis (FMEA); six sigma (zero defect); PDCA (plan-do-check-act) or Deming cycle. However, there is no consensus on the way to estimate the impact of service quality on financial performance (Zeithaml *et al.*, 1996; Rust *et al.*, 1995). The relationship between these two variables is neither straightforward nor simple (Zahorik and Rust, 1992). Research on the direct relationship between customer satisfaction and profitability has revealed mixed results ranging from positive to no effect (Christopher *et al.*, 1998; Zeithaml, 2000; Jones and Sasser, 1995). The findings lack in depth analysis and fail to answer questions like: How will service quality attribute be paid off (return on investment)? Or, how much should the company invest in service quality to maximise profitability?

Fornell and Wernerfelt, (1988); Rust and Zahorik, (1993); Zahorik and Rust, (1992), said there are two approaches for addressing these questions: offensive marketing and defensive marketing. Such approaches do not have their roots in either industrial or service marketing but have emerged from the traditional consumer goods marketing

(Storbacka *et al.*, 1994). Offensive marketing focuses on acquiring new customers and increase customers' transactions (purchase frequency), whereas defensive marketing is focused on minimising customer switching behaviour. This thesis evaluates the defensive impact of service quality through customer retention in order to measure the financial impact of service quality for the firm.

According to Fornell and Wernerfelt (1987, 1988); Reichheld and Sasser (1990); Anderson and Sullivan (1990); Grönroos, (1990), the basic assumption is that there is a direct and strong relationship between service quality attributes and customer behaviours, for instance; repurchase intention. The assumption is based on the idea that customer satisfaction can be predicted and assessed as the difference between perception and expectation. Therefore, if the service is performed poorly, then the difference between customer perception and expectation will be negative or the customer will be dissatisfied. If the difference is positive, a customer will be satisfied or desired. Moreover, this relationship is relied upon the assumption that the relationship between service quality attributes and customer satisfaction is linear and asymmetric.

Goodman and associates (1995) in their work said most customer satisfaction programs, the relationship between service attributes performance and customer satisfaction is assumed linear and symmetric. However, there are some other studies that explain the non-linear and asymmetric relationships, For example, Mitall and Baldasare (1996) in health care; Danaher (1998) in airline industry; Mittal, Ross and Baldasare (1998) in automotive industry; Bolton and Lemon (1999) in entertainment, and Kumar (1998) in business-to-business marketing that explain the relationship between performance of service attributes and customer satisfaction.

Research reveals that there is a significant difference between the key drivers of customer satisfaction and dissatisfaction (Shiba *et al.*, 1993; Dutka, 1993; Gale, 1994; Oliver, 1997).

According to two-factor theory of Herzberg (1959), job satisfaction factors can be classified into two groups: 'motivators' (increase job satisfaction) and 'hygiene factors' (prevent dissatisfaction). Two-factor theory has also been adopted in marketing theory, where multi-attribute models are used to understand the construct of customer satisfaction. These models imply that service attributes do not have the same importance from customer perspective. In the context of customer satisfaction, the impact of low attribute-level performance on overall satisfaction is greater than attributes with high performance (Mittal *et al.*, 1998). This relationship has explained through prospect theory (Kahneman and Tversky, 1979) which describes how individuals form decisions and react to losses and gains.
However, later studies developed the three-factor theory (e.g., Anderson and Mittal, 2000; Matzler and Sauerwein, 2002).

As a result, service and product attributes fall into three groups: basic, performance and exciting attributes (the three-factor theory). The theory originally was developed by Kano (1984) based on Herzberg" s two-factor theory.

2.2.3 Customer Satisfaction (CS)

According to Blanchard and Galloway, (1994); Heskett *et al.*, (1990) relative to the expectation (Zeithaml *et al.*, (1990), customer satisfaction is the result of a customer's perception of the service quality.

Moreover, Looy et al. (2003) defines customer satisfaction as:

"The customer's feeling regarding the gap between his or her expectations towards a company, product or service and the perceived performance of the company, product or service."

Both the service management and marketing literature suggest that there is a strong relationship between customer satisfaction, customer behavioural intentions (e.g., switching and word-of-mouth) and, in turn, profitability (Yi, 1990). By improving product and service attributes performance, customer satisfaction level should increase (Mittal *et al.*, 1998; Wittink and Bayer, 1994) which, in turn, lead to greater customer retention (Zeithaml *et al.*, 1996; Anderson 1994). Accordingly, improved customer retention generates more profit (Anderson and Mittal, 2000). Despites it importance, there seems to be little experimental research that quantifies the complex relationships.

Customer satisfaction can be interpreted as an overall evaluation of service quality attributes or service attribute performance (Fornell *et al.*, 1996; Johnson and Fornell, 1991; Boulding *et al.*, 1993). Several studies discussed the relationship between two constructs of service attribute performance and overall customer satisfaction (Anderson and Sullivan, 1993; Oliva *et al.*, 1995; Oliver, 1993; Mittal *et al.*, 1998). It is argued that the relationship in most cases is nonlinear and asymmetric. More importantly, there is a strong relationship between customer satisfaction and customer future intentions (e.g. retention) and profitability (Anderson and Sullivan, 1993; Bearden and Teel, 1983; Boulding et al., 1993; Oliver, 1980; Yi, 1990; Rust *et al.*, 1994). Such comprehensive approaches to model the customer relationship profitability are lacking, as most studies have only focused on discrete aspects of the conceptual framework.

2.2.4 Customer Retention (CR)

According to Reichheld and Sasser, (1990); Reichheld *et al.*, (2000), since 1990s the subject of customer satisfaction and customer retention, and their relationship with company's financial performance has become the core of attention for many managers. By interpreting customer behaviours like retention to profit, firms move closer to the inter-dependent variable – profitability. In addition, the marketing domain has increasingly shifted from transactional approach (the value of an individual sale) to relationship marketing approach (the value of long-term relationships and repeat purchases). Table 2.1 presents the shift from transactional marketing to relationship marketing. More important, relationship marketing acknowledges that existing and new customers require different strategies.

Table 2.1: Transaction approach and relationship approach (Adopted from Peck *et al.* 2000, p. 44)

Characteristics	Transactions focus	Relationships focus
Focus	Obtaining new customers	Customer retention
Orientation	Service features	Customer value
Timescale	Short	Long
Customer service	Little emphasis	High emphasis
Customer commitment	Limited	High
Customer contact	Limited	High
Quality	An operations concern	The concern of all

Research in this area revealed that there is an asymmetric and non linear relationship between customer satisfaction and customer retention. Even though, customer dissatisfaction may have a greater impact on retention than customer satisfaction. It should be noticed that a number of factors such as type of industry, market competition, switching costs and risk factors may change the dynamics between customer satisfaction and retention (ACSI).

Retention and defection are like two sides of the same coin. Retention rate can be defined as the average likelihood that a customer repurchases product/service from the same firm. The defection or churning rate is defined as the average likelihood that a customer switches or defects from the company to another company, see Equations 2.1 and 2.2.

$$\text{Retention rate (\%)} = 1 - \frac{1}{Average\ Lifetime\ \ Duration} \qquad (2.1)$$

Average defection rate (%) = 1 – Average retention rate (2.2)

Lowering customer switching rates can be profitable to companies. Research confirmed that retaining customers is a more profitable strategy than acquisition of new customers.

Further, Reichheld and Sasser (1990) emphasis on zero customer defections (churning) as an overall performance:

“Ultimately, defections should be a key performance measure for senior management and a fundamental component of incentive systems. Managers should know the company’s defection rate, what happens to profits when the rate moves up or down, and why defections occur.”

Anderson and Sullivan, (1990); Reichheld and Sasser (1990) studied the financial impact of customer retention based on two assumptions. First, acquiring new customers is more expensive than retaining existing customers as it involves advertising, promotion and start-up operating expenses. New customers, therefore, are more likely to be unprofitable for a period of time after acquisition. Second, existing customers are more likely to generate more profit to companies through cross-selling and word-of-mouth.

A study from Rose (1990) reveals that a customer that retain with company minimum 10 years is on average three times more profitable than a customer with 5 years customer history.

2.2.5 Customer Loyalty (CL)

Marketing literature uses a wide range of terms to describe loyalty and methods to measure it. Andreassen and Lindestad (1998) defined loyalty as "an intended behaviour caused by the service and operationalised loyalty as a repurchase intention and willingness to provide positive word-of-mouth".

Jones and Sasser (1995) according to their research found customer satisfaction as the key element in securing customer loyalty. Customer loyalty has been described in service management and marketing literature. The service management literature defines loyalty as the behaviour that can be seen in various forms such as relationship continuance, cross-selling, up-selling, and word of mouth or customer referral (recommendation). This type of behaviours increase profitability through enhances revenues reduced costs to obtain new customers and retained existing customers, and lower customer-price sensitivity.

According to Jacoby and Kyner, (1973), marketing literature has defined customer loyalty into distinct ways. The first defines customer loyalty as an attitude which indicates an individual's overall attachment to a product, service, or brand. The second defines loyalty as behaviour that can be evaluated in form of repurchase, word of mouth, and increasing the scale and scope of a relationship. However, the behavioural view of loyalty is similar from both service management and marketing point of view. In this thesis, we examine the behavioural rather than attitudinal loyalty (word of mouth). This approach is intended to, first, to include behavioural loyalty in the conceptualisation of customer loyalty that has been linked to customer retention (switching intention) and satisfaction, and second, to make the demonstrated service quality attributes- customer satisfaction-retention-loyalty relationship providing managers and decision makers interested in customer behaviours linked to firm performance.

Despite of several studies into customer loyalty, there is no consensus on the most appropriate way to measure loyalty. Existing studies in customer loyalty can be classified into three groups regardless of definition, measurement, and limitation. These three groups are: (1) loyalty as repeat purchase and word of mouth behaviour (Liljander and Strandvik, 1993), (2) loyalty as a combined composite of repeat patronage and attitudinal component (Dick and Basu, 1994), and (3) a psychological prospect of loyalty (Czepiel, 1990). In this study, customer loyalty is defined as customer word of mouth (WOM) behaviour. Jones and Sasser (1995) discuss that WOM is one of the most important factors in acquiring new customers.

Despite the benefits that accrue from WOM, many organisations can not yet link the service quality-customer satisfaction to WOM. This is due to the fact that satisfaction plays as a mediating attitude between service quality attributes and customers‟ word of mouth. More importantly, customer retention is not the same as customer loyalty. Customer retention rate is measured on a period-by-period basis and it is used as an indication of customer switching behaviour or intention, whereas customer loyalty has a much stronger theoretical meaning. If a customer is loyal toward a service or a brand, he or she has a positive emotional or psychological disposition towards this brand. Customers might continue to purchase a particular brand but this may be purely out of convenience or inertia. In this case, a customer may be retained, but not necessarily stay loyal to the product or service.

2.3. Marketing or Business Intelligence

According to Lee and Trim, (2006), companies need to develop and sustain long-term working relationship with their customers. In doing so, companies need a systematic process of gathering, analysing, supplying and applying information about the external market and internal environment. As a result, marketing or business intelligence plays a significant role in the formulation and implementation of plans to achieve this goal. Marketing intelligence supports the decision-making process by providing external (e.g., customer needs) and internal data from the environment (e.g., employee loyalty).

Cornish (1997) defined marketing intelligence as:

"the process of acquiring and analysing information in order to understand the market (both existing and potential customers) to determine the current and future needs and preferences, attitudes and behaviour of the market; and to assess changes in the business environment that may affect the size and nature of the market in the future."

In reality, most businesses rely on conjecture to evaluate the efficiency of their processes, whereas it is hard to make decisions without objective about how to improve business performance. As a result, the analytical result of customer value has received lots of attentions as a force for competitive differentiation.

According to analyst firm IDC (2006), the business intelligence market is a $20 billion market. Business intelligence has changed dramatically since its inception in the early 1990s.

2.4 The Link between CRM and Database Marketing

Since the significant transformation in areas of information technology (IT) and the internet, and the improvement in flexible manufacturing and outsourcing practices, understating individual customer needs has become a key determinant of a company's profitability.

This shift in marketing direction can be viewed in the definition of marketing that was updated by the American Marketing Association (2004), to be:

"Marketing is an organisational function and a set of processes for creating, communicating, and delivering value to customers and for managing customer relationships in ways that benefit the organisation and its stakeholders."

Therefore, marketing plays an important role in aligning company's business processes and practices with customers' demand. Traditionally, database marketing provides valuable information about customers by identifying and analysing different segments of customer population. This provides the opportunity for firms to increasingly disaggregate the levels of database marketing to ultimately reach their customers. Thus, Customer Relationship Management applies database marketing techniques at the customer level to strengthen company-customer relationships.

The shift from transactional marketing to relational marketing has dramatically raised the importance of evaluation of the long-term economic value of a customer for the company. The concept of customer value refers to the present value of the future cash flows attributed to the customer relationship. Customer value is the economic value of the customer relationship to the company. Use of customer value as a marketing metric tends to redirect the forms of strategic planning towards long-term customer relationship, rather than maximising short-term sales.

2.5 Customers as Decision Makers

The main objective of modern companies involves measuring the quality of customer relationship rather than track product releases to project profit and the number of transactions. Customers are not concerned with the amount of profit they are generating for the company; they rather expect the company to meet their needs. In other words, a customer cares about the quality of the relationship he has with the company.

According to Yastrow (2007), “relationships have become powerful differentiators.” More importantly, he argues that companies should enhance personal relationships with their customers.

According to Rust and Zahorik, (1993), the chain of impact of the performance of service attributes on customer satisfaction, and consequently its impact on customer retention and loyalty, leading to profitability. However, there is a lack of studies investigating the relationship between customer perception and customer future intentions, i.e. purchase volume, length of association and word-of-mouth. Such analysis helps managers to estimate customer migration, and assign resources accordingly.

2.6 Customer Value

Kurma and Reinartz, (2005) did a research on customer value and found that, in order to implement long-term strategy, the management needs to know how the value of a customer evolves over time. To do so, corresponding control measures must be put in place. Lifetime value (LTV) is the general term used to describe the long-term economic value of a customer. In simple terms, customer value implies the fact that each customer has a value over his/her lifetime with a firm. Estimating, however, the lifetime of a customer by itself requires sophisticated modelling, as it involves prediction of the probability of retention. More importantly, the inputs of the lifetime value can change subject to nature of product or service, data availability, and analysis capability. Therefore, the formulation should be adapted based on the type of industry and company attributes. For example, contractual relationship such as mobile phone subscription needs a different formulation vis a vis non-contractual relationship such as the airline industry.

In theory, customer value represents the amount of profit generated from each customer, and therefore it should be willing to spend money to acquire or retain each customer. However, calculating customer value is very difficult due to its complexity and the uncertainty surrounding customer relationships. In order to calculate customer value, the following parameters are required:

- Churn rate: is the percentage of customers who end their relationship with a company in a given period. Therefore, one minus the churn rate is the retention rate.

- Discount rate: is the cost of capital used to discount future revenue from a customer.

- Retention cost: is the amount of money has to be spent in a given period to retain an existing customer.

- Period: is the length of customer relationship decided to be analysed (one year is the most commonly used period). Customer lifetime value is a multi-period calculation (for example; 3-7 years).

- Periodic revenue: is the amount of revenue generated by a customer in the period.

- Profit margin: is the difference between revenue and costs, even though this may be reflected as a percentage of gross or net profit.

Using the analytical result of customer value evaluation, the marketing department should target the customer that has the highest likelihood to be profitable to the company. The customer value-based approach brings the following benefits to the company:

1. Increased rate of investment (ROI)
2. Increase in acquisition and retention of profitable customers
3. Decrease in costs

2.7 Customer Segmentation

Due to an ever increasing number of competitors, reduction in customer switching costs and consequent customer retention, the competition to acquire more customers has intensified among companies. The organisation needs to prioritise its customers in order to create the capabilities, processes and infrastructure to meet their demands. Without segmentation, differences in customer needs might never be recognised.

Customer segmentation is a process of classifying customers into a number of smaller groups, or market segments based on the characteristics or responses of customers in those segments. This approach helps managers to denitrify the most attractive segments and to develop an appropriate strategy for winning and retaining high value customers.

Bounsaythip and Rinta-Runsala (2001) define segmentation as:

“Customer segmentation is a term used to describe the process of dividing customers into homogeneous groups on the basis of shared or common attributes (habits, tastes, etc.).”

According to Ahn et al., (2003), the needs of diverse customers in the modern business environment cannot be met by mass traditional marketing strategy. Segmentation theory categorises customers and markets into different clusters or groups with similar needs and/or characteristics that are likely to exhibit similar behaviours. Therefore, segmentation is an essential element for customer relationship management (CRM) system.

Wedel and Kamakura (1997) classified segmentation parameters into two groups:

(1) the general variables that include the customer demographics and lifestyles, and

(2) the product specific variables such as customer purchasing behaviours.

Kamakula, (1998) refers to Customer segmentation as the process of classifying customers into different groups of customers. It enables viewing the entire database in a single picture, thus allowing the firm to treat customers differently according to class and pursue marketing that is suitable to each class. Studying customer profitability reveals that there is not always a positive correlation between customer revenue and customer profitability (Kaplan and Narayanan, 2001). Customers from different segments contribute differently to financial performance. In other words, some customers bring more income to the firm than the others.

Foster et al. (2001) states that “each dollar of revenue does not contribute equally to net income”.

Keiningham *et al.* (2005) cited that “while improving revenue for profitable clients does indeed improve profitability, exactly the opposite occurs for unprofitable clients’’. As a result, customers’ profitability level has an essential influence to net income. Customer segmentation can be viewed as a tactic to prioritise customers by their value, to the company. For example, in some scenarios, a small proportion of customers bring the most profit to the company. A study from Banc One of Columbus, Ohio, reveals that 20 per cent of their customers provide all of the bank’s profit, while the rest, 80%, only cost money (McDougall et al., 1997). Therefore, different segments should be approached by different strategies (Elsner et al., 2004).

2.8 Costumer Activity Measurement

Customer behaviours are meaningless unless it translates into a measurable metrics. In reality, companies balance the cost of an initiative against the service attribute (e.g., reduced waiting in the queue) instead of measuring the cost against the increase in, for instance, customer satisfaction (and finally how increased satisfaction will impact profits). The problem is that some benefits, while appearing to be objectively significant, may have only a limited effect on customer behaviour.

Moutinho and Smith, (2000) studied and found that, unless a company realises the cost versus benefits of increased customer outcomes (satisfaction, retention and loyalty), the effort to implement a new strategy like new technology may be a waste of capital. More interestingly there is evidence in the literature that there have been attempts to describe the relationship between these constructs, nevertheless, these descriptions are by no means fully established.

It is found that the link between customer behaviours and profitability is not nearly as straightforward as usually proposed. As a result, this study aims to provide an objective means to explain the relationship between service quality attributes and customer behaviours.

2.9 Competitive Advantage

Parasuraman, (1997) in his work said in a highly competitive market, the shortest route to differentiation is through the development of brands and active promotion to both intermediaries and final consumers. In the long run, however, branding, targeting and positioning would all be much more effective if the supplier had some tangible advantage to offer consumers (Baker, 1993). This is evident in the banking industry, where many banks are providing more or less the identical products for nearly the same price. Unless a bank can extend its product quality beyond the core service with additional and potential service features and value, it is unlikely to gain a sustainable competitive advantage (Chang, Chan, and Leck, 1997). Thus, the most likely way to both retain customers and improve profitability is by adding value via a strategy of differentiation (Baker, 1993) while increasing margins through higher prices.

Ennew and Binks, (1996); Woodruff, (1997) said today's customers do not just buy core quality products or services; they also buy a variety of added value or benefits. This forces the service providers such as banks to adopt a market orientation approach that identifies consumer needs and designs new products and redesigns current ones. Further, competitive pressures then push other financial service firms to actively target consumer segments by integrating service quality, brand loyalty, and customer retention strategies (Ennew and Binks, 1996).

2.10 Corporate Image

Harwood, (2002) said consumers have more choices for their financial needs than ever before. Technology, globalisation, increased competition and increased consumer mobility have dramatically changed the way people bank. Many financial institutions are looking at branding techniques to differentiate themselves.

Harwood (2002) argued that branding, as a tool to build image, is critical in the banking industry where all firms offer about the same kinds of products. Hence, it is critical that banks have a comprehensive knowledge of customers' values, attitudes, needs and perceptions of various services the bank offers and the image which customers have of the bank itself (Kaynak, 1986a, 1986b).

Accordingly, bankers must be able to build and manage their bank's image in order to clearly define the differences between their bank and its competitors.

Bharadwaj et al. (1993) argue that services are highly intangible and are, therefore, high in experience and credence qualities. As a consequence, brand reputation is important as a potential competitive advantage. Alvarez (2001) proposed that logic is no longer enough to sell the benefits of an intangible product or service, especially with commodity products and skeptical consumers. This situation calls for emotion or image to change the perception of the audience in any real or profound way (Alvarez, 2001). Furthermore, both Marthur (1988) and Gronroos (1984) proposed image as an alternative to product differentiation.

2.11 Switching Barriers

Fornell, (1992) in his work said switching barriers have been used as marketing strategies to make it costly for customers to switch to another organisation. Such barriers include search costs, transaction costs, learning costs, loyal customer discounts and emotional costs. These barriers provide disincentives for the customer to leave the current organization.

Curasi and Kennedy (2002) have shown that customer satisfaction does not predict the continuation of the relationship. High switching costs are an important factor binding the customer to the service organisation. Even with relatively low levels of satisfaction, the customer continues to patronise the service provider because repurchasing is easier and more cost effective than searching for a new provider or sampling the services of an unknown provider (Curasi and Kennedy, 2002). Other than switching costs, cross-selling is another critical variable driving customer retention. Cross-selling is the bank's effort to sell as many different products and services as they can to a particular customer (Daniell, 2000). One aspect of loyalty is the impact of cross-selling, which forms a critical element in increasing revenue. Profitability could, as a consequence, be threatened not only by loss of market share but also by diminished opportunities for cross-selling (Jones and Farquhar, 2003). Furthermore, the more products or services you sell to a customer, the less likely it is that they will sever the relationship (Daniell, 2000).

CHAPTER 3

METHODOLOGY

3.1 Binary Logistic Regression Analysis

Despite the similarity between linear regression and logistic regression, linear regression can not be applied to a situation in which the dependent variable is categorical or dichotomous. The linearity assumption of linear regression will be violated when the dependent variable is dichotomous (Barry, 1993). Since the probability of an event must lie between 0 and 1, it is impractical to model probability with linear regression technique, because linear regression model allows the dependent variable to take values greater than 1 or less than 0. One solution for this issue is to transform the data using the logarithmic transformation (Berry and Feldman, 1985).

The method is useful for situations in which you need to predict the presence of a characteristic or outcome based on values of a set of predictor variables.

3.2 Logistic Regression

3.2.1 The Logistic function

The logistic function $f(z) = \dfrac{1}{1+e^{-z}}$, describes the mathematical form on which the logistic model is based. The value of this function have been plotted as z varies from $-\infty$ to $+\infty$. The plot is shown in Figure 3.1.

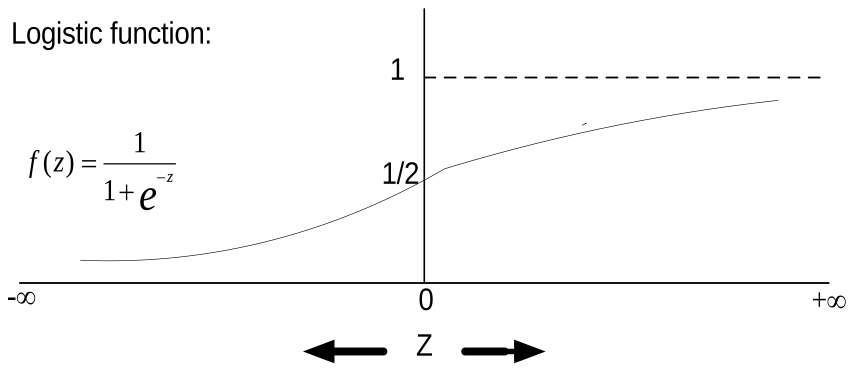


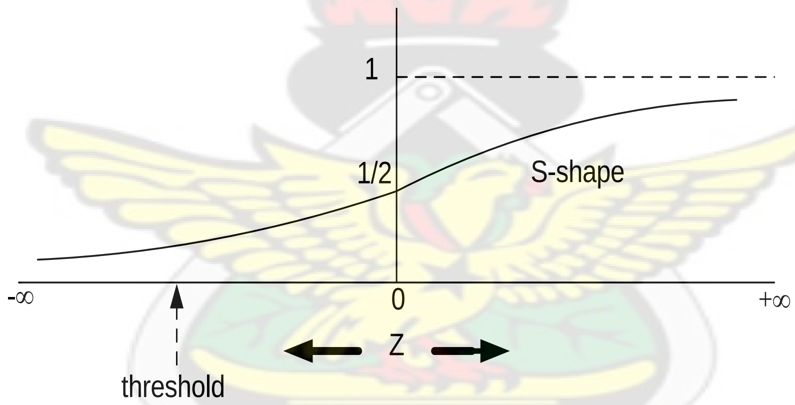
Figure 3.1 Graph depicting the nonlinear nature of the logistic regression function

Binary data usually result from a nonlinear relationship between z and f (z) . A fixed change in z often has less impact when f (z) is near 0 or 1 than when f(z) is near 0.5. In practice, nonlinear relationship between f (z) and z are often monotonic, with f (z) increasing continuously or f (z) decreasing continuously as z increases. The s-shape curves in Figures 3.1 is typical.

The logistic model is designed to describe a probability, which is always some number between 0 and 1. In switching intentions term, such a probability gives the risk of an individual defecting. The logistic model, therefore, is set up to ensure that whatever estimate of risk we get, it will always be some number between 0 and 1. Thus, for the logistic model, we can never get a risk estimate either above 1 or below 0. This is not always true for linear regression models, and this is why the logistic model is often the first choice when a probability is to be estimated.

Another reason why the logistic model is preferred derives from the shape of the logistic function. As the graph of figure 3.1 depict, if we start at z = -∞ and moves to the right, then as z increases, the value of f (z) hovers close to zero for a while, then starts to increase dramatically toward 1, and finally levels off around 1 as z increases towards +∞. The result is an elongated, S – shaped picture.

The S-shape of the logistic function appeals to managers if the variable z is viewed as representing an index that combines contributions of several risk factors, and *f* (z) represents the risk for a given value of z. Then, the S-shape of *f* (z) indicates that the effect of z on an individual's risk is minimal for low z's until some threshold is reached (Figure 3.2). The risk then rises rapidly over a certain range of intermediate z values, and then remains extremely high around 1 once z get large enough. This threshold idea is thought by managers to apply to a variety of diverse conditions.



3.2 Graph depicting z = index of combined risk factors of a logistic function

3.2.2 The Logistic Model

To obtain the logistic model from the logistic function, we write z as the linear sum of risk factors of the form;

$Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_i X_i$, where the X_i's are independent variables of interest and the β_i are constant terms representing unknown parameters.

In essence, then, z is an index that combines the X's. We now substitute the linear sum expression for z in the right-hand side of the formula for f (z) to get the expression

$$f(z)=\frac{1}{1+e^{-z}}=\frac{1}{1+e^{-(\alpha+\sum\beta_i x_i)}}$$ for i ranging from 1 to k.

$P(z)=f(z)$ is the probability of customer switching intention; α is a constant, β is the estimated coefficients, X_i's are the independent variables. From the expression the probability of switching behaviour increases with a unit increase in the independent variable when a coefficient of independent variable is positive. In this research work the logistic regression technique is used to construct a model to predict and classify customer data.

3.2.3 Logit Transformation

The logit transformation, denoted as logit P(X), is give by:

logit p(x) $= In\left[\frac{p(x)}{1-p(x)}\right]$, where $p(x)=\frac{1}{1+e^{-(\alpha+\sum\beta_i x_i)}}$

This transformation allows us to compute a number, called logit P(x), for an individual with independent variable given by X.

The logit transformation is obtained as follows:

Given $p(x)=\frac{1}{1+e^{-(\alpha+\sum\beta_i x_i)}}$

$$1-p(x)=1-\frac{1}{1+e^{-(\alpha+\sum\beta_i x_i)}}$$

$$=\frac{e^{-(\alpha+\sum\beta_i x_i)}}{1+e^{-(\alpha+\sum\beta_i x_i)}}$$

$$\frac{p(x)}{1-p(x)} = \frac{\dfrac{1}{1+e^{-(\alpha+\sum\beta_i x_i)}}}{\dfrac{e^{-(\alpha+\sum\beta_i x_i)}}{1+e^{-(\alpha+\sum\beta_i x_i)}}}$$

$$= e^{(\alpha+\sum\beta_i x_i)}$$

$$In\left[\frac{p(x)}{1-p(x)}\right] = In\left[e^{(\alpha+\sum\beta_i x_i)}\right]$$

Logit (p(x)) $= \alpha + \sum \beta_i X_i$

Hence, in logit form,

$$= \alpha + \sum \beta_i X_i$$

Whereas f (z) must fall in the (0,1) range, the logit can be any real number. The real numbers are also the range for the linear predictors $X_i's$.

3.3 Measurement Requirement

Binary logistic regression analysis requires that the dependent variable be dichotomous. Binary logistic regression analysis also requires that the independent variable be metric or dichotomous.

If an independent variable is nominal level and not dichotomous, the logistic regression procedure in SPSS has an option to dummy code the variable.

3.4 Assumptions

Logistic regression does not make any assumption of normality, linearity, and homogeneity of variance for the independent variables.

Because it does not impose these requirements, it is preferred to other types of analysis such as the discriminant analysis when the data does not satisfy these assumptions.

3.5 Sample Size Considerations

Sample size calculation for logistic regression is a complex problem, but based on the work of Peduzzi et al. (1996) the following guideline for a minimum number of cases to include in the study is suggested.

Let p be the smallest of the proportion of negative or positive cases in the population and k the number of covariates (the number of independent variables), then the minimum number of cases to include is:

$$N = \frac{10k}{p}$$

Alternatively, using a guideline provided by Hosmer and Lemeshow (Hosmer and Lemeshow, 2000), the minimum number of cases per independent variable is 10, with a preferred ratio of 20 to 1.

3.6 Methods For Including variables

There are three methods available for including variables in the regression equation:

1. the simultaneous method in which all independents are included at the same time,
2. the hierarchical method in which control variable are entered in the analysis before the predictors whose effects we are primarily concerned with and
3. the stepwise method in which variable are selected in the order in which they maximize the statistically significant contribution to the model.

For all methods, the contribution to the model is measured by the model likelihood ratio statistic (LR), a statistical measure of the fit between the dependent and independent variables.

3.7 Method For Including Variables

3.7.1 Maximum Likelihood Estimation

Logistic regression uses maximum likelihood estimation to compute the coefficients for the logistic regression equation.

Under weak regularity conditions, such as the parameter space having fixed dimension with true value falling in its interior convex, maximum likelihood (ML) estimation has desirable properties.

- it has large-sample normal distributions
- it has asymptotically consistent; as $n \to \infty$, $\hat{\beta} \xrightarrow{p} \beta$
- it has asymptotically efficient, producing large-sample standard errors not greater than those from other estimation procedures (Agresti, 2000).

Maximum likelihood estimation is one of several alternative approaches that statisticians have developed for estimating the parameters in a mathematical model. Another well-known and popular approach is least squares (LS) estimation. ML estimation and least square estimation are different approaches that give the same results for classical linear regression analyses when the dependent variable is assumed to be normally distributed. When compared to least squares, the ML method can be applied in the estimation of complex nonlinear as well as linear model. In particular, because the logistic model is a nonlinear model, ML estimation is the preferred estimation method for logistic regression. ML estimation requires no restrictions of any kind on the characteristics of the independent variable. Thus, when using ML estimation, the independent variable can be normal, ordinal, or interval.

To describe the ML procedure, we introduce the likelihood function (L). Given the data, the likelihood function, L, is the probability of those data, treated as a function of the unknown parameters in the model and, thus can alternatively be denoted as L(β), where β denotes the collection of unknown parameters being estimated in the model. In matrix terminology, the collection β is referred to as a vector; its components are the individual parameters being estimated in the model, denoted here as $\beta_1, \beta_2 \ldots \beta_q$, where q is the number of individual components.

The likelihood function L(β) represent the joint probability or likelihood of observing the data that have been collected. The term 'joint probability' means a probability that combines the contributions of all the subjects in the study. Thus,

$$L(\beta \mid x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots x_n \mid \beta) = \prod_{i=1}^{n} f(x_i \mid \beta).$$

For convenience, we work with the logarithm of the likelihood function, called the loq-likelihood. Thus there are two alternative ML approaches that can be used to estimate the parameters in a logistic model. These are the unconditional (L_u) method and the conditional (L_c) method. In making the choice between unconditional and conditional ML approaches, the researcher needs to consider the number of parameters in the model relative to the number of subjects under study. In general, unconditional ML estimation is preferred if the number of parameters in the model is small relative to the number of subjects. In contrast, condition of ML estimate is preferred if the number of parameters in the model is large relative to the subjects. As shown below, the ML estimation formula for the unconditional $\left(L_u\right)$ approach directly describes the joint probability of the study data as the product of the joint probability for the cases (switching persons) and the joint probability for the non – cases (non – switching persons) $L_u = \prod_{l=1}^{m} p(x_l) \prod_{l=m+1}^{n} [1 - p(x_i)]$; where

$$m = cases\,(denoted\ as\ x_{1,}\ x_{2,\ ...,}\ x_m), n - m = noncases\,(denoted\ as\ x_{m+1,}\ x_{m+2,...,}\ x_n),$$

The logistic model, $P(X) = \frac{1}{1+e^{-(a+\sum \beta_i x_i)}}$

We use products by assuming that we have independent observations on all subjects. The probability of obtaining the data for lth case is given by $P(x_l)$, where P(X) is the logistic model formula for individual X. The probability of the data for lth non – case is given by $1 - p(x_l)$.

When the logistic model formula involving the parameters is substituted into the likelihood expression above, the formula shown here is obtained.

$$L_u = \frac{\prod_{l=1}^{n} \exp(\alpha + \sum_{i=1}^{k} \beta_i x_i l)}{\prod_{l=1}^{n} [1 + \exp(\alpha + \sum_{i=1}^{k} \beta_i x_i l)]}$$

This expression for the likelihood function L is a function of the unknown parameters and the β_i

The equivalent formula for the conditional (L_c) approach is given below:

$$L_u = \frac{\prod_{l=1}^{n} \exp(\sum_{i=1}^{k} \beta_i x_i l)}{\sum_u [\prod_{l=1}^{n} \exp(\sum_{i=1}^{k} \beta_i x_i l)]}.$$

It must be noted however, that the formula for the likelihood functions for both the unconditional and conditional ML approaches are quite complex mathematically. The formulae are, however built into SPSS. The program does the heavy calculations of forming the likelihood function internally and maximizing this function to obtain the ML solutions.

Once the likelihood function has been determined, the method of maximum likelihood chooses that estimator of the set of unknown parameters, β, which maximizes the likelihood function $L(\beta)$. The estimator is given by $\widehat{\beta} = \arg\max \widehat{l}\,(\beta \setminus x_1, x_2, ..., x_n), \widehat{\beta} \in \beta$ and its components are $\widehat{\beta}_1, \widehat{\beta}_2, ..., \widehat{\beta}_q$.

In general, maximizing the likelihood function $L(\beta)$ is equivalent to maximizing the natural log of $L(\beta)$, which is computationally easier. The components of θ are then found as solutions of equations of partial derivatives: $\frac{\partial InL(\beta)}{\partial \beta_j} = 0, j = 1, 2, ..., q$

Each equation is stated as the partial derivative of the log of the likelihood function with respect to $\beta_j = 0$, where β_j is the jth individual parameter. If there are q parameters in total, then the above set of equations is a set of q equations in q unknowns. These equations must then be solved iteratively.

3.7.1.1 Likelihood Equations

If we let $x_i = (x_{i1}, ..., x_{ip})$ denote p explanatory variables in the set of i values with, $i = 1, 2, ..., N$ then the logistic regression model $P(X) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i x_i)}}$, regarding α as a regression parameter with unit coefficient, is $P(x_i) = \frac{1}{1 + e^{-(\sum_{j=1}^{P} \beta_j x_{ij})}}$

When more than one observation occur at a fixed x_i value, it is sufficient to record the number of observations n_i and the number of successes. We then let y_i refer to this success count rather than to an individual binary response. Then $Y_1, ..., Y_N$ are independent binomial functions.

$$\prod_{i=1}^{N} P(x_i)^{y_i} [1 - P(x_i)]^{n_i - y_i} = \left\{ \prod_{i=1}^{N} Exp \left[\log \left(\frac{P(x_i)}{1 - P(x_i)} \right)^{y_i} \right] \right\} \left\{ \prod_{i=1}^{N} [1 - P(x_i)]^{n_i} \right\}$$

$$=\left\{Exp\left[\sum_i y_i \log\frac{p(x_i)}{1-p(x_i)}\right]\right\}\left\{\prod_{i=1}^{N}\left[1-P(x_i)\right]^{n_i}\right\}.$$

Therefore, for the logistic regression model $P(x_i)=\dfrac{1}{1+e^{-\left(\sum_{j=1}^{P}\beta_j x_{ij}\right)}}$, the ith logit is $\left(\sum_{j=1}^{P}\beta_j X_{ij}\right)$, so the exponential term in the last expression equals $\exp\left[\sum_j y_i\left(\sum_j \beta_j x_{ij}\right)\right]=\exp\left[\sum_j\left(\sum_j y_i x_{ij}\right)\beta_j\right]$. Also, since $\left[1-P(x_i)\right]=\left[1+\exp\left(\sum_j \beta_j x_{ij}\right)\right]^{-1}$, the log likelihood equals $L(\beta)=\sum_j\left(\sum_j y_i x_{ij}\right)\beta_j-\sum_j n_i\log\left[1+\exp\left(\sum_j \beta_j x_{ij}\right)\right]$. This depends on the binomial counts only through the sufficient statistics $\left\{\sum_j y_i\ x_{ij},\ j,\ldots,p\right\}$.

The likelihood equations result from setting $\partial L(\beta)/\partial(\beta)=0$.

Since $\dfrac{\partial L(\beta)}{\partial(\beta_j)}=\sum_j y_i x_{ij}-\sum_j n_i\,x_{ij}\dfrac{\exp\left(\sum_k \beta_k x_{ik}\right)}{1+\exp\left(\sum_k \beta_k x_{ik}\right)}$, the likelihood equations are $\sum_j y_i x_{ij}-\sum_j n_i\widehat{P}_i x_{ij}=0,\ j=1,\ldots,P$, where $\widehat{P}_i=\exp\left(\sum_k \widehat{\beta}_k x_{ik}\right)/\left[1+\exp\left(\sum_k \widehat{\beta}_k x_{ik}\right)\right]$ is the ML estimate of $P(x_i)$. The equations are nonlinear and iterative solution. Let X denote the $N\times P$ matrix of values of (x_{ij}). The likelihood equations have form $X'y=X'\widehat{\mu}$, where $\widehat{\mu}=n_i\widehat{P}_i$.

3.7.1.2 Newton – Raphson Iterative Method

The Newton – Raphson method is an iterative method for solving nonlinear equations whose solutions determines the point at which a function takes its maximum. It begins with an initial guess for the solution. It obtains a second guess by approximating the function to be maximized in a neighborhood of the initial guess by a second – degree polynomial and then finding the location of that polynomial's maximum value. It then approximates the function in a neighborhood of the second guess by another second – degree polynomial, and the third guess is the location of its maximum. In this manner, the method generates a sequence of guesses. These converge to the location of the maximum when the function is suitable and / or the initial guess is good. In more detail, here is how Newton –Raphson determines the value $\hat{\beta}$ at which a function $L(\beta)$ is maximized.

Let $\mu' = (\partial L(\beta)/\partial\beta_1, \partial L(\beta)/\partial\beta_2, \ldots, \partial L(\beta)/\partial\beta_j)$

Let H denote the matrix having entries $h_{ab} = \partial^2 L(\beta)/\partial\beta_a\,\partial\beta_b$, called the Hessian matrix. Let $u^{(t)}$ and $H^{(t)}$ be u and H evaluated at calculated value $\beta^{(t)}$, at the iteration t. Step t in the iterative process $(t = 0, 1, 2, \ldots)$ approximates $L(\beta)$ near $\beta^{(t)}$ by the terms up to the second order in its Taylor series expansion,

$$L(\beta) = L(\beta^{(t)}) + \mu^{(t)}(\beta - \beta^{(t)}) + \left(\frac{1}{2}\right)(\beta - \beta^{(t)})\ H^{(t)}(\beta - \beta^{(t)}).$$

Solving $\partial L(\beta)/\partial\beta = \mu^{(t)} + H^{(t)}(\beta - \beta^{(t)}) = 0$, for β yield the next guess. That guess can be expressed as $\beta^{(t+1)} = \beta^{(t)} - (H^{(t)})^{-1}\mu^{(t)}$, assuming that $H^{(t)}$ is nonsingular. Iterations proceed until changes in $L(\beta^{(t)})$ between successive cycles are sufficiently small. The ML estimator is the limit of $\beta^{(t)}$ as $t \to \infty$ (Walker and Duncan, 1967).

For many models, $L(\beta)$ has concave shape and β is the point at which

$$\frac{\partial Inl(\beta)}{\partial\beta_j} = 0,\ j = 1, 2, \ldots, q.$$

The ML estimate is then the solution of likelihood equation $\dfrac{\partial Inl(\beta)}{\partial\beta_j} = 0.$

The overall measure of how well the model fits is given by the likelihood value, which is similar to the residual or error sum of squares value for multiple regression. A model that fits the data well will have a small likelihood value. A perfect model would have a likelihood value of zero. Maximum – likelihood estimation is an iterative procedure that successively tries to work to get closer and closer to the correct answer.

3.8 Statistical Inference Using Maximum Likelihood Techniques

Once ML estimates have been obtained, these estimates can now be used to make statistical inferences concerning the intention (exposure) – switching relationships study.

The maximized likelihood value, $L(\beta)$, which is the numerical value of the likelihood function L when ML estimates are substituted for their corresponding parameter values can be used to illustrate how statistical inferences are made.

We consider the following three hypothetical models, each written in *logit* form:

Model 1: logit $P_1(X) = \alpha + \beta_1 X_1 + \beta_2 X_2$

Model 2: logit $P_2(X) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

Model 3: logit $P_3(X) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_3 + \beta_5 X_2 X_3$

Model 1 involves two variables X_1 and X_2. Model 2 contains these same two variables and a third variable X_3. Model 3 contains the same three X's as in model 2 plus two additional variables, which are the product terms X_1, X_3 *and* X_2, X_3. Let $\widehat{L}_1, \widehat{L}_2$ *and* $\widehat{L}_3$ denote the maximized likelihood values based on fitting models 1, 2, and 3, respectively. The fitting may be done either by unconditional or conditional methods, depending on which method is more appropriate for the model and data set being considered. Since a model fits the data better when it has more parameters, it follows that $\widehat{L}_1 \leq \widehat{L}_2 \leq \widehat{L}_3$.

This relationship among the $\widehat{L}_s$ is similar to the property in classical multiple linear regression analyses that the more parameters a model has, the higher is R square statistic for the model. In other words, the maximized likelihood value $\widehat{L}$ is similar to R square, in that the higher the $\widehat{L}$, the better the fit.

It follows from algebra that if $\widehat{L}_1 \leq \widehat{L}_2 \leq \widehat{L}_3$ then the same inequality relationship holds for the natural logarithms of these $\widehat{L}_s$, that is , $In\widehat{L}_1 \leq In\widehat{L}_2 \leq In\widehat{L}_3$.

However, if we multiply each log of $\widehat{L}$ by -2, then the inequalities switch around so that $-2\,In\widehat{L}_3 \leq -2\,In\widehat{L}_2 \leq -2\,In\widehat{L}_1$.

The statistics $-2\,In\widehat{L}_1$ is called the log likelihood statistic for model 1, and similarly, the other two statistics are the log likelihood statistics for their respective models. These statistics are important because they are used to test hypotheses about parameters in the model using what is called a likelihood ratio test.

3.9 The Likelihood Ratio Test

Statisticians have shown that the difference log likelihood statistics for two models $(-2\,In\,L_1 - (-2\,In\,L_2))$,one of which is a special case of the other, has an approximate Chi-square distribution in large samples. Such a test statistic is called a likelihood ratio statistic (LR). The degrees of freedom (df) for this chi-square test are equal to the difference between the number of parameters in the two models.

In general, the likelihood ratio statistic requires the identification of two models to be compared, one of which is a special case of the other. The larger model called the full model and the smaller model called the reduced model; that is, the reduced model is obtained by setting certain parameters in the full model equal to zero. Given for example, two models:

Model 1 : logit $P_1(x) = \alpha + \beta_1 X_1 + \beta_2 X_2$

Model 2: logit $P_2(x) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

Model 1 is a special case of model 2 by setting $\beta_3 = 0$

The set of parameters in the full model that is set equal to zero specify the null hypothesis being tested.

H_0: Parameters in full model =0;

df=number of parameters set equal to zero

Correspondingly, the degrees of freedom for the likelihood ratio test are to the number of parameters in the larger model that must be set equal to zero to obtain the smaller model.

3.10 How The LR Test Works

Algebraically, $LR = -2InL_1 - (-2InL_2) = -2In\frac{\hat{L}_1}{\hat{L}_2}$

If the additional variable X_3, makes an extremely large contribution to the risk of switching over that already contributed by X_1 and X_2 then, it follows that the maximized likelihood value L_2 is much larger than the maximized likelihood value L_1

If L_2 is much larger than L_1, then $\frac{L_1}{L_2} \approx 0$ Now the natural log of any fraction between 0 and 1 is a negative number. As this fraction

Approaches 0, $In\frac{\widehat{L}_1}{\widehat{L}_2} \approx In(o) = -\infty$

$$\Rightarrow LR = -2In\frac{\widehat{L}_1}{\widehat{L}_2} \approx \infty$$

Thus, the LR for a highly significant X_3 variable is large and positive and approaches ∞.

In contrast, if the additional variable, X_3 make no contribution whatsoever to risk of disease over and above that contributed by X_1 and X_2, this would mean that $\widehat{L}_1 \approx \widehat{L}_2$.

$$\Rightarrow \frac{\widehat{L}_1}{\widehat{L}_2} \approx 1$$

$$\Rightarrow LR \approx -2In(1) = -2 \times 0 = 0$$

Thus, the LR for a highly non-significant X_3 variable is approximately 0. In summary, the likelihood ratio statistic, regardless of which two models are being compared, yields a value that lies between 0, when there is extreme non-significance, and ∞, when there is extreme significance $(0 \leq LR \leq \infty)$. The LR test does this by comparing the log likelihoods of the two models, if this difference is statistically significant, then the less restrictive model (the one with more variables) is said to fit the data significantly better than the more restrictive model.

3.11 Wald Test

A Wald test is used to test the statistical significance of each coefficient (β) in the model. A

Wald test calculates a Z statistic, which is: $Z = \frac{\widehat{\beta}}{SE}$

This Z value is then squared, yielding a Wald statistic with a chi – square distribution. However, several authors have identified problems with the use of the Wald statistics .Menard (1995) warns that for large coefficients, standard error is inflated, lowering the Wald statistic (chi-square) value. Agresti (1996) states that likelihood-ratio test is more reliable for small sample sizes than the Wald test.

3.12 Overall Model Fit

The *null model* -2 Log Likelihood is given by -2 * In (L_0) where L_0 is the likelihood of obtaining the observations if the independent variables had no effect on the outcome.

The *full model* -2 Log Likelihood is given by -2* In (L) where L is the likelihood of obtaining the observations with all independent variables incorporated in the model.

The difference of these two yields a Chi – Sqaure statistic which is a measure of how well the independent variables affect the outcome or dependent variable.

If the P –value for the overall model fit statistic is less than the conventional 0.05 then there is evidence that at least one of the independent variables contributes to the prediction of the outcome.

3.13 Interpreting Parameters In Logistic Regression

3.13.1 Odds

For a probability of success, the odds are defined to be $\frac{p(x)}{1-p(x)}$. The odds are nonnegative, with $\frac{p(x)}{1-p(x)} > 1$ when a success is more likely than a failure. The quantity $\frac{p(x)}{1-p(x)}$, Whose log value gives the logit, describes the odds for developing the disease for a person with independent variable specified by X, where P(x) denotes the probability of the event of interest. In its simplest form, an odd is the ratio of the probability that some event will occur and the probability that the same event will not occur.

The logit form of the logistic model, shown again here, gives an expression for the log odds of developing the disease for an individual with a specific set of X's.

$$\textit{logit odds for individual characterized by X,} = In\left[\frac{p(x)}{1-p(x)}\right]$$

$$= \alpha + \sum \beta_i x_i$$

Where $\frac{p(x)}{1-p(x)}$ describes risk in logistic model for individual characterized by X.

The ratio of two odds is called the odds ratio.

3.13.2 Odds Ratio

3.13.2.1 Definition

The odds ratio for a predictor is defined as the relative amount by which the odds of the outcome increase (O.R greater than 1.0) or decrease (O.R. less than 1.0) when the value of a predictor variable is increased by 1.0 units. In other words, (odd for PV + 1) /(odds for PV), when PV is the value of the predictor variable.

Any odds ratio, by definition, is a ratio of two odds, written for example as $odds_2 / odds_1 = \frac{p(x_2)}{1-p(x_2)} / \frac{p(x_1)}{1-p(x_1)}$ in which the subscripts indicate two individuals or two groups of individual being compared. For joint distributions with cell probabilities (D_{ij}), the equivalent definition for the odds in row i is $\frac{D_{i1}}{D_{i2}}\; i = 1,2$. Then the odds ratio is $\frac{D_{11}/D_{12}}{D_{21}/D_{22}} = \frac{D_{11}/D_{22}}{D_{12}/D_{21}}$.

More generally, when we describe an odds ratio, the two groups being compared can be defined in terms of X_i, which denotes a general collection of x variables, from 1 to k $X_2 = (X_{21}, X_{22}, \dots, X_{2k})$

$X_1 = (X_{11}, X_{12}, \dots, X_{1k})$

Let X_2 denote the collection of $X's$ that specify group 2 and let X_1 denote the collection of $X's$ that specify group 1.

Notationally, to distinguish the two groups X_2 and X_1 in an odds ratio, we write

$$OR_{X_2 X_1} = \frac{odds}{odds}\ for\ \frac{x_2}{x_1}$$

We now apply the logistic model to this expression to obtain a general odds ratio formula involving the logistic model parameters.

3.13.2.2 Derivation of OR Formula

Given a logistic model of the general form $p(x) = \frac{1}{1+e^{-(\alpha+\sum \beta_i x_i)}}$, we write the odds for group 2 as

$$\frac{p(x_2)}{1-p(x_2)} = \frac{\frac{1}{1+e^{-(\alpha+\sum \beta_i x_{2i})}}}{1-\frac{1}{1+e^{-(\alpha+\sum \beta_i x_{2i}}}}$$

$$= \quad e^{(\alpha+\sum \beta_i x_{2i})}$$

and the odds group1 as,

$$= \frac{p(x_1)}{1-p(x_1)} = \frac{\frac{1}{1+e^{-(\alpha+\sum \beta_i x_{1i})}}}{1-\frac{1}{1+e^{-(\alpha+\sum \beta_i x_{1i}}}}$$

$$= \quad e^{(\alpha+\sum \beta_i x_{1i})}$$

To get an odds ratio, we then divide the first odds by the second odds. The result is an expression for the odds ratio written in terms of the two risks $P(X_2)$ and (X_1) . That is, for

$$OR_{X_2,X_1} = \frac{e^{(\alpha+\sum\beta_i x_{2i})}}{e^{(\alpha+\sum\beta_i x_{1i})}}$$

$$= e^{\sum_{i=1}^{k}\beta_i(x_{2i}-x_{1i})}$$

We thus have a general exponential formula for the odds ratio from a logistic model comparing any two groups of individuals, as specified in terms of X_2 and X_1. We can give an equivalent alternative to our OR formula by using the algebraic rule that says that the exponential of a sum is the same as the product of the exponentials of each terms in the sum. That is,

$$e^{\sum_{i=1}^{k}\beta_i(x_{2i}-x_{1i})=e^{\beta 2(x21-x11)}e^{\beta 3(x22-x12)\cdots}e^{\beta k(x2k-x1k)}}$$

We alternatively write this expression using the product symbol Π, where Π is a mathematical notation which denotes the product of a collection of terms. Thus, using algebraic theory, we obtain the alternative formula for OR as:

$$\prod_{i=1}^{k} e^{\beta_i(x_{2i}-x_{1i})} = e^{\beta_2(x_{21}-x_{11})}e^{\beta_3(x_{22}-x_{12})}...e^{\beta_k(x_{2k}-x_{1k})}$$

Thus the product formula for OR tells us that, when the logistic model is used, the contribution of the variables to the odds ratio is multiplicative.

3.13.2.3 Properties Of The Odds Ratio.

The odds ratio can equal any nonnegative number. The condition $\frac{p(x_2)}{1-p(x_2)} = \frac{p(x_1)}{1-p(x_1)}$ and hence (when all cell probabilities are positive) $\beta = 1$ corresponds to independence of X and Y.

When $1 < \beta < \infty$, subjects in row 1 are more likely to move a success than subjects in row 2, that is, $D_1 > D_2$. When $0 < \beta < 1$, $D_1 < D_2$ When one cell has zero probability, β equals 0 or ∞. Values of θ farther from 1 in a given direction represent stronger association. Two values represent the same association, but in opposite directions, when one is the inverse of the other. For inference, we use log β independence corresponds log β = 0. The log odds ratio is symmetric about this value - reversal of rows or of columns results in a change in its sign. Two values for log β that are the same except for sign represent the same strength of association.

3.13.3 Regression Coefficients (β_i)

The regression coefficients are the coefficients $b_0, b_1, b_2, \ldots, b_i$ of the regression equation:

$$\text{logit}\,(p) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \ldots + b_k X_k$$

An independent variable with a regression coefficient not significantly different from 0 (p< 0.05) can be removed from the regression model. If P < 0.05 then the variable contributes significantly to the prediction of the outcome variable.

The logistic regression coefficients show the change (increase when $b_i > 0$ decrease when $b_i < 0$) in the predicted log odds of having the characteristic of interest for a one-unit change in the independent variables.

Given the models logit $P(x) = \alpha + \sum \beta_i X_i$, the parameter β_i refers to the effect of X_i on the log odds that D= 1 while controlling the other X_j. Exponentiating both sides shows that the odds are an exponential function of X_i. This provides a basic interpretation for the magnitude of β_i . For instance $e^{(\beta_i)}$ is the multiplicative effect on the odds of a l- unit increase in X_i, at fixed levels of other X_j. The sign of β_i determines whether P(x) is increasing or decreasing as x increase. The rate of climb or descent increase as $|\beta|$ increases; as $\beta \to 0$, the curve straightens to a horizontal straight line. When $\beta_i = 0$, D is independent of X. For quantitative X_i with $\beta > 0$, the curve for P(x) has the shape of the cumulative distribution function (cdf) of the logistic distribution. Since the logistic density is symmetric P(x) approaches 1 at the same rate that it approaches 0.

3.13.4 The Intercept (α)

As a simple example, consider what the logit becomes when all the X's are 0. To compute this, we need to work with the mathematical formula, which involves the unknown parameters and the X's.

If we plug in 0 for all the X's in the formula, we find that the logit of P(x) reduce simply to α. Because we have already seen that any logit can be described in term of an odds, we can interpret this result to give some meaning to the parameter α.

This interpretation for α is more appealing: to describe it as the log of the background, or baseline, odds. By background odds, we mean the odds that would result for a logistic model without any X's at all. The form of such a model is $(X) = \frac{1}{1+e^{-\alpha}}$. we might be interested in this model to obtain a baseline risk or odds estimate that ignores all possible predictor variables. Such an estimate can serve as a starting point for comparing other estimates of risk or odds when one or more X's are considered.

3.13.5 The P -Value

There are several different ways of estimating the P-value. The Wald chi-square is fairly popular, but it may yield inaccurate results with small sample sizes. The likelihood ratio method may be better. It uses the difference between the probability of obtaining the observed results under the logistic model and the probability of obtaining the observed results in a model with no relationship between the independent and dependent variables. A P of 5% or less is the generally accepted point at which to reject the null hypothesis. With a P- value of 0.05 we can say with a 95% probability of being correct that the variable is having some effect, assuming our model is specified correctly. When a model has a high p- value (p-value > 0.05) there is a very good change that model is not significant and should not be used.

CHAPTER 4

DATA ANALYSIS AND RESULTS

4.1 Data Collection

This thesis utilized quantitative survey for data collection (Table 4.2). Data were collected by face to face interview and analysed by standard statistical technique (SPSS) to establish relationships between variables. The research survey instrument was a self administered questionnaire shown in appendix A. In order to have consistent responses, respondents were selected from one region out of the ten regions of Ghana. The questionnaire was distributed among customers of the bank. The sample consists of 250 respondents. From this sample, 50.4% of the respondents were aged between 30 to 40 years. Different groups in terms of age, occupation and location were interviewed.

4.2 Summary of Data

4.2.1 Customer's Demographics and Background Information

Data was classified into Age group, Gender, Educational level and Income level. Each group was categorised as shown in the second column of Table 4.1.

Table 4.1 Demographics of Respondents

Demograhic	Category	Frequency	Valid Percent	Cumulative Percent
Age Group				
18 – 30 years old	1	64	25.6	25.6
31 – 40 years old	2	62	24.8	50.4
41 – 50 years old	3	30	12	62.4
51 – 60 years old	4	32	12.8	75.2
61 – 70 years old	5	47	18.8	94
Above 70 years old	6	15	6	100
Total		**250**	**100**	
Gender				
Male	1	123	49.2	49.2
Female	2	127	50.8	100
Total		**250**	**100**	
Educational Level				
Postgraduate Degrees	5	33	13.2	13.2
Bachelor Degrees	4	107	42.8	56
Diploma	3	73	29.2	85.2
Trade Qualification	2	6	2.4	87.6
High School Qualification	1	31	12.4	100
Total		**250**	**100**	
Income Level				
Less than ghs 100	1	18	7.2	7.2
ghs (100 - 499)	2	80	32	39.2
ghs (500 - 999)	3	92	36.8	76
ghs (1000 - 1999)	4	34	13.6	89.6
Above ghs 2000	5	26	10.4	100
Total		**250**	**100**	

Table 4.2 shows the extract of data used in the analysis. Columns 2 to 5 give details of the customer demographics, whiles columns 7 to 15 give details of the independent (predictor) variables and column 16 gives the details of the dependent variables.

Table 4.2 Results of the questionnaire agenda

	Customer's Demographic					Independent Variables									Dependent Variable
id	age	sex	edu	Income	length of stay	Accuracy of banking records	Accuracy of transaction	Access to electronic transaction	Quality of staff	Efficiency of customer service	Physical appearance of the branches	Convinience of branch locations	Bank's effort on new pdt & services	Pricing	Likelihood of staying
1	2	1	4	5	5	1	1	0	1	1	1	1	1	0	1
2	4	2	1	5	5	0	0	1	0	1	1	1	1	1	1
3	4	1	5	3	3	1	1	1	0	0	0	0	0	0	1
4	3	1	5	5	5	0	0	0	1	1	0	1	1	0	0
5	5	1	2	2	3	0	1	1	1	0	1	0	1	1	1
6	5	2	1	2	5	1	1	0	0	1	1	1	1	1	1
7	4	1	4	3	4	0	0	1	1	1	1	1	0	1	1
8	1	1	4	2	2	1	1	0	0	0	0	1	1	0	0
9	4	1	4	5	5	0	0	1	1	1	1	1	1	1	1
10	2	1	1	2	3	0	1	1	0	1	1	1	1	1	1
. . .															. . .
241	2	1	3	2	3	1	1	0	1	1	1	1	1	0	0
242	1	2	3	2	1	1	1	1	0	0	1	1	0	0	0
243	1	2	4	2	1	0	1	0	0	0	0	0	0	1	0
244	1	2	4	1	1	1	1	1	1	1	1	1	1	0	1
245	1	2	4	3	1	1	1	1	0	1	1	0	0	0	1
246	1	2	4	2	3	0	0	1	1	0	1	1	1	0	1
247	3	1	3	3	4	1	1	1	1	1	1	1	1	0	1
248	1	1	3	3	3	1	1	0	1	1	0	0	0	0	1
249	1	2	4	3	3	1	1	1	1	0	1	1	1	1	0
250	1	2	5	3	2	1	1	0	1	0	1	1	0	0	1

4.3 Computation

Standard statistical software (SPSS) version 16, Toshiba (Windows Vista Home Premium) 3GB RAM, 184GB operating system were use to analyse and establish relationships between variables. The data were run for three times to check for wrong data entries with four and twenty iterations being the lower and highest respectively. The expected tests were;

i) Variables in the Equation (showing B-values, Standard errors, Wald, Degree of freedom (df), Significance of values (sig), Exp(B), Confidence Interval (C.I) for Exp(B) and the predicted probabilities),

ii) Omnibus test of model co-efficient,

iii) Hosmer & Lemeshow test,

iv) Model summary and

v) Classification table.

4.4 Results

4.4.1 The Impact Of Customer Satisfaction On Customer Switching Intentions

Table 4.3 shows the result of logistic regression analysis using SPSS. The significance value of the Wald statistics for each independent variable indicates that overall customer attributes can project customer switching intentions ($P<0.05$).

Table 4.3 Logistic regression estimates of the impact of customer attributes on customer switching behaviours.

Variables in the Equation

	B	S.E.	Wald	Df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
							Lower	Upper
ABR	-.176	.414	.181	1	.670	.838	.372	1.888
AoT	.161	.408	.157	1	.692	1.175	.528	2.614
AET	.300	.308	.954	1	.032	1.350	.739	2.468
SDS	-.183	.332	9.303	1	.048	.833	.434	1.597
EoCS	.977	.341	8.198	1	.004	2.656	1.361	5.184
PAoB	.314	.354	.785	1	.376	1.369	.684	2.740
CoBL	.189	.351	.290	1	.590	1.208	.607	2.406
P	.814	.347	5.509	1	.019	.443	.225	.874
NP	.129	.305	.179	1	.672	1.138	.625	2.071
Constant	.140	.350	.160	1	.689	1.150		

Table 4.3 column six determines the significant predictor variables at 0.05 level of significance. These variables are;

AET – Access to Electronic Transactions

SDS – The Staff who delivered the Service

EoCS – The Efficiency of Customer Service

P - Pricing

Thus the logistic function is given by equation 4.1a below:

P (Switching intentions) = $\frac{1}{1+\ell^{-(0.14+0.300AET-0.183SDS+0.977EoCS+0.814P)}}$ **4.1a**

In logit form ;

logit(P)(switching) = $0.140+0.300AET-0.183SDS+0.977EoCS+0.814P$ **4.1b**

Table 4.4 is the baseline classification table of the dependent variable (likelihood of staying) when the independent variables are not considered. It indicates the percentage of retention of 65.6%.

Table 4.4 Classification Table

Observed			Predicted: Likelihood of staying No	Predicted: Likelihood of staying Yes	Percentage Correct
Step 0	Likelihood of staying	No	0	86	.0
		Yes	0	164	100.0
	Overall Percentage				65.6

Table 4.5, the Omnibus Tests of Model Coefficients gives us an overall indication of how well the model performs, with predictors entered into the model. This is referred to as a 'goodness of fit' test. This set of results gave a highly significant value of 0.045 (the **Sig.** value should be less than .05). The chi-square value, in the result is 17.22 with 9 degrees of freedom.

Table 4.5 Omnibus Tests of Model Coefficients

	Chi-square	Df	Sig.
Step	17.220	9	.045
Block	17.220	9	.045
Model	17.220	9	.045

The Hosmer and Lemeshow test of Table 4.6 indicates a confirmation of omnibus test for the model. The significance should be greater than 0.05. It is the most reliable test of model fit available in SPSS.

The model Chi-square value is 6.766 with a significance level of 0.562 larger than 0.05, therefore indicating support for the model.

Table 4.6 Hosmer and Lemeshow Test

Chi-square	Df	Sig.
6.766	8	.562

Table 4.7 gives the Model Summary which provides information about the usefulness of the model. The Cox & Snell R Square and the Nagelkerke R Square values provide an indication of the amount of variation in the dependent variable explained by the model (from a minimum value of 0 to a maximum of approximately 1). These are described as pseudo R square statistics. 6.7% and 9.2% of the variability is explained by the set of variables.

Table 4.7 Model Summary

-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
304.607	.067	.092

Estimation terminated at iteration number 4 because parameter. parameter estimates changed by less than .001

Table 4.8 is the classification table which gives us an indication of how well the model is able to predict the correct category (switching/no switching) for each attribute. The model correctly classified 67.2 percent of attributes overall (sometimes referred to as the percentage accuracy in classification, PAC)

The sensitivity of the model is the percentage of the customers that has the characteristic of interest (e.g switching) that ha been accurately identified by the model (the true positive).

The model correctly classified 92.7 percent of customer who did not have a problem switching. The specificity of 18.6 percent is the percentage of customers without characteristic of interest (switching) that is correctly identified (true negatives).

The positive predictive value $= \left(\frac{152}{152+70}\right) = \frac{152}{222} = 68.5\%$, indicating that of the customers predicted to have a no switching problem.

The negative predictive value $= \left(\frac{16}{16+12}\right) = \frac{16}{28} = 57.1\%$ indicates customers who predicted to have a switching problem.

Table 4.8 Classification Table

Observed		Predicted: Likelihood of staying No	Predicted: Likelihood of staying Yes	Predicted: Percentage Correct
Likelihood of staying	No	16	70	18.6
	Yes	12	152	92.7
Overall Percentage				67.2

a. The cut value is .500

Fig 4.1 is a histogram of the predicted probabilities of a customer switching intention. It shows all of the cases in of customers may intention to churn on the left-hand side (close to 0), and all the cases for which customers intend to stay on the right-hand side (close to 1.)

The points clustered in the centre of the plot presenting a probability of >0.5 that the customer may churn. However, for these cases there is little more than 50:50 chances that the data are correct predicted.

Figure 4.1 Predicted Probabilities of a customer switching

```
                Observed Groups and Predicted Probabilities

       40 +                                                                       Y                            +
          |                                                                       Y                            |
          |                                                                       Y                            |
F         |                                                                       Y                            |
R      30 +                                                                       Y                            +
E         |                                                                       Y                            |
Q         |                                                                       Y Y                          |
U         |                                                                       Y Y                          |
E      20 +                                                            Y          Y Y                          +
N         |                                                            Y   Y      Y Y                          |
C         |                                                            Y   Y      Y Y                          |
Y         |                                                        Y   Y   Y      Y Y                          |
       10 +                                                     Y  Y   N   Y  Y   Y YY  Y    Y                 +
          |                                                   Y Y  Y   N   NY Y   Y YY  Y    Y Y               |
          |                                         N   N   Y  NYN  N   N   NY N   N NY  N  Y Y N               |
          |                                     YN  N   N Y Y  NNN  NN  NY  NY NY NN NY  N  N NYNYY             |
Predicted  ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:    0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:   NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY

           Predicted Probability is of Membership for Yes
           The Cut Value is .50
           Symbols: N - No
                    Y - Yes

           Each Symbol Represents 2.5 Cases.
```

4.4.2 Prediction of Customer satisfaction, Customer's demographic on Customer switching intentions

The study found that customer satisfaction and customer demographic really affect customer future intentions such as switching. The combined effect on switching is;

$$\text{P(Switching intentions)} = \frac{1}{1+\ell^{-(0.14+0.300AET-0.183SDS+0.977EoCS+0.814P-0.10AGE)}} \qquad \textbf{4.2a}$$

In logit form;

$$\text{logitP(Switching)} = \alpha + \beta_1 AET + \beta_2 SDS + \beta_3 EoCS + \beta_4 P + \beta_5 AGE \qquad \textbf{4.2b}$$

Illustration

Suppose a logistic model involving the variables AET (0,1), SDS (0,1), EoCS (0,1), P (0,1), and AGE (35yrs) is fit to a set of data. Suppose further that the estimated coefficients of each of the variables in the model are given by the table below:

Table 4.9 showing table of values for predicted variables

VARIABLE	COEFFICIENT
CONSTANT	0.140
AGE	-0.100
AET	0.300
SDS	-0.183
EoCS	0.977
P	0.814

From equation 4.4a, the probability of switching assuming the highest satisfaction level ie P (switching) = 1.

$$P(\text{Switching} = 1) = \frac{1}{1+\ell^{-(0.14+0.300AET-0.183SDS+0.977EoCS+0.814P-0.100AGE)}}$$

$$= \frac{1}{1+\ell^{-(0.14+0.300(1)-0.183(1)+0.977(1)+0.814(1)-0,10(35)}}$$

$$= \frac{1}{1+\ell^{-(2.048-3.5)}}$$

$$= 0.1897$$

This indicates that there is probability that 18.97% of customers, with the given characteristics are likely to switch from the bank.

4.5 Discussion

The result (18.97%) churning has implications for management strategies and customer service policy. The banking industry is facing an accelerated rate of churning among customers since over its market has reached to maturity stage. With intensified competition, business as usual relationships with customers may not be an effective way to improve customer retention in the future. By using this methodology, the bank set up different strategies for different customer segments to develop and promote various services instead of uniform strategies for all customers.

4.5.1 Effect of B – Values in relations to customer's intentions to stay

The B–values tell the direction of the relationship (which factors increase the likelihood of a Yes answer and which factors decrease it). Negative B–value indicates that an increase in the independent variable score will result in a decreased probability of the case recording a score of 1 in the dependent variable.

From table 4.3 the staff who delivered service showed a negative B – value (-0.183). This indicates that, Customers are satisfied with the quality of service provided by the Banks staff, and are less likely to switch.

For the other three (3) significant categorical variable (Access to electronic transactions, the efficiency of customer service and pricing), the B values are positive. This suggests those customers who have a problem with Access to electronic transactions, the efficiency of customer service and pricing are more likely to say Yes to the question whether they will stay with the bank.

4.5.2 Effect of Exp B - Odds Ratios (OR) to customer's intentions to stay

According to Tabachnick and Fidell (2001), the odds ratio is 'the increase (or decrease if the ratio is less than one) in odds of being in one outcome category when the value of the predictor increases by one unit.

From table 4.3 The odds or risk of customer answering yes, they will stay with the bank, is 1.350 times higher for a customer who has a problem of access to electronic transactions than for those who do not worry about access to electronic transactions, all other factors being equal.

The Quality of staff (staff who delivered the service) is also a significant predictor, according to the Sig. value ($p = 0.048$). The odds ratio for this variable however is 0.833, a value less than 1. This indicates that, the more customers receive exciting service the less he/she will have a problem staying with the bank. For every extra service received (OK + 1) the risk of having problem staying with the bank decreases by a factor of 0.833, all other factors being equal.

The Efficiency of customer service, is also a significant predictor, with Sig. value ($p = 0.04$). The risk for this variable however is 2.656. This means that the risk of a customer staying with the bank is 2.656 times higher for poorly delivered customer service than for customers who do not have problem with the efficiency of customer service.

Lastly, pricing is showed a significant predictor, according to the Sig. value ($p = 0.019$). The odds ration for this variable, however is 0.443, a value less than 1. This indicates that, the risk of as customer staying with the bank for low pricing (fee charges and account maintenance cost) is 2.2573 times higher than those customers who have problem of high pricing.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusions

The research revealed that Age, Access to electronic transactions, Quality of staff (the staff who delivered the service), the efficiency of customer service, and pricing are statistically significant in the prediction of customer switching.

The study tested for the interrelationship between customer demographics, customer satisfaction, and customer intention of switching and predicted a switching rate of 18.97%.

$$P(\text{ Switching} = 1) = \frac{1}{1+\ell^{-(2.048-3.5)}} = 0.1897$$

Since the results of this study are based on consumers' perceptions only, future research should investigate the congruence between consumers' and service providers' perceptions. This will help the industry to better understand whether both consumers and banks have the same perceptions regarding issues relevant to retention.

5.2 Recommendations

The research revealed that retention rate of customers of Barclays Bank Ghana is decreasing and if not managed will affect its sustainability and profitability. The following are recommended to guide managers when developing service strategies;

- Staff training and servicing support (systems) to improve consumers' positive experiences while interacting with the bank.

- In order to retain younger customers, Barclays Bank of Ghana should introduce new products or services that young consumers' value most, ie online services linked to mobile networks for easy banking with cell phones.
- Future research should investigate the congruence between consumers' and service providers' perceptions.

REFERENCES

Abratt, R and Russell, J. (1999). Relationship Marketing in Private Banking South Africa. The International Journal of Bank Marketing, 17(1):5-6.

Anderson, E.W. and Mittal, V. (2000), "Strengthening the satisfaction-profit chain", Journal of Service Research, 3(2): 107.

Anderson, E.W. and Sullivan, M.W. (1990), ''Customer satisfaction and retention across firms'', TIMS College Conference on Services Marketing, Nashville, TN.

Anderson, E.W. and Sullivan, M.W. (1993), ''The antecedents and consequences of customer satisfaction for firms'', Marketing Science, 12(2): 125-143.

Anderson, E.W., Fornell, C. and Lehman, D.R. (1994), ''Customer satisfaction, market share and profitability: findings from Sweden'', Journal of Marketing, 58(2): 112-122.

Anderson, E.W., Fornell, C. and Mazvancheryl, S.K. (2004), ''Customer satisfaction and shareholder value'', Journal of Marketing, 68(4):172-85.

Andreassen, J.C. and Lindestad, B (1998), "Customer loyalty and complex services", International Journal of Service and Management, 9(1): 7-23.

Berry, L.L. (1983), "Relationship marketing", Emerging Perspectives in Services Marketing, AMA, Chicago, IL.

Berry, L.L. (1995), "Relationship marketing of services growing interest, emerging perspectives", Journal of the Academy of Marketing Science, 23(4): 236-245.

Boulding, W., Kalra, A., Staelin, R. and Zeithaml, V. (1993), "A dynamic process model of service quality: from expectations to behavioural intentions", Journal of Marketing Research, 30:7-27.

Busacca, B. and Padula, G. (2005), "Understanding the relationship between attribute performance and overall satisfaction", Marketing Intelligence and Planning, 23(6):543-561.

Colgate, M., Stewart, K and Kinsella, R. (1996). Customer Defection: A Study of the Student Market in Ireland. *The International Journal of Bank Marketing,*

Cornish, S.L. (1997), "Product innovation and the spatial dynamics of market intelligence: does proximity to markets matter?", Economic Geography, 73(2):143-165.

Czepiel, J.A. (1990), "Managing relationships with customers: a differentiating philosophy of marketing", Service Management Effectiveness: 299-323. San Francisco: Jossey-Bass Publishers.

Dawkins, P. and Reichheld, F. (1990), "Customer retention as a competitive weapon", Directors on Boards, 14:42-47.

Dawkins, P. M and Reichheld, F. F. (1990). Customer Retention as a Competitive Weapon. Directors and Boards, 14:4.

Dick, A.S. and Basu, K. (1994), "Customer loyalty: toward an integrated conceptual framework", Journal of Service Industry Management, 4:1.

Dutka, A. (1993), AMA Handbook for Customer Satisfaction, American Marketing Association, Lincoln, IL.

Eckerson, Wayne (2003), "Evaluating BI toolsets", TDWI.

Edvardsson, B., Johnson, M.D., Gustafsson, A. and Strandvik, T. (2000), "The effects of satisfaction and loyalty on profits and growth: products versus services", Total Quality Management and Business Excellence, 11: 7.

Eggert, A. and Fassot, G. (2001), "Begriffsabgrenzung und Anspruchspektrum des eCRM", In Eggert A. and Fassot, G. (eds.), eCRM: electronic customer relationship management. Stuttgart: Schaffer-Poeschel Verlag.

Emerson, C.J. and Grimm, C.M. (1998), ''The relative importance of logistics and marketing customer service: A strategic perspective'', Journal of business Logistics, 19(1): 17-32.

Fisher, A. (2001). Winning the Battle for Customers. *Journal of Financial Services Marketing, 6*(1):77-84.

Fornell, C. and Wernerfelt, B. (1987), "Defensive marketing strategy by customer complaint management: a theoretical analysis", Journal of Marketing Research, 24: 337-346.

Fornell, C. and Wernerfelt, B. (1988), "A model for customer compliant management", Marketing Science, 7: 271-286.

Fornell, C., Johnson, M.D., Anderson, E.W., Cha, J. and Bryant, B. (1996), "The American customer satisfaction index: description, findings, and implications", Journal of Marketing, 60(4): 7-18.

Fornier, S. (1994), "A consumer-based relationship framework for strategic brand management", Published PhD dissertation, University of Florida.

Gale, B.T. (1994), "Managing Customer Value", The Free Press, New York, NY.

Ganesan, S. (1994), "Determinants of long-term orientation in buyer-seller relationships", Journal of Marketing, 58 : 1-19.

Grönroos, C. (1990), Service Management and Marketing, Lexington Books, Toronto, ONT.

Grönroos, C. (1994), "From marketing mix to relationship marketing towards a paradigm shift in marketing", Management Decision, 32(2): 4-32.

Grönroos, C. (2007), "Service management and marketing", 3rd ed., Chichester, UK: Wiley.

Gruca, T.S. and Rego, L.L. (2005), "Customer satisfaction, cash flow, and shareholder value", Journal of Marketing, 69(3): 115-30.

Gummesson, E. Total relationship marketing, Elsevier, 3rd edition, Oxford: 2008.

Hallowell, R. (1996), The relationship of customer satisfaction, customer loyalty, and profitability: an empirical study', International Journal of Service Industry Management, 7(4): 27-42.

Healy, T. J. (1999). Why You Should Retain Your Customers. *America's Community Banker*, *8*(9): 22-26.

Heskett, J.L., Jones, T.O., Loveman, G., Sasser Jr., W.E. and Schlesinger, L.A. (1994), "Putting the service-profit chain to work", Harvard Business Review.

Hosmer, D.W. and Lemeshow, S. (1989), Applied logistic regression. New York: Wiley.

Hull, L. (2002). Foreign-owned Banks: Implications for New Zealand's Financial Stability. Discussion Paper Series, DP2002(5).

Jacoby, J. and Kyner, D.B. (1973), "Brand loyalty vs. repeat purchasing behaviour", Journal of Marketing Research, 1:9.

Johnson, M.D. and Fornell, C. (1991), "A framework for comparing customer satisfaction across individuals and product categories", Journal of Economic Psychology, 12(2): 267-286.

Jones, T.O. and Sasser Jr., W.E. (1995), "Why satisfied customers defect", Harvard Business Review, 73: 6.

Kano, N. (1984), "Attractive quality and must be quality", The Journal of Japanese Society for Quality Control, 14(2): 147-156.

Kaynak, E. and Kucukemiroglu, O. (1992). Bank and Product Selection: Hong Kong. *The International Journal of Bank Marketing, 10*(1): 3-17.

Kano, N., Seraku, N., Takahashi, F., Tsuji, S. (1984), "Attractive quality and must-be quality, English translation of the article: Miryoku-teki Hinshitu to Atarima Hinshitu. Hinshitu", the Journal of the Japanese Society for Quality Control (April), 14(2): 39-48.

Kotler, P. (1988), “Marketing Management: Planning, Analysis and Control, Prentice-Hall, Englewood Cliffs, NJ.

Kotler, P.: Paper presented at the trustees meeting of the Marketing Science Institute, November 1990, Boston, referred to in Marketing Science Institute Review (Spring 1991).

Kumar, V. and Reinartz, W.J. ‚Customer Relationship Management: A Database Approach’, John Wiley and Sons, Inc.

Liljander, V. and Strandvik, T. (1993), “Estimating Zones of tolerance in perceived service quality and perceived service value”, International Journal of Service Industry Management, 4(2): 28.

Looy, B.V., Gemmel, P. and Dierdonck, R.V. Services Management: an integrated approach, 2nd Edition, Prentice Hall, Essex: 2003

Lusch, R.F and Vargo, S.L. (2006), “Toward a service-dominant logic of marketing: dialog, debate, and directions”, New York: M.E. Sharp.

Manrodt, K.B. and Davis, F.W. (1993), ‘‘The evolution to service response logistics’’, International Journal of Physical Distribution and Logistics Management, 23(5): 56-64.

Marple, M and Zimmerman, M. (1999). A Customer Retention Strategy. *Mortgage Banking*, *59*(11):45-50.

Matzler, K. and Sauerwein, S. (2002), ‘‘the factor structure of customer satisfaction: an empirical test of the importance grid and the penalty-reward-contrast analysis’’, The international Journal of Service Industry Management, 13(4):314-322.

Matzler, K., Bailom, F., Hinterhuber, H.H., Renzl, B. and Pichler, J. (2004), ‘‘The asymmetric relationship between attribute-level performance and overall customer satisfaction: a reconsideration of the importance-performance analysis’’, Industrial Marketing Management, 33:271-277.

Mendzela, E. (1999). Managing Customer Risk. *The CPA Journal, 69*(6):56-59.

Mittal, V., William, T.R. and Patrick, M.B. (1998), “The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions”, Journal of Marketing, 62: 33-47.

Morgan, R.M. and Hunt, S.D (1994), “The commitment-trust theory of relationship marketing”, Journal of Marketing, 58: 20-38.

Mousavi, A., Adl, P., Rakowski, R.T., and Gunasekaran, A. (2001), “Customer Optimisation Route and Evaluation (CORE)”, International Journal of Computer Integrated Manufacturing, 14(2):236 – 243.

Oliver, R.L. (1980), “A cognitive model of the antecedents and consequences of satisfaction decisions”, Journal of Marketing Research, 17:460-469.

Oliver, R.L. (1997), “Satisfaction: A Behavioural Perspective on the Customer”, McGraw-Hill, Boston, MA.

Page, M., Pitt, L and Berthon, P. (1996). Analysing and Reducing Customer Defection. *Long Range Planning, 29*(6):821-824.

Parasuraman, A. (1988). Reflections on Gaining Competitive Advantage through Customer Value. *Journal of the Academy of Marketing Science, 25*(2): 154-161.

Parasuraman, V., Zeithaml, A. and Berry, L.B. (1985), “A conceptual model of service quality and its implications for future research”, Journal of Marketing, 49:41-50.

Payne, A. and Frow, P. (2005), “A strategic framework for customer relationship management”, Journal of Marketing, 69: 167-176.

Peck, H., Adrian, P., Christopher, M. and Clark, M, Relationship Marketing: strategy and implementation, Butterworth Heinemann, Oxford: 2000.

Porter, C. (1993), ‘The marketing strategy letter’ :14.

Raaij, E.M. (2005), "The strategic value of customer profitability analysis", Marketing Intelligence and Planning, 23(4) : 372-381.

Ralston, R.W. (1996), "Model maps out sure to growth in marketplace", Marketing News, 30(11): 12.

Reichheld, F. F and Kenny, D. (1990). The Hidden Advantages of Customer Retention. Journal of Retail Banking, 7(4):19-23.

Reichheld, F. F and Sasser, W. E. (1990). Zero Defections: Quality comes to Services. *Harvard Business Review:*105-111.

Reichheld, F.F. and Sasser, W.E. (1990), "Zero defections: quality comes to services", Harvard Business Review : 105-111.

Reichheld, F.F., Markey, R.G. and Hopton, C. (2000), "The loyalty effect-the relationship between loyalty and profits", European Business Journal, 12(3): 134.

Rose, S. (1990), "The coming revolution in credit cards", Journal of Retail Banking, 12 : 17-19.

Rust, R.T. and Oliver, R.L. (1994), ''Service quality: new directions in theory and practice'', SAGE Publications, Inc, New York,1:19.

Rust, R.T. and Zahorik, A.J. (1993), "Customer satisfaction, customer retention, and market share", Journal of Retailing, summer 69(2): 193-215.

Rust, R.T. and Zahorik, A.J. (1995), "Return on quality (ROQ): marketing service quality financially accountable", Journal of Marketing, 59: 58-70.

Shemwell, D.J., Cronin, J.J. and Bullard, W.R. (1994), "Relational Exchange in Services: an empirical investigation of ongoing customer service-provider relationships", International Journal of Service Industry Management, 5(3): 57-68.

Shiba, S., Graham, A. and Walden, D. (1993), "A new American TQM, Four Practical Revolutions in Management", Productivity Press, Portland, OR.

Spector, P.E. (1992), “Summated rating scale construction: an introduction”, Sage, Newbury Park, CA.

Storbacka, K., Strandvik, T. and Grönroos, C. (1994), “Managing customer relationship for profit: the dynamics of relationship quality”, International Journal of Service Industry Management, 5(5): 21-38.

Varva, T.G. (1992), “Aftermarketing: How to keep customer for life through relationship marketing”, Business One Irwin, Burr Ridge, IL.

Vesset, D., Wilhide, K. and McDonough, B. (2006), ‘Worldwide business analytics software 2006-2010 forecast and 2005 vendor shares’, IDC.

Wittink, D.R. and Bayer, L.R. (1994), ‘The measurement imperative’, Marketing Research, 6(4): 14-23.

Yastrow, S. “We: The ideal customer relationship”, SelectBooks Inc., NY (2007).

Yi, (1990), ‘A critical review of consumer satisfaction’, American Marketing Association, Chicago, IL : 68-123.

Zahorik, A.J. and Rust, R.T. (1992), ‘Modelling the impact of service quality on profitability: a review’, in advances in Service Marketing and Management,1 : 247-76.

Zeithmal, V.A. (2000), ‘Service quality, profitability, and the economic worth of customers: what we know and what we need to learn’, Academy of Marketing Science, 28: 1.

Zeithmal, V.A., Berry, L.L. and Parasuraman, A. (1996), “The behavioural consequences of service quality”, Journal of Marketing, 60(2): 31-46.

APPENDIX A

Questionnaire Agenda

Dear Respondent;

In this survey, I aim to measure behavioural variable, in the banking industry, which significantly affects profitability. The survey should not take long to complete (max 5min). Most questions can be answered with a tick, but there are opportunities for you to add your own comments.

Please supply the following information;

Age

☐18-30 yrs ☐ 31-40 yrs ☐41-50 yrs ☐ 51-60 yrs ☐61-70 yrs ☐Above 70yrs

☐Gender ☐Male ☐ Female

Educational Level

☐Postgraduate ☐ Bachelor Degree ☐ Diploma ☐ Trade qualification
☐High school qualification

Income Level per month (in Ghana Cedis)

☐ Less 100 ☐ 100-499 ☐500-999 ☐ 1000-1999 ☐2000 Above

Bank of respondent and length of stay
-How long have you stayed with the bank

☐ less 1yr ☐ 1-2 yrs ☐ 2-3 yrs ☐ 3-4 yrs ☐ 4-5 yrs ☐ 5 yrs above

-Likelihood of staying with the bank into the foreseeable future

☐ Very unlikely ☐ Very likely

Please rank your bank based on;
(1.Strongly disagree 2. Disagree 3.Reasonable 4. Agree 5.Strongly agree)

Consumers are satisfied with …	**1**	**2**	**3**	**4**	**5**
Accuracy of banking records					
Accuracy of transactions					
Access to electronic transactions					
The staff who deliver the service					
The efficiency of customer service					
Physical appearance of the branches					
Convenience of branch locations					
The bank's effort to inform customers about new product and services					
Pricing					

APPENDIX B

Table A **SPSS OUTPUT FOR CUSTOMER SATISFACTION**

	B	S.E.	Wald	Df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
							Lower	Upper
ABR	-.176	.414	.181	1	.670	.838	.372	1.888
AoT	.161	.408	.157	1	.692	1.175	.528	2.614
AET	.300	.308	.954	1	.032	1.350	.739	2.468
SDS	-.183	.332	9.303	1	.048	.833	.434	1.597
EoCS	.977	.341	8.198	1	.004	2.656	1.361	5.184
PAoB	.314	.354	.785	1	.376	1.369	.684	2.740
CoBL	.189	.351	.290	1	.590	1.208	.607	2.406
P	.814	.347	5.509	1	.019	.443	.225	.874
NP	.129	.305	.179	1	.672	1.138	.625	2.071
Constant	.140	.350	.160	1	.689	1.150		

APPENDIX C

Table B **SPSS OUTPUT FOR CUSTOMER DEMOGRAPHICS**

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Age			9.668	5	.085			
	age(1)	.480	.460	2.872	1	.090	2.181	.885	5.376
	age(2)	-.100	.593	2.511	1	.013	.391	.122	1.249
	age(3)	-.393	.593	.439	1	.507	.675	.211	2.157
	age(4)	41.193	1.906E4	.000	1	.998	7.758E17	.000	.
	age(5)	20.976	2.321E4	.000	1	.099	1.288E9	.000	.
	sex(1)	-.264	.329	.648	1	.421	.768	.403	1.462
	Edu			11.066	4	.026			
	edu(1)	-21.401	1.348E4	.000	1	.999	.000	.000	.
	edu(2)	-.386	.596	.420	1	.517	.680	.212	2.184
	edu(3)	-1.388	.611	5.160	1	.023	.250	.075	.827
	edu(4)	-.269	.714	.142	1	.706	.764	.189	3.094
	income			11.216	4	.024			
	income(1)	-.671	.689	.950	1	.330	.511	.133	1.971
	income(2)	.279	.650	.184	1	.668	1.322	.369	4.731
	income(3)	-1.425	.789	3.258	1	.071	.241	.051	1.130
	income(4)	-.470	.879	.286	1	.593	.625	.112	3.498
	Lfs			14.526	5	.013			
	lfs(1)	-.224	.566	.157	1	.692	.799	.264	2.422
	lfs(2)	-.708	.476	2.216	1	.137	.493	.194	1.251
	lfs(3)	1.458	.667	4.777	1	.029	4.299	1.163	15.898
	lfs(4)	.646	.555	1.355	1	.244	1.907	.643	5.658
	lfs(5)	19.828	4.019E4	.000	1	1.000	4.086E8	.000	.
	Constant	1.703	.878	3.761	1	.052	5.492		

a. Variable(s) entered on step 1: age, sex, edu, income, lfs.

APPENDIX D

RESULTS OF QUESTIONNAIRE AGENDA