# Statistical Methods in Customer Relationship Management

# Statistical Methods in Customer Relationship Management

**V. Kumar**

*Georgia State University, USA*

**J. Andrew Petersen**

*University of North Carolina at Chapel Hill, USA*

WILEY

A John Wiley & Sons, Ltd., Publication

This edition first published 2012


*Registered office*
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ,
United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.



*Library of Congress Cataloging-in-Publication Data:*

Kumar, V., 1957–
Statistical methods in customer relationship management / V. Kumar, J. Andrew Petersen.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-119-99320-9 (cloth)
1. Customer relations–Management. 2. Customer relations–Management—Statistical models.
I. Petersen, J. Andrew. II. Title.
HF5415.5.K864 2012
658.8′12015195—dc23

2012012175

A catalogue record for this book is available from the British Library.

ISBN: 978-1-119-99320-9

Set in 10/12pt Times Roman by Thomson Digital, Noida, India

*Dedicated with Love*
*To my parents, Patta & Viswanathan and uncle Kannan,*
*Other family members – Prita, Anita, Rohan and Aparna, and*
*My in-laws Dr. Lalitha and Ramamurthy*

*– V. Kumar*

*Dedicated with Love*
*To Katie: my wife, sweetheart, and the mother of our two*
*wonderful children Alexa and William*

*– J. Andrew Petersen*

# Contents

# Preface

Companies invest millions of dollars in Customer Relationship Management (CRM) systems and strategies. The primary objectives of these systems are to (a) acquire profitable customers, (b) retain profitable customers, (c) prevent profitable customers from migrating to competition, and (d) winning back 'lost' profitable customers. These four objectives collectively lead to increasing the profitability of an organization. While most firms recognize the benefits of establishing CRM systems and strategies, not all firms have been successful in their CRM implementations. So, why did they fail?

Traditional marketing theory and practice have always recommended that a company focus on expanding the customer base will lead to increased profitability. But, what about retaining the customers? Additionally, what about preventing customers from churning and winning back 'lost' customers? This book, at its core, explores these topics that are an integral part of the customer management process. Information and understanding these topics will contribute towards an enhanced financial performance of companies. Companies such as IBM, Zappos, Continental Airlines, Henkel, and Hewlett-Packard have understood the importance and relevance of all these four customer management topics as they apply to their organizations, thereby becoming leaders in their respective domains through the implementation of CRM programs. In order to understand the managerial relevance of these four customer management topics and successfully implement CRM systems, it is important to understand the engines that drive these systems – the quantitative models.

This book focuses on the quantitative and modeling aspects of customer management strategies that lead to future firm profitability. The book stresses on developing an understanding of the statistical models used in CRM applications as the guiding concept for profitable customer management. To understand and explore the functioning of models used in CRM applications, this book traces the management strategies throughout a customer's tenure with a firm. Specifically, the book will review five sets of models that will facilitate effective customer management strategy development and CRM implementation. They are:

1. *Models for customer acquisition.* A customer's tenure with a firm starts with his/her acquisition by the firm. Here, the firm's decisions include identifying the right customers to acquire, forecasting the number of new customers, and the response of promotional campaigns, among others.

2. *Models for customer retention.* During the customer's tenure with the firm, the firm would be interested in retaining this customer for a longer period of time. This calls for investigating the role of trust and commitment with the firm, metrics for customer satisfaction, and the role of loyalty and reward programs, among others. Additionally, decisions on who will buy, what the customers will buy, when they will buy, and how much they will buy are explored through a discussion of the various models used in the literature.

3. *Integrated models for acquiring and retaining customers.* This set of models focuses on the small subset of the literature that has linked customer acquisition and retention. Establishing this link is crucial as it has important implications for customer profitability and optimal allocation of marketing resources between acquisition and retention.

4. *Models for customer churn.* In modeling customer attrition, some of the important decisions made by a firm include identifying whether the customer will churn or not, and if so what will be the probability of the customer churning, and when.

5. *Models for customer win-back.* Customer win-back refers to 'reacquiring' customers after the customer has terminated the relationship with the firm. Here, it is important to understand what CRM models and strategies can do to bring the customer back to the firm in both B2C and B2B settings.

The above-mentioned five sets of models – customer acquisition, customer retention, customer acquisition and retention, customer churn, and customer win-back – form the core of this book. Apart from covering these models, the book will also investigate the need for such CRM models, review the implementation of these models, and look into the future of these models.

# 1 Need for this book

A review of the literature on CRM thus far shows the sheer volume, variety, and range of models and the research problems addressed. For academics and practitioners interested and/or involved in the area of CRM, this range becomes too wide to cover and may therefore lead to the development of ineffective CRM systems.

The foremost objective of this book is to serve as a guide to all the models in the CRM literature by classifying them into four sections: customer acquisition, customer retention, customer churn, and customer win-back. These four sections also happen to be the stages a customer goes through during his/her tenure with a firm. By traversing the path of a customer's lifetime with a firm, this book will enlighten and educate the readers on all the models involved in the four stages and help them build effective CRM systems.

## 2 Supplements to the book

These are as follows:

1. The chapters pertaining to the five sets of CRM models will contain sample datasets to explain the working of the models described. These chapters will contain the SAS codes and SAS outputs for the sample datasets at the end of the respective chapters, and the sample datasets used (in Excel format) will be uploaded to the book's companion website located at `www.wiley.com/go/customer_relationship`.

2. Downloadable PowerPoint presentations are available for all chapters via the text's website.

## 3 Organization of the book

The book adopts a model-based approach toward CRM. It illustrates and reviews the quantitative and modeling aspects needed to understand and implement CRM strategies. The modeling techniques presented here form the foundation for designing and implementing strategic marketing decisions. A brief description of the chapters contained in this book is as follows:

- *Chapter 1: Introduction.* This chapter provides an overview and functioning of a CRM system. It introduces key concepts and metrics needed to understand and implement CRM models. It also describes the process of building and developing CRM models.

- *Chapter 2: Need for CRM models.* This chapter outlines the need for CRM models by emphasizing the developments thus far in the CRM literature. Further, this chapter discusses the overall structure of CRM models, and their uses and benefits from a business/marketing standpoint, rather than a technical one. The framework adopted in this book will be whether the models are deterministic or stochastic, and whether they are discrete or continuous. Thus all the models discussed in the book should fall into one of the five chapters that follow.

- *Chapter 3: Models for customer acquisition.* This chapter starts the discussion on the first stage of CRM – acquiring new customers. The objectives of customer acquisition modeling include identifying the right customers to acquire, predicting whether customers will respond to company promotion campaigns, forecasting the number of new customers, and examining the short- and long-run effects of marketing and other business variables on customer acquisition. The model specifications that are covered in this chapter include logit, probit, Tobit, linear regression, log-linear, vector autoregression, hazard function, and decision calculus. Besides covering the modeling aspects of customer acquisition, this chapter also reviews a large number of studies that focus on the effects of marketing variables as drivers or predictors of customer acquisition.

- *Chapter 4: Models for customer retention.* After customers have been acquired, this chapter talks about the second step of CRM – retaining customers. Customer retention strategies are used in both in contractual (where customers are bound by contracts such as cell (mobile) phone subscription or magazine subscription) and in non-contractual settings (where customers are not bound by contracts such as grocery purchases or apparel purchases). In providing the models for customer retention, the primary objectives of this chapter include examining the factors influencing customer retention, predicting customers' propensity to stay with the company or terminate the relationship, and predicting the duration of the customer–company relationship. In examining these factors on retention, this chapter reviews model specifications that include logit, probit, Tobit, multinomial logit, hazard function, discrete-hazard function, proportional hazards function, discrete proportional hazards function, random intercepts, Pareto/NBD, shifted-beta geometric distribution, log-normal, negative binomial, Poisson, linear regression, system of regressions, and deterministic methods.

- *Chapter 5: Integrated models for acquiring and retaining customers.* While most studies consider customer acquisition and retention as independent processes, only a few have linked them together. This chapter focuses on studies which seek to model the two processes simultaneously. Researchers have used several metrics in their studies, including acquisition spending per solicited prospect and the retention spending per customer, potential revenue streams for retention and acquisition, customers' probability to engage in a relationship and the duration of customers' relationship with the firm, and the time that elapses before a prospective customer acquires a particular service and the subsequent duration for which the customer retains service before dropping it. To understand the integrated approach to model acquisition and retention, this chapter reviews model specifications such as probit, Tobit, decision calculus, and simulation techniques. Therefore, the chapter focuses on models that help companies explore the optimal way of allocating resources on customer acquisition and retention, and the variables impacting the trade-off between customer acquisition and retention.

- *Chapter 6: Models for customer churn.* This chapter focuses on preventing customer churn, which is an important function in the CRM process of any firm. Specifically, the chapter describes customer churn models that focus on areas such as (a) modeling churn with time-varying covariates, (b) analyzing the mediation effects of customer status and partial defection on customer churn, (c) modeling churn using two cost-sensitive classifiers, (d) dynamic churn modeling using time-varying covariates, (e) factors inducing service switching, (f) antecedents of switching behavior, and (g) impact of price reductions on switching behavior. In understanding the above-mentioned topics in customer churn, this chapter reviews model specifications such as binomial logit, time-series regression, logistic regression, hierarchical logistic regression, hazard function, proportional hazards function, neural networks, decision trees, Markov models, and fuzzy logic.

- *Chapter 7: Models for customer win-back.* This chapter deals with the fourth component of CRM called customer win-back, reacquisition, reactivation, or regain – an area that lacks research interest. Specifically, the chapter will cover several models and topics that are applied in customer win-back such as (a) a conceptual approach for winning back customers, (b) adopting a lifetime value framework as a basis for customer win-back, (c) optimal pricing strategies for recapture of lost customers, and (d) a model for the perceived value of a win-back offer. The model specifications reviewed in this chapter include split-hazard function and ANOVA design.

- *Chapter 8: Implementing CRM models.* This chapter provides the implementation details about CRM models, as they apply to B2B and B2C companies. Specifically, the critical things required for an effective implementation of CRM models are data, technology enablement, a skilled workforce, and a relevant strategy to be followed based on the model outputs. This chapter also touches upon the financial benefits of building and implementing CRM models. Case studies highlighting the key factors for successful implementation are discussed.

- *Chapter 9: The Future of CRM.* This chapter discusses future directions with respect to quantitative model building and development. The chapter highlights emerging areas and topics such as social media, mobile marketing, and customized campaigns that are prime areas for future research. The popular commercial uses and expected future development trends are also identified for each of the three marketing methods. This chapter also highlights the gaps in current research and advocates areas in need of better models.

## Acknowledgments

We wish to thank Alan Zhang and Bharath Rajan for their assistance and contribution in the preparation of this text. We would also like to thank our colleagues in various universities for giving us valuable suggestions in developing this book. We owe additional thanks to Renu for proofreading the manuscript.

# 1

# Customer relationship management

## 1.1 Introduction

Henkel, a European multinational corporation that operates in three business areas (home care, personal care, and adhesive technologies), also operates in a highly competitive, fast-moving consumer-goods industry that also has a global outreach. Some of the key operational challenges include low product margins on its products and a lack of direct customer contact, among other region-specific challenges. Henkel has managed to overcome these challenges by implementing customer relationship management(CRM) practices. Management at Henkel realized the importance of identifying and understanding the needs of individual high-value customers, in order to target, establish, develop, and retain long-lasting relationships with customers. Through these practices, Henkel has actively pursued the development of strong relationships with its customers, and at the same time increased its profitability [1].

Just like Henkel, many corporations are increasingly adopting CRM as a means to forge their competitive advantage – the ability to understand individual customer needs, and therefore to manage their marketing efforts more efficiently. Such firms are also under tremendous pressure to adjust quickly to rapid changes in the marketplace with regard to the customer, technology and marketing functions. Customers are becoming not only more value-conscious, but also less loyal and less tolerant of low service levels. Consequently, markets are becoming more fragmented, making differentiation more difficult and competition more intense. These changes are driving companies to be customer-centric, and shifting their marketing functions from product-based to customer-based ones. At the same time, the exponential growth in data storage technology has made it possible for firms to process a much more

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

substantial amount of customer-level information. All of these changes have had a significant influence on the rapid growth, increasing the awareness and adoption of CRM worldwide.

While most firms recognize the benefits of adopting CRM practices, not all firms have been successful in their CRM implementations. We believe that having the right approach to CRM planning is critical to a firm's success. Over the years, while technology has played a key role in the success of CRM implementation, it is but one component of CRM implementation. An important part of CRM is identifying the different types of customers and then developing specific strategies for interacting with each customer. Examples of such strategies are developing better relationships with *profitable* customers, not *loyal* customers. That means locating and attracting customers who will be profitable, and finding appropriate strategies for unprofitable customers, which could mean eventually terminating the relationship with customers who are causing the firm to lose money.

In this book, we discuss CRM and its related strategies from a modeling perspective, with a specific focus on methodologies that can be used to obtain insights on customer metrics within a company's own customer database. To help understand the discussion on CRM models and methodologies, we start this first chapter by providing an overview of CRM and its components. First, we present a formal definition of CRM and discuss the relevant concepts in CRM, such as customer value and customer databases. Next, we explain what is required to implement CRM strategies from a manager's perspective. In essence, managers ultimately want to evaluate their firms' marketing performances based on quantifiable indicators called customer metrics. Managers usually have to go through several data-processing steps before being able to generate the customer metrics they want. In this section, we focus on the database sources, the impact of technology on the implementation process of CRM, as well as providing a list of common customer metrics used to evaluate managerial performance. Finally, we introduce the role of analytical methodologies to help managers with the necessary levers/drivers to maximize the performance metrics.

## 1.2 What is CRM?

As the conceptualizations of CRM have evolved significantly, there are various definitions of CRM depending on the perspectives looked at. An important concept in CRM is *customer value*. Customer value is essentially the financial value of the customer relationship to the firm. It can be expressed in terms of contribution margin or net profit. Customer value is widely used by firms to evaluate their marketing efforts. However, it is a general term which does not refer to any specific time or duration. A better term that gives managers an idea of how the value of a client has evolved over time is c*ustomer lifetime value* (CLV). CLV refers to the net economic value of a customer to a firm over his/her entire lifetime (three years in most cases – but it can depend on the length of the average purchase cycle of customers) with the company [2].

By definition, CRM refers to the practice of collecting, storing, and analyzing customer-level information, and incorporating the results into the decision-making process of a firm [3]. This also involves automating, enhancing, and integrating core business processes such as production, operations, sales, marketing, and finance, among others. The power of CRM lies in its adaptability to further the performance of any individual activity of the business, or even the entire business as a whole.

Companies are now beginning to design and implement CRM initiatives from a CLV perspective so that it becomes easier for them to measure the impact of CRM activities. Apart from evaluating the value of customers, such an approach to CRM activities also provides a basis for the competitive advantage of a firm: the customer-centric organization. By improving customer satisfaction, and customer loyalty, CRM helps firms acquire and retain profitable customers, and reactivate dormant customers. The ultimate goal of CRM is to maximize the lifetime value of each individual customer to the firm, thereby increasing firm profitability. In this regard, most of the CRM initiatives can be attributed to one of the following four categories:

1. *Customer acquisition.* Customer acquisition is the process of acquiring new customers, the foundation step of the whole CRM process.

2. *Customer retention.* Customer retention is the process of keeping and developing relationships with the customers after the company acquires them.

3. *Customer churn.* Customer churn, sometimes referred to as customer attrition, is the process of managing the rate of existing customers leaving a firm.

4. *Customer win-back.* Customer win-back is the process of reacquiring the customers that have left a firm through customer churn.

Now that we have reviewed the reason for adopting CRM practices, in the next section we will discuss the essential inputs needed for any CRM implementation.

## 1.3 What is needed to implement CRM strategies?

The implementation of a CRM strategy is an ongoing process of developing and executing a series of small projects aimed at satisfying the business needs and enhancing the value proposition to customers. In this section, we focus on three essential ingredients needed to implement CRM strategies from a modeling perspective: database, technology, and metrics.

### 1.3.1 Database

The database is the core of any CRM planning. Companies gather information to store, analyze, and make marketing decisions based on the results of data analysis. This section provides a basic overview of the categories of databases and the sources from which data can be collected.

#### 1.3.1.1 Categories of databases

There are several types of databases and various ways to categorize them. This can be done according to firms' main business function, information contents, underlying marketing activities, or database technology. As the focus of this book is on data modeling, we look at the following two types of databases in detail: the transaction-related database and the customer database.

- *Transaction-related database.* This database refers to all the information associated with the transactions that customers have made. Examples of this type of information are:
  - What transactions have the customers conducted?
  - What type of product was purchased?
  - How frequent is this type of product purchased by the customer?
  - How much was spent in the transaction?
- *Customer database.* This database is essentially a collection of information about a firm's customers. In general, the following information may be included in customer databases:
  - Basic information: name, address, ZIP (post) code, and telephone number.
  - Demographic information: age, gender, marital status, education, number of people in household, income, and so on.
  - Psychographic information: values, activities, interests, preferences, and so on.
  - Other relevant information: inquiries and referrals, satisfaction, loyalty.

Over a period of time, databases will begin to comprise prospects who have yet to be acquired, along with active and inactive customers. Information on prospects and active and inactive customers are useful to marketers and should be included in customer databases. While data from active customers help marketers learn what has been done well, data from inactive customers help marketers to identify what needs to be improved, and data from prospects who were not acquired show the effectiveness of acquisition efforts and the type of customer the firm has a hard time acquiring. For inactive customers, the following additional information would be important to document:

- How long have the customers been inactive?
- How long have they been active?
- What was their purchasing pattern when they were active?

- How much did they spend?
- How were they initially acquired?
- Why are they inactive?

For prospects who were not acquired, the following information would be important to document:

- How much was spent on the prospects?
- Is the profile of the prospects that were not acquired different from the profile of prospects that were acquired?
- What types of prospects should the company target in the future?
- Why did these prospects choose not to adopt?

### 1.3.1.2 Sources of databases

Managers acquire databases from two main sources: primary data sources and secondary data sources. Figure 1.1 shows a concise summary of available data sources as primary and secondary data sources.

Primary data are original data collected firsthand by the focal firm that are not available or cannot be derived from any other sources. Primary data collection is usually conducted in-house in the forms of experiments or survey methods such as questionnaires, interviews, or observations. Although primary data collection is a costly and time-consuming process, it is sometimes necessary for managers to obtain primary data if the required data cannot be obtained elsewhere or if the

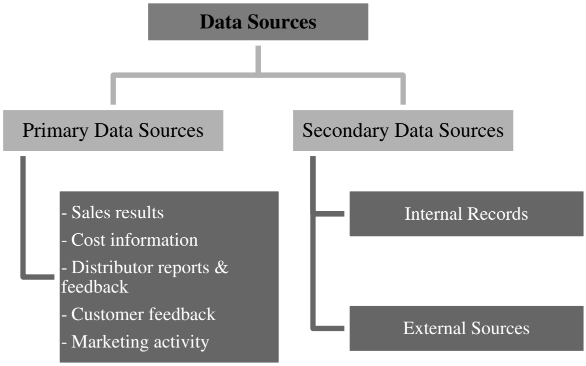



*Figure 1.1 Sources of data. (Adapted from Aaker, D. A., Kumar, V., Day, G., and Leone, R. P. (2010)* Marketing Research, *10th Edition. John Wiley & Sons, Inc., Hoboken, NJ.)*

reliability of those data cannot be determined even when they can be obtained from other sources.

On the other hand, secondary data are data that have already been made available or published in any form. There are two types of secondary data sources: internal records and external sources. Information from internal records is the primary information that the firm obtains directly from its daily business operations (e.g., sales results, cost information, etc.), from customer feedback, or from its marketing activities. This internal information usually comes from various departments within the firm, such as the internal marketing research department, sales analysis group, accounting department, or corporate strategic planning unit. Information from external sources is the secondary information obtained from non-internal sources. There are three main external sources:

1. *Published data sources.* These data are made available in either electronic or printed form from official sources such as government organizations, trade associations, periodicals, newspapers, books, annual reports, and private studies.

2. *Standardized sources of marketing data sources.* Besides published data sources, managers can look at information available from a wide variety of sources from retail stores, warehouses, scanner-based systems, and so on that help provide the managers with a full picture of the market situation of a product category or brand.

3. *Internet.* The Internet is an important and significant source of secondary information. Given the rapid growth of social media activities, managers are interested in both official and non-official information obtained from customers' activities on the Internet, such as customer feedback, reviews about the firm's products and services, etc.

The use of databases for collecting, storing, and analyzing customer data has been crucial for innovations in the CRM process. Nevertheless, technology improvements have been a key driver in making database innovations and other CRM processes accessible, user-friendly, and affordable for firms.

### 1.3.2 Technology

An important factor that drives CRM development in its current stage is the rapid growth of technology. CRM implementation thus has evolved into a user-friendly, flexible, low-cost, and high-tech process. In particular, the three main components of CRM technologies, namely, customer touch points, CRM applications, and data storage technology, have gone through significant improvements. Customer touch points have moved away from the traditional face-to-face interaction between customers and salespeople. With the introduction of Voice over Internet Protocol (VoIP) technology, speech recognition technology, and social networking applications, interactions with customers can be in various forms of Web-based (e-mail, web sites, Facebook, Twitter, etc.) and phone-based (telesales, automatic voice recognition systems, etc.) interactions rather than a physical interpersonal interaction. Further, with the widespread use of the Internet and the growth of PDAs and smart phones, CRM applications are now offered in many forms, such as traditional ERP (Enterprise Resource Planning) systems and mobile or Web-based online portals.

All these developments and enhancements have two key implications for CRM analysts in the area of modeling and data analysis. First, as more data become available, the ways of obtaining them have also increased tremendously. This has given rise to creative ways of collecting customer data and adding more data points about a particular customer, thereby creating a more complete picture of the customer. Second, while massive databases add to the knowledge and resource pool of the organization, they also pose significant modeling challenges regarding the use of appropriate data to glean relevant managerial insights.

### 1.3.3 Metrics

The old adage 'You cannot manage what you cannot measure' is most appropriate where metrics are concerned. Metrics help companies track and assess their performance and, more importantly, evaluate the returns on their CRM initiatives. In the process of implementing CRM, managers have to deal with a huge amount of data with the ultimate goal of evaluating managerial performances based on the value that each individual customer brings to the firm. In order to record and quantify those evaluations, managers need a set of indicators that measure customer values. Metrics perform this role.

The benefits of developing and using metrics are significant to companies. Some of the key benefits that accrue to the firm are: (a) tighter control over business processes and CRM activities, (b) means to measure changes in revenues, costs, and profits, (c) benchmarks and targets to attain certain levels of performance, (d) measures on return on investment (ROI), (e) aid in the acquisition and retention, preventing churn, and assisting win-back of profitable customers, and (f) realigning marketing resources to maximize customer value.

There are two broad categories of metrics, brand-level and customer-level. Brand-level metrics are metrics that measure the brand's competitiveness in the market, such as market share, customer equity, sales growth, and so on. Customer-level metrics break down those brand-level metrics to the individual customer, such as acquisition cost per customer, size of wallet, and so on. When combined, brand-level and customer-level metrics give managers a complete picture of how the firm or the brand fares in the market, as well as how its customer needs differ on an individual level, and how to leverage these differences to enhance the overall competitiveness of the firm.

Table 1.1 presents some commonly used metrics at both brand-level and customer-level. At this stage, the table is meant to provide a general view of the types of CRM metrics available. In the subsequent chapters, we will delve further with detailed discussions about these metrics.

Table 1.1 Popular CRM metrics.

| Metric | Definition | Use of metric |
|---|---|---|
| 1. Market share | The percentage of a firm's sales to the sales of all firms in a given market | Brand-level |
| 2. Sales growth | The increase or decrease in sales volume or sale value in a given period compared to that in the previous period | Brand-level |
| 3. Acquisition rate | The proportion of prospects converted to customers | Brand-level |
| 4. Acquisition cost | The acquisition spending of a focal firm per prospect acquired | Brand-level and customer-level |
| 5. Retention rate | The average likelihood that a customer *makes a repurchase* from the focal firm in period $t$, given that this customer has purchased in the last period $t-1$ | Brand-level and customer-level |
| 6. Defection rate | The average likelihood that a customer *defects* from the focal firm in period $t$, given that this customer has purchased in the last period $t-1$ | Brand-level and customer-level |
| 7. Survival rate | The ratio of customers who continue to remain as customers (survive) until a period $t$ from the beginning of observing these customers | Brand-level |
| 8. Average lifetime duration | The average duration customers continue to remain as customers | Brand-level |
| 9. P-active | The probability of a customer making a repurchase (being active) in a given period | Customer-level |
| 10. Win-back rate | The ratio of acquisition of customers who had been lost in an earlier period | Brand-level |
| 11. Share-of-wallet | The ratio of total sales of all customers of the focal firm in a product category to the total spending of those customers in the product category across all different firms | Brand-level and customer-level |
| 12. Size of wallet | The total spending of a customer on a product category across all different firms | Customer-level |
| 13. Share of category requirement | The ratio of the sales volumes of a particular product category of the focal firm or brand to the total sales volumes of the product category in the market<br>Also considered the market share of a firm or a brand with respect to a particular product category | Brand-level and customer-level |

(*Continued*)

Table 1.1 (*Continued*).

| Metric | Definition | Use of metric |
|---|---|---|
| 14. Past customer value | The gross contribution of a customer when adjusted for the time value of money | Customer-level |
| 15. RFM value | RFM stands for Recency, Frequency, and Monetary value:<br>– Recency indicates the most recent purchase date of a customer<br>– Frequency measures how often a customer purchases from the firm<br>– Monetary value measures the average per transaction spending of a customer | Customer-level |
| 16. Customer lifetime value | The total discounted contribution margins of a customer (excess of recurring revenues over recurring costs to the focal firm) over a specific time period | Customer-level |
| 17. Customer equity | The total lifetime value of all customers of the focal firm | Brand-level |

The important thing for a company to remember is that determining which metric(s) to measure and manage should depend on how each metric relates to the desired short-term or long-term outcome. If the metric(s) chosen cannot be quantifiably related to desired outcome measures such as profitability and shareholder value, the metric(s) are not generally worth measuring and managing.

## 1.4 Analytical methods

Data, technology, and performance metrics combine to provide interesting insights into both the brand and the customer. However, in order to maximize the performance metrics, one has to understand the drivers/levers affecting them. For example, if a firm wants to maximize CLV, one proven strategy is to optimally allocate the marketing resources to each customer. However, this would not be possible if there were no statistical model created to link marketing activities to CLV. Thus, statistical models provide the necessary function of creating the linkage between performance metrics and the possible drivers/levers. In this book, we create many linkages through statistical models to help acquire, retain, and win back profitable customers as well as identifying customer propensity to churn. With the advent of higher computing power, the use of statistical tools is becoming a common phenomenon in all organizations.

## 1.5 Conclusion

Up till now we have seen the basics of CRM and the role of databases, technology, and metrics in implementing CRM initiatives. While the connection between the components for ensuring CRM success is fairly straightforward, the challenge for marketers is in generating the metrics. With advances in data collection and database technologies, identifying the correct approach to employ for generating metrics could determine the success or failure of a CRM campaign. So how can marketers generate metrics that can help them in managerial decision making?

Using statistical methods and models, it is possible to reduce volumes of data to easy-to-use metrics that can help in evaluating business performance. The following chapters discuss in detail various statistical methods used in designing CRM campaigns involving customer acquisition, customer retention, preventing customer churn, and customer win-back. The models discussed in these chapters to address specific marketing challenges can then be used in comprehensive metrics to measure marketing and business performance.

## References

1. Kumar, V. and Reinartz, W.J. (2012) *Customer Relationship Management – Concept, Strategy & Tools*, Springer, Berlin.
2. Kumar, V. (2008) *Managing Customers for Profit*, Prentice Hall, Upper Saddle River, NJ.
3. Aaker, D.A., Kumar, V., Day, G., and Leone, R.P. (2010) *Marketing Research*, 10th ed., John Wiley & Sons, Inc., Hoboken, NJ.

# 2

# CRM in action

## 2.1 Introduction

A survey over the past few years would reveal that successful companies have been the ones that have placed emphasis on and modeled their business practices around customer relationships. This is true for companies across industries, geographical markets, and product offerings. Companies such as Zappos, IBM, and Continental Airlines have become leaders in their respective domains by implementing CRM programs in their organizations. By leveraging the combined power of database, technology, and statistical models, companies have been able to model their entire marketing process around customer relationships – in most cases, at the individual customer level. Gartner Research has predicted that the healthy growth of CRM will continue through 2012, when total worldwide CRM software revenue will reach $13.3 billion. Further, the market for CRM software in North America is projected to reach $7.6 billion in 2012 [1]. With so many resources being directed toward CRM activities, it is important to understand this investment from a managerial viewpoint.

Traditional marketing theory and practice have always advised managers to identify ways to enhance revenues and maximize profits by expanding the customer base. While this may be a viable strategy, it does not always work. For instance, in mature industries and mature markets, customer acquisition may not hold the key to better financial performance. Although we can assume that maximizing profitability is the goal of the firm, we cannot assume that simply maximizing key measures such as customer acquisition and customer retention is the most efficient way to accomplish this goal. What firms must realize is that higher acquisition rates and retention rates, for example, do not necessarily result in higher profitability [2]. Key customer metrics such as acquisition, retention, churn, and win-back certainly are essential for establishing a profitable CRM strategy; this does not mean,

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

however, that 'maximizing' each individual metric is the correct recipe for success. Although identifying these metrics is certainly a crucial step in achieving successful CRM implementation, managers cannot stop there. Where these metrics truly become valuable is in answering questions that determine the correct approach in implementing them. Existing modeling approaches suggest that implementing specific and tailored strategies for key customer metrics yields a greater impact on customer decisions and can therefore lead to higher profitability. Managers need help in developing these specific strategies in each of the four steps of the customer–firm relationship life cycle: acquisition, retention, churn, and win-back.

Using a powerful metric such as CLV, firms can address marketing issues with greater accuracy. This metric will help companies develop important CRM strategies in each of the four steps of customer–firm relationship that will help enhance firm performance. Figure 2.1 illustrates the strategies, known as the Wheel-of-Fortune.

The Wheel-of-Fortune strategies illustrated in the figure provide answers on one or more of the four tasks pertaining to customer acquisition, customer retention, customer churn, and customer win-back. To generate such targeted strategies, it is essential to develop appropriate statistical models that can help managers generate the necessary insights. The following sections highlight the importance of understanding acquisition, retention, churn, and win-back, and how they are integrated into the CRM playbook of an organization. Following this brief introduction, the subsequent chapters in this book delve deeper into the model development process in each of these four areas.

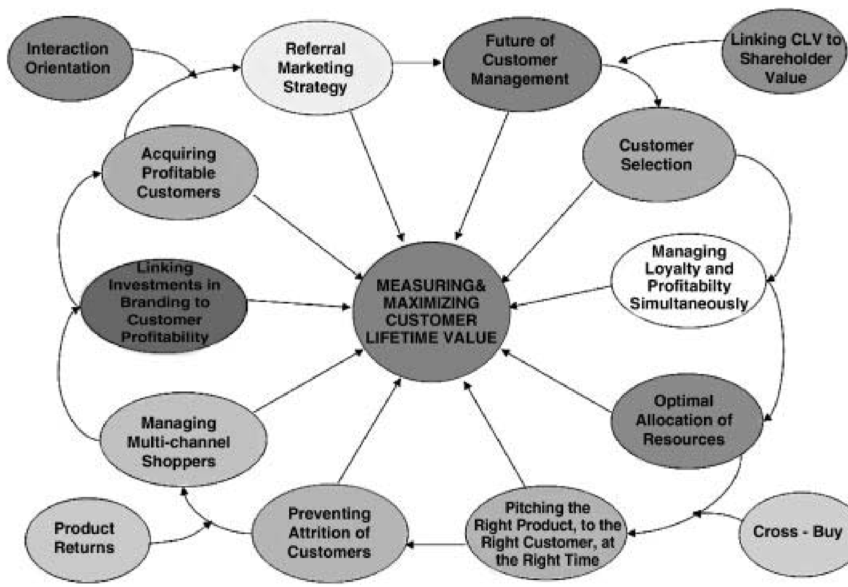



*Figure 2.1 Wheel-of-Fortune strategies. (Adapted and updated from Kumar, V. (2008)* Managing Customer for Profit. *Prentice Hall, Upper Saddle River, NJ.)*

## 2.2 The importance of customer acquisition

Customer acquisition is at the fore for any organization when designing CRM campaigns. Over the past century, we have seen the evolution of customer acquisition techniques spurred by technological advancements and customer heterogeneity. Whether we are talking about improvements in data collection capabilities, data storage capacities, or the ability to analyze data collected from customer behavior and profiles, these improvements have paved the way toward increasingly intricate and focused acquisition techniques. From mass-level acquisition techniques like Ford's Model T 'universal car,' to segment-level acquisition strategies such as Kellogg's Special K weight-loss brand, and finally to one-to-one acquisition such as Dell PC customization, companies have had a growing ability and need to tailor their offerings as well as their acquisition methods to individual customers.

From a traditional manager's perspective, a successful and positive customer acquisition strategy involves attaining the highest possible customer acquisition rate by implementing mass-level strategies. Any combination of mass marketing (radio, billboards, etc.) and direct marketing (telemarketing, mail, e-mail, etc.) would be implemented in order to target 'eligible' customers rather than 'interested' ones. This meant searching for customers who, regardless of their purchasing habits or particular tastes, could not be ruled out of using the product based on general product usage criteria. A CRM strategy that had as its goal to sell diapers would target new/young parents through mass marketing channels such as parenting magazines, promotions at podiatrist offices, or even home furniture magazines. The mass-level acquisition approach could also include more general methods that would target people who could potentially be young parents based on demographics (age, marital status, gender, etc.) [3].

Companies such as Tide (owned by Proctor & Gamble), McDonald's, Coca-Cola, and General Motors most epitomized this strategy during the second half of the twentieth century. This was partly due to the fact that, for example, an advertiser in the 1960s could reach 80% of US women by airing on CBS, NBC, and ABC at the same time [4]. Kraft Food and Beverage Company recently implemented mass-level acquisition when it bought Cadbury Schweppes in 2010 for the purpose of increasing its general brand range and broadening its customer base [5]. However, despite the long-standing success of mass-level acquisition, changes in the firm–customer relationship have over the years required a change in acquisition philosophy. A new approach to CRM pertaining to customer acquisition, dubbed 'cultivating customers,' is gaining ground. This is a conscious move from mass marketing of products to one that is focused on the end consumer [6]. This is a direct result of increases in data collection and storage capabilities that have uncovered layer upon layer of customer differentiation. Differentiating and segmenting with regards to demographic, psychographic, or purchasing power-related characteristics became more affordable and possible, and eventually became necessary in order to keep up with competing firms.

Segment-level acquisition ushered in a new philosophy in customer acquisition: customers are different, and so we should begin to cater to those differences.

Although segment-level acquisition did not take this theory to the extent that one-to-one customer acquisition has, it certainly acknowledged a growing trend of subsets or groups of customers within a larger target market. Being able to collect, store, and analyze customer data in more practical, affordable, and detailed ways has made all of this possible. As firms have become more capable and committed with data analyses, offerings have become more specific, thus increasing the amount of choice for customers. This has in turn spurred customers to expect more choice and customization in their purchases. This continuous firm–customer interaction has consistently shaped segment-level marketing practices in the process to better understand customers.

The emergence of segment-level acquisition has manifested itself in many ways. From a demographic standpoint, companies can now differentiate between customer groups who live in certain geographic environments or areas. But companies also segment different customer groups based on psychographic characteristics such as hobbies and leisure activities, or based on purchasing-related characteristics that take into account customers' financial positions. These segmentation techniques are often combined in the process of customer acquisition, exemplified by Tata's launch of their Nano car. On a purchasing-power level, Tata was appealing to the Indian lower–middle class by offering the most affordable car ever made. The customer segment that was targeted with the launch of the Nano was a large contingent of two-wheeler-riding Indians who with this offering could upgrade their method of transportation to a car. It is here that that Tata incorporated a psychographic element, by targeting customers for whom the status symbol of owning a vehicle rather than a two-wheeled bike or scooter was of high importance. By enticing these customers from both a pricing and status symbol standpoint, Tata targeted a subset of customers to whom this offering would appeal (the rise in metropolitan traffic shortly after the release of the Nano can be somewhat attributed to people abandoning two-wheelers in favor of this affordable car) [7].

The CRM strategies of insurance companies and mobile phone companies are also great examples of segment-level acquisition, as their business models are based on the concept of providing different packages for different groups of people. With life insurance companies, for example, companies tailor their packages using almost every segmentation method. Whether it is demographic (using age, gender, health) or purchasing-power related, firms can offer a wide array of policies based on specific and relevant customer information. Mobile phone companies build their business models based on a similar concept that involves family plans, individual plans, and different packages that offer customers different frequency use of texting, calling, and data to find the plan most appropriate for their usage.

But as segmenting capabilities due to data mining improvements have improved over time, segment-level acquisition has yielded an overwhelming number of segments for firms to keep track of. It is through the continued improvements and innovations in data collection, storage, and analysis that acquisition has moved toward one-to-one acquisition. This one-to-one acquisition has reached a level of specificity that segment-level did not offer. The acquisition process now occurs between firms and individual concepts from start to finish. The emergence of modeling-based CRM has been a cornerstone in this trend.

Although many of the same segmentation criteria from segment-level acquisition are present in one-to-one acquisition, they are taken to a new depth of detail for each customer. Where one-to-one acquisition has truly taken flight is in e-commerce and online interactions with customers. Companies' online portals such as Amazon.com, Nytimes.com, and Yahoo.com have offered users the ability to personalize everything from the content they view to the way that they view it. Amazon.com, for example, uses collaborative filtering in order to recommend music or books to users that it would like [8]. The online radio company Pandora bases its entire product on a similar concept of personalization. Users create radio stations based on an artist or song that they like, and Pandora will play music similar to that song or artist which users can signal 'thumbs up' or 'thumbs down' in order to continue fine-tuning the station to their liking. PC manufacturer Dell and sportswear giants Adidas and Nike have all engaged in one-to-one acquisition techniques where customers are granted the opportunity to tailor their products according to their own taste. In the case of Adidas's launch of its Adizero soccer boots, it offered customers the ability to personalize every aspect of the Adizero while offering suggestions as its database took note of the customers' tendencies.

Customer acquisition is an important step for companies in developing a successful and comprehensive CRM strategy. Identifying the right customers to acquire, predicting customer response to promotional activities, and understanding the long-term effects of marketing on customer acquisition are key components in the acquisition process. In order to maximize the efficiency of each one of these steps, firms must use the tested/proven metrics and models for analyzing the various components of customer acquisition.

In Chapter 3, we will discuss the different models and metrics that should be used by firms in their respective business environments. We will take an in-depth look at the modeling-based approach to customer acquisition for both B2B firms and B2C firms, highlighting the differing modeling components necessary for successful customer acquisition in each. Additionally, we will look at the different models needed for contractual and non-contractual customer acquisition relationships within B2B and B2C respectively. Some of the models discussed will include: (a) differences between customers acquired through promotions and those acquired through regular means, (b) effect of marketing activities and shipping and transportation costs on acquisition, (c) impact of the depth of price promotions, and (d) differences in the impact of marketing-induced and word-of-mouth customer acquisition on customer equity.

## 2.3 The significance of customer retention

The increases in customer data collection, storage, and analysis have impacted not only the customer acquisition process, but also customer retention. This is evident from that fact that since the early 1960s, companies have changed their focus from short-term acquisition and transactions to long-term relationships and CLV [9]. In fact, retention studies indicate that for every 1% improvement in customer retention rate, a firm's value increases by 5% [10]. However, as important as customer retention may be in adopting a CRM strategy, it is important to note that many factors must be taken into account during the decision and process of customer retention.

Who to retain can often be a difficult question to answer. This is because the cost of retaining some customers can exceed their future profitability and thus make them unprofitable customers. When to engage in the process of customer retention is also an important component. As a result, firms must monitor their acquired customers appropriately to ensure that their customer loyalty is sustained for a long period of time. Finally, identifying how much to spend on a customer is arguably the most important piece of the customer retention puzzle. It is very easy for firms to over-communicate with a customer and spend more on his/her retention than the customer will ultimately give back to the firm in value.

Identifying who to retain has become more intricate with advances in data modeling techniques. In effect, data availability decides the way in which a firm measures an acquired customer's future profitability, and, correspondingly, determines whether or not that customer is worth the money and effort required to retain him/her. Mobile carrier Sprint Nextel engaged in this process in a rather peculiar and different way by firing 1000 customers in 2007. The firm had determined that around 1000 of its customers had become unprofitable. That is, Sprint was spending more on retaining these customers than their level of purchase contributions. As a result, Sprint let go of these 1000 unprofitable customers. In order to minimize backlash for this unorthodox and abrupt cancellation of service, Sprint applied a credit to those customers' accounts that brought their credit to zero [11]. This is one of many examples of companies actively engaging in measuring whether or not a customer is worth retaining. In helping to make such decisions, the CLV metric can provide better guidance than any other customer-level metric.

When to engage in the activity of retaining a customer can be a highly misunderstood and undervalued component in customer retention. Monitoring a customer's purchasing and attitudinal behavior is vital in understanding when a firm should aggressively and actively pursue retention of that customer. This is important for two reasons. First, firms can often lose sight of a customer's loyalty and lose its profitable customers, thereby creating undue financial stress. Second, monitoring customer behavior allows the firm to identify the attitudinal changes in that customer. This is important because understanding the attitudinal changes of a customer with regard to the firm's brand advises the firm on how and when to be aggressive in its retention strategies for that particular customer.

Determining how much to spend on a customer is an important assessment involved in identifying who and when to retain. Innovations in statistical modeling now allow firms to measure a customer's future value and profitability to the firm, which makes it easier to make decisions on how much to spend compared to the future value. This logical approach to customer retention calls for data pertaining to several aspects of customer transactions over a period of time, and advanced modeling techniques. This analysis also accounts for a customer's responsiveness to retention efforts, as this determines the method of communicating the intention to retain and the related costs attached to it.

Researchers and managers alike are placing more importance on the study of customer retention and its impact on company profits. In both B2B and B2C firms, model-based approaches are becoming increasingly available, and thus necessary, in both contractual and non-contractual relationships. In Chapter 4, we will introduce and discuss in detail the various metrics and model-based approaches being applied to customer retention. The chapter will delve into models that include: (a) explaining customer retention or defection, (b) predicting the continued use of the service relationship through the customer's expected future use and overall satisfaction with the service, (c) renewal of contracts using dynamic modeling, (d) modeling the probability of a member lapsing at a specific time using survival analysis, (e) modeling the duration of relationship using the negative binomial distribution (NBD)/Pareto model and the proportional hazard model, (f) use of loyalty and reward programs for retention, and (g) assessing the impact of a reward program and other elements of the marketing mix.

However, despite the importance of customer retention as a single link in the chain of CRM, it is necessary for managers to take a holistic and comprehensive view of the CRM process. One of the biggest mistakes that managers currently make is in viewing customer acquisition and retention as separate processes. Although most studies assess acquisition and retention separately, we propose several models and metrics through which managers and firms can link the two together in order to improve their CRM process. In Chapter 5, we will propose several models linking the acquisition and retention processes that include: (a) use of nonlinear programming to optimize budget allocation between acquisition and retention spending, (b) jointly modeling the acquisition and retention of a membership available to existing members of an organization, and (c) allocation and distribution of marketing spending between customer acquisition and retention across communication channels.

## 2.4 The impact of customer churn

Customer churn can be difficult to identify and can be harmful to a firm's brand image and profitability when it goes unnoticed. But with the great strides that have been made in technology, statistical modeling, and metrics, it is much easier for firms to achieve higher profitability from reduced customer churn [12]. But what many firms fail to realize is that the majority of customers who are in the churn stage will not complain or voice their concerns. A study on this found that an estimated 4% of customers in the churn stage will actually voice their opinions, with the other 96% lost without voicing their discontent. Further, about 91% of the lost customers will never be won back [13].

Customer churn contains several key measures and questions that must be answered in order to take a holistic view of the process. Engaging in active monitoring of acquired and retained customers is the most crucial step in being able to determine which customers are likely to churn. Determining who is likely to churn is an essential step. This is possible by monitoring customer purchase behavior, attitudinal response, and other metrics that help identify customers who feel underappreciated or underserved. Customers who are likely to churn do demonstrate 'symptoms' of their dissatisfaction, such as fewer purchases, lower response to marketing communications, longer time between purchases, and so on. The collection of customer data is therefore crucial in being able to identify and capture such 'symptoms' and that would help in analyzing the retention behavior and the choice of communication medium.

Understanding who to save among those customers who are identified as being in the churn phase is again a question of cost vs. future profitability. As a general rule, the best candidates to save from churning are those that are likely to contribute more profits in the future. However, the CLV-based churn models described in later chapters provide more direction on this aspect of churn behavior. Other significant questions include: (a) When are the customers likely to defect? (b) Can we predict the time of churn for each customer? (c) When should we intervene and save the customers from churning? (d) How much do we spend on churn prevention with respect to a particular customer? These are some of the questions that the churn models seek to answer. These answers can provide directions and help with strategy development on profitable churn management.

A firm's ability to identify and address a customer early in the churn phase has a significant impact on the firm's profitability. Harnessing the available models and metrics related to customer loyalty and customer propensity to switch is an important step that managers must take in the customer churn phase of CRM. In Chapter 6, we discuss customer churn management and how firms can benefit from the use of models and metrics to prevent profitable customers from churning. We will explore different models that include: (a) modeling churn with time-varying covariates, (b) analyzing the mediation effects of customer status and partial defection on customer churn, (c) modeling churn using two cost-sensitive classifiers, (d) dynamic churn modeling using time-varying covariates, (e) factors inducing service switching, (f) antecedents of switching behavior, and (g) impact of price reductions on switching behavior.

## 2.5 The benefits of customer win-back

No matter what the efforts, firms must recognize that 100% customer retention is impossible. There are factors that are simply out of a firm's control that can affect a customer and influence him/her to leave. However, customer win-back has proven to be a valuable channel through which firms can regain and even improve on previous customer profitability. Strauss and Friege define customer win-back as 'Rebuilding the relationship with customers who explicitly quit the business relationship' [14]. BellSouth offers a great example of successful customer win-back. When it was losing around 21 000 small-business lines per month during 2001, it implemented customer win-back strategies and managed to regain 26 000 customers per month in 2003. Bellsouth based this successful customer win-back strategy on package and price bundling offers to customers. This strategy gave customers the ability to customize their package by adding desired options (e.g., long-distance calling) and immediately obtain the lowest possible bundle price for that combination of services. As CEO Duane Ackerman noted, 'the extended customer life more than offsets bundle discounts' [15]. In order for firms to achieve similar success to that of BellSouth, it is important to go through several steps in the customer win-back process. First, the firm must identify which customers have defected, and who among them can still be won over. Second, it is important for the firm to identify and understand what it needs to offer the lost customers in order to secure their return to the firm. Finally, firms must measure the cost associated with winning back the customer.

Identifying the right customers to win back depends on factors such as (a) the interests of the customers to reconsider their choice of quitting, (b) the product categories that would interest the customers, and (c) the stage of customer life cycle, among others. Although reacquiring lost customers may be a hard sell, research by Marketing Metrics has found that firms still have a 20–40% chance of selling to lost customers vs. only 5–20% of selling to new prospects [16]. It is therefore important for firms to implement the necessary data modeling techniques to identify and pursue lost customers who still exhibit a strong tendency to buy.

Understanding what to offer customers in winning them back is an important step in the win-back process. A customer who defected because of a lack of a personal touch from the firm may not necessarily respond favorably to impersonal marketing communication. A great example of a firm identifying the reasons for a defection, and addressing it accordingly with an appropriate offer, can be seen in the case of Toronto Raptors' win-back of their season ticket holders after a NBA lockout. The Raptors, in association with Maple Leaf Sports and Entertainment (MLSE), used a CRM technology called 'Ventriloquist Express' which enables a personal voice recording to reach customers' mobile phones. The Raptors head coach Butch Carter left a voicemail for all former 4000 season ticket holders, apologizing for the lockout and promising a season of success. At that point in the message, the fan would have the option of listening to a second personal message, asking for a callback from the organization, or ending the message. The response was so strong that the Raptors implemented the same technique for non-season-ticket holders to great effect. Improvements in data mining and innovative analytic technology have also enabled companies to use data mining for win-back purposes. Merrill Lynch, power tool maker Positec, and Coastal Contacts have also used analytics for the customer win-back [17].

Measuring the cost of win-back is as important as determining who to win back and what to offer them. The cost of win-back, much like the cost of retention or churn, is juxtaposed with the customer's future profitability and value to the firm. Answering all of the above questions is an essential part in developing the right win-back strategy. However, as shown in greater detail in the coming chapters, developing the necessary models and metrics for obtaining exact quantifications decides the most appropriate strategy to win back a customer.

Customer win-back completes the set of the four major inputs in the CRM process (with customer acquisition, retention, and churn being the other three). Customer win-back is a critical component within the overall CRM strategy as it deals with customers, but there has been less research in this field than in the previous three. Chapter 7 will delve into the benefits and uses of models and metrics in the customer win-back process. Through the use of these models and metrics, we will show how managers in firms can understand the importance of regaining lost profitable customers in order to have a comprehensive and profitable CRM strategy. Some of the models explored will include: (a) adopting a lifetime value framework as a basis for customer win-back, (b) optimal pricing strategies for recapture of lost customers, and (c) a model for the perceived value of a win-back offer.

## 2.6 Conclusion

The need for CRM models in the four areas is vital to being able to create well-defined, consistent, and successful CRM strategies. Building the right customer relationships that benefit the firm in the long term is not guesswork. Advances in modeling-based CRM and customer data processes are huge assets for the firm that uses them efficiently and rigorously. Our goal in this book is to provide an overview and in-depth analysis of the available models and metrics in each phase within the CRM process in order to provide a holistic view of the benefits that the model-based approach has on CRM practices. In the following chapters, we will discuss each of the four areas in the CRM processes and provide CRM models that will benefit the firm in maximizing profitability.

## References

1. Beal, B. (2008) Gartner: CRM spending looking up. *SearchCRM*. http://searchcrm.techtarget.com/news/1311658/Gartner-CRM-spending-looking-up (accessed April 24, 2012).
2. Kumar, V. (2008) *Managing Customer for Profit*, Prentice Hall, Upper Saddle River, NJ.
3. Berson, A., Smith, S., and Thearling, K. (2000) *Building Data Mining Applications for CRM*, McGraw-Hill, New York.
4. Bianco, A. (2004) The vanishing mass market. www.businessweek.com (accessed July 12, 2004).
5. Farrell, G. (2010) Kraft boosted by Cadbury purchase. *Financial Times*, www.ft.com (accessed August 6, 2010).
6. Rust, R.T., Moorman, C., and Bhalla, G. (2010) Rethinking marketing. *Harvard Business Review*, Reprint R1001F.
7. Kurczweski, N. (2011) Tata Nano seeks footing as world's cheapest car. *Edmunds Auto Observer*, www.autoobserver.com (accessed February 15, 2011).
8. Arora, N., Dreze, X., Ghose, A. *et al.* (2008) Putting one-to-one marketing to work: personalization, customization, and choice. *Marketing Letters*, **19**(3), 305–321.
9. Rust, R., Lemon, K., and Zeithaml, V. (2004) Return on marketing: using customer equity to focus marketing strategy. *Journal of Marketing*, **68**(1), 109–127.

10. Manzur, E., Olavrireta, S., and Farias, P. (2008) Customer retention and price matching: the AFPs case. *Journal of Business Research*, **61**(6), 691–696.
11. Nelson, R. (2007) Sprint may cancel your service if you call customer service too often. http://www.gadgetell.com/tech/comment/sprint-may-cancel-your-service-if-you-call-customer-service-to-often/ (accessed July 6, 2007).
12. Winer, R.S. (2001) A framework for customer relationship management. *California Management Review*, **34**, 89–105.
13. Thompson, B. (2005) The loyalty connection: secrets to customer retention and increased profits. *CustomerThink Corporation*. http://www.rightnow.com/briefcase-files/PDFs/The_Loyalty_Connection__Secrets_to_Customer_Retention_and_Increased_Profits.pdf (accessed April 24, 2012).
14. Stauss, B. and Friege, C. (1999) Regaining service customers. *Journal of Service Research*, **1**(4), 347–361.
15. Caruso, J. (2003) Carriers go bonkers over bundling. *Network World*. http://www.networkworld.com/columnists/2003/1013edit.html (accessed October 13, 2003).
16. Griffin, J. and Lowenstein, M.W. (2001) *Customer Winback: How to Recapture Lost Customers-And Keep Them Loyal*, Jossey-Bass, San Francisco.
17. Griffin, J. (2002) Scoring big with customer win-back: leveraging the 'we want you back' opportunity. *Marketing Times Online*. http://www.wiley.com/WileyCDA/WileyTitle/productCd-0787946672.html (accessed March 2001).

# 3

# Customer acquisition

## 3.1 Introduction

As many companies have transited from brand-centric to customer-centric, the management of customer relationships has become especially important. Much research attention has also transferred from the product life cycle to the customer life cycle. Along the customer life cycle, the essential processes of CRM can be separated into four circulatory steps: customer acquisition, customer retention, customer churn, and customer win-back. Customer acquisition, the first step, is the foundation of the whole CRM process and is also a cornerstone in the development of the business of a company. The statement that acquiring a new customer is several times more costly than retaining an existing customer has clearly shown the importance of a successful customer acquisition program. And, without sustained customer acquisition there will eventually be no customers left for the company to try and retain. Despite the significance of customer acquisition, research on this topic is still not sufficient. This chapter seeks to review and summarize the limited research on customer acquisition and provide guidelines on how a firm can use its own prospect and customer data to aid in successful and profitable customer acquisition.

In the past 15 years, researchers have paid attention to the key issues which occur in the customer acquisition process. Questions that have been considered, as shown in Figure 3.1, include:

- How likely is it that the prospects will respond to our acquisition promotion?
- How many new customers can we acquire in this campaign?
- How many orders will each of our newly acquired customers place?

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

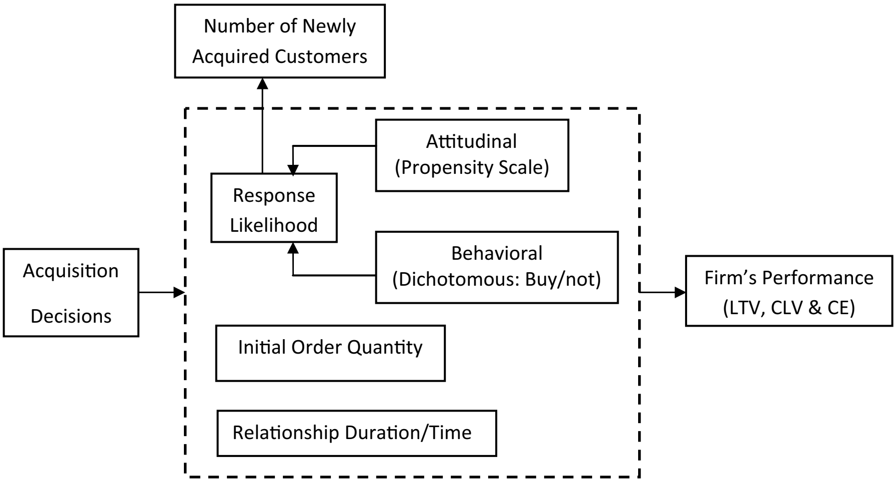



*Figure 3.1 Issues addressed in customer acquisition models (LTV, Lifetime Value; CE, Customer Equity).*

- How do the marketing variables, such as shipping fee, word-of-mouth referral, and promotion depth, influence prospects' response behavior?
- How long will the newly acquired customers stay with our companies?
- How much profit or value will this acquisition campaign bring to our companies?

Many studies have been conducted and many models have been developed to try and solve these questions. Table 3.1 lists a representative set of studies that have considered these issues and accounted for many of the problems that might occur in the model-building process. To provide a comprehensive understanding of how to model customer acquisition, we will review the issues in the studies one by one along with the related modeling techniques (Table 3.2). We will also provide empirical examples at the end of each subsection to demonstrate how to apply this knowledge to a representative sample of customers from a B2B firm.

Due to the differences in how the data can be collected and modeled, the first question that often needs to be answered is whether the company has customers that purchase in a non-contractual or contractual manner. In many instances this will determine the type of statistical model that needs to be used in order to gain any insights from the data.

In non-contractual business settings, such as for retailing firms and catalog companies, customers can freely split their spending across several firms and make purchases from these firms either occasionally/irregularly or frequently/regularly. The key difference from a modeling perspective is that in all non-contractual situations, the firm never actually observes when a customer is no longer a customer. Thus, once the customers make the first purchase from the company, it is uncertain how

Table 3.1 Summary of select empirical studies modeling customer acquisition.

| Studies | Probability/ no./time | LTV/ CLV/CE | Conceptual/ empirical | Selection bias correction | Endogeneity correction | Censoring accounted | Explanatory variables | Industry type |
|---|---|---|---|---|---|---|---|---|
| Wangenheim and Bayón (2007) | Yes | Yes | Empirical | No | No | No | WOM and moderators | Energy market |
| Lix *et al.* [1] | Yes | No | Empirical | No | No | No | Demographics and life-style variables | Retailer |
| Lewis [2] | Yes | No | Empirical | No | Yes | No | Shipping, pricing and WOM | Online retailer |
| Villanueva *et al.* [3] | Yes | Yes | Empirical | No | Yes | No | Acquisition channels and their interactions | Internet log-in |
| Lewis [4] | Yes | Yes | Empirical | No | No | Yes | Discount depth and RFM | Online retailer |
| Hansotia and Wang [5] | Yes | Yes | Empirical | No | No | Yes | Historical transactions | Motor club membership |
| Thomas [6] | Yes | No | Empirical | No | No | Yes | Historical transactions | Membership |
| Reinartz *et al.* [7] | Yes | Yes | Empirical | Yes | No | Yes | Customer actions | High-tech manufacturer |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | and characteristics and firm actions | |
| Thomas *et al.* [8] | Yes | Yes | Empirical | Yes | No | Yes | Customer actions and characteristics and firm actions | B2B firm, pharma, and catalog retailer |
| Blattberg and Deighton [9] | No | Yes | Conceptual | No | No | No | — | — |
| Berger and Bechwati [10] | No | Yes | Conceptual | No | No | No | — | — |

WOM, Word Of Mouth; RFM, Recency, Frequency and Monetary value.

Table 3.2 Review of customer acquisition models.

| Research interest | Specification | Estimation | Representative studies |
|---|---|---|---|
| Probability of being acquired | Logit | MLE | Hansotia and Wang [5] |
| | | | Wangenheim and Bayón (2007) |
| | Probit | MLE | Reinartz, Thomas, and Kumar (2005) |
| | Linear Regression | OLS | Lix, Berger, and Magliozzi (1995) |
| | Log-linear | MLE | Lix, Berger, and Magliozzi (1995) |
| No. of newly acquired customers | Systems of linear regression | 3SLS | Lewis [2] |
| | Vector autoregression | | Villanueva, Yoo, and Hanssens (2008) |
| Initial order quantity | Systems of linear regression | 3SLS | Lewis [2] |
| Duration/time | Hazard function | | Lewis [4] |
| | | | Schweidel, Fader, and Bradlow (2008) |
| Firm's performance (LTV, CLV, and CE) | Tobit | MLE | Lewis (2006) |
| | | | Hansotia and Wang [5] |
| | | | Reinartz, Thomas, and Kumar (2005) |
| | | | Thomas, Reinartz, and Kumar (2004) |
| | Vector autoregression | | Villanueva, Yoo, and Hanssens (2008) |
| | Decision calculus | Deterministic | Blattberg and Deighton [9] |
| | Decision calculus | Deterministic | Berger and Bechwati [10] |

MLE, Maximum Likelihood Estimation; OLS, Ordinary Least Squares; 3SLS, Three-Stage Least Squares.

often they will buy again and for how long they will repeat-purchase from that company. Thus, companies in non-contractual business settings need to consider the probability that new customers will be attracted to buy the first time, how many orders they will bring to the companies over time, and how much profit they will generate over their lifetimes.

In contractual business settings, such as newspaper subscriptions and telecommunications services, customers usually pay their subscription or service fees monthly or yearly and companies know when customers leave the relationship. Once the customers are acquired to join the membership or sign a contract, companies have a guarantee of continuous future cash flow for a period of time. Thus, companies in contractual business settings are concerned most with the expected duration or time of the relationship with customers, mainly because the duration of the relationship is often a key driver of future customer profitability. The key difference here from a modeling perspective is that in all contractual situations, the firm does observe when a customer is no longer a customer since the regular cash flows end. Because of the unique characteristics of non-contractual and contractual business settings, researchers need to build different models to study the issues in customer acquisition in contractual versus non-contractual settings.

We will provide additional discussion in each section where the contractual nature of the situation drives the model selection decision.

### 3.1.1 Data for empirical examples

In this chapter we will be providing a description of the key modeling frameworks that attempt to answer each key research question raised at the beginning of the chapter. We will also be providing at least one empirical example at the end of each subsection which will show how sample data can be used to answer these key research questions. For all the empirical examples in this chapter we provide a dataset titled ‘Customer Acquisition.’ In this dataset you will find a representative sample of 500 prospects from a typical B2B firm. The data include the following variables which will be used in some combination for each of the subsequent analyses:

| Variable | |
|---|---|
| *Customer* | *Customer number (from 1 to 500)* |
| *Acquisition* | *1 if the prospect was acquired, 0 otherwise* |
| *First_Purchase* | *Dollar value of the first purchase (0 if the customer was not acquired)* |
| *CLV* | *The predicted customer lifetime value score. It is 0 if the prospect was not acquired or has already churned from the firm (000s)* |
| *Duration* | *The time in days that the acquired prospect has been or was a customer, right-censored at 730 d* |
| *Censor* | *1 if the customer was still a customer at the end of the observation window, 0 otherwise* |
| *Acq_Expense* | *Dollars spent on marketing efforts to try and acquire that prospect* |

| | |
|---|---|
| *Acq_Expense_SQ* | *Square of dollars spent on marketing efforts to try and acquire that prospect* |
| *Industry* | *1 if the prospect is in the B2B industry, 0 otherwise* |
| *Revenue* | *Annual sales revenue of the prospect's firm (in millions of dollars)* |
| *Employees* | *Number of employees in the prospect's firm* |
| *Ret_Expense* | *Dollars spent on marketing efforts to try and retain that customer* |
| *Ret_Expense_SQ* | *Square of dollars spent on marketing efforts to try and retain that customer* |
| *Crossbuy* | *The number of categories the customer has purchased* |
| *Frequency* | *The number of times the customer purchased during the observation window* |
| *Frequency_SQ* | *The square of the number of times the customer purchased during the observation window* |

These data will be used for each of the examples presented at the end of each of the sections. These examples will cover the topics of response probability, number of newly acquired customers, initial order quantity, duration/time, and firm's performance.

## 3.2 Response probability

The first issue in customer acquisition is to model the probability of prospects being acquired. Hansotia and Wang [5] argued that prospects' response likelihood varies based on their profiles and the promotional materials they received. The authors adopted a logistic regression to model prospects' probability of response and used prospect profile variables as predictors. In logistic regression, because we can only observe whether the prospects respond or not, a latent response variable $y^*$ indicating unobserved utility is assumed. Thus, $y$ is usually defined such that

$$\begin{aligned} y_i^* &= \beta x_i + \varepsilon_i \\ y_i &= \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \le 0 \end{cases} \end{aligned} \tag{3.1}$$

where $y_i$ = the acquisition of customer $i$ (1 = acquired, 0 = not acquired) and $x_i$ = a vector of covariates affecting the acquisition of customer $i$. The probability that the prospect responds is given by

$$\Pr(y=1) = \Pr\left(y_i^* > 0\right) = \Pr(\varepsilon_i > -\beta x_i). \tag{3.2}$$

For a logistic regression, $\varepsilon_i$ has a logistic distribution, with mean 0 and variance equal to $\pi^2/3$. The cumulative distribution function of $\varepsilon_i$ is expressed as

$$F_\varepsilon(\beta x_i) = \frac{1}{1+\exp(-\beta x_i)} \tag{3.3}$$

and

$$\Pr(y=1) = 1 - \Pr(\varepsilon_i \leq -\beta x_i) = 1 - F_\varepsilon(-\beta x_i) = \frac{1}{1+\exp(-\beta x_i)}. \tag{3.4}$$

Setting the estimated value of $\Pr(y=1)$ to $\hat{p}$, when $\beta$ is estimated by $\hat{\beta}$, we have

$$\hat{p} = \Pr(y=1) = \frac{1}{1+\exp(-\beta x_i)}. \tag{3.5}$$

Equation 3.5, the probability of response $p$, is estimated by the log-odds function as the well-known logistic regression model

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{\beta}X. \tag{3.6}$$

To present how to estimate the logit model, we adopt Franses and Paap's (2001) introduction of the MLE method for the logit model. A brief introduction of MLE is provided in Appendix appA. The likelihood function for the logit model is the product of the choice probabilities over the $i$ individuals, that is

$$L(\beta) = \prod_{i=1}^{N} (\Lambda(X_i\beta))^{y_i}(1-\Lambda(X_i\beta))^{1-y_i}, \tag{3.7}$$

where $\Lambda(X_i\beta)$ is the cumulative distribution function according to the standardized logistic distribution, and the log-likelihood is

$$l(\beta) = \sum_{i=1}^{N} y_i \log\Lambda(X_i\beta) + \sum_{i=1}^{N}(1-y_i)\log(1-\Lambda(X_i\beta)). \tag{3.8}$$

Due to fact that

$$\frac{\partial\Lambda(X_i\beta)}{\partial\beta} = \Lambda(X_i\beta)(1-\Lambda(X_i\beta))X_i', \tag{3.9}$$

the gradient is given by

$$G(\beta) = \frac{\partial l(\beta)}{\partial\beta} = -\sum_{i=1}^{N}(\Lambda(X_i\beta))X'_i + \sum_{i=1}^{N} X'_i y_i, \tag{3.10}$$

and the Hessian matrix is given by

$$H(\beta) = \frac{\partial^2 l(\beta)}{\partial\beta\partial\beta'} = -\sum_{i=1}^{N}\Lambda(X_i\beta)(1-\Lambda(X_i\beta))X'_i X_i. \tag{3.11}$$

By using prospects' profile variables as explanatory variables in the logistic regression model, the authors estimated an equation to predict each prospect's probability of responding to the acquisition offer. The estimated model can then be used to understand prediction accuracy within the same dataset (i.e., in-sample prediction) or score prospects not in the dataset for prediction (i.e., out-of-sample prediction). Prospects then can be sorted based on the predicted probability for the purpose of selection and optimal resource allocation.

Wangenheim and Bayón [11] argued that customer satisfaction influenced the number of WOM referrals which had an impact on customer acquisition. They asserted that the reception of a WOM referral has an increased marginal effect on the likelihood of a prospect to purchase. And they managed to answer two interesting questions related to customer acquisition: first, whether prospects' purchase likelihood is a function of WOM referrals; and, second, whether the characteristics of the WOM referrals resource including source expertise and similarity influence the probability that received WOM induces a purchase behavior. Since the authors conducted empirical studies in the energy market where the number of competitors is relatively small, they used customers' switching behavior as the dependent variable in modeling customer acquisition. To answer the first question, they only included the variable indicating the reception of WOM referral as the independent variable and used a binary logistic regression model to examine the effect that the independent variable has on the dependent variable:

$$prob(y = 1) = \frac{\exp(\beta \cdot x)}{1 + \exp(\beta \cdot x)} \tag{3.12}$$

where $y$ indicated the switching behavior, and the binary independent variable $x$ indicated whether WOM referral had been received. Thus, the effect of WOM referral on switching was obtained by subtracting the switching probability of a customer who had not received WOM $(x = 0)$ from that of a customer who had received WOM $(x = 1)$. To answer the second question, the authors only included the sample of prospects who had received WOM referral in the modeling procedure and also used a logit model to estimate whether switching behavior is a function of WOM referral resource characteristics, source expertise and similarity. which were included as independent variables, conditioning on receiving WOM referral:

$$prob(y = 1|x = 1) = \frac{\exp(\alpha \cdot z)}{1 + \exp(\alpha \cdot z)} \tag{3.13}$$

where $z$ indicated the two independent variables, source expertise and source similarity.

A binary logit model is not the only model that can be used for response probabilities. The type of model depends on the assumption of the distribution of the error term. Referring back to Equation 3.1, if the error term $\varepsilon$ has a logistic distribution, the corresponding model is a logit model; if the error term has a standard normal distribution, the corresponding model is a probit model. In this case the cumulative distribution function of $\varepsilon$ for a probit model is

$$F(x_i\beta) = \Phi(x_i\beta) = \int_{-\infty}^{x_i\beta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz. \quad (3.14)$$

The probit model is also estimated by the MLE method. Following Franses and Paap (2001), the relevant likelihood function is given by

$$L(\beta) = \prod_{i=1}^{N} (\Phi(X_i\beta))^{y_i} (1 - \Phi(X_i\beta))^{1-y_i}, \quad (3.15)$$

and the corresponding log-likelihood function is

$$l(\beta) = \sum_{i=1}^{N} y_i \log \Phi(X_i\beta) + \sum_{i=1}^{N} (1 - y_i) \log(1 - \Phi(X_i\beta)). \quad (3.16)$$

Differentiating $l(\beta)$ with respect to $\beta$ gives

$$G(\beta) = \frac{\partial l(\beta)}{\partial \beta} = -\sum_{i=1}^{N} \frac{y_i - \Phi(X_i\beta)}{\Phi(X_i\beta)(1 - \Phi(X_i\beta))} \phi(X_i\beta) X'_i, \quad (3.17)$$

and the Hessian matrix

$$H(\beta) = \frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'} = \sum_{i=1}^{N} \frac{\phi(X_i\beta)^2}{\Phi(X_i\beta)(1 - \Phi(X_i\beta))} X'_i X_i. \quad (3.18)$$

Similar to the logit model, the probit model is often used to model binary response variables, especially in cases where there is a desire to estimate a two-stage model. The probit model, while more complicated to estimate due to its lack of a closed-form solution, is theoretically appealing in a two-stage modeling framework due to its standard normal distribution. This makes it useful when the second stage is a least squares regression with a normally distributed error term. For example, Reinartz *et al.* [7] linked customer acquisition and relationship duration together using a probit two-stage least squares model. These authors used a probit model to determine the selection or acquisition process as shown by Equation 3.1 and predicted each prospect's response probability. The estimated probability was then included in the duration model as an independent variable to account for the interaction between acquisition and duration.

As the prospects' response likelihood is not always behaviorally observed, attitudinal propensity scale is also used to measure prospects' response intention. One might argue that attitudinal propensity measures are not perfectly correlated with behaviors, but researchers sometimes still use attitudinal dependent variables considering that attitude is usually positively related with behavior. Lix *et al.* [1] used two dichotomous attitudinal dependent variables, one indicating whether joined the membership (attitudinal) and the other one indicating whether used direct marketing channels (behavioral). The authors also used two interval-scale dependent variables, indicating prospects' attitudinal propensity toward pro-environment and US products. The authors included 150 independent variables containing prospects' demographic and life-style information in the regression and log-linear analysis. A brief introduction to log-linear analysis is provided in Appendix B. As Hansotia and Wang (1995), they used the estimated model to score samples in a holdout dataset and measured the effectiveness of their estimated models.

Whether it is the case that a logit or probit framework is used to model response probability, the output of the model is quite useful for determining which customer the firm is likely to acquire. In addition, the results of the binary choice model can provide the drivers of customer acquisition which can be useful for managers to make decisions in future customer acquisition campaigns.

### 3.2.1 Empirical example: Response probability

One of the key questions we want to answer with regard to customer acquisition is whether we can determine which future prospects have the highest likelihood of adoption. To do this we first need to know which past prospects were acquired and which were not. In the dataset provided for this chapter we have a binary variable which identifies whether or not a prospect was acquired by the firm (and hence became a customer) and a set of drivers which are likely to help explain a customer's decision to adopt. At the end of this example you should be able to identify the following:

1. The drivers of customer acquisition likelihood.
2. The parameter estimates from the response probability model.

A B2B firm wants to improve the acquisition rate of customers and reduce the acquisition spending on prospects by better understanding which prospects are most likely to adopt. A random sample of 500 prospects (some of whom became customers) was taken from a recent prospect list. The information we need for our model includes the following list of variables:

| | |
|---|---|
| **Dependent variable** | |
| *Acquisition* | *1 if the prospect was acquired, 0 otherwise* |
| **Independent variables** | |
| *Acq_Expense* | *Dollars spent on marketing efforts to try and acquire that prospect* |
| *Acq_Expense_SQ* | *Square of dollars spent on marketing efforts to try and acquire that prospect* |
| *Industry* | *1 if the prospect is in the B2B industry, 0 otherwise* |
| *Revenue* | *Annual sales revenue of the prospect's firm (in millions of dollars)* |
| *Employees* | *Number of employees in the prospect's firm* |

In this case, we have a binary dependent variable (*Acquisition*) which tells us whether the prospect did adopt (= 1) or did not adopt (= 0). We also have five independent variables we believe will be drivers of adoption. First, we have how many dollars the firm spent on each prospect (*Acq_Expense*) and the squared value of that variable (*Acq_Expense_SQ*). We want to use both the linear and squared term since we expect that for each additional dollar spent on the acquisition effort for a given prospect, there will be a diminishing return to the value of that dollar. Second, since the focal firm of this example is a B2B firm, the other three variables are firmographic variables of the prospects. These include whether the prospect sells to B2B (= 1) or B2C (= 0) customers (*Industry*), how much (in millions) that the prospect firms makes in annual revenue (*Revenue*), and how many employees the prospect firm has (*Employees*).

First, we need to model the probability that a prospect will adopt. Since our dependent variable (*Acquisition*) is binary, we select a logistic regression using the modeling framework as described earlier in the chapter (see Equation 3.1). We could also select a probit model and in general achieve the same results. In this case the $y$ variable is *Acquisition* and the $x$ variables represent the five independent variables in our database. When we run the logistic regression we get the following result:

| Variable | Estimate | Standard error | $p$-value |
|---|---|---|---|
| *Intercept* | −26.206 | 3.537 | <0.0001 |
| *Acq_Expense* | 0.067 | 0.010 | <0.0001 |
| *Acq_Expense_SQ* | −0.000 04 | 0.000 008 | <0.0001 |
| *Industry* | 0.033 | 0.389 | 0.9326 |
| *Revenue* | 0.032 | 0.012 | 0.0070 |
| *Employees* | 0.005 | 0.001 | <0.0001 |

As we can see from the results, four of the five independent variables are significant at a $p$-value of 5% or better with only *Industry* being statistically non-significant. First, this means that acquisition expense has a positive, but diminishing effect (*Acq_Expense* $> 0$ and *Acq_Expense_SQ* $< 0$) on acquisition likelihood. Second, it suggests that a prospect who is B2B (vs. B2C) will not matter in terms of acquisition likelihood, all else being equal. Third, the higher the *Revenue* that the prospect has, the more likely the prospect will be acquired. And, finally, the more *Employees* the prospect has, the more likely the prospect will be acquired.

It is also important to understand exactly how changes in the drivers of acquisition likelihood are likely to lead to either increases or decreases in acquisition likelihood. To do this we need to determine the odds ratio for each of the parameter estimates. Since we are dealing with a logistic regression, this means that we are interested in the log-odds ratio. For example, for *Revenue* = x,

$$\text{Odds}(\text{Acquisition}|\text{Revenue} = \text{x}) = \exp(-26.206 + 0.032\text{x}),$$

and, for *Revenue* = x + 1,

$$\text{Odds}(\text{Acquisition}|\text{Revenue} = \text{x} + 1) = \exp(-26.206 + 0.032(\text{x} + 1)).$$

By dividing the second equation by the first we get

$$\frac{\text{Odds}(\text{Acquisition}|\text{Revenue} = \text{x} + 1) = \exp(-26.206 + 0.032(\text{x} + 1))}{\text{Odds}(\text{Acquisition}|\text{Revenue} = \text{x}) = \exp(-26.206 + 0.032\text{x})}.$$

We then simplify the equation to get the following:

$$\frac{\text{Odds}(\text{Acquisition}|\text{Revenue} = \text{x} + 1)}{\text{Odds}(\text{Acquisition}|\text{Revenue} = \text{x})} = \exp(0.032) = 1.033$$

When we compute the log-odds ratio for each of the statistically significant variables we get the following results for an increase in 1 unit of the independent variable.

| Variable | Log-odds ratio |
|---|---|
| *Acq_Expense* | exp(0.06696−0.00008***Acq_Expense*) |
| *Revenue* | 1.033 |
| *Employees* | 1.005 |

We gain the following insights from the log-odds ratios. With regard to *Acq_Expense*, we see that the odds ratio is dependent on the level of *Acq_Expense*. This is due to the fact that we include both the level and squared terms for *Acq_Expense*. For example, if we usually spend \$500 on a given prospect, by spending \$501 we should see an increase in the likelihood of acquisition by exp(0.06696 −0.00008*500) = exp(0.0296) = 1.027. This means that by increasing our spending from \$500 to \$501, we should see an increase in acquisition likelihood by 2.7%. It is also important to note that this will vary depending on the initial level of *Acq_Expense*. With regard to *Revenue*, we see that for each increase in *Revenue* by \$1 million the acquisition likelihood should increase by 3.3%. Finally with regard to *Employees*, we see that for each increase in *Employees* by 1 person the acquisition likelihood should increase by 0.5%.

As a result we now know how changes in acquisition expense and changes in prospect characteristics are likely to either increase or decrease our likelihood of acquisition. This information can provide significant insights to managers who are charged with determining the optimal amount of resources to spend on acquisition.

## 3.2.2 How do you implement it?

To implement the logistic regression in this example we used the PROC Logistic feature in SAS. While we did use SAS to estimate the model, many other statistical packages are capable of estimating a logistic regression including (but not limited to) SPSS, MATLAB, and GAUSS.

## 3.3 Number of newly acquired customers and initial order quantity

Companies will always want to acquire as many customers as possible and acquire those customers who will bring large orders for profit maximization. Companies usually perform certain marketing activities to encourage newly acquire customers to buy more the first time and in subsequent purchases. The operation of shipping fees, such as free shipping for large orders or normal shipping fees charged for small orders, is one common way to influence the purchase behavior of potential customers. Lewis [2] examined the data from an online retailing company which experimented with multiple shipping fee schedules for 502 days, including large order size incentives and penalties. The author wanted to examine the effects shipping fees on order incidence, order size, shipping revenues, and customer acquisition simultaneously. Thus, Lewis developed a system of linear regressions to account for the possible correlation among the dependent variables. Among these regressions equations, the author modeled the number of newly acquired customers and the average order size for new customers in the following two equations:

$$CA_{cq} = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_3 + \beta_5 x_4 + \beta_6 x_5 + \varepsilon_1, \tag{3.19}$$

where $CA_{cq}$ denotes the number of newly acquired customers, $x_1$ to $x_5$ denote the explanatory variables, including shipping, pricing, and customer base terms, and

$$Amount(new) = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_3 + \beta_5 x_4 + \beta_6 x_5 + \varepsilon_2, \tag{3.20}$$

where $Amount(new)$ denotes the average order size for new customers, and $x_1$ to $x_5$ denote the shipping variables and pricing variable. And, to account for the possible correlation between the equations and the possible endogeneity of explanatory variables, the author used three-stage least squares in the estimation and used the lagged values of each quantity as instruments.

Besides shipping fees, Villanueva *et al.* [3] argued that marketing channels influence the customer acquisition process and the value that newly acquired customers will bring to the company. These authors developed two functions to measure how much new customers acquired through different acquisition channels, including WOM and marketing activities, contribute to firm performance. One function is called ‘the value-generating function,’ which links newly acquired customers’ contributions to the firm’s equity growth, and the other function is called ‘the acquisition response function,’ which expresses the interactions between marketing spending and the number of acquisitions. The authors argued that it was important to incorporate all the indirect effects of each acquisition on customer equity, not just each customer’s direct financial contribution to the company. As the estimated variables are treated as potentially endogenous, the authors used a three-variable vector autoregression (VAR) modeling technique:

$$\begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \end{pmatrix} + \sum_{l=1}^{p} \begin{pmatrix} a_{11}^l & a_{12}^l & a_{13}^l \\ a_{21}^l & a_{22}^l & a_{23}^l \\ a_{31}^l & a_{32}^l & a_{33}^l \end{pmatrix} \begin{pmatrix} x_{t-l} \\ y_{t-l} \\ z_{t-l} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{pmatrix}, \tag{3.21}$$

where $x$ denotes the number of customers acquired through the firm's marketing actions, $y$ denotes the number of customers acquired through WOM, and $z$ denotes the firm's performance. The subscript $t$ denotes time, $p$ denotes the lag order of the model, and $(e_{1t,}\ e_{2t},\ e_{3t})'$ are white-noise disturbances distributed as $N(0, \Sigma)$. The direct effects are captured by $a_{31}$, $a_{32}$; the feedback effects are captured by $a_{13}$, $a_{23}$; the cross-effects are captured by $a_{12}$, $a_{21}$; and the reinforcement effects are captured by $a_{11}$, $a_{22}$, $a_{33}$. The authors used impulse response functions (IRFs) to estimate the long-term effect of an unexpected shock in one variable on the other variables in the system. Dekimpe and Hanssens (2004) give a review of VAR models under the persistence modeling framework and we provide the related introduction in Appendix C.

Both of the previous examples we have discussed deal with cross-sectional time series data, or snapshots of cross-sections of customers that are repeatedly observed over time. There is a great benefit to having cross-sectional time series data – mainly because the effects that are uncovered by the modeling frameworks are persistent over time across customers. However, it might be the case that one of two limitations are present that do not enable you to use cross-sectional time series data. First, it is possible that the data you have is not repeatedly measured over time, that is, it is just cross-sectional. In this case you cannot use a time series modeling framework such as a VAR model. Second, it may be the case that there is no variation over time in any of your key drivers. For instance, if you always charge the same shipping costs over time, repeatedly measuring shipping costs in multiple time periods makes no sense. Thus, we also need to have a method available for instances when we only have cross-sectional data.

When only cross-sectional data are available it is rather straightforward to predict the number of newly acquired customers and the expected initial order quantity. For the number of newly acquired customers, all that we need to do is to determine the expected response probability of each prospect, as outlined in the previous subsection, and aggregate the prediction across all prospects. We can then compare the actual number of acquired customers to the predicted number of acquired customers.

For the initial order quantity we can run a two-stage regression, similar to the two-stage least squares regression as outlined in Reinartz *et al.* [7] where the acquisition probability is modeled using a probit model and the initial order quantity is modeled as a least squares regression that is conditional on the prospect being acquired. The result of this two-stage model is a prediction of expected initial order quantity conditional on acquisition probability that can be applied to any future prospect.

Thus, regardless of whether the data you have are cross-sectional or cross-sectional time series, this section describes various models that can be used for both number of newly acquired customers and initial order quantity.

### 3.3.1 Empirical example: Number of newly acquired customers

Besides being able to predict whether or not we are likely to acquire a prospect as we did in the previous example, we are also interested in determining how well our response probability model does in helping us accurately predict the total number of customers we are likely to acquire and specifically which prospects we are most likely to acquire. Thus, at the end of this example we should be able to do the following:

1. Predict the number of prospects we are likely to acquire.

2. Determine the accuracy of our prediction.

The information we need for this prediction includes the following list of variables:

| | |
|---|---|
| **Dependent variable** | |
| *Acquisition* | *1 if the prospect was acquired, 0 otherwise* |
| **Independent variables** | |
| *Acq_Expense* | *Dollars spent on marketing efforts to try and acquire that prospect* |
| *Acq_Expense_SQ* | *Square of dollars spent on marketing efforts to try and acquire that prospect* |
| *Industry* | *1 if the prospect is in the B2B industry, 0 otherwise* |
| *Revenue* | *Annual sales revenue of the prospect's firm (in millions of dollars)* |
| *Employees* | *Number of employees in the prospect's firm* |

In this case we only have cross-sectional data, that is, data from a single snapshot in time. Thus, we need to use the estimates we obtained from the response probability model in the first example to help us determine the predicted probability that each prospect will adopt. To do this we use the parameter estimates from the response probability model and values for the $x$ variables to predict whether a customer is likely to be acquired. For a logistic regression we must apply the proper probability function as noted earlier in the chapter (see Equation 3.5):

$$\mathrm{P}(Acquisition = 1|X\beta) = \frac{1}{1 + \exp(-X\beta)}.$$

For example, for Customer 9, when we input the statistically significant variables into the $X\beta$ computation we get $\mathrm{P}_{[\text{Customer} = 9]}(\text{Acquisition} = 1|X\beta) = 0.06$ or 6%. Once we do this for each of the customers we can then decide which prospect we believe we are likely to acquire given our acquisition spending and on each prospect's characteristics.

Next, we need to create a cutoff value to determine at which point we are going to divide the prospects into the two groups – predicted to acquire and predicted not to acquire. There is no rule that explicitly tells us what that cutoff number should be. Often by default we select 0.5 since it is equidistant from 0 and 1. However, it is also reasonable to check multiple cutoff values and choose the one that provides the best predictive accuracy for the dataset. By using 0.5 as the cutoff for our example, any prospect whose predicted probability of acquisition is greater than or equal to 0.5 is classified as predicted to acquire, and the rest are predicted as not to acquire. To determine the predictive accuracy we compare the predicted to the actual acquisition values in a $2 \times 2$ table. For our sample of 500 we get Table 3.3.

As we can see from the table, our in-sample model accurately predicts 90.6% of the prospects who chose not to adopt (183/202) and 91.6% of the prospects who chose to adopt (273/298). This is a significant increase in the predictive capability of a random guess model[1] which would be only 58.4% accurate for this dataset. Since our model is significantly better than the best alternative, in this case a random guess model, we determine that the predictive accuracy of the model is good. If there are other benchmark models available for comparison, the 'best' model would be the one which provides the highest accuracy of both the prediction to acquire and not to acquire, or in other words the prediction would provide the highest sum of the diagonal. In this case the sum of the diagonal is 456 and it is accurate 91.2% of the time (456/500).

### 3.3.2 How do you implement it?

The implementation of this example was carried out using a SAS Data step and the Freq procedure. First we computed the predicted response probabilities for each prospect using the coefficients from the response probability example and scored the prospects as predicted to acquire (1) if the value was higher than the cutoff (0.5), and predicted not to acquire (0) if the value was lower than the cutoff. Next we ran PROC Freq and created a $2 \times 2$ table which compared the actual and predicted acquisition values to determine how well our model predicted the number of newly acquired customers.

Table 3.3 Predicted versus actual acquisition.

| | | Actual acquisition | | |
|---|---|---|---|---|
| | | 0 | 1 | *Total* |
| Predicted acquisition | 0 | *183* | *19* | ***202*** |
| | 1 | *25* | *273* | ***298*** |
| | *Total* | ***208*** | ***292*** | ***500*** |

[1] A random guess model would do the following. First, it would determine which bucket acquire or not acquire has more prospects in it. In this case 292 prospects were acquired versus 208 who were not. Then it would predict that all prospects would be acquired and it would be accurate 58.4% of the time (292/500).

### 3.3.3 Empirical example: Initial order quantity

Many firms have realized that it is not sufficient to merely focus on just trying to acquire as many customers as possible without any concern for the value that the customer is likely to provide. Research in marketing has shown that the initial order value can be a valuable predictor in a customer's future value to the firm – or at the least justify the amount of money that is spent on customer acquisition. Thus, it can be useful to understand the drivers of initial order value and in turn be able to predict each prospect's expected initial order value. At the end of this example we should be able to do the following:

1. Determine the drivers of initial order quantity (value).
2. Predict the expected initial order quantity for each prospect.
3. Determine the predictive accuracy of the model.

The information we need for this model includes the following list of variables:

| | |
|---|---|
| **Dependent variables** | |
| *Acquisition* | *1 if the prospect was acquired, 0 otherwise* |
| *First_Purchase* | *Dollar value of the first purchase (0 if the customer was not acquired)* |
| **Independent variables** | |
| *Lambda(λ)* | *The computed inverse Mills ratio from the acquisition model* |
| *Acq_Expense* | *Dollars spent on marketing efforts to try and acquire that prospect* |
| *Acq_Expense_SQ* | *Square of dollars spent on marketing efforts to try and acquire that prospect* |
| *Industry* | *1 if the prospect is in the B2B industry, 0 otherwise* |
| *Revenue* | *Annual sales revenue of the prospect's firm (in millions of dollars)* |
| *Employees* | *Number of employees in the prospect's firm* |

Again, we need to note that we are dealing with cross-sectional data. We see from the data requirement that in order to determine the drivers of initial order quantity we need to have two dependent variables: *First_Purchase* and *Acquisition*. This is due to the fact that expected initial order quantity is derived from the following equation:

$$\mathrm{E}(\mathit{Initial\ Order\ Quantity}) = \mathrm{P}(\mathit{Acquisition} = 1)$$
$$^{*}\mathrm{E}(\mathit{First\_Purchase} | \mathit{Acquisition} = 1).$$

This equation shows us that the expected initial order quantity is a function of the probability that the prospect will be acquired multiplied by the expected value of a purchase given that the prospect was acquired. If we were to merely run a regression with *First_Purchase* as the dependent variable and ignore the probability that the prospect will be acquired, we would get biased estimates due to a potential sample selection bias.

Sample selection bias is a problem that is common in many marketing problems and has to be statistically accounted for in many modeling frameworks. In this case the prospect has a choice of whether or not to be acquired before making a purchase. If we were to ignore this choice we would bias the estimates from the model and we would have fewer precise predictions for the value of *First_Purchase*. To account for this issue we need to be able to predict the value for both the probability of *Acquisition* (similar to what we have done for the first empirical example) and the expected value of *First_Purchase* given that the prospect is expected to be acquired. One important consideration to note is that we cannot just run two models independently since there is likely to be a correlation between the error terms of the two models. Thus, we need to use a modeling framework that can simultaneously estimate the coefficients of the two models, or at least account for the correlation between *First_Purchase* and *Acquisition*. To do this we use a two-stage modeling framework similar to that described earlier in this chapter and found in Reinartz *et al.* [7].

The first model for *Acquisition* will be set up using the same equation as for the response probability example (see Equation 3.1). The only difference here is that instead of using a logistic regression we will be using a probit model to estimate the coefficients. The main reason for this lies in the error term of the probit model which is distributed normal with a mean of 0 and a standard deviation of 1. The fact that the probit model and the OLS regression model (which we will be using for *First_Purchase*) are both normally distributed allows us to more easily estimate them in a two-stage framework.

Once we estimate the probit model we need to create a new variable, $\lambda$, which will represent the correlation in the error structure across the two equations. This variable, also known as the sample selection correction variable, will then be used as an independent variable in the *First_Purchase* model to remove the sample selection bias in the estimates. To compute $\lambda$ we use the following equation, also known as the inverse Mills ratio:

$$\lambda = \frac{\phi(X'\beta)}{\Phi(X'\beta)}.$$

In this equation $\phi$ represents the normal probability density function, $\Phi$ represents the normal cumulative density function, $X$ represents the value of the variables in the *Acquisition* model, and $\beta$ represents the coefficients derived from the estimation of the *Acquisition* model.

Finally, we want to estimate a regression model for *First_Purchase* and include the variable $\rho$ as an additional independent variable. This is done in a straightforward manner with the following equation:

$$First\_Purchase = \gamma'\alpha + \mu\lambda + \varepsilon.$$

In this case *First_Purchase* is the value of the initial order quantity, $\gamma$ is the matrix of variables used to help explain the value of *First_Purchase*, $\alpha$ are the coefficients for the independent variables, $\mu$ is the coefficient of the inverse Mills ratio, $\lambda$ is the inverse Mills ratio, and $\varepsilon$ is the error term.

When we estimate the two-stage model, we get the following parameter estimates for each of the two equations:

| Acquisition model | Estimate | *First_Purchase* model | Estimate |
|---|---|---|---|
| *Intercept* | −15.134 | *Intercept* | −62.261 |
| *Acq_Expense* | 0.039 | *Acq_Expense* | 0.705 |
| *Acq_Expense_SQ* | −0.000 02 | *Acq_Expense_SQ* | −0.0008 |
| *Industry* | 0.051[a] | *Industry* | −2.764[a] |
| *Revenue* | 0.018 | *Revenue* | 3.026 |
| *Employees* | 0.003 | *Employees* | 0.254 |
| | | *Lambda* ($\lambda$) | 19.101 |

[a]Denotes not significant at $p < 0.05$.

We gain the following insights from the results. We see that $\lambda$ is positive and significant. We can interpret this to mean that there is a potential selection bias problem since the error term of our selection equation is correlated positively with the error term of our regression equation. We also see that all other variables of the *First_Purchase* model are significant with the exception of *Industry*, meaning that we have uncovered many of the drivers of initial order quantity. We find that *Acq_Expense* is positive with a diminishing return, as noted by the positive coefficient on *Acq_Expense* and the negative coefficient on *Acq_Expense_SQ*. We also find that two of the firm characteristic variables are positive (*Revenue* and *Employees*) showing that firms with higher annual revenue and firms with more employees tend to have larger initial order quantities.

Our next step is to predict the value of *First_Purchase* to see how well our model compares to the actual values. We do this by starting with the equation for expected initial order quantity at the beginning of this example:

$$\mathrm{E}(\textit{Initial Order Quantity}) - \mathrm{P}(\textit{Acquisition} - 1)^{*}\mathrm{E}(\textit{First\_Purchase} | \textit{Acquisition} - 1) \\ = \Phi(X'\beta)^{*}(\gamma'\alpha + \mu\lambda).$$

In this case $\Phi$ is the normal cumulative distribution function (CDF), $X$ is the matrix of independent variable values from the *Acquisition* equation, $\beta$ is the vector of parameter estimates from the *Acquisition* equation, $\gamma$ is the matrix of independent variables from the *First_Purchase* equation, $\alpha$ is the vector of parameter estimates from the *First_Purchase* equation, $\mu$ is the parameter estimate for the inverse Mills ratio, and $\lambda$ is the inverse Mills ratio. Once we have predicted the *First_Purchase* value for each of the prospects, we want to compare this to the actual value from the database. We do this by computing the mean absolute deviation (MAD) and mean absolute percent error (MAPE). The equations are as follows:

$$\mathrm{MAD} = \mathrm{Mean}\{\text{Absolute Value}[\mathrm{E}(\textit{Initial Order Quantity}) - \textit{First\_Purchase}]\}$$

$$\mathrm{MAPE} = \mathrm{Mean}\{\text{Absolute Value}[(\mathrm{E}(\textit{Initial Order Quantity}) \\ - \textit{First\_Purchase})/\textit{First\_Purchase}]\}.$$

We find for the acquired customers that MAD = 51.96 and MAPE = 18.69%. This means that on average each of our predictions of *First_Purchase* deviates from the actual value by $51.96 or 18.69% (based on a Mean(*First_Purchase*) = $372.47). If we were instead to use the mean value of *First_Purchase* ($372.47) as our prediction for all prospects (this would be the benchmark model case), we would find that MAD = 127.17 and MAPE = 135.48%. Thus, our model does a significantly better job of predicting the value of initial order quantity than the benchmark case.

### 3.3.4 How do you implement it?

In this example we used a two-stage least squares approach with a probit model for acquisition and a least squares regression for the initial order quantity. We used multiple procedures in SAS to implement this model. First we used PROC logistic with a probit link function to estimate the model of customer acquisition. Next we used a SAS Data step to compute the inverse Mills Ratio using the output of the probit model. Finally we ran an OLS regression using PROC Reg and added the inverse Mills ratio as an additional variable. While we did use SAS to implement this modeling framework, programs such as SPSS can be used as well.

## 3.4 Duration/time

In contractual settings, the time customers spent with a company is critical to the maintenance and development of a business. Lewis [4] argued that customer acquisition promotions influence the duration of customer–company relationships. In the newspaper subscription industry, the author used survival analysis to model the relationship between discount depth and the length of time as a subscriber. The author adopted accelerated failure time models (Kalbfleish and Prentice, 1980) estimated as shown in the following specification:

$$\log(t_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \sigma\varepsilon_i, \tag{3.22}$$

where $\varepsilon_i$ is a random disturbance term, $x_i$ and $x_i^2$ are the discount and discount squared, and $\beta_s$ and $\sigma$ are parameters to be estimated. The author varied the restrictions on $\varepsilon$ and $\sigma$ and specified various forms of the hazard function. The author used the exponential model as a baseline model, which assumes that the error term $\varepsilon$ has a standard extreme-value distribution and that $\sigma$ is equal to 1. The author also used two generalized gamma models: one included only the discount as an independent variable, and the other included also the quadratic form of discount.

Kalbfleish and Prentice (1980) gave a brief introduction of the accelerated failure time model as below. Suppose that $Y = \log T$ is related to the covariate $z$ via a linear model $Y = z\beta + W$, where $W$ is an error variable with density $f$. Exponentiation gives $T = \exp(z\beta)T'$ where $T' = \exp(w) > 0$ has hazard function $\lambda_0(t')$, say, that is independent of $\beta$. It follows that the hazard function for $T$ can be written in terms of this baseline hazard $\lambda_0(\cdot)$ according to

$$\lambda(t;z) = \lambda_0(t\exp(-z\beta))\exp(-z\beta). \tag{3.23}$$

The survival function is

$$F(t;z) = \exp\left[-\int_0^{t\exp(-z\beta)} \lambda_0(u)\,du\right] \tag{3.24}$$

and the density function is the product of Equations 3.15 and 3.16. To estimate the accelerated failure time model, the MLE method is usually used. Fanses and Paap (2001) introduced the estimation procedure of the accelerated failure time model for the Weibull distribution. We provide the reference in Appendix D.

Schweidel *et al.* [12] argued that time until acquisition, which is defined as the time that elapses before a prospective customer acquires a particular service, influences the duration that customers stay with the company. Using data from telecommunication services, the authors used a customer-level bivariate timing model for the time until acquisition of customers who eventually acquired service in the observation period. The authors assumed a parametric distribution for the probability of acquiring service at time $t_A$:

$$f(t_A|\Theta) = S_A(t_A - 1|\Theta) - S_A(t_A|\Theta) \tag{3.25}$$

where $S_A(t_A|\Theta)$ is the survival function of the parametric distribution, $t_A = 1, 2, \ldots, T$, and $T$ is the length of the observation period. In the empirical studies, the authors considered three sets of possible baseline hazard specifications for the acquisition processes: the Weibull, log-logistic and expo-power distribution. The hazard function and survival function of these three baseline distributions are listed in Table 3.4. We provide an introduction to survival analysis in Appendix G.

Table 3.4 Hazard function and survival function of three baseline distributions.

| Baseline distribution | Hazard function | Survival function |
|---|---|---|
| Weibull | $h(t\|\gamma,\alpha) = \gamma\alpha t^{\alpha-1}$ | $S(t\|\gamma,\alpha) = \exp(-\gamma t^{\alpha})$ |
| Log-logistic | $h(t\|\gamma,\alpha) = \frac{\gamma\alpha(\gamma t)^{\alpha-1}}{1+(\gamma t)^{\alpha}}$ | $S(t\|\gamma,\alpha) = \frac{1}{1+(\gamma t)^{\alpha}}$ |
| Expo-power | $h(t\|\gamma,\alpha,\theta) = \gamma\alpha t^{\alpha-1}\exp(\theta t^{\alpha})$ | $S(t\|\gamma,\alpha,\theta) = \exp(\gamma/\theta)^{(1-\exp(\theta t^{\alpha}))}$ |

Source: Schweidel, Fader, and Bradlow (2008).

### 3.4.1 Empirical example: Duration/time

Customer acquisition is of key importance to many firms. However, many firms also want to look beyond just the point of acquisition and answer questions such as, 'How long will a newly acquired customer still be a customer?' So in this example we will try to uncover the drivers of new customer duration. We expect that customer duration will be a function of the customer's exchange characteristics and firmographic information. But, we also want to see if the acquisition effort by the firm affects each new customer's duration with the firm. At the end of this example we should be able to:

1. Determine the drivers of new customer duration.
2. Predict the duration of each new customer.

The information we need for this model includes the following list of variables:

| | |
|---|---|
| **Dependent variables** | |
| *Duration* | *The time in days that the acquired prospect has been or was a customer, right-censored at 730 d* |
| **Independent variables** | |
| *Acq_Expense* | *Dollars spent on marketing efforts to try and acquire that prospect* |
| *Acq_Expense_SQ* | *Square of dollars spent on marketing efforts to try and acquire that prospect* |
| *Ret_Expense* | *Dollars spent on marketing efforts to try and retain that customer* |
| *Ret_Expense_SQ* | *Square of dollars spent on marketing efforts to try and retain that customer* |
| *Crossbuy* | *The number of categories the customer has purchased* |
| *Frequency* | *The number of times the customer purchased during the observation window* |
| *Frequency_SQ* | *The square of the number of times the customer purchased during the observation window* |
| *Industry* | *1 if the prospect is in the B2B industry, 0 otherwise* |
| *Revenue* | *Annual sales revenue of the prospect's firm (in millions of dollars)* |
| *Employees* | *Number of employees in the prospect's firm* |
| *Censor* | *1 if the customer was still a customer at the end of the observation window, 0 otherwise* |

In this case the sample data are from a B2B firm that actually observes when a customer churns, that is, a contractual setting. We see from the list of variables that we have one dependent variable (*Duration*) which is right-censored. Our observation window for the data is only two years (730 days) after the prospects were acquired. Thus, we only observe a customer leaving the firm if it happens before the end of the second year. As a result, the data for *Duration* is right-censored at 730 days – meaning that any customer whose *Duration* value is 730 in the data table has yet to leave the firm at the end of the observation window. We also have 10 independent variables which we hope will explain the variation in each new customer's duration with the firm.

These independent variables include the money the firm spent on acquisition (*Acq_Expense* and *Acq_Expense_SQ*), the money the firm spent on customer retention (*Ret_Expense* and *Ret_Expense_SQ*), customer exchange characteristics (*Crossbuy*, *Frequency*, and *Frequency_SQ*), and firmographic variables (*Industry*, *Revenue*, and *Employees*).

To determine the drivers of customer duration we use the accelerated failure time model similar to Equation 3.22 earlier in this chapter. In this case we have the following equation:

$$\ln(Duration) = X'\beta + \sigma\varepsilon,$$

where ln(*Duration*) is the natural logarithm of the time the customer was active with the firm, $X$ is a matrix of the 10 independent variables, $\beta$ is a vector of coefficients, $\sigma$ is the scale parameter, and $\varepsilon$ is the error term. We estimate the model by assuming that $\varepsilon$ follows a Weibull distribution – common among accelerated failure time models.[2] We estimate the model only with the 292 prospects that were acquired as customers. We get the following results from the estimation:

| Variable | Estimate | $p$-value |
|---|---|---|
| *Intercept* | 2.837 | <0.0001 |
| *Acq_Expense* | 0.007 | <0.0001 |
| *Acq_Expense_SQ* | −0.000 01 | <0.0001 |
| *Ret_Expense* | 0.001 | <0.0001 |
| *Ret_Expense* | −0.000 000 04 | 0.017 |
| *Crossbuy* | 0.098 | <0.0001 |
| *Frequency* | 0.111 | <0.0001 |
| *Frequency_SQ* | −0.001[a] | 0.173 |
| *Industry* | 0.524 | <0.0001 |
| *Revenue* | 0.012 | <0.0001 |
| *Employees* | 0.0001 | <0.0001 |
| *Scale* | 0.138 | |
| *Shape* | 7.252 | |

[a]Denotes not significant at $p < 0.05$.

We see that each of the variables in the model is significant to at least a level of 1% with the exception of *Frequency_SQ*. In addition we see that for both *Acq_Expense* and *Ret_Expense* there is a positive, but diminishing, effect on *Duration*. This suggests that the more the firm spends, up to a threshold, on both acquisition

[2] There are many other distributions which can be used including the exponential, gamma, normal, and log-normal among others. It is often worthwhile to try several distributions to see if the model fit and prediction improve.

and retention, the longer the customer is likely to stay with the firm. While we expect that spending on retention efforts is directly linked to the duration of a customer, we also see here that spending on acquisition also helps determine the duration of a customer. We see for *Crossbuy* that the more products the customer has purchased, the longer the expected *Duration*. We see a positive effect for *Frequency* with a negative, but not significant, squared term. This suggests that customers who do not purchase very often are more likely to churn than customers who purchase at a moderate to high frequency. We see that customers who are B2B have higher *Revenue*, and more *Employees* are more likely to have a longer *Duration*. Finally, we obtain both a *Scale* and a *Shape* parameter which will help us with the prediction of *Duration* for each of the customers.

Next, we want to see how well we are predicting *Duration* for the customers. Since we do not actually observe *Duration* for the customers who have yet to churn, we cannot validate our results on those customers. We can predict the time until expected churn for the customers who have yet to churn, but we can only test to see how well our results accurately predict *Duration* for the customers who have already churned (unless we used a holdout sample where we actually knew when the customers churned beyond two years).

So, in order to test the predictive accuracy of the model on the 157 non-censored customers who have already churned, we need to have an equation to predict the value of *Duration*. We start by referring back to the original equation we estimated:

$$\ln(Duration) = X'\beta + \sigma\varepsilon.$$

We just need to recognize that $\sigma$ is the scale parameter (0.138) and $\varepsilon$ is derived from the 50% percentile of the Weibull distribution, or $\ln(-\ln(1-p)) = \ln(-\ln(1-0.5)) = -0.367$. Then to compute *Duration* we just need take the inverse logarithm of the right hand side. We get the following:

$$Duration = \exp(X'\beta + 0.138^* - 0.367) = \exp(X'\beta - 0.051).$$

Now we need to compare the actual *Duration* values for the 157 customers who churned during the observation window to the predicted values of *Duration*. We find a MAD of 45.97 days and a MAPE of 13.88%. We can compare this to the benchmark case where our prediction of *Duration* is the mean of the non-censored values, which is 333 days. We get a MAD of 170.77 days and a MAPE of 171.92%. We see that our model does significantly better than the benchmark model in predicting the *Duration* of customers in our database.

### 3.4.2 How do you implement it?

We implemented this model using the PROC Lifereg procedure in SAS where the dependent variable was right-censored. It is also possible to estimate this model and other accelerated failure time models using other programs such as STATA and R, among others.

## 3.5 Firm's performance (LTV, CLV, and CE)

Making customer acquisition selection decisions on the basis of the response probability, initial order quantity, and duration is not enough. Companies should select prospects to acquire based on their lifetime contribution, which can be termed as LTV, CLV, or CE. Hansotia and Wang [5] in the last scenario of their studies considered that prospects' LTV varied due to their profiles and the promotion packages they received. These authors modeled the present value of revenue (PVR) and estimated LTV by subtracting the present value of cost of goods and the estimated marketing and operation costs for that customer. A common issue that occurs in the estimation of CLV is that observations may be right-censored. Some of the customers may have already lapsed in the observation period while others may still use the product or service beyond the observation period. Estimating the lifetime value of the customers who are still active over the observation period by OLS regression may underestimate their lifetime contribution to the companies. Hansotia and Wang [5] used the Tobit model to estimate the PVR for customers. The Tobit model assumes that there is a latent variable $y_i^*$ which linearly depends on $x_i$ via a vector of parameters $\beta$. And there is a normally distributed error $u_i$ to capture the random fluctuation of the relationship. For the right-censored situation in which censoring occurs above a value $y_L$, the Tobit model is

$$\begin{aligned} y_i &= \begin{cases} y_i^* & \text{if } y_i^* < y_L \\ y_L & \text{if } y_i^* \geq y_L \end{cases} \\ y^* &= \beta x_i + u_i,\ u_i \sim N(0, \sigma^2). \end{aligned} \tag{3.26}$$

The authors did not restrict the latent variables having a normal distribution but chose from a set of distributions including the gamma, log-normal, logistic, and Weibull distributions. In their studies, the authors presented the likelihood function, which is the probability of observing the sample values, as

$$L = \prod_{i=1}^{M} \{\text{Prob}[PVR = \textit{observed value}]\}^S \{\text{Prob}[PVR > \textit{observed value}]\}^{1-S} \tag{3.27}$$

where $S = 1$ if observation $i$ is uncensored, 0 otherwise. The authors estimated the Tobit model using the method of maximum likelihood by the Lifereg procedure in SAS. We provide the introduction and estimation of the type 1 Tobit model in Appendix E. For the other types of Tobit models, readers should refer to Amemiya (1985).

In non-contractual settings, customers can terminate the relationship with companies without any notice. Thus, it is not easy to identify whether the customers have lapsed or not and whether the observation is censored or not. Reinartz *et al.* [7] considered that the account is still active and the duration is right-censored if the expected time until the next purchase exceeds the recency since last purchase.

Using the data from a B2B high-tech manufacturer, Reinartz *et al.* [7] adopted a standard right-censored Tobit model to estimate the duration with independent variables including the estimated response probability. These authors further estimated customer profitability with a standard right-censored Tobit model using a set of exogenous variables and including the predicted duration and response probability. Similar to Reinartz *et al.* [7], Thomas *et al.* [8] developed an ARPRO (Acquisition, Retention, and PROfitability) model which uses a standard right-censored Tobit model to estimate and predict the total long-term customer profitability as a function of both the customer's predicted relationship duration and acquisition probability. The authors found that decreasing marketing spending for a B2B firm and catalog retailer and increasing spending to customers for a pharmaceutical firm would lead to increases in total customer profitability.

In an online grocery retailing industry, Lewis [4] separated customers' purchase activities in the observation period into three quarters, each quarter's activities expressed as RFM measures. The author used a Tobit model to estimate the three-quarter activity as a function of second-quarter RFM scores. The estimated model was then used to predict the fourth-quarter activity and the customer asset value was then calculated as a total of observed expenditures and the predicted fourth-quarter activity.

Blattberg and Deighton [9] proposed to find the optimal level of acquisition spending based on the maximization of first-year value. These authors solved this issue using a tool called decision calculus, an approach to decision making in which managers break down a complex problem into smaller and simpler problems which are then incorporated into a model to solve the original complex problem. To calculate the optimal level of acquisition spending that maximizes first-year value, the authors asked managers four questions to determine the acquisition expenditure per prospect $\$A$, the acquisition rate obtained as a result of that expenditure $a$, the ceiling rate (the highest possible number of customers the company could reasonably acquire in a given time period), and the margin generated by a customer in a year $\$m$. The authors then drew a curve of the actual acquisition probability from the equation

$$a = ceiling\,rate \times [1 - \exp(-k_1 \times \$A)] \tag{3.28}$$

where $k$ is a constant that controls the steepness of the curve. Next, the authors calculated the net contribution from acquiring a prospect in the first year $(a\$m - \$A)$ and drew the second curve showing how the average value of acquired customers in the first year varies with spending on acquisition. Combing the two curves, the authors found the optimal level of acquisition spending at the peak of the second curve where acquisition should stop.

The fundamental question encountered in managing customer acquisition is to model the probability of being acquired for customers. The logit or probit model is the most common method used since they are by nature used to model binary outcomes, such as being acquired or not. In addition, logit or probit models are used to examine the effect of different customer characteristics and purchase behavior on the probability. In some circumstances, researchers may want to model several acquisition metrics together, such as the number of newly acquired customers and the initial order quantity, and systems of linear regression are able to complete the task. In addition, it is possible that the acquisition metrics, such as the number of customers acquired through promotion and through networking, and company profitability, are correlated and variant across time, and VAR is a suitable modeling method to account for the correlation and time effects. Furthermore, researchers may want to model the duration of newly acquired customers and determine the effects that various explanatory variables have on the time to defect. Since most of the duration data contain censored observations, OLS regression would provide biased estimates. A classic parametric method for survival data is the accelerated failure time model, which can be estimated by the method of maximum likelihood. The variation of the accelerated failure time model, depending on the assumption of its disturbance distribution, such as the Weibull, exponential, gamma, log-logistic and log-normal distribution, is often used in duration modeling. Customer acquisition leads to firm profitability, which is the metric researchers have a strong interest to model. CLV is often used to represent firm profitability achievement and capability, and researchers often have to encounter censored observations while modeling CLV. Tobit models are essentially suitable methods in the modeling and have gained extensive popularity. Besides stochastic methods, researchers have used deterministic methods, such as decision calculus, to determine the optimal level of acquisition spending for profit maximization.

### 3.5.1 Empirical example: Firm's performance

The final step of the analysis is to determine whether the customers that were acquired are profitable. We also want to know whether we can determine the drivers of the customer profitability to see if future acquisition efforts can help lead to acquire a larger number of profitable customers. At the end of this example we should be able to:

1. Determine the drivers of customer profitability.

2. Determine the predictive accuracy of the customer profitability model.

The information we will need for this model includes the following list of variables for the 292 prospects that became customers ($Acquisition = 1$):

| | |
|---|---|
| **Dependent variables** | |
| *Censor* | *1 if the customer was still a customer at the end of the observation window, 0 otherwise* |
| *CLV* | *The predicted customer lifetime value score. It is 0 if the prospect was not acquired or has already churned from the firm (000s)* |
| **Independent variables** | |
| *Acq_Expense* | *Dollars spent on marketing efforts to try and acquire that prospect* |

| | |
|---|---|
| *Lambda(λ)* | *The computed inverse Mills ratio from the acquisition model* |
| *Acq_Expense_SQ* | *Square of dollars spent on marketing efforts to try and acquire that prospect* |
| *Ret_Expense* | *Dollars spent on marketing efforts to try and retain that customer* |
| *Ret_Expense_SQ* | *Square of dollars spent on marketing efforts to try and retain that customer* |
| *First_Purchase* | *Dollar value of the first purchase (0 if the customer was not acquired)* |
| *Crossbuy* | *The number of categories the customer has purchased* |
| *Frequency* | *The number of times the customer purchased during the observation window* |
| *Frequency_SQ* | *The square of the number of times the customer purchased during the observation window* |
| *Industry* | *1 if the prospect is in the B2B industry, 0 otherwise* |
| *Revenue* | *Annual sales revenue of the prospect's firm (in millions of dollars)* |
| *Employees* | *Number of employees in the prospect's firm* |

In this case we have two dependent variables. One dependent variable, called *Censor*, helps us to delineate between those customers who are still customers and those customers who have already left the firm. Our second dependent variable is called *CLV*, or customer lifetime value. For the sake of this example the CLV score (in thousands) for each of the acquired customers has been predicted for you. Since CLV is a forward-looking measure (i.e., it measures the expected future profitability of a customer), the only customers with a CLV score are those who are still customers with the firm. These are the customers who meet the following two criteria – they have a value of 1 for *Acquisition* and a value of 1 for *Censor*.

Similar to the example for initial purchase quantity, we need to take into account both the probability that the customer is still a customer (*Censor*) and the expected future profitability of the customer (*CLV*). To do this we get the following equation:

$$\mathrm{E}(Future\ Profitability) = \mathrm{P}(Censor = 1)^{*}\mathrm{E}(CLV|Censor = 1).$$

In this case we also recognize that we will have to account for potential sample selection bias and estimate the model in a two-stage framework. Again, similar to the modeling framework for initial order quantity, we first model *Censor* as a probit using all of the independent variables. Then, we compute the inverse Mills ratio for those customers who have $Censor = 1$. Finally, we run an OLS regression with *CLV* as the dependent variable only for the 135 customers who are still customers. We use all the independent variables listed and the inverse Mills ratio (λ). See the initial order quantity example for a more detailed description of the steps. We get the following results:

| Censor model | Estimate | CLV model | Estimate |
|---|---|---|---|
| *Intercept* | −32.151 | *Intercept* | −2.613 |
| *Acq_Expense* | 0.070[a] | *Acq_Expense* | 0.009 |

| | | | |
|---|---|---|---|
| *Acq_Expense_SQ* | −0.000 1 | *Acq_Expense_SQ* | −0.000 01 |
| *Ret_Expense* | 0.006 | *Ret_Expense* | 0.004 |
| *Ret_Expense_SQ* | 0.000 000 5[a] | *Ret_Expense_SQ* | −0.000 000 7 |
| *First_Purchase* | −0.002[a] | *First_Purchase* | 0.003 |
| *Crossbuy* | 0.758 | *Crossbuy* | 0.203 |
| *Frequency* | 0.852 | *Frequency* | 0.155 |
| *Frequency_SQ* | −0.004[a] | *Frequency_SQ* | −0.005 |
| *Industry* | 4.588 | *Industry* | 0.566 |
| *Revenue* | 0.072 | *Revenue* | 0.008 |
| *Employees* | 0.002[a] | *Employees* | 0.0002[a] |
| | | *Lambda* ($\lambda$) | 0.174[a] |

[a]Denotes not significant at $p < 0.05$.

We gain the following insights from the results. We see that $\lambda$ is positive, but not significant. Thus we cannot interpret this to mean that there is not a selection problem, only that it is unlikely given that the error term of our selection equation is not correlated with the error term of our regression equation. We also see that with the exception of *Employees*, the remaining variables in the *CLV* model are all significant at $p < 0.05$. This means that we have uncovered many of the key drivers of *CLV* for this set of customers. We find that *Acq_Expense* and *Ret_Expense* are positive with a diminishing return, as noted by the positive coefficient on the level term and the negative coefficient on the squared term. This suggests that the more you spend on both acquisition and retention efforts, up to a threshold, the higher the customer's expected lifetime value. For *First_Purchase* we see that there is a positive coefficient suggesting that the higher the initial purchase amount, the higher the expected customer lifetime value. For *Crossbuy* we see that there is a positive coefficient suggesting that the more products a customer purchases, the higher the expected lifetime value. We find that *Frequency* is positive with a diminishing return, as noted by the positive coefficient on the level term and the negative coefficient on the squared term. This suggests that customers who purchase consistently at a moderate rate over their tenure are most likely to have the highest lifetime value. Then in terms of firmographic variables we find that customers who are B2B (*Industry*) and customers who have higher *Revenue* are more likely to have a higher lifetime value than customers who are not in the B2B *Industry* and have a lower *Revenue*.

Our next step is to predict *CLV* for each of the customers and see if our predictions are accurate. We do this by starting with the equation for expected future profitability at the beginning of this example:

$$\begin{aligned}\mathrm{E}(Future\ Profitability) &= \mathrm{P}(Censor = 1)^{*}\mathrm{E}(CLV \mid Censor = 1)\\ &= \Phi(X'\beta)^{*}(\gamma'\alpha + \mu\lambda).\end{aligned}$$

In this case $\Phi$ is the normal CDF, $X$ is the matrix of independent variable values from the *Censor* equation, $\beta$ is the vector of parameter estimates from the *Censor* equation, $\gamma$ is the matrix of independent variables from the *CLV* equation, $\alpha$ is the vector of parameter estimates from the *CLV* equation, $\mu$ is the parameter estimate for the inverse Mills ratio, and $\lambda$ is the inverse Mills ratio. Once we have predicted the *CLV* value for each of the prospects we want to compare this to the actual value from the database.

We do the comparison for the *CLV* values of the 135 customers where $Censor = 1$, that is, the customers that are still active with the firm. We find for these data that the MAD is 0.29 (or \$290) and the MAPE is 4.5%. For our benchmark model we use the mean of 6.58 as our predicted value for each customer. In this case we get a MAD of 0.71 (or \$710) and a MAPE of 11.05%. We see that our *CLV* model provides a significant better prediction of *CLV* than the benchmark model. This shows that identifying the drivers of CLV can significantly help a firm understand which customers are most likely to be profitable in the future.

### 3.5.2 How do you implement it?

Similar to the example for initial order quantity, we use a two-stage modeling framework. We first use a binary probit model using PROC Logistic and a probit link function to model the probability that a customer has already quit. Next, we use a SAS Data step to compute the inverse Mills ratio. Finally, we use PROC Reg in SAS to determine the drivers of CLV. Other statistical programs such as SPSS can be used to run these two-stage least squares models.

## 3.6 Chapter summary

The purpose of this chapter was to explore the current models for customer acquisition and provide some empirical examples as to how firms can apply this knowledge. Customer acquisition is the first key step in the CRM process. We have also shown that when firms engage in optimal prospect selection by understanding the drivers of customer acquisition and the ‘right’ amount of customer acquisition effort by understanding the relationship between marketing spending and customer value, the result can generate significant customer and firm profitability.

## Customer acquisition – SAS code

```
/* Import Data - Library: statcrm */
proc import out = statcrm.customer_acquisition
      datafile = "C:\_Your_Data_Location_\Statistics in CRM\Customer
Acquisition\Customer Acquisition.xls" dbms = excel replace;
    range = "'Customer Acquisition Data$'"; getnames = yes; mixed = no;
scantext = yes; usedate = yes; scantime = yes;
run; quit;
/* Response Probability */
proc logistic data = statcrm.customer_acquisition descending;
model acquisition = acq_expense acq_expense_sq industry revenue
employees/rsquare;
output out = statcrm.response_probability_pred p = pred;
run; quit;
/* Number of Acquired Customers */
data statcrm.response_probability_pred1;
set statcrm.response_probability_pred;
if pred >= 0.5 then pred_acq = 1; else pred_acq = 0;
act_acq = acquisition;
run; quit;
proc freq data = statcrm.response_probability_pred1;
table pred_acq * act_acq; run; quit;
/* Initial Order Quantity */
proc logistic data = statcrm.customer_acquisition descending;
model acquisition = acq_expense acq_expense_sq industry revenue
employees/link = probit;
output out = statcrm.first_purchase_probit xbeta = xb_probit;
run; quit;
data statcrm.first_purchase_imr;
set statcrm.first_purchase_probit;
imr_acquisition = (pdf('Normal',xb_probit))/(probnorm(xb_probit));
run; quit;
proc reg data = statcrm.first_purchase_imr;
model first_purchase = acq_expense acq_expense_sq industry revenue
employees imr_acquisition;
output out = statcrm.first_purchase_imr1 p = xbeta;
where first_purchase > 0;
run; quit;
data statcrm.first_purchase1;
set statcrm.first_purchase_imr1;
pred_fp = probnorm(xb_probit)*(xbeta);
ad = abs(first_purchase - pred_fp);
ape = ad/first_purchase;
ad1 = abs(first_purchase - 372.47);
ape1 = ad1/first_purchase;
run; quit;
proc sql; select mean(ad) as mad, mean(ape) as mape,
mean(ad1) as random_mad, mean(ape1) as mape1
from statcrm.first_purchase1; quit;
/* Duration - Time */
proc lifereg data = statcrm.customer_acquisition;
```

```
model duration*censor(1) = acq_expense acq_expense_sq ret_expense
ret_expense_sq crossbuy frequency frequency_sq industry revenue
employees;
where acquisition = 1;
output out = statcrm.duration xbeta = xb p = pred sres = resid;
run; quit;
data statcrm.duration1;
set statcrm.duration;
pred_duration = exp(xb+0.138*(log(-log(1-0.5))));
ad = abs(duration - pred_duration);
ad1 = abs(duration - 333.3165);
run; quit;
proc sql; select mean(duration) from statcrm.duration1 where
acquisition = 1 and censor = 0; quit;
proc sql; select mean(ad) as mad, (mean(ad/duration)) as mape,
mean(ad1) as random_mad, (mean(ad1/duration)) as mape1
from statcrm.duration1 where acquisition = 1 and censor = 0; quit;
/* Firm Performance */
proc logistic data = statcrm.customer_acquisition descending;
model censor = acq_expense acq_expense_sq ret_expense ret_expense_sq
first_purchase crossbuy frequency frequency_sq industry revenue
employees/link = probit;
output out = statcrm.clv1a xbeta = xb_censor;
where acquisition = 1;
run; quit;
data statcrm.clv1b;
set statcrm.clv1a;
imr_censor = pdf('Normal',xb_censor)/probnorm(xb_censor);
run; quit;
proc reg data = statcrm.clv1b;
model clv = acq_expense acq_expense_sq ret_expense ret_expense_sq
first_purchase
   crossbuy frequency frequency_sq industry revenue employees
imr_censor;
where clv > 0;
output out = statcrm.clv1c p = pred_clv;
run; quit;
data statcrm.clv1d;
set statcrm.clv1c;
ad = abs(pred_clv - clv);
ad1 = abs(pred_clv - 6.58);
run; quit;
proc sql; select mean(clv) from statcrm.clv1d where censor = 1; quit;
proc sql; select mean(ad) as mad, (mean(ad/clv)) as mape,
mean(ad1) as random_clv, (mean(ad1/clv)) as mape1
from statcrm.clv1d where acquisition = 1 and duration = 730; quit;
```

# Customer acquisition – SAS output

```
                  The SAS System
                The LOGISTIC Procedure
              Model Information
   Data Set                       STATCRM.CUSTOMER_ACQUISITION
   Response Variable              Acquisition               Acquisition
   Number of Response Levels      2
   Model                          binary logit
   Optimization Technique         Fisher's scoring

                 Number of Observations Read     500
                 Number of Observations Used     500

                     Response Profile
              Ordered                             Total
                Value       Acquisition       Frequency
                    1            1                  292
                    2            0                  208
            Probability modeled is Acquisition = 1.

                   Model Convergence Status
            Convergence criterion (GCONV = 1E-8) satisfied.

                   Model Fit Statistics
                                         Intercept
                             Intercept         and
         Criterion                Only  Covariates
         AIC                   680.968     206.464
         SC                    685.183     231.751
         -2 Log L              678.968     194.464

   R-Square     0.6205     Max-rescaled     R-Square     0.8354

         Testing Global Null Hypothesis: BETA = 0
   Test                  Chi-Square          DF         Pr > ChiSq
   Likelihood Ratio        484.5043           5             <.0001
   Score                   307.1902           5             <.0001
   Wald                     90.1336           5             <.0001

         The SAS System

                The LOGISTIC Procedure
        Analysis of Maximum Likelihood Estimates
                                        Standard          Wald
Parameter          DF     Estimate        Error    Chi-Square     Pr > ChiSq
Intercept           1     -26.2062       3.5373       54.8855         <.0001
Acq_Expense         1       0.0672       0.0104       42.0956         <.0001
Acq_Expense_SQ      1     -0.00004     7.989E-6       26.8366         <.0001
Industry            1       0.0327       0.3868        0.0072         0.9326
Revenue             1       0.0319       0.0118        7.2657         0.0070
Employees           1      0.00496     0.000735       45.5539         <.0001
```

```
                  Odds Ratio Estimates
                         Point           95% Wald
Effect                 Estimate     Confidence       Limits
Acq_Expense              1.070          1.048         1.091
Acq_Expense_SQ           1.000          1.000         1.000
Industry                 1.033          0.484         2.205
Revenue                  1.032          1.009         1.057
Employees                1.005          1.004         1.006

 Association of Predicted Probabilities and Observed Responses
      Percent Concordant    97.5    Somers' D    0.950
      Percent Discordant     2.5    Gamma        0.950
      Percent Tied           0.0    Tau-a        0.463
      Pairs                60736    c            0.975

    The SAS System

                The FREQ Procedure
             Table of pred_acq by act_acq
          pred_acq   act_acq
          Frequency,
          Percent ,
          Row Pct ,
          Col Pct ,     0,     1, Total
          fffffffff^ffffffff^ffffffff
               0 ,   183 ,   19 ,  202
                 , 36.60 , 3.80 , 40.40
                 , 90.59 , 9.41 ,
                 , 87.98 , 6.51 ,
          ffffffff^ffffffff^ffffffff^
               1 ,     25,    273 ,   298
                 ,   5.00 , 54.60 , 59.60
                 ,   8.39 , 91.61 ,
                 , 12.02 , 93.49 ,
          ffffffff^ffffffff^ffffffff^
          Total       208      292     500
                   41.60   58.40  100.00

      The SAS System

                 The LOGISTIC Procedure
                   Model Information
Data Set                     STATCRM.CUSTOMER_ACQUISITION
Response Variable            Acquisition            Acquisition
Number of Response Levels    2
Model                        binary probit
Optimization Technique       Fisher's scoring

          Number of Observations Read      500
          Number of Observations Used      500
```

Response Profile

| Ordered Value | Acquisition | Total Frequency |
|---|---|---|
| 1 | 1 | 292 |
| 2 | 0 | 208 |

Probability modeled is Acquisition = 1.

Model Convergence Status

Convergence criterion (GCONV = 1E-8) satisfied.

Model Fit Statistics

| Criterion | Intercept Only | Intercept and Covariates |
|---|---|---|
| AIC | 680.968 | 204.122 |
| SC | 685.183 | 229.410 |
| -2 Log L | 678.968 | 192.122 |

Testing Global Null Hypothesis: BETA = 0

| Test | Chi-Square | DF | Pr > ChiSq |
|---|---|---|---|
| Likelihood Ratio | 486.8459 | 5 | <.0001 |
| Score | 307.1902 | 5 | <.0001 |
| Wald | 107.5310 | 5 | <.0001 |

The SAS System

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| Intercept | 1 | -15.1339 | 1.8623 | 66.0378 | <.0001 |
| Acq_Expense | 1 | 0.0389 | 0.00543 | 51.3430 | <.0001 |
| Acq_Expense_SQ | 1 | -0.00002 | 4.16E-6 | 33.5373 | <.0001 |
| Industry | 1 | 0.0513 | 0.2154 | 0.0568 | 0.8117 |
| Revenue | 1 | 0.0177 | 0.00661 | 7.1859 | 0.0073 |
| Employees | 1 | 0.00285 | 0.000403 | 50.0781 | <.0001 |

Association of Predicted Probabilities and Observed Responses

| | | | |
|---|---|---|---|
| Percent Concordant | 97.5 | Somers' D | 0.950 |
| Percent Discordant | 2.5 | Gamma | 0.950 |
| Percent Tied | 0.0 | Tau-a | 0.463 |
| Pairs | 60736 | c | 0.975 |

The SAS System

The REG Procedure

Model: MODEL1

Dependent Variable: First_Purchase First_Purchase

| | |
|---|---|
| Number of Observations Read | 292 |
| Number of Observations Used | 292 |

```
                        Analysis of Variance

                             Sum of           Mean
 Source            DF       Squares          Square       F Value      Pr > F
 Model              6       6743424         1123904       1212.34     <.0001
 Error            285        264211       927.05463
 Corrected Total  291       7007635
      Root MSE               30.44757         R-Square      0.9623
      Dependent Mean        372.46868         Adj R-Sq      0.9615
      Coeff Var               8.17453
                        Parameter Estimates
                                          Parameter Standard
Variable          Label             DF     Estimate     Error  t Value  Pr > |t|
Intercept         Intercept          1    -62.26094  48.96509   -1.27   0.2046
Acq_Expense       Acq_Expense        1      0.70484   0.13030    5.41   <.0001
Acq_Expense_SQ    Acq_Expense_SQ     1 -0.00076743 0.00009221  -8.32   <.0001
Industry          Industry           1     -2.76418   3.70605   -0.75   0.4564
Revenue           Revenue            1      3.02602   0.11146   27.15   <.0001
Employees         Employees          1      0.25406   0.00476   53.33   <.0001
imr_acquisition                      1     19.10122   9.65693    1.98   0.0489

           The SAS System
                    mad     mape random_mad   mape1
                ffffffffffffffffffffffffffffffffffffffff
                51.95896 0.186859  127.1712 1.354829

           The SAS System
                       The LIFEREG Procedure
                         Model Information
    Data Set              STATCRM.CUSTOMER_ACQUISITION
    Dependent Variable                     Log(Duration)    Duration
    Censoring Variable                            Censor      Censor
    Censoring Value(s)                                 1
    Number of Observations                           292
    Noncensored Values                               157
    Right Censored Values                            135
    Left Censored Values                               0
    Interval Censored Values                           0
    Name of Distribution                         Weibull
    Log Likelihood                           51.102184898

                Number of Observations Read        292
                Number of Observations Used        292

                            Fit Statistics
              -2 Log Likelihood                -102.204
              AIC (smaller is better)           -78.204
              AICC (smaller is better)          -77.086
              BIC (smaller is better)           -34.083
       Algorithm converged.
```

Type III Analysis of Effects

| Effect | DF | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|
| Acq_Expense | 1 | 52.0446 | <.0001 |
| Acq_Expense_SQ | 1 | 223.6009 | <.0001 |
| Ret_Expense | 1 | 179.5271 | <.0001 |
| Ret_Expense_SQ | 1 | 3.6092 | 0.0575 |
| Crossbuy | 1 | 213.8306 | <.0001 |
| Frequency | 1 | 113.1688 | <.0001 |
| Frequency_SQ | 1 | 4.6705 | 0.0307 |
| Industry | 1 | 424.5219 | <.0001 |
| Revenue | 1 | 214.4291 | <.0001 |
| Employees | 1 | 10.4581 | 0.0012 |

The SAS System

The LIFEREG Procedure

Analysis of Maximum Likelihood Parameter Estimates

| Parameter | DF | Estimate | Standard Error | 95% Confidence Limits | | Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|---|
| Intercept | 1 | 2.8692 | 0.3053 | 2.2709 | 3.4675 | 88.34 | <.0001 |
| Acq_Expense | 1 | 0.0068 | 0.0009 | 0.0049 | 0.0086 | 52.04 | <.0001 |
| Acq_Expense_SQ | 1 | -0.0000 | 0.0000 | -0.0000 | -0.0000 | 223.60 | <.0001 |
| Ret_Expense | 1 | 0.0011 | 0.0001 | 0.0009 | 0.0012 | 179.53 | <.0001 |
| Ret_Expense_SQ | 1 | -0.0000 | 0.0000 | -0.0000 | 0.0000 | 3.61 | 0.0575 |
| Crossbuy | 1 | 0.0992 | 0.0068 | 0.0859 | 0.1125 | 213.83 | <.0001 |
| Frequency | 1 | 0.1198 | 0.0113 | 0.0977 | 0.1419 | 113.17 | <.0001 |
| Frequency_SQ | 1 | -0.0014 | 0.0006 | -0.0027 | -0.0001 | 4.67 | 0.0307 |
| Industry | 1 | 0.5378 | 0.0261 | 0.4866 | 0.5889 | 424.52 | <.0001 |
| Revenue | 1 | 0.0121 | 0.0008 | 0.0105 | 0.0137 | 214.43 | <.0001 |
| Employees | 1 | 0.0001 | 0.0000 | 0.0000 | 0.0001 | 10.46 | 0.0012 |
| Scale | 1 | 0.1387 | 0.0080 | 0.1239 | 0.1552 | | |
| Weibull Shape | 1 | 7.2123 | 0.4154 | 6.4425 | 8.0741 | | |

```
The SAS System
               ffffffff
               327.3949
The SAS System
        mad    mape random_mad  mape1
    ffffffffffffffffffffffffffffffffffff
    45.42903 0.138056   165.937 1.723717
```

The SAS System

The LOGISTIC Procedure

Model Information

| | | |
|---|---|---|
| Data Set | STATCRM.CUSTOMER_ACQUISITION | |
| Response Variable | Censor | Censor |
| Number of Response Levels | 2 | |
| Model | binary probit | |
| Optimization Technique | Fisher's scoring | |

| | |
|---|---|
| Number of Observations Read | 292 |
| Number of Observations Used | 292 |

Response Profile

| Ordered Value | Censor | Total Frequency |
|---|---|---|
| 1 | 1 | 135 |
| 2 | 0 | 157 |

Probability modeled is Censor = 1.

Model Convergence Status

Convergence criterion (GCONV = 1E-8) satisfied.

Model Fit Statistics

| Criterion | Intercept Only | Intercept and Covariates |
|---|---|---|
| AIC | 405.139 | 66.965 |
| SC | 408.816 | 111.086 |
| -2 Log L | 403.139 | 42.965 |

Testing Global Null Hypothesis: BETA = 0

| Test | Chi-Square | DF | Pr > ChiSq |
|---|---|---|---|
| Likelihood Ratio | 360.1737 | 11 | <.0001 |
| Score | 183.0936 | 11 | <.0001 |
| Wald | 28.7364 | 11 | 0.0025 |

The SAS System

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| Intercept | 1 | -32.1509 | 13.5467 | 5.6327 | 0.0176 |
| Acq_Expense | 1 | 0.0695 | 0.0420 | 2.7354 | 0.0981 |
| Acq_Expense_SQ | 1 | -0.00009 | 0.000036 | 6.3828 | 0.0115 |
| Ret_Expense | 1 | 0.00597 | 0.00279 | 4.5733 | 0.0325 |
| Ret_Expense_SQ | 1 | 4.555E-7 | 9.229E-7 | 0.2436 | 0.6216 |
| First_Purchase | 1 | -0.00170 | 0.00784 | 0.0472 | 0.8279 |
| Crossbuy | 1 | 0.7579 | 0.2159 | 12.3223 | 0.0004 |
| Frequency | 1 | 0.8515 | 0.2529 | 11.3406 | 0.0008 |
| Frequency_SQ | 1 | -0.00353 | 0.0115 | 0.0948 | 0.7581 |
| Industry | 1 | 4.5876 | 1.0493 | 19.1152 | <.0001 |
| Revenue | 1 | 0.0718 | 0.0298 | 5.8001 | 0.0160 |
| Employees | 1 | 0.00183 | 0.00208 | 0.7773 | 0.3780 |

Association of Predicted Probabilities and Observed Responses

| | | | |
|---|---|---|---|
| Percent Concordant | 99.7 | Somers' D | 0.993 |
| Percent Discordant | 0.3 | Gamma | 0.993 |
| Percent Tied | 0.0 | Tau-a | 0.496 |
| Pairs | 21195 | c | 0.997 |

The SAS System

The REG Procedure

Model: MODEL1

Dependent Variable: CLV CLV

Number of Observations Read 135

```
              Number of Observations Used     135
                   Analysis of Variance
                              Sum of         Mean
Source              DF        Squares       Square    F Value     Pr > F
Model               12      106.94992      8.91249      60.14    <.0001
Error              122       18.07983      0.14820
Corrected Total    134      125.02974
       Root MSE             0.38496     R-Square       0.8554
       Dependent Mean       6.57930     Adj R-Sq       0.8412
       Coeff Var            5.85110
```

Parameter Estimates

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| Intercept | Intercept | 1 | -2.61302 | 0.93166 | -2.80 | 0.0059 |
| Acq_Expense | Acq_Expense | 1 | 0.00870 | 0.00286 | 3.04 | 0.0029 |
| Acq_Expense_SQ | Acq_Expense_SQ | 1 | -0.00001041 | 0.00000265 | -3.93 | 0.0001 |
| Ret_Expense | Ret_Expense | 1 | 0.00356 | 0.00040352 | 8.82 | <.0001 |
| Ret_Expense_SQ | Ret_Expense_SQ | 1 | -6.76812E-7 | 1.117292E-7 | -6.06 | <.0001 |
| First_Purchase | First_Purchase | 1 | 0.00259 | 0.00108 | 2.40 | 0.0181 |
| Crossbuy | Crossbuy | 1 | 0.20338 | 0.02013 | 10.11 | <.0001 |
| Frequency | Frequency | 1 | 0.15521 | 0.03415 | 4.55 | <.0001 |
| Frequency_SQ | Frequency_SQ | 1 | -0.00485 | 0.00181 | -2.69 | 0.0082 |
| Industry | Industry | 1 | 0.56644 | 0.07845 | 7.22 | <.0001 |
| Revenue | Revenue | 1 | 0.00817 | 0.00370 | 2.21 | 0.0289 |
| Employees | Employees | 1 | 0.00016932 | 0.00028115 | 0.60 | 0.5481 |
| imr_censor | | 1 | 0.17380 | 0.13503 | 1.29 | 0.2005 |

```
        The SAS System
                   ffffffff
                   6.579302

      The SAS System

    mad    mape                random_clv       mape1
      ffffffffffffffffffffffffffffffffffffffffffff
  0.289925           0.045168            0.7087606.110484
```

# References


1. Lix, T.S., Berger, P.D., and Magliozzi, T.L. (1995) New customer acquisition: prospecting models and the use of commercially available external data. *Journal of Direct Marketing*, **9**(4), 8–18.
2. Lewis, M. (2006b) The effect of shipping fees on customer acquisition, customer retention, and purchase quantities. *Journal of Retailing*, **82**(1), 13–23.
3. Villanueva, J., Yoo, S., and Hanssens, D.M. (2008) The impact of marketing-induced versus word-of-mouth customer acquisition on customer equity growth. *Journal of Marketing Research*, **45**(1), 48–59.
4. Lewis, M. (2006a) Customer acquisition promotions and customer asset value. *Journal of Marketing Research*, **43**(2), 195–203.

5. Hansotia, B.J. and Wang, P. (1997) Analytical challenges in customer acquisition. *Journal of Direct Marketing*, **11**(2), 7–19.
6. Thomas, J.S. (2001) A methodology for linking customer acquisition to customer retention. *Journal of Marketing Research*, **38**(2), 262–268.
7. Reinartz, W., Thomas, J.S. and Kumar, V. (2005) Balancing acquisition and retention resources to maximize customer profitability. *Journal of Marketing*, **69**(1), 63–79.
8. Thomas, J.S., Blattberg, R.C., and Fox, E.J. (2004) Recapturing lost customers. *Journal of Marketing Research*, **41**(1), 31–45.
9. Blattberg, R.C. and Deighton, J. (1996) Manage marketing by the customer equity test. *Harvard Business Review*, **74**(4), 136–144.
10. Berger, P.D. and Bechwati, N.N. (2001) The allocation of promotion budget to maximize customer equity. *Omega*, **29**(1), 49–61.
11. Wangenheim, F. and Bayón, T. (2007) The chain from customer satisfaction via word-of-mouth referrals to new customer acquisition. *Journal of the Academy of Marketing Science*, **35**(2), 233–249.
12. Schweidel, D.A., Fader, P.S., and Bradlow, E.T. (2008) A bivariate timing model of customer acquisition and retention. *Marketing Science*, **27**(5), 829–843.

# 4

# Customer retention

## 4.1 Introduction

After customers have been acquired, it is important to discuss the second step of CRM – retaining customers. Customer retention strategies are used in both contractual (where customers are bound by contracts such as mobile phone subscription or magazine subscription) and non-contractual settings (where customers are not bound by contracts, such as grocery purchases or apparel purchases). Reichheld and Sasser (1990) stated that a 5% improvement in customer retention can cause an increase in profitability between 25 and 85%, in terms of net present value, depending upon the industry. Since then, companies have constantly allocated resources on customer retention management and researchers have put much emphasis on studying customer retention.

Research on customer retention has two main streams. One stream is interested in investigating the effects various marketing variables on customer retention, which in turn influences a firm's performance. The other stream is interested in building econometric and statistical models to estimate or predict the customer retention decisions from both the customer and company prospective. Figure 4.1 shows an integrated framework which describes the various relationships examined in different studies across many industries. These industries have included telecommunications, financial services, hairdressing, restaurants, and retailing, among others. The common thread of these relationships is the mindset that (a) increased product and service quality lead to increased customer satisfaction, (b) increased customer satisfaction leads to increased customer retention (which is mediated by relationship quality such that a higher relationship quality positively enhances the link between satisfaction and retention), and (c) increased customer retention leads to increased firm performance.

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

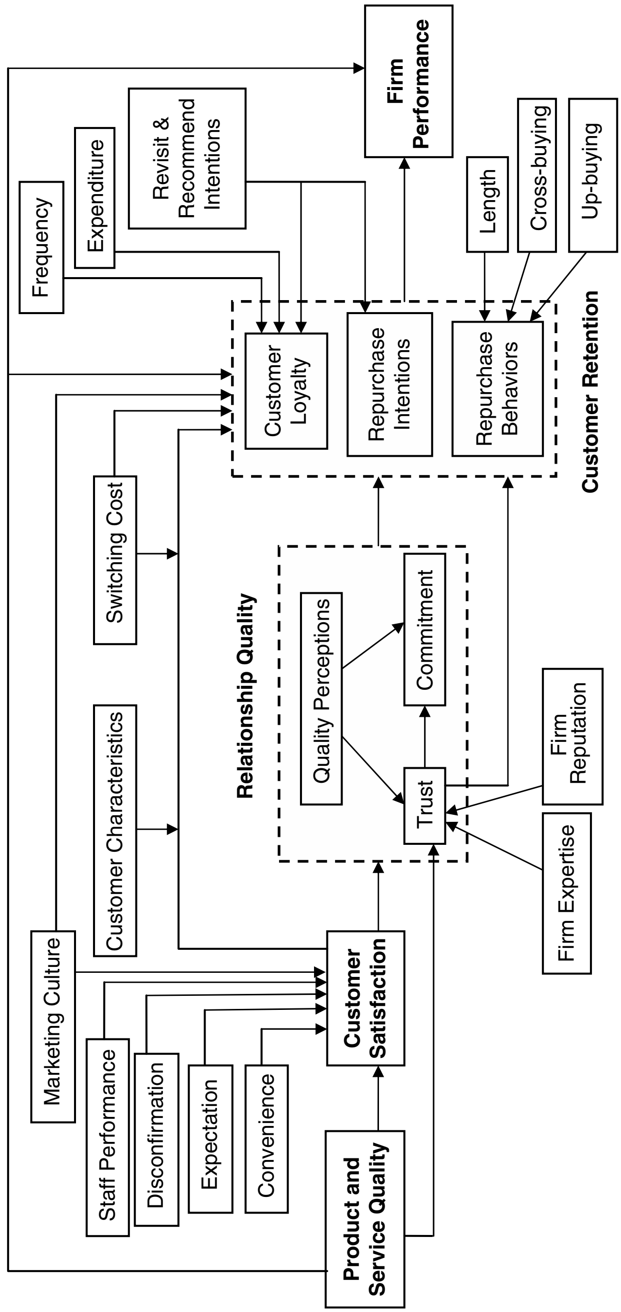



*Figure 4.1* *Integrated relationships addressed in customer retention.*

Besides understanding the general linkages between product and service quality and firm performance, it is also important to understand the antecedents of each key construct which help drive the value of the construct. For instance, research has shown that several different factors can influence customer satisfaction level including staff performance, disconfirmation, expectation, convenience, and marketing culture (which can also influence customer retention directly). Furthermore, research has shown that customer characteristics and switching costs are two factors most commonly examined as moderating links between customer satisfaction and customer retention.

To examine the effects of various factors highlighted in Figure 4.1 on customer retention, the most commonly used method is confirmatory factor analysis. Confirmatory factor analysis is a multivariate statistical method used to test the understanding researchers have of the constructs of interest. A strong empirical or conceptual foundation is necessary for building up a factor model. Usually, researchers propose a set of hypotheses to explain the phenomenon of interest and confirmatory factor analysis is used to test these hypotheses. For example, one might hypothesize that satisfied customers tend to buy more from the company. Then the researcher would use confirmatory factor analysis to create latent values for customer satisfaction and customer retention, often using surveys of customers, to test this hypothesis. The most commonly used model-fitting procedure for confirmatory factor analysis is maximum likelihood estimation. Statistical software like AMOS, LISREL, EQS, and SAS can be used for confirmatory factor analysis.

The second stream of research on customer retention studies issues (Figure 4.2) involved in decisions made by managers on current customers. There are often several key questions that managers are interested in answering after a customer has been acquired. These include:

- Will the recently acquired customer repurchase or not in the future?
- What will be the lifetime duration of the customer (i.e., when will the customer churn)?

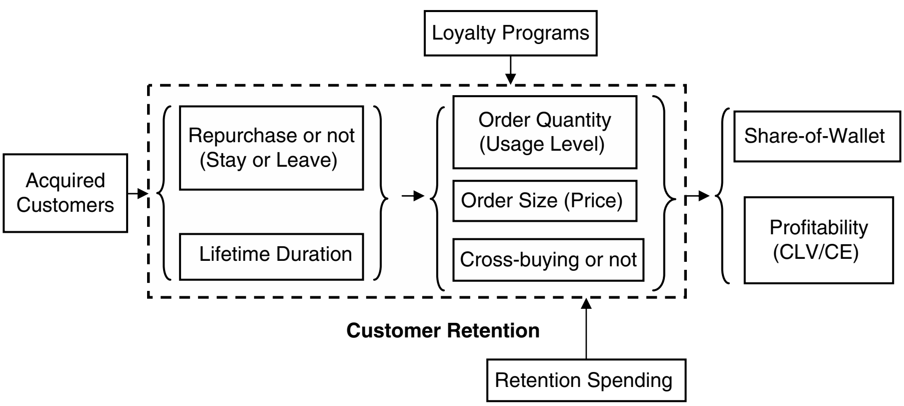



*Figure 4.2 Issues addressed in customer retention modeling.*

- Given the customer is going to repurchase:
  - How many items is that customer going to purchase?
  - How much is that customer likely to spend?
  - Will that customer purchase in multiple product categories?
- Does the customer mainly purchase from my firm (high share-of-wallet) or from many different firms (low share-of-wallet)?
- What is the long-term impact of this customer's purchase behavior on firm value?

Many different research studies have been conducted and many models have been developed to try and answer these questions. Table 4.1 contains a representative set of studies that have considered these issues and accounted for many of the problems that might occur in the model-building process. To provide a comprehensive understanding of how to model customer retention, we will review the issues in the studies one by one along with the related modeling techniques. We will also provide empirical examples at the end of each subsection to demonstrate how to apply this knowledge to a representative sample of customers from a B2C firm.

Similar to customer acquisition models, the first question that needs to be answered when considering model selection is whether the firm's customers purchase in a contractual versus non-contractual manner. In most instances this will determine the type of statistical model that needs to be used in order to gain any insights from the data.

### 4.1.1 Data for empirical examples

In this chapter we will be providing a description of the key modeling frameworks that attempt to answer each key research question raised at the beginning of the chapter. We will also be providing at least one empirical example at the end of each subsection which will show how sample data can be used to answer these key research questions. For all the empirical examples in this chapter we provide a dataset titled 'Customer Retention' which is broken down into two related data tables. In this dataset you will find two tables of data that include a representative sample of 500 customers from a typical B2C firm, where all the customers are from the same cohort. In this case, the cohort consists of a random sample of 500 customers who all made their first purchase with the firm in quarter 1. In the first data table we provide the transaction information for each customer over the course of 12 quarters. Thus, the data table here consists of 6000 rows (500 customers $\times$ 12 quarters) and 8 columns. In the second data table we provide the demographic information for each customer. Thus, the data table here consists of 500 rows (500 customers) and 6 columns.

Table 4.1 Review of customer retention models.

| Research interest | Specification | Estimation | Representative studies |
|---|---|---|---|
| Repurchase or not (stay or leave) | Logit | MLE | Meer [1] |
| | | | Seo, Ranganathan, and Badad [2] |
| | | | Lemon, White, and Winer (2002) |
| | | | Bolton, Kannan, and Bramlett (2000) |
| | | | Kivetz, Urminsky, and Zheng (2006) |
| | | | Lewis [3] |
| | Multinomial logit | MLE | Leenheer, Heerde, Bijmolt, and Smidts (2007) |
| | Probit | MLE | Verhoef [4] |
| | | | Verhoef and Donkers [5] |
| | Hazard | MLE | Bhattacharya [6] |
| | Discrete hazard | MCMC[a] | Borle, Singh, and Jain (2008) |
| | Proportional hazard | MLE | Schweidel, Fader, and Bradlow [7] |
| | Discrete proportional hazard | MLE | Kivetz, Urminsky, and Zheng (2006) |
| | Neural networks | Bayesian | Basens, Viaene, Poel, Vanthienen, and Dedene (2002) |
| | Random intercepts | MLE | Bolton, Lemon, and Bramlett (2006) |
| Duration | Negative binomial/Pareto | Method of Moment | Reinartz and Kumar [8] |
| | Proportional hazard | MLE | Meyer-Waarden [9] |
| | | | Bolton [10] |
| | Shifted-beta geometric distribution | MLE | Fader and Hardie [11] |
| | Tobit | MLE | Reinartz, Thomas, and Kumar (2005) |
| | | | Thomas [12] |

*(continued)*

Table 4.1 (*Continued*)

| Research interest | Specification | Estimation | Representative studies |
|---|---|---|---|
| Order quantity (usage level) | Tobit | MLE | Bolton, Kannan, and Bramlett (2000) |
| | | | Bolton and Lemon [13] |
| | | | Kivetz, Urminsky, and Zheng (2006) |
| | Log-normal | MCMC | Borle, Singh, and Jain (2008) |
| | Negative binominal | MLE | Zhang, Dixit, and Friedmann (2010) |
| | Poisson | MLE | Anderson and Simester [14] |
| | System of regressions | 3SLS | Lewis [15] |
| Order size (price) | Linear regression | OLS | Anderson and Simester [14] |
| | System of regressions | 3SLS | Lewis [15] |
| Cross-buying or not | Probit | MLE | Verhoef and Donkers [5] |
| Share-of-wallet | Tobit | MLE | Verhoef [4] |
| Profitability | Tobit | MLE | Reinartz, Thomas, and Kumar (2005) |
| | Deterministic | N/A | Blattberg and Deighton [16] |
| | | | Berger and Nasr-Bechwati [17] |

[a] MCMC = Markov Chain Monte Carlo.

The first data table (labeled *Transactions*) includes the following variables, which will be used in some combination for each of the subsequent analyses:

| Variable | |
|---|---|
| *Customer* | *Customer number (from 1 to 500)* |
| *Quarter* | *The quarter (from 1 to 12) when the transactions occurred* |
| *Purchase* | *1 when the customer purchased in the given quarter, 0 if no purchase occurred in that quarter* |
| *Order_Quantity* | *The dollar value of the purchases in the given quarter* |
| *Crossbuy* | *The number of different categories purchased in a given quarter* |
| *Ret_Expense* | *Dollars spent on marketing efforts to try and retain that customer in the given quarter* |
| *Ret_Expense_SQ* | *Square of dollars spent on marketing efforts to try and retain that customer in the given quarter* |

The second data table (labeled *Demographics*) includes the following variables which will be used in some combination for each of the subsequent analyses:

| Variable | |
|---|---|
| *Customer* | *Customer number (from 1 to 500)* |
| *Gender* | *1 if the customer is male, 0 if the customer is female* |
| *Married* | *1 if the customer is married, 0 if the customer is not married* |
| *Income* | *1 if income < \$30 000*<br>*2 if \$30 001< income < \$45 000*<br>*3 if \$45 001 < income < \$60 000*<br>*4 if \$60 001 < income < \$75 000*<br>*5 if \$75 001 < income < \$90 000*<br>*6 if income > \$90 001* |
| *First_Purchase* | *The value of the first purchase made by the customer in quarter 1* |
| *Loyalty* | *1 if the customer is a member of the loyalty program, 0 if not* |
| *Share-of-Wallet (SOW)* | *The percentage of purchases the customer makes from the given firm given the total amount of purchases across all firms in that category* |
| *CLV* | *The discounted value of all expected future profits, or customer lifetime value* |

These two data tables will be used for each of the examples presented at the end of each of the sections. These examples will cover the topics of repurchase or not, order quantity, order size, cross-buying, share-of-wallet (SOW), and profitability (CLV).

## 4.2 Repurchase or not (stay or leave)

### 4.2.1 Will a customer repurchase?

Companies spend significant resources in acquiring potential customers and one of their biggest concerns is whether these newly acquired customers just make a first-time purchase and leave, or make subsequent purchases and stay with companies for an extended period of time. Repurchase or not is usually modeled as a binary outcome where either a repurchase occurs (1) or does not occur (0). The most commonly used method to model this binary outcome is logistic regression. Geoffrey (2006) built a logit model to predict who became an active customer using online activity and surfing behavior as explanatory variables. The author collected clickstream data to track each mouse-click of customers and to analyze customer online behavior. Lemon *et al.* [18] conducted a study in the television entertainment service subscription industry and estimated customers' keep (repurchase) or drop (churn) decisions. As the study was conducted in a monthly contractual setting, the behavior of the renewal of contract was directly observed and coded (0/1). These authors adopted logistic regression to model the decision of whether to remain in the service relationship as a function of expected future use and satisfaction with the service. Lewis [3] conducted two studies, one in the non-contractual setting of online retailing and the other in the contractual setting of newspaper subscription, to investigate the effects of acquisition promotion discount depth on repeat purchasing and on renewal of contract using logistic regression.

A direct question that managers will ask is whether the newly acquired customers will repurchase or not. This is a binary classification problem which can be tackled by many statistical analyses such as logistic and probit regressions. Baesens *et al.* [19] adopted neural networks (NNs) to solve the problem and used a Bayesian learning paradigm during NN training. These authors used RFM as explanatory variables and compared the predictive results to logistic regression, and linear and quadratic discriminant analysis. In addition, Baesens *et al.* (2004) conducted another studies using Bayesian network classifiers to investigate whether newly acquired customers will increase or decrease their future spending from initial purchase information. The authors first adopted a linear regression to estimate the slope of the life cycle of customers based on their historical contributions. The estimated slope was then discretized into a binary variable (positive/negative) to represent increasing or decreasing spending. This binary variable was used as the dependent variable and customers' past transaction activities were taken as independent variables. In addition to econometric and statistical techniques, we suggest that researchers can adopt technology from machine learning and artificial intelligence to assist customer retention decision modeling.

Bolton *et al.* [20] conducted a study to investigate the factors that might influence the firm's service contract renewal decision. These authors modeled these decisions as a function of service quality and price and argued that firms assess the value of contracts renewal based on their prior service experiences under the old contract. In their model specification process, the authors considered that it was essential to account for the intrafirm association and potential heterogeneity since different firms with different characteristics and demands might assess the value of new contracts differently. Thus, they adopted a random intercept model in their analysis. Following Bolton *et al.*'s [20] specification, the probability that a firm ($N$) renews a contract ($c$) can be modeled as

$$\Pr(renew\ contract\ c|firm\ N) = \Pr(U_{nc} > 0) \tag{4.1}$$

where

$$U_{nc} = V_{nc} + \varepsilon_{nc}, \tag{4.2}$$

$$V_{nc} = \mu_n + \beta' x_{nc}, \tag{4.3}$$

in which the variables associated with the fixed parameters are denoted by vector $x_{nc}$. The explanatory variables include both contract-level and firm-level variables. The random intercepts $\mu_n$ were assumed to follow a univariate normal distribution across firms (with mean $\mu$ and estimated variance $\sigma^2$). In this way, the amount of intrafirm correlation can be captured by the variance of the random intercept. The authors further assumed that the error term $\varepsilon_{nc}$ is an independently and identically distributed extreme value. A complementary log–log model is defined as

$$\Pr(retention\ of\ contract\ c\ by\ firm\ n) = 1 - \exp[-\exp(V_{nc})]. \tag{4.4}$$

The random intercepts model is estimated by marginal maximum likelihood estimation, utilizing a Fisher-scoring solution. We provide an introduction to the random intercept model and its estimation in Appendix J.

### 4.2.2 When will a customer no longer repurchase?

Another important question concerning repurchase behavior is when a customer is likely to leave. Bhattacharya [6] investigated the hazard of lapsing of customers in a paid membership context. The data the author used came from an art museum and contained members' joining, affiliation, and helping characteristics. The author used survival analysis in the analysis because it is able to model the timing and occurrence of events. The dependent variable in the study was the hazard of lapsing and is defined as

$$h_{it} = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \tag{4.5}$$

where $h_{it}$ is the instantaneous probability of member $i$ lapsing at time $t$ and $\Pr(\cdot)$ is the probability of an event between time $t$ and $t + \Delta t$, given that the member is in the sample at risk at time $t$. In Bhattacharya's [6] study, the hazard rate of lapsing from origin state $j$ (i.e., being a member) to destination state $k$ (i.e., lapsing) can be described as

$$h_{ik}(t) = \exp[\beta_{ik} X_{ik}(t-1)] \tag{4.6}$$

where $X$ is a vector of lagged independent variables observed at time $t-1$. The hazard model is usually estimated using maximum likelihood techniques. An introduction to survival analysis and the estimation are provided in Appendix G.

In a study by Kivetz *et al.* [21], the authors adopted a discrete-time model in which the hazard model likelihood is decomposed into probabilities of purchase within given time intervals. The full discretized survival function is expressed as a function of the baseline hazard function $h(t)$, time-varying covariates $X_t$, and estimated covariate coefficients $\beta$:

$$S(t, X_t) = \exp\left[-\sum_{u=1}^{t} \exp(X_u \beta) \int_{u-1}^{u} h(w)\, dw\right] \tag{4.7}$$

In the study, the authors decomposed the survival function into day-specific components where the dependent variable is the probability of purchase on a given day, conditional on no purchase having yet occurred:

$$\begin{aligned} \Pr(t, X_t) &= 1 - \frac{S(t, X_t)}{S(t-1, X_{t-1})} = 1 - \exp\left[-\exp(X_t \beta) \int_{t_l}^{t_u} h(u)\, du\right] \\ &= 1 - \left[\frac{S(t_u)}{S(t_l)}\right] \exp(X_t \beta). \end{aligned} \tag{4.8}$$

Seetharaman and Chintagunta (2003) give the following likelihood function which can be maximized to estimate the parameters of the discrete-time proportional hazards model at the individual level:

$$L = \prod_{v=1}^{T} \Pr(v, X_v)^{\delta_v *} [1 - \Pr(v, X_v)]^{1-\delta_v}, \tag{4.9}$$

where in the authors' context, $\delta_v$ is an indicator variable (1/0) that takes the value 1 if the product is purchased by the household on shopping trip $v$ and 0 otherwise, and $\Pr(v, X_v)$ is the household's probability of purchasing the product on shopping trip $v$, given by Equation 4.8. An introduction to the discrete-time hazard models is also provided in Appendix H.

In Borle *et al.*'s (2008) study, the authors adopted a discrete-hazard approach to model the hazard of lifetime $h(LIFE_{hi})$ for customer $h$, which is the risk of leaving in the $i$th spell (probability that the customer will leave the company without making the $i$th purchase after having made the $(i-1)$th purchase):

$$h(LIFE_{hi}) = [1 + \exp(-\delta_{hi})]^{-1} \tag{4.10}$$

where $\delta_{hi}$ is specified as follows:

$$\delta_{hi} = \delta_h + \delta_i + \delta_{1h} \log lagTIME_{hi} + \delta_{2h} \log lagAMNT_{hi} + \delta_3 GENDER_h \tag{4.11}$$

where $\delta_i = \delta' i + \delta'' i^2 + \delta''' i^3$ and $i$ indexes the purchase occasion. This third-order polynomial expression addressed non-stationarity across purchase occasions. The authors also allow for a heterogeneity structure over the coefficients for the lagged variables as follows:

$$\delta_{1h} \sim Normal\left(\bar{\delta}_1, \tau_3^2\right) \tag{4.12}$$

$$\delta_{2h} \sim Normal\left(\bar{\delta}_2, \tau_4^2\right). \tag{4.13}$$

The authors also estimated the defection model jointly with the interpurchase time and the purchase amount models together by assigning appropriate prior distributions to the parameters to be estimated and using a Markov chain Monte Carlo (MCMC) sampling algorithm.

Schweidel, Fader, and Bradlow [7] argued that after customers have been acquired, they churn following a parametric distribution. Incorporating time-varying covariates into the retention modeling, these authors adopted a proportional hazards regression using a baseline hazard function, $h_R(t_R|\Phi)$. As in the acquisition modeling, three sets of possible baseline hazard specifications were considered for the retention process: the Weibull, log-logistic and expo-power distribution. The survival function for the retention modeling was

$$S_R(t_R|\beta, \Phi, X(t)) = \exp\left[-\sum_{v=1}^{t_R} \exp(\beta X(t))\left(\int_{v-1}^{v} h_R(u|\Phi)\, du\right)\right] \tag{4.14}$$

where $\beta$ is the impact of the time-varying marketing activities, denoted $X(t)$. The duration of service retention $d_h$ is distributed as

$$g(t_R|\beta, \Phi, X(t)) = S_R(t_R - 1|\beta, \Phi, X(t)) - S_R(t_R|\beta, \Phi, X(t)) \tag{4.15}$$

where $t_R = 1, 2, \ldots, T$. An introduction to the proportional hazards model is provided in Appendix I.

### 4.2.3 Empirical example: Repurchase or not (stay or leave)

One of the key questions we want to answer with regard to customer retention is whether we can determine which customers have the highest likelihood of repurchase. To do this we first need to know which current customers actually made additional purchases after their initial first purchase. In the dataset provided for this chapter we have a binary variable which identifies whether or not a customer purchased in a given time period, in this case quarter. We also provide a set of drivers which are likely to help explain a customer's decision to repurchase. At the end of this example you should be able to do the following:

1. Identify the drivers of customer repurchase behavior.
2. Interpret the parameter estimates from the repurchase model.
3. Predict the number of repeat purchases by customers.
4. Determine the predictive accuracy of the repurchase model.

A B2C firm wants to improve the repurchase rate of customers and reduce the retention spending on customers by better understanding which customers are most likely to repurchase in a given time period. A random sample of 500 customers from a single cohort was taken from the customer database. The information we need for our model includes the following list of variables:

| | |
|---|---|
| **Dependent variable** | |
| *Purchase* | *1 when the customer purchased in the given quarter, 0 if no purchase occurred in that quarter* |
| **Independent variables** | |
| *Lag_Purchase* | *1 if the customer purchased in the previous quarter, 0 if no purchase occurred in the previous quarter* |
| *Avg_Order_Quantity* | *The average dollar value of the purchases in all previous quarters* |
| *Ret_Expense* | *Dollars spent on marketing efforts to try and retain that customer in the given quarter* |
| *Ret_Expense_SQ* | *Square of dollars spent on marketing efforts to try and retain that customer in the given quarter* |
| *Gender* | *1 if the customer is male, 0 if the customer is female* |
| *Married* | *1 if the customer is married, 0 if the customer is not married* |
| *Income* | *1 if income < \$30 000*<br>*2 if \$30 001< income < \$45 000*<br>*3 if \$45 001 < income < \$60 000*<br>*4 if \$60 001 < income < \$75 000*<br>*5 if \$75 001 < income < \$90 000*<br>*6 if income > \$90 001* |
| *First_Purchase* | *The value of the first purchase made by the customer in quarter 1* |
| *Loyalty* | *1 if the customer is a member of the loyalty program, 0 if not* |

In this case, we have a binary dependent variable (*Purchase*) which tells us whether the customer did purchase (= 1) or did not purchase (= 0) in a given quarter. We also have 10 independent variables that we believe will be drivers of repurchase behavior.

We believe that transaction behavior in the past is likely to explain future purchase behavior. As a result we use several lagged operationalizations of current variables as independent variables in this example. First, we have whether or not the customer purchased in the last quarter (*Lag_Purchase*). This variable can be obtained by taking the lagged value of the purchase indicator variable, noting that one observation will be lost for each customer for each lag that is taken. In this case we are only using a one-period lag. Second, we have the average past order quantity (*Avg_Order_Quantity*). In this case the value for average order quantity is the mean of the *Order_Quantity* variable in all quarters before the current time period. Third, we have how many dollars the firm spent on each customer (*Ret_Expense*) in each time period and the squared value of that variable (*Ret_Expense_SQ*). We want to use both the linear and squared terms since we expect that for each additional dollar spent on the retention effort for a given customer, there will be a diminishing return to the value of that dollar. Finally, since the focal firm of this example is a B2C firm, the other five variables are demographic and static variables of the customers. These include the *Gender* of the customer, whether the customer is *Married*, the *Income* of the customer, the value of the customer's first purchase (*First_Purchase*), and whether the customer is a member of the loyalty program (*Loyalty*).

First, we need to model the probability that a customer will purchase in a given time period. Since our dependent variable (*Purchase*) is binary, we select a logistic regression to estimate the model. We could also select a probit model and in general achieve the same results. In this case the *y* variable is *Purchase* and the *x* variables represent the nine independent variables in our database. When we run the logistic regression we get the following result:

| Variable | Estimate | Standard error | *p*-value |
|---|---|---|---|
| *Intercept* | −4.988 | 0.224 | <0.0001 |
| *Lag_Purchase* | 2.145 | 0.010 | <0.0001 |
| *Avg_Order_Quantity* | 0.018 | 0.001 | <0.0001 |
| *Ret_Expense* | 0.107 | 0.010 | <0.0001 |
| *Ret_Expense_SQ* | −0.002 | 0.0002 | <0.0001 |
| *Gender* | −0.202 | 0.093 | 0.030 |
| *Married* | −0.059 | 0.108 | 0.587 |
| *Income* | 0.132 | 0.032 | <0.0001 |
| *First_Purchase* | 0.001 | 0.001 | 0.226 |
| *Loyalty* | 0.268 | 0.100 | 0.008 |

As we can see from the results, seven of the nine independent variables are significant at a *p*-value of 5% or better with only *Married* and *First_Purchase* being statistically non-significant. First, this means that *Lag_Purchase* has a positive effect on current purchase, that is, customers who made a purchase in the previous quarter are more likely to make a purchase in the current quarter. Second, since the coefficient on *Avg_Order_Quantity* is positive and statistically significant, this means that customers who in the past have spent more on average are also more likely to purchase in the current time period. Third, we find a positive, but a diminishing, return on the effect of retention spending (*Ret_Expense*) on purchasing in the same quarter since the coefficient on *Ret_Expense* is positive and the coefficient on *Ret_Expense_SQ* is negative. Fourth, we find a positive effect for females (negative coefficient on *Gender*) meaning that females are generally more likely to purchase than males. Fifth, we find a positive income effect suggesting that customers who have a higher income are more likely to purchase in the current quarter. Finally, since the coefficient on *Loyalty* is positive this suggests that customers who are members of the loyalty program are more likely to purchase in the current quarter.

It is also important to understand exactly how changes in the drivers of repurchase likelihood are likely to lead to either increases or decreases in repurchase likelihood. To do this we need to determine the odds ratio for each of the parameter estimates. Since we are dealing with a logistic regression, this means that we are interested in the log-odds ratio. For example, for *Lag_Purchase*, we want to know the change in repurchase likelihood when *Lag_Purchase* = 0 and when *Lag_Purchase* = 1. For *Lag_Purchase* = 0, we get the following:

$$\text{Odds}(\text{Purchase}|\text{Lag\_Purchase} = 0) = \exp(-4.988 + 2.145(0))$$

and, for *Lag_Purchase* = 1,

$$\text{Odds}(\text{Purchase}|\text{Lag\_Purchase} = 1) = \exp(-4.988 + 2.145(1)).$$

By dividing the second equation by the first we get:

$$\frac{\text{Odds}(\text{Purchase}|\text{Lag\_Purchase} = 1) = \exp(-4.988 + 2.145(1))}{\text{Odds}(\text{Purchase}|\text{Lag\_Purchase} = 0) = \exp(-4.988)}$$

We then simplify the equation to get the following:

$$\frac{\text{Odds}(\text{Purchase}|\text{Lag\_Purchase} = 1)}{\text{Odds}(\text{Purchase}|\text{Lag\_Purchase} = 0)} = \exp(2.145) = 8.542$$

When we compute the log-odds ratio for each of the statistically significant variables we get the following results for an increase in 1 unit of the independent variable. For the case of categorical variables such as *Gender*, the log-odds ratio is merely $\exp(\beta_{\text{variable}})$.

| Variable | Log-odds ratio |
|---|---|
| *Lag_Purchase* | 8.542 |
| *Avg_Order_Quantity* | 1.018 |
| *Ret_Expense* | (0.105-0.004***Ret_Expense*) |
| *Gender* | 0.817 |
| *Income* | 1.141 |
| *Loyalty* | 1.307 |

We gain the following insights from the log-odds ratios. With regard to *Lag_ Purchase*, we see that a customer who purchased in the previous quarter is 854.2% more likely to purchase in the current quarter than a customer who did not purchase in the previous quarter. With regard to *Avg_Order_Quantity*, we see that for every increase in \$1, the probability of purchase in the current quarter increases by 1.8%. With regard to *Ret_Expense*, we see that the odds ratio is dependent on the level of *Ret_Expense*. This is due to the fact that we include both the level and squared terms for *Ret_Expense*. For example, if we usually spend \$15 on a given customer, by spending \$16 we should see an increase in the likelihood of purchase by $\exp(0.105–0.004^{*}16) = \exp(0.041) = 1.041$. This means that by increasing our spending from \$15 to \$16, we should see an increase in purchase likelihood by 4.2%. And it is important to note that this will vary depending on the initial level of *Ret_Expense*. With regard to *Gender*, we see that customers who are males are 18.3% less likely to purchase in a given period than females. With regard to *Income*, we see that for each increase in *Income* level by 1 the purchase likelihood should increase by 14.1%. Finally, with regard to *Loyalty*, we see that by being a member of the loyalty program the probability of purchase in a given quarter is 30.7% higher than a customer who is not in the loyalty program.

Now that we have determined the drivers of repurchase behavior by customers we need to use the output of the model to determine our model's predictive accuracy. To do this we need to use the estimates we obtained from the repurchase model to help us determine the predicted probability that each customer will repurchase. We use the parameter estimates from the repurchase model and values for the $x$ variables for each customer in each time period to predict whether a customer is likely to purchase in that time period. For a logistic regression we must apply the proper probability function as noted earlier in the chapter:

$$\mathrm{P}(\text{Repurchase} = 1|X\beta) = \frac{1}{1 + \exp(-X\beta)}.$$

Once we compute the probability of repurchase, we need to create a cutoff value to determine at which point we are going to divide the customers into the two groups – predicted to purchase and predicted not to purchase. There is no rule that explicitly tells us what that cutoff number should be. Often by default we select 0.5 since it is equidistant from 0 and 1. However, it is also reasonable to check multiple cutoff values and choose the one that provides the best predictive accuracy for the dataset. By using 0.5 as the cutoff for our example, any customer whose predicted probability of repurchase is greater than or equal to 0.5 is classified as predicted to purchase and the rest are predicted not to purchase. To determine the predictive accuracy we compare the predicted to the actual repurchase values in a $2 \times 2$ table. For our sample of 500 customers over 11 quarters (we drop quarter 1 since all customers purchased in quarter 1) we get Table 4.2.

As we can see from the table, our in-sample model accurately predicts 88.4% of the customers who chose not to purchase at a given time period (2494/2822) and 87.7% of the customers who chose to purchase (2348/2678). This is a significant

Table 4.2 Predicted versus actual repurchase.

| | | Actual repurchase | | |
|---|---|---|---|---|
| | | 0 | 1 | *Total* |
| Predicted repurchase | 0 | *2494* | *330* | ***2824*** |
| | 1 | *328* | *2348* | ***2676*** |
| | *Total* | ***2822*** | ***2678*** | ***5500*** |

increase in the predictive capability of a random guess model[1] which would be only 51.3% accurate for this dataset. Since our model is significantly better than the best alternative, in this case a random guess model, we determine that the predictive accuracy of the model is good. If there are other benchmark models available for comparison, the 'best' model would be the one that provides the highest accuracy of both the prediction to purchase and not to purchase, or in other words the prediction would provide the highest sum of the diagonal. In this case the sum of the diagonal is 4842 and it is accurate 88.0% of the time (4842/5500).

As a result we now know how changes in retention expense, past customer transactions, and customer characteristics are likely to either increase or decrease our likelihood of repurchase. And we also know that these drivers do a good job in helping us predict whether a customer is going to repurchase or not. This information can provide significant insights to managers who are charged with determining the optimal amount of resources to spend on retention efforts.

### 4.2.4 How do you implement it?

To implement the logistic regression in this example we used the PROC Logistic feature in SAS. To determine predictive accuracy we carried out a SAS Data step and the Freq procedure. While we did use SAS to estimate the model and determine predictive accuracy, many other statistical packages are capable of estimating a logistic regression including (but not limited to) SPSS, MATLAB, and GAUSS.

## 4.3 Lifetime duration

The biggest concern in customer duration modeling is to predict whether customers will repurchase or not in a non-contractual setting, such as in the retailing industry, and whether customers will renew a contract or not in a contractual setting, such as in the telecommunications industry. The estimation of duration is also essential for the calculation of CLV, which is an important metric for customer selection and

[1] A random guess model would do the following. First, it would determine which bucket acquire or not acquire has more customers in it. In this case 2678 customers purchased versus 2822 who did not. Then it would predict that all customers would not purchase and it would be accurate 51.3% of the time (2822/5500).

optimal marketing resource allocation. Logit and probit models are still the most popular and simple modeling methods ever used. The multinomial logit model is another similar method when researchers are to model customers' repurchase choice among several alternatives, such as different stores. To model the timing and occurrence of leaving, survival analysis which models the hazard of lapsing is by nature a suitable modeling method. Survival analysis gives the instantaneous probability of customers lapsing at time *t,* which is also called the hazard rate, conditional on non-lapsing having occurred. In some cases, researchers consider that a customer's shopping trips occur at discrete-time points, such as days or weeks, and the discrete-time hazard model will be used to model the probability that the purchase event will occur at discrete time $t$. The proportional hazards model, which does not specify any functional form for the baseline hazard, is another widely used technique in duration data modeling. This model is chosen because of its attractive features, such as the relative risk-type measure of association, no parametric assumptions, and the use of the partial likelihood function. In some circumstances, researchers consider that estimating customers' or firms' choices of contract renewal without accounting for the intrafirm association and potential heterogeneity will provide biased estimates. Random effect models, such as a random intercept model which allows intercepts not to be constant but to follow a distribution, are used in such situations. Besides these traditional statistical methods, methods from machine learning fields, such as NNs or Bayesian network classifiers, have been used to solve the binary classification problem, repurchase or not.

How long customers will stay with the company, or the duration of customers' lifetime, is another important question in customer retention modeling. In non-contractual settings, companies do not directly observe when customers defect and hence cannot decide which customers are more likely to be active or inactive. The negative binomial distribution (NBD)/Pareto model gives the probability that a customer with a particular observed transaction history is still alive at time $T$ since trial. Hazard models, such as proportional hazards models, can also estimate the duration of customers' lifetimes. A newly proposed model, the shifted-beta geometric (sBG) distribution which can fit some flexible function of time to the observed data, is able to project the survivor function beyond the observed time horizon.

In a non-contractual setting, in the catalog retailing industry, Reinartz and Kumar [8] analyzed the profitability of long-life customers (long relationship duration) versus short-life customers (short relationship duration). Since customers in the study were not bound by any contract, these authors did not actually observe when a customer defection occurred and thus did not observe each customer's lifetime duration. As a result, the authors adopted a methodology which predicts the probability that the customer is still actively engaged in the relationship with the firm. In this case they used a NBD/Pareto model. This model was proposed by Schmittlein, Morrison, and Colombo (1987) and further developed by Schmittlein and Peterson (1994). The key result of the NBD/Pareto model is an answer to the question: which individual customers are more likely to represent active or inactive customers? The outcome of the NBD/Pareto model is the probability that a customer with a particular observed transaction history is still alive at time $T$ since the first purchase occurred. Schmittlein and Peterson give the desired probability for $\alpha > \beta$ as

$$P[Alive \mid \gamma, \alpha, s, \beta, x, t, T] = \left(1 + \frac{s}{\gamma + x + s}\left\{\left(\frac{\alpha + T}{\alpha + t}\right)^{\gamma + x}\left(\frac{\beta + T}{\alpha + t}\right)^{s} F[a_1, b_1; c_1; z_1(t)] - \left(\frac{\beta + T}{\alpha + T}\right)^{s} F[a_1, b_1; c_1; z_1(T)]\right\}^{-1}\right) \tag{4.16}$$

where

$$a_1 = \gamma + x + s, \quad b_1 = s + 1, \quad c_1 = \gamma + x + s + 1, \quad z_1(y) = (\alpha - \beta)/(\alpha + y);$$

$F(a_1, b_1; c_1; z_1)$ is the Gauss hypergeometric function; $\gamma$, $\alpha$, $s$, and $\beta$ are model parameters; $x$ is the number of purchases since the first purchase; $t$ is the time since the most recent transaction occurred; and $T$ is the time since the first purchase occurred. The corresponding probabilities for $\beta > \alpha$ and $\alpha = \beta$ are then derived by Schmittlein and Peterson (1994, p. 65). Given that the outcome of the NBD/Pareto model is a continuous probability estimate, Schmittlein and Peterson's (1994) model is extended by transforming the continuous $P(Alive)$ estimate into a dichotomous alive/dead measure. Knowing a person's 'time of birth' and given a specified probability level, the authors can approximate when a customer is deemed to have left the relationship. They can then estimate the lifetime duration from birth $t_0$, until the data associated with the cutoff threshold, $t_{cutoff}$. A key assumption of this model is that the time when the customer made the purchase is known. The dataset that Reinartz and Kumar [8] used fulfilled this requirement and the observations were not left-censored.

Similarly in a paper by Fader, Hardie, and Lee (2005), the authors proposed an alternative to the NBD/Pareto model which requires the same input values ($x$, $t$, and $T$) and generally the same outcome, but is much less computationally burdensome to implement. The model proposed is the beta geometric negative binomial distribution or BG/NBD model. The key benefit of this approach is that it can be implemented using Microsoft Excel.

In a study of retailing stores by Meyer-Waarden [9], one of the basic variables used was customer defection (coded as a binary variable (0/1)). In this case the situation was non-contractual, but the author created a heuristic to use in order to determine which customers should be considered no longer active. A customer was considered no longer a customer when the time between the last purchase in a given store and the end of the observation period was greater than four times the average interpurchase time for that same point of sale. The author included two variables in a proportional hazards model: a positive random variable $T$ that denoted the lifetime and a binary variable for whether the defection event occurs. Thus in the proportional hazards model, the survival function $S(t)$, which denotes the likelihood that the customer will not to have left a given store by time $t$, is as follows:

$$S(t) = \Pr(T = t) = 1 - F(t) = 1 - \Pr(T < t) \tag{4.17}$$

where $f(t)$, the probability density function, denotes the likelihood that a customer will defect at moment $t$. It is calculated as the product of the survival function $S(t)$ and the hazard function $h(t)$:

$$f(t) = \lim[\Pr(\mathrm{t} < T < t + dt)] = h(t)S(t), \tag{4.18}$$

where $h(t)$ is the hazard function and denotes the likelihood that defection occurs at duration time $t$, given that it has not occurred in the duration time $[0, t]$. It also represents the ratio between $f(t)$ and $S(t)$:

$$h(t) = \Pr\left(\frac{t \leq T \leq t + dt}{T > t}\right) = \frac{f(t)}{[1 - F(t)]} = \frac{f(t)}{S(t)}. \tag{4.19}$$

In Bolton [10], $h(t)$ is considered as the conditional likelihood that service termination occurs at duration time $t$, given that it has not occurred in the duration time $[0, t]$. Bolton [10] denoted $h(t|X)$ as the hazard rate for a customer $i$ with specific characteristics captured by the vector $X$ (such as different levels of overall satisfaction). The hazard rate was assumed to take the form

$$h(t|X) = h_0(t)\exp(\beta' x_{it}) \tag{4.20}$$

where $h_0(t)$ is the baseline hazard function that estimates longitudinal effects and the effects of independent variables on hazard rate. The parameter estimates of the proportional hazards model are obtained by maximizing the partial likelihood which is given by

$$L(\beta) = \prod_{Y_i Uncensored} \frac{\exp(X_i\beta)}{\sum_{Y_j \geq Y_i} \exp(X_i\beta)}. \tag{4.21}$$

The hazard ratios and confidence intervals can be estimated by the maximum likelihood method. The author modeled the lifetime duration for one of the existing stores as

$$S(t) = [S_0(t)]^p, \quad \text{where} \quad p = \exp(bx). \tag{4.22}$$

Meyer-Waarden [9] included as explanatory variables dummy variables for loyalty cards, the distance of the household from the stores, and the SOW for each store as the focal store.

While survival analysis is widely used in estimating customers' lifetime duration in contractual settings, Fader and Hardie [11] suggested using a methodology that is able to project the survivor function beyond the observed time horizon. These authors proposed an approach, the sBG distribution, which can fit some flexible function of time to the observed data. Following Fader and Hardie [11], the proposed model is based on two assumptions. First, it assumes that an individual remains a customer of the firm with constant retention probability $1-\theta$. This also means that the duration, $T$, is characterized by the (shifted) geometric distribution with probability mass function and survivor function

$$P(T=t|\theta)=\theta(1-\theta)^{t-1}, \quad t=1,2,3,\ldots \tag{4.23}$$

$$S(t|\theta)=(1-\theta)^{t}, \quad t=1,2,3,\ldots. \tag{4.24}$$

Second, heterogeneity in $\theta$ follows a beta distribution with probability density function (PDF)

$$f(\theta|\alpha,\beta)=\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}, \quad \alpha,\beta>0,$$

where $B(\cdot,\cdot)$ is the beta function. The beta distribution is a flexible distribution that is bounded between zero and one characterizing heterogeneity in the churn probabilities. Based on the value of parameters $\alpha$ and $\beta$, the churn probabilities could be 'U-shaped,' homogeneous, 'J-shaped,' or 'reverse-J-shaped.' To compute the probability that a customer fails to renew his/her contract at the end of period $t$ or survives beyond period $t$, $P(T=t)$ and $S(t)$, the authors took the expectation over the beta distribution that characterizes the cross-sectional heterogeneity in $\theta$ to arrive at the expressions for a randomly chosen individual:

$$P(T=t|\alpha,\beta)=\frac{B(\alpha+1,\beta+t-1)}{B(\alpha,\beta)}, \quad t=1,2,\ldots \tag{4.25}$$

$$S(t|\alpha,\beta)=\frac{B(\alpha,\beta+t)}{B(\alpha,\beta)}, \quad t=1,2,\ldots. \tag{4.26}$$

The sBG probabilities are then calculated by using the following forward recursion formula from $P(T=1)$ :

$$P(T=t)=\begin{cases}\dfrac{\alpha}{\alpha+\beta}, & t=1\\[2ex] \dfrac{\beta+t-2}{\alpha+\beta+t-1}P(T=t-1), & t=2,3,\ldots.\end{cases} \tag{4.27}$$

The authors proposed the sBG distribution as the model for the duration of customer relationships in a discrete-time contractual setting, where transactions can occur only at fixed points in time. This model can be implemented in a simple Excel spreadsheet.

Since the data that researchers have usually have right-censored observations, the Tobit model is a commonly used method in lifetime duration estimation. After correcting the selection bias problem using Heckman's (1979) two-step procedure, Reinartz, Thomas, and Kumar (2005) forecasted the expected relationship duration for each customer with a standard right-censored Tobit model. Thomas [12] also adopted a standard right-censored Tobit model to capture the length of time a customer spends with a firm.

### 4.3.1 Empirical example: Lifetime duration

One of the key questions we want to answer with regard to lifetime duration is whether we can determine which customers have the highest likelihood of being active in the future. In a non-contractual setting this means estimating the probability that the customer is currently active given his/her past purchase history. In the case of a contractual setting this means estimating the expected lifetime of the customers who have yet to defect given the historical information about all customers in the past (including those who have already defected). This is often done using an accelerated failure time or proportional hazards model. Since the data provided for this chapter represent a non-contractual setting (i.e., we do not observe customer defection), our goal is to determine the probability that a customer is active given that customer's past purchase history. To do this we need to have information on the transaction behavior of each customer including the timing of the first purchase, the timing of the most recent purchase, and the number of transaction which have occurred during the observation window. In the dataset provided for this chapter we have a detailed description of the transaction history for each customer. We just need to compute the values for each of the three required variables. At the end of this example you should be able to identify the following:

1. The probability that a customer is active at the end of quarter 12.

A B2C firm wants to improve its ability to identify customers who are likely to still be actively engaged in a relationship with the firm. A random sample of 500 customers from a single cohort was taken from the customer database. The information we need for our model includes the following list of variables:

| Variables | |
|---|---|
| $x$ | *The number of transactions by a given customer over all time periods. Here we assume that it is the sum of the variable Purchase where customers at most made 1 purchase per quarter* |
| $t_x$ | *This is the time of the last transaction, that is, the last quarter where Purchase = 1* |
| $T$ | *The total time between the first purchase and the end of the observation window, that is, 12 quarters for all customers* |

In this case we do not have a dependent variable since we do not actually observe customer defection. Instead, we use attributes of the customer transaction information to form probabilities of a customer being active. In this case we require the number of transactions a customer has had over the observation window ($x$). To simplify this case we assume that customers only purchase at most once from any quarter. Thus, when $Purchase = 1$, we observe one transaction. Second, we require the last time when a customer had a transaction with the firm ($t_x$). In this case it is the last quarter where $Purchase = 1$. Finally, we require the time when the customer has been a customer. Since this is a cohort of customers who all made their first purchases in quarter 1, they all receive a value of 12 for $T$.

Once we have created our data table, we use the BG/NBD framework as described by Fader, Hardie, and Lee (2005) to estimate the parameters $r$, $\alpha$, $a$, and $b$ using the following likelihood function:

$$L(r, \alpha, a, b|x, t_x, T) = \frac{B(a, b + x + 1)}{B(a, b)} \frac{\Gamma(r + x)\alpha^r}{\Gamma(r)(\alpha + T)^{r+x}} + \frac{B(a + 1, x)}{B(a, b)} \frac{\Gamma(r + x)\alpha^r}{\Gamma(r)(\alpha + t_x)^{r+x}} \tag{4.28}$$

where $B(\cdot)$ is the beta distribution function, $\Gamma(\cdot)$ is the gamma distribution function, $a$ and $b$ are the parameters of the beta distribution function, $r$ and $\alpha$ are the parameters of the gamma distribution function, and $x$, $t_x$, and $T$ are the data from the firm. We estimate the model using Solver in Excel to obtain the parameter estimates for $a$, $b$, $\alpha$, and $r$. We get the following result:

| Parameter | Estimate |
|---|---|
| *R* | 126.537 |
| $\alpha$ | 159.864 |
| *a* | 0.512 |
| *b* | 3.329 |
| *Log-likelihood* | −4676.1 |

Once we obtain the parameter estimates from the model we then need to estimate the probability that each customer in the database is still active at the end of the 12th quarter, or $P(Alive)$. To do this we need to construct the equation for $P(Alive)$. Given that we used the BG/NBD framework, it follows that

$$P(Alive|r, \alpha, a, bx, t_x, T) = 1/\left[1 + \frac{a}{b + x}\left(\frac{\alpha + T}{\alpha + t_x}\right)^{r+x}\right]. \tag{4.29}$$

Using this equation we are able to solve for the probability that each customer is active. We summarize the results in the following table:

| | $P(Alive)$ |
|---|---|
| *Minimum* | 0.2% |
| *Average* | 50.1% |
| *Maximum* | 96.8% |

The results of this table show that the average customer in the database has a 50.1% chance of being active given the purchase histories of all the customers. Given the low minimum (0.2%) and the relatively high maximum (96.8%), the results also show that there is quite a bit of variation in the probability that a customer is active, meaning that it is highly likely that different customers will respond quite differently to retention efforts from the firm given their likelihood of still being actively engaged in a relationship with the firm.

As a result we now know the probability that each customer is active at the end of the 12th quarter. This information can provide significant insights to managers who are charged with determining the optimal amount of resources to spend on retention efforts in the future by only targeting those customers who are likely to still be engaged in a relationship with the firm.

### 4.3.2 How do you implement it?

To implement this example we used the Solver function in Excel. While we did use Excel to estimate the model, many other statistical packages are capable of estimating this model including (but not limited to) MATLAB, GAUSS, and R.

## 4.4 Order quantity and order size

### 4.4.1 How much (in $) will a customer order?

Loyalty programs are designed to increase the level of customer loyalty and encourage customers to make more purchases or use more services. Bolton *et al.* [22] investigated whether loyalty programs increase customers' satisfaction and their usage levels of products/services. Since the data used had customers' transactions, these authors directly modeled the number of transactions during a certain period using a Tobit model. Bolton and Lemon [13] modeled customers' usage level in a dynamic model. These authors argued that there is some minimum unobserved level of usage (threshold) associated with use of the service at all, say $c$. The amount of usage actually observed can be written as

$$\begin{gathered} Usage_{t+1} = \\ Usage\ Value^*_{t+1} \text{ if } Usage\ Value^*_{t+1} > c, \text{ and} = 0 \text{ if } Usage\ Value^*_{t+1} \leq c, \\ Usage\ Value_{t+1} = F(x_{1t}, x_{2t}, Z). \end{gathered} \tag{4.30}$$

The independent variables in the future usage equation are overall satisfaction, price, and a vector of cross-sectional economic variables. This specification is easily captured by a Tobit model. Following Borle *et al.* [23], the authors assumed that the amount expended by customer $h$ on purchase occasion $i$, $AMNT_{hi}$, followed a log-normal process

$$\log AMNT_{hi} \sim Normal(\mu_{hi}, \sigma^2), \tag{4.31}$$

where the $\mu_{hi}$ parameter is

$$\mu_{hi} = \mu_h + \mu_i + \mu_{1h} \log lagAMNT_{hi} + \mu_2 GENDER_h, \tag{4.32}$$

where $\mu_i = \mu' i + \mu'' i^2$. The coefficient $\mu_2$ specifies the impact of gender on purchase amounts and the coefficient $\mu_i$ captures a nonlinear trend in the purchase amounts across purchase occasions. The coefficient $\mu_{1h}$ specifies the impact of lagged dollars spent on future amounts expended. The authors allowed this parameter to vary across customers as

$$\mu_{1h} \sim Normal\left(\bar{\mu}_1, \tau_2^2\right) \tag{4.33}$$

and jointly estimated the purchase amount with the interpurchase time and the customer defection model using a MCMC sampling algorithm.

### 4.4.2 How many items will a customer order?

Anderson and Simester [14] argued that the number of units ordered can possibly be zero, a situation which should not be considered as censoring and truncation. The authors thus considered that the Tobit model is not appropriate for quantity modeling and they proposed that the number of units purchased is a count measure, following a Poisson distribution. They used a Poisson model to specify the number of units ordered by customer $i(Q_i)$ as

$$Prob(Q_i = q) = \frac{\exp(-\lambda_i)\lambda_i^q}{q!}, \quad q = 0, 1, 2, \ldots, \tag{4.34}$$

where $Q_i$ is assumed drawn from a Poisson distribution with parameter $\lambda_i$, $\ln(\lambda_i) = \beta x_i$, and $x_i$ include the RFM measures and a dummy variable indicating whether customer $i$ received the promotion. We provide an introduction to the Poisson regression model in Appendix J.

In modeling count event, the data are often overdispersed, meaning conditional variance exceeding conditional mean. Since Poisson regression assumes equidispersion (equality of mean and variance), it is too restrictive so that negative binomial regression is often used as an alternative modeling method. Zhang *et al.* [24] adopted negative binomial regression to model the number of brand purchases in a given time period. In their study, the number $X$ of items of brand $H$ that household $i$

purchased during a one-year period is assumed to follow a Poisson distribution with a mean purchase rate of $\lambda_i$, which is determined by a set of explanatory variables in the negative binomial regression (NBR) model. $\lambda_i$ is parameterized as

$$\lambda_i = \exp(X'_i\beta)\exp(\varepsilon_i) \tag{4.35}$$

where $X_i$ denotes the explanatory variables, such as brand $H$'s price, share of advertisements in the category, share of displays in the category, customer loyalty, household size, and household income, and $\varepsilon_i$ is the error term that is assumed to follow a gamma distribution in the NBR specification. An introduction to the NBR model of Cameron and Trivedi (1998) is provided in Appendix K.

### 4.4.3 What is the average order size?

Marketing activities may influence the order size, such as the average unit price, by existing customers. Researchers have investigated such effects by linear regression. Anderson and Simester [14] adopted a multivariate regression (OLS) to analyze the effect of discount depth on the average price of products purchased by existing customers. These authors included RFM measures, promotion dummy variables, and the average price of products by existing customers in previous purchases. Lewis [15] modeled the average order size of existing customers including shipping fee variables, pricing variables, and coupon promotion dummy variables as explanatory variables. To account for the possible endogenous bias in the systems of equations, the author adopted a three-stage least squares estimation in modeling order incidence, order size, and net shipping contribution.

### 4.4.4 Empirical example: Order quantity

Many firms have realized that it is not sufficient to merely focus on just trying to get a customer to repurchase. The firm should also focus on how much value that purchase is likely to provide. Research in marketing has shown that the order value can be a valuable predictor in a customer's future value to the firm – or at the least justify the amount of money that is spent on customer retention efforts. Thus, it can be useful to understand the drivers of order quantity and in turn be able to predict each prospect's expected order quantity given an order is likely to occur. At the end of this example we should be able to do the following:

1. Determine the drivers of order quantity (value).
2. Predict the expected order quantity for each customer.
3. Determine the predictive accuracy of the model.

The information we need for this model includes the following list of variables:

| | |
|---|---|
| **Dependent variables** | |
| *Purchase* | *1 when the customer purchased in the given quarter, 0 if no purchase occurred in that quarter* |
| *Order_Quantity* | *The dollar value of the purchases in the given quarter* |
| **Independent variables** | |
| *Lambda* ($\lambda$) | *The computed inverse Mills ratio from the acquisition model* |
| *Lag_Purchase* | *1 if the customer purchased in the previous quarter, 0 if no purchase occurred in the previous quarter* |
| *Avg_Order_Quantity* | *The average dollar value of the purchases in all previous quarters* |
| *Ret_Expense* | *Dollars spent on marketing efforts to try and retain that customer in the given quarter* |
| *Ret_Expense_SQ* | *Square of dollars spent on marketing efforts to try and retain that customer in the given quarter* |
| *Gender* | *1 if the customer is male, 0 if the customer is female* |
| *Married* | *1 if the customer is married, 0 if the customer is not married* |
| *Income* | *1 if income < \$30 000*<br>*2 if \$30 001< income < \$45 000*<br>*3 if \$45 001 < income < \$60 000*<br>*4 if \$60 001 < income < \$75 000*<br>*5 if \$75 001 < income < \$90 000*<br>*6 if income > \$90 001* |
| *First_Purchase* | *The value of the first purchase made by the customer in quarter 1* |
| *Loyalty* | *1 if the customer is a member of the loyalty program, 0 if not* |

We see from the data requirement that in order to determine the drivers of order quantity we need to have two dependent variables: *Purchase* and *Order_Quantity*. This is due to the fact that expected order quantity is derived from the following equation:

$$\mathrm{E}(\mathit{Order\ Quantity}) = \mathrm{P}(\mathit{Purchase} = 1)^{*}\mathrm{E}(\mathit{Order\_Quantity}|\mathit{Purchase} = 1).$$

This equation shows us that the expected order quantity is a function of the probability that the customer will purchase in the given quarter multiplied by the expected value of a purchase given that the customer made the purchase. If we were to merely run a regression with *Order_Quantity* as the dependent variable and ignore the probability that the customer will make a purchase, we would get biased estimates due to a potential sample selection bias.

Sample selection bias is a problem that is common in many marketing problems and has to be statistically accounted for in many modeling frameworks. In this case the customer has a choice as to whether or not to purchase before deciding how much to purchase. If we were to ignore this choice we would bias the estimates from the model and we would have less precise predictions for the value of *Order_Quantity*. To account for this issue we need to be able to predict the value for both the probability of *Purchase* (similar to what we have done for the first empirical example in this chapter) and the expected value of *Order_Quantity* given that the customer is expected to make a purchase. One important consideration to note is that we cannot just run two models independently since there is likely to be a correlation between the error terms of the two models. Thus, we need to use a modeling framework that can simultaneously estimate the coefficients of the two models, or at least account for the correlation between *Order_Quantity* and *Purchase*. To do this we use a two-stage modeling framework similar to that described earlier in this chapter and found in Reinartz *et al.* [25].

The first model for *Purchase* will be set up using the same equation as for the repurchase probability example earlier in the chapter. The only difference here is that instead of using a logistic regression we will be using a probit model to estimate the coefficients. The main reason for this lies in the error term of the probit model which is normally distributed with a mean of 0 and a standard deviation of 1. The fact that the probit model and the OLS regression model (which we will be using for *Order_Quantity*) are both normally distributed allows us to more easily estimate them in a two-stage framework.

Once we estimate the probit model we need to create a new variable, λ, which will represent the correlation in the error structure across the two equations. This variable, also known as the sample selection correction variable, will then be used as an independent variable in the *Order_Quantity* model to remove the sample selection bias in the estimates. To compute λ we use the following equation, also known as the inverse Mills ratio:

$$\lambda = \frac{\phi(X'\beta)}{\Phi(X'\beta)}.$$

In this equation $\phi$ represents the normal PDF, $\Phi$ represents the normal cumulative density function, $X$ represents the value of the variables in the *Purchase* model, and $\beta$ represents the coefficients derived from the estimation of the *Purchase* model.

Finally, we want to estimate a regression model for *Order_Quantity* and include the variable ρ as an additional independent variable. This is done in a straightforward manner in the following equation:

$$Order\_Quantity = \gamma'\alpha + \mu\lambda + \varepsilon.$$

In this case *Order_Quantity* is the value of the order quantity in the given time period, $\gamma$ is the matrix of variables used to help explain the value of *Order_Quantity*, $\alpha$ are the coefficients for the independent variables, $\mu$ is the coefficient on the inverse Mills ratio, λ is the inverse Mills ratio, and ε is the error term.

When we estimate the two-stage model, we get the following parameter estimates for each of the two equations:

| Repurchase model | Estimate | Order quantity model | Estimate |
|---|---|---|---|
| *Intercept* | −2.783 | *Intercept* | −68.630 |
| *Lag_Purchase* | 1.231 | *Lag_Purchase* | 50.769 |
| *Avg_Order_Quantity* | 0.010 | *Avg_Order_Quantity* | 0.816 |
| *Ret_Expense* | 0.059 | *Ret_Expense* | 1.361 |
| *Ret_Expense_SQ* | −0.001 | *Ret_Expense_SQ* | −0.037 |
| *Gender* | −0.114 | *Gender* | 11.377 |
| *Married* | −0.025[a] | *Married* | −2.469[a] |
| *Income* | 0.072 | *Income* | 11.037 |
| *First_Purchase* | 0.001[a] | *First_Purchase* | 0.232 |
| *Loyalty* | 0.152 | *Loyalty* | 6.005 |
| | | *Lambda* ($\lambda$) | 35.168 |

[a]Denotes not significant at $p < 0.05$.

We gain the following insights from the results. We see that $\lambda$ is positive and significant. We can interpret this to mean that there is a potential selection bias problem since the error term of our selection equation is correlated positively with the error term of our regression equation. We also see that all other variables of the order quantity model are significant with the exception of *Married*, meaning that we have likely uncovered many of the drivers of order quantity.

We find that *Lag_Purchase* is positive, suggesting that customers who purchased in the previous quarter are more likely to spend more in the current quarter. We find that *Avg_Order_Quantity* is also positive, suggesting that the higher the average past order values of the customer, the higher the current order value. We find that *Ret_Expense* is positive with a diminishing return, as noted by the positive coefficient on *Ret_Expense* and the negative coefficient on *Ret_Expense_SQ*. This means that marketing efforts to retain and build relationships with the customer do cause the customer to purchase more, to a point. Then, after the threshold is reached, marketing efforts actually decrease the value of the purchase on average. This is likely due to the fact that overly contacting customers can often strain the relationship between the customer and firm. We find that that four of the customer characteristic variables are positive (*Gender*, *Income*, *First_Purchase*, and *Loyalty*) suggesting that customers who are male, have a higher income, have a higher first purchase value, and are members of the loyalty program tend to have larger order quantities.

Our next step is to predict the value of *Order_Quantity* to see how well our model compares to the actual values. We do this by starting with the equation for expected order quantity at the beginning of this example:

$$\begin{aligned} \mathrm{E}(\mathit{Order\ Quantity}) &= \mathrm{P}(\mathit{Purchase} = 1)^{*}\mathrm{E}(\mathit{Order\_Quantity}|\mathit{Purchase} = 1) \\ &= \Phi(X'\beta)^{*}(y'\alpha + \mu\lambda). \end{aligned}$$

In this case $\Phi$ is the normal CDF distribution, $X$ is the matrix of independent variable values from the *Purchase* equation, $\beta$ is the vector of parameter estimates from the *Purchase* equation, $\gamma$ is the matrix of independent variables from the *Order_Quantity* equation, $\alpha$ is the vector of parameter estimates from the *Order_Quantity* equation, $\mu$ is the parameter estimate for the inverse Mills ratio, and $\lambda$ is the inverse Mills ratio. Once we have predicted the *Order_Quantity* value for each of the customers, we want to compare this to the actual value from the database. We do this by computing the mean absolute deviation (MAD) as follows:

$$\text{MAD} = \text{Mean}\{\text{Absolute Value}[\text{E}(\textit{Order Quantity}) - \textit{Order\_Quantity}]\}.$$

We find for the acquired customers that MAD = 54.77. This means that on average each of our predictions of *Order_Quantity* deviates from the actual value by \$54.77. If we were to instead use the mean value of *Order_Quantity* (\$129.10) across all customers across quarters 2 through 12 (we drop quarter 1 due to the lagged nature of many of the independent variables in both the repurchase model and the order quantity model) as our prediction for all prospects (this would be the benchmark model case), we would find that MAD = 133.71, or \$133.71. As we can see, our model does a significantly better job of predicting the value of initial order quantity than the benchmark case.

### 4.4.5 How do you implement it?

In this example we used a two-stage least squares approach with a probit model for acquisition and a least squares regression for the initial order quantity. We used multiple procedures in SAS to implement this model. First we used PROC Logistic with a probit link function to estimate the model of customer purchase behavior. Next we used a SAS Data step to compute the inverse Mills ratio using the output of the probit model. Finally we ran an OLS regression using PROC Reg and added the inverse Mills ratio as an additional variable. While we did use SAS to implement this modeling framework, programs such as SPSS can be used as well.

## 4.5 Cross-buying

Cross-selling is the most common technique companies use to increase the repeat purchase incidence, quantity, and revenues to companies. After companies have built a certain level of loyalty with existing customers, these customers are more likely to cross-buy from the company. Verhoef and Donkers [5] studied the effect of acquisition channels on cross-buying. The acquisition channels examined included mass media, direct marketing, the Internet, personal selling, intermediaries, and word of mouth. For customers who were retained, cross-buying was coded as a binary variable. These authors modeled cross-buying in a probit model and acquisition channels were included as dummy variables.

### 4.5.1 Empirical example: Cross-buying

Another key question we want to answer with regard to customer retention is whether we can determine which customers have the highest likelihood of cross-buying in multiple categories. To do this we first need to know which current customers actually purchased in multiple categories when they made a purchase. In the dataset provided for this chapter we have a variable *Crossbuy* which identifies how many categories of products a customer purchased in a given time period. We also provide a set of drivers which are likely to help explain a customer's decision to cross-buy. At the end of this example you should be able to do the following:

1. Identify the drivers of customer cross-buying behavior.
2. Interpret the parameter estimates from the cross-buying model.
3. Predict whether a customer is likely to cross-buy or not.
4. Determine the predictive accuracy of the cross-buying model.

A B2C firm wants to understand which customers are most likely to cross-buy in a given time period. This is important to know since many studies have shown that customers who purchase across multiple categories are more likely to be more profitable than customers who purchase across fewer categories. A random sample of 500 customers from a single cohort was taken from the customer database. The information we need for our model includes the following list of variables:

| | |
|---|---|
| **Dependent variable** | |
| *Crossbuy* | *The number of product categories a customer purchased during a given quarter* |
| **Independent variables** | |
| *Lag_Purchase* | *1 if the customer purchased in the previous quarter, 0 if no purchase occurred in the previous quarter* |
| *Lag_Crossbuy* | *The number of product categories a customer purchased during the previous quarter* |
| *Avg_Order_Quantity* | *The average dollar value of the purchases in all previous quarters* |
| *Ret_Expense* | *Dollars spent on marketing efforts to try and retain that customer in the given quarter* |
| *Ret_Expense_SQ* | *Square of dollars spent on marketing efforts to try and retain that customer in the given quarter* |
| *Gender* | *1 if the customer is male, 0 if the customer is female* |
| *Married* | *1 if the customer is married, 0 if the customer is not married* |
| *Income* | *1 if income < \$30 000*<br>*2 if \$30 001< income < \$45 000*<br>*3 if \$45 001 < income < \$60 000*<br>*4 if \$60 001 < income < \$75 000*<br>*5 if \$75 001 < income < \$90 000*<br>*6 if income > \$90 001* |
| *First_Purchase* | *The value of the first purchase made by the customer in quarter 1* |
| *Loyalty* | *1 if the customer is a member of the loyalty program, 0 if not* |

In this case, we have a discrete dependent variable (*Crossbuy*) which tells us how many categories a customer purchased in a given quarter. In order to understand the likelihood of cross-buying behavior we need to transform this discrete variable into a binary variable. We do this by setting *CB* equal to 1 when *Crossbuy* $> 1$ and *CB* equal to 0 otherwise. We also have nine independent variables which we believe will be drivers of repurchase behavior.

We believe that transaction behavior in the past is likely to explain a customer's cross-buying behavior. As a result we use several lagged operationalizations of current variables as independent variables in this example. First, we have whether or not the customer purchased in the last quarter (*Lag_Purchase*). We also have a variable for the number of product categories a customer purchased last quarter (*Lag_Crossbuy*).These two variables can be obtained by taking the lagged value of the purchase/cross-buy variable, noting that one observation will be lost for each customer for each lag that is taken. In this case we are only using a one-period lag for both variables. Second, we have the average past order quantity (*Avg_Order_Quantity*). In this case the value for average order quantity is the mean of the *Order_Quantity* variable in all quarters before the current time period. Third, we have how many dollars the firm spent on each customer (*Ret_Expense*) in each time period and the squared value of that variable (*Ret_Expense_SQ*). We want to use both the linear and squared terms since we expect that for each additional dollar spent on the retention effort for a given customer, there will be a diminishing return to the value of that dollar. Finally, since the focal firm of this example is a B2C firm, the other five variables are demographic and static variables of the customers. These include the *Gender* of the customer, whether the customer is *Married*, the *Income* of the customer, the value of the customer's first purchase (*First_Purchase*), and whether the customer is a member of the loyalty program (*Loyalty*).

First, we need to model the probability that a customer will cross-buy in a given time period. Since our dependent variable (*CB*) is binary, we select logistic regression to estimate the model. We could also select a probit model and in general achieve the same results. In this case the $y$ variable is *CB* and the $x$ variables represent the 10 independent variables in our database. We must also only select those occasions when a purchase occurred since we are interested in whether or not a customer cross-buys conditional on a purchase occasion. This gives us a total of 2678 observations for this sample. When we run the logistic regression we get the following result:

| Variable | Estimate | Standard error | $p$-value |
|---|---|---|---|
| *Intercept* | −7.641 | 0.357 | <0.0001 |
| *Lag_Purchase* | 0.137 | 0.169 | 0.4157 |
| *Lag_Crossbuy* | 0.162 | 0.044 | 0.0002 |
| *Avg_Order_Quantity* | 0.012 | 0.001 | <0.0001 |
| *Ret_Expense* | 0.038 | 0.010 | <0.0001 |
| *Ret_Expense_SQ* | −0.001 | 0.0002 | 0.0004 |
| *Gender* | 0.260 | 0.010 | 0.0072 |

| | | | |
|---|---|---|---|
| *Married* | 0.181 | 0.111 | 0.103 |
| *Income* | 0.233 | 0.035 | <0.0001 |
| *First_Purchase* | 0.024 | 0.001 | <0.0001 |
| *Loyalty* | 0.258 | 0.108 | 0.0165 |

As we can see from the results, 8 of the 10 independent variables are significant at a $p$-value of 5% or better with only *Lag_Purchase* and *Married* being statistically non-significant. First, this means that *Lag_Crossbuy* positively affects current cross-buy, that is, customers who purchased across more categories in the previous quarter are more likely to cross-buy in the current quarter. Second, since the coefficient on *Avg_Order_Quantity is* positive and statistically significant, this means that customers who in the past have spent more on average are also more likely to cross-buy in the current time period. Third, we find a positive, but a diminishing, return on the effect of retention spending (*Ret_Expense*) on cross-buying in the same quarter since the coefficient on *Ret_Expense* is positive and the coefficient on *Ret_Expense_SQ* is negative. Fourth, we find that males are more likely to cross-buy than females. Fifth, we find a positive income effect suggesting that customers who have a higher income are more likely to cross-buy in the current quarter. Sixth, we find that customers who had a higher *First_Purchase* are more likely to cross-buy. Finally, since the coefficient on *Loyalty* is positive, this suggests that customers who are members of the loyalty program are more likely to cross-buy in the current quarter.

It is also important to understand exactly how changes in the drivers of cross-buy likelihood are likely to lead to either increases or decreases in cross-buy likelihood. To do this we need to determine the odds ratio for each of the parameter estimates. Since we are dealing with logistic regression, this means that we are interested in the log-odds ratio. For example, for *Lag_Crossbuy*, we want to know the change in cross-buy likelihood when *Lag_Crossbuy* = x and when *Lag_Crossbuy* = x + 1. For *Lag_Crossbuy* = x, we get the following:

$$\text{Odds}(\text{Cross-buy}|\text{Lag\_Crossbuy} = 0) = \exp(-7.641 + 0.162(\text{x}))$$

and, for *Lag_Purchase* = 1,

$$\text{Odds}(\text{Cross-buy}|\text{Lag\_Crossbuy} = 1) = \exp(-7.641 + 0.162(\text{x} + 1)).$$

By dividing the second equation by the first we get

$$\frac{\text{Odds}(\text{Cross-buy}|\text{Lag\_Crossbuy} = \text{x} + 1) = \exp(-7.641 + 0.162(\text{x} + 1))}{\text{Odds}(\text{Cross-buy}|\text{Lag\_Crossbuy} = \text{x}) = \exp(-7.641 + 0.162(\text{x}))}.$$

We then simplify the equation to get the following:

$$\frac{\text{Odds}(\text{Cross-buy}|\text{Lag\_Crossbuy} = \text{x} + 1)}{\text{Odds}(\text{Cross-buy}|\text{Lag\_Crossbuy} = \text{x})} = \exp(0.162) = 1.176.$$

When we compute the log-odds ratio for each of the statistically significant variables we get the following results for an increase in 1 unit of the independent variable. For the case of categorical variables such as *Gender*, the log-odds ratio is merely $\exp(\beta_{\text{variable}})$.

| Variable | Log-odds ratio |
|---|---|
| *Lag_Purchase* | 1.176 |
| *Avg_Order_Quantity* | 1.012 |
| *Ret_Expense* | (0.037−0.002* *Ret_Expense*) |
| *Gender* | 1.297 |
| *Income* | 1.262 |
| *First_Purchase* | 1.024 |
| *Loyalty* | 1.294 |

We gain the following insights from the log-odds ratios. With regard to *Lag_Crossbuy*, we see that each additional category a customer purchased in the previous quarter makes the customer 17.6% more likely to cross-buy in the current quarter. With regard to *Avg_Order_Quantity*, we see that for every increase in \$1, the probability of cross-buying in the current quarter increases by 1.2%. With regard to *Ret_Expense*, we see that the odds ratio is dependent on the level of *Ret_Expense*. This is due to the fact that we include both the level and squared terms for *Ret_Expense*. For example, if we usually spend \$15 on a given customer, by spending \$16 we should see an increase in the likelihood of cross-buy by $\exp(0.037-0.002^{*}16)=\exp(0.005)=1.005$. This means that by increasing our spending from \$15 to \$16, we should see an increase in cross-buy likelihood by 0.5%. And, it is important to note that this will vary depending on the initial level of *Ret_Expense*. With regard to *Gender*, we see that males are 29.7% more likely to cross-buy than females. With regard to *Income*, we see that for each increase in *Income* level by 1 the cross-buy likelihood should increase by 26.2%. With regard to *First_Purchase*, we see that for each increase in \$1, the likelihood of cross-buy increases by 2.4%. Finally, with regard to *Loyalty*, we see that by being a member of the loyalty program the probability of cross-buy in a given quarter is 29.4% higher than a customer that is not in the loyalty program.

Now that we have determined the drivers of cross-buying behavior by customers we need to use the output of the model to determine our model's predictive accuracy. To do this we need to use the estimates we obtained from the cross-buy model to help us determine the predicted probability that each customer will cross-buy. We use the parameter estimates from the cross-buy model and values for the $x$ variables for each customer in each time period to predict whether a customer is likely to purchase in that time period. For a logistic regression we must apply the proper probability function as follows:

$$P(\text{Crossbuy}=1|X\beta)=\frac{1}{1+\exp(-X\beta)}.$$

Once we compute the probability of cross-buy, we need to create a cutoff value to determine at which point we are going to divide the customers into the two groups – predicted to cross-buy and predicted not to cross-buy. There is no rule that explicitly tells us what that cutoff number should be. Often by default we select 0.5 since it equidistant from 0 and 1. However, it is also reasonable to check multiple cutoff values and choose the one that provide the best predictive accuracy for the dataset. By using 0.5 as the cutoff for our example, any customer whose predicted probability of cross-buy is greater than or equal to 0.5 is classified as predicted to cross-buy and the rest are predicted not to cross-buy. To determine the predictive accuracy we compare the predicted to the actual cross-buy values in a $2 \times 2$ table. For our sample of 500 customers over 11 quarters and 2678 purchase occasions (we drop quarter 1 since all customers purchased in quarter 1) we get Table 4.3.

As we can see from the table, our in-sample model accurately predicts 74.4% of the customers who chose not to cross-buy on a given purchase occasion (980/1317) and 77.4% of the customers who chose to cross-buy on a given purchase occasion (1053/1361). This is a significant increase in the predictive capability of a random guess model[2] which would be only 50.8% accurate for this dataset. Since our model is significantly better than the best alternative, in this case a random guess model, we determine that the predictive accuracy of the model is good. If there are other benchmark models available for comparison, the 'best' model would be the one which provides the highest accuracy of both the prediction to cross-buy and not to cross-buy, or in other words the prediction would provide the highest sum of the diagonal. In this case the sum of the diagonal is 2033 and it is accurate 75.9% of the time (2033/2678).

As a result we now know how changes in retention expense, past customer transactions, and customer characteristics are likely to either increase or decrease our likelihood of cross-buy. And we also know that these drivers do a good job in helping us predict whether a customer is going to cross-buy or not. This information can provide significant insights to managers who are charged with determining which customers are most likely to cross-buy.

Table 4.3 Predicted versus actual cross-buy.

| | | Actual cross-buy | | |
|---|---|---|---|---|
| | | 0 | 1 | *Total* |
| Predicted cross-buy | 0 | *980* | *308* | ***1288*** |
| | 1 | *337* | *1053* | ***1390*** |
| | *Total* | ***1317*** | *1361* | ***2678*** |

[2] A random guess model would do the following. First, it would determine which bucket cross-buy or not cross-buy has more instances in it. In this case 1361 cross-buy occasions happened versus 1317 that did not. Then it would predict that all customers would cross-buy and it would be accurate 50.8% of the time (1361/2678).

### 4.5.2 How do you implement it?

To implement the logistic regression in this example we used the PROC Logistic feature in SAS. To determine predictive accuracy we carried out a SAS Data step and the Freq procedure. While we did use SAS to estimate the model and determine predictive accuracy, many other statistical packages are capable of estimating a logistic regression including (but not limited to) SPSS, MATLAB, and GAUSS.

## 4.6 SOW

Loyalty programs and direct mailings are two ways that companies use to manage customer relationships. The goal is to build close relationships with customers and thus enhance customers' relationship perceptions. Researchers have investigated the effects of companies' CRM efforts on customer retention and customer share development. Verhoef [4] considered customer retention as a binary outcome, and customer share of customer $i$ for supplier $j$ in category $k$ at time $t$ was defined as number of services purchased in category $k$ at supplier $j$ at time $t$ divided by number of services purchased in category $k$ from all suppliers at time $t$. The author first modeled customer retention with a probit model for all customers and then modeled customer share with a regression model for the customers who remained with the company. The author used Heckman's (1976) two-step procedure to correct for sample selection bias by incorporating the inverse Mills ratio obtained from the probit model. For both the probit and regression models, the author included past purchase behavior, customer relationship perceptions, and loyalty program dummy variables as independent variables. For the regression model, the difference between the logs of customer share at $T_1$ and $T_0$ was the dependent variable denoting the customer share development.

Another important question related to repurchasing behavior and SOW was addressed by Leenheer *et al.* [26]. These authors investigated whether loyalty programs really do enhance behavioral loyalty (i.e., repeated purchase behavior). They also investigated the effect of these loyalty programs on a customer's SOW in a supermarket chain. Since the authors argued that the SOW a customer has for a store depends on its attraction to customers, the actual modeling task was to build an attraction model. In this case the SOW of a given customer was defined as

$$SOW_{is} = \frac{A_{is}}{\sum_{S}^{s=1} A_{is}}. \tag{4.36}$$

Since there were more than two stores in the study, the attraction to a store $A_{is}$ could not be modeled by binary logistic regression. Instead the authors chose to model the outcome using a multinomial logit model. This type of model is chosen when the dependent variable is nominal and has more than two categories. As an example, let us assume we have $j$ categories and the first category is defined as the reference category. Then the log-odds ratio of the $j$th category is defined as

$$\ln \frac{P(Y_i = j)}{P(Y_i = 1)} = \alpha_j + \sum_{k=1}^{K} \beta_{jk} X_{ik} = D_{ji}. \tag{4.37}$$

The predicted probability for categories $j = 2, \ldots, J$ relative to the reference category is

$$P(Y_i = j) = \frac{\exp(D_{ji})}{1 + \sum_{q=2}^{M} \exp(D_{qi})} \tag{4.38}$$

and, for the reference category, the predicted probability is

$$P(Y_i = 1) = \frac{1}{1 + \sum_{q=2}^{M} \exp(D_{qi})}. \tag{4.39}$$

Again, the multinomial logit model can be estimated by the MLE method. An introduction to the model from Greene's [27] econometric textbook is provided in Appendix appF. In Leenheer *et al.*'s (2007) study, the explanatory variables include household and store characteristics and loyalty program membership. These authors argued that an attraction model cannot take zeros but SOW can be zero if customers do not choose to buy from the store, so that in this case the authors used a probit model to estimate a selection variable to account for the correlated error terms of the attraction and SOW models. In addition, the authors argued that there might be a self-selection problem that loyal customers are more likely to enroll in a loyalty program leading to a potential endogeneity of loyalty program membership. To account for endogeneity, the authors adopted a two-stage least squares (2SLS) procedure using several instrumental variables to correct for the estimation bias of loyalty program membership.

In another study related to SOW, Cooil *et al.* (2007) used data from the Canadian banking industry to try and predict how changes in a customer's satisfaction would lead to changes in that customer's SOW at the given bank. These authors used a two-level latent class regression model to uncover the effect of changes in satisfaction on changes in SOW which were moderated by several demographic and situational characteristics. This two-level latent class regression was able to allow for household-level random effects within the latent class structure.

### 4.6.1 Empirical example: SOW

Besides understanding a customer's purchase patterns from a given firm, many firms also want to understand how a customer spreads purchases in a given category across all firms. Research has shown that understanding a customer's SOW with a given firm can help understand the likelihood that a customer is going to repurchase from a given firm and inevitably the customer's long-term value to the firm. Thus, it can be useful to understand the drivers of SOW and in turn be able to predict each customer's expected SOW. At the end of this example we should be able to do the following:

1. Determine the drivers of SOW.

2. Predict the expected SOW for each customer.

3. Determine the predictive accuracy of the model.

The information we need for this model includes the following list of variables:

| | |
|---|---|
| **Dependent variables** | |
| *Share-of-Wallet (SOW)* | *The percentage of purchases the customer makes from the given firm, given the total amount of purchases across all firms in that category* |
| **Independent variables** | |
| *Purchase_Rate* | *The average value for purchases across all 12 quarters* |
| *Avg_Order_Quantity* | *The average dollar value of the purchases in all 12 quarters* |
| *Avg_Crossbuy* | *The average value for cross-buy across all 12 quarters* |
| *Avg_Ret_Expense* | *Average dollars spent on marketing efforts to try and retain that customer in all 12 quarters* |
| *Avg_Ret_Expense_SQ* | *Square of average dollars spent on marketing efforts to try and retain that customer in all 12 quarters* |
| *Gender* | *1 if the customer is male, 0 if the customer is female* |
| *Married* | *1 if the customer is married, 0 if the customer is not married* |
| *Income* | *1 if income < \$30 000*<br>*2 if \$30 001< income < \$45 000*<br>*3 if \$45 001 < income < \$60 000*<br>*4 if \$60 001 < income < \$75 000*<br>*5 if \$75 001 < income < \$90 000*<br>*6 if income > \$90 001* |
| *First_Purchase* | *The value of the first purchase made by the customer in quarter 1* |
| *Loyalty* | *1 if the customer is a member of the loyalty program, 0 if not* |

In this case we have a limited dependent variable (*SOW*) which falls on the continuum between 1 and 100. The minimum is this case is 1% since all the customers in our database made at least one purchase with the given firm, and the maximum is 100% since all the customers in the database could potentially purchase these products from only this firm. Thus, we need to account for this bounded dependent variable. In this case we can use a variation of the Tobit model we used in previous examples. In the standard Tobit model case we have the situation where the lower bound of the dependent variable is defined, usually at 0, and the upper bound of the Tobit model is infinite. However, in this case we need to accommodate both the lower bound censoring and the upper bound censoring, where the lower bound is 1 and the upper bound is 100. Thus, we have the following definition for SOW:

$$SOW_i = \begin{cases} 100 & \text{if } SOW_i^* \geq 100 \\ SOW_i^* & \text{if } 1 < SOW_i^* < 100 \\ 1 & \text{if } SOW_i^* \leq 1. \end{cases}$$

In this case the SOW for customer $i$ is only truly observed when the value is between 1 and 100. When the value of the SOW is 1 or 100, we only observe the censored value of SOW. To estimate this model we need to be able to handle observation-by-observation censoring using the following log-likelihood:

$$LL = \sum\nolimits_{i\in[1<SOW<100]} \ln\left[\phi\left(\frac{SOW - X'\beta}{\sigma}\right)/\sigma\right] \\ + \sum\nolimits_{i\in[SOW=100]} \ln\left[\Phi\left(-\frac{100 - X'\beta}{\sigma}\right)\right] + \sum\nolimits_{i\in[SOW=100]} \ln\left[\Phi\left(\frac{1 - X'\beta}{\sigma}\right)\right]$$

where $SOW$ is the SOW of a given customer, $X$ is the matrix of independent variables, $\beta$ is the vector of coefficients of the independent variables, $\phi$ denotes the normal PDF, $\Phi$ denotes the normal CDF, and $\sigma$ is the estimated standard error. Our objective then is to maximize the log-likelihood function through the estimation of the coefficients ($\beta$) and the standard error of the equation ($\sigma$). We get the following results when we estimate the model:

| Variable | Estimate | Standard error | $p$-value |
|---|---|---|---|
| *Intercept* | −1.738 | 3.711 | 0.6395 |
| *Purchase_Rate* | 7.846 | 6.279 | 0.2115 |
| *Avg_Order_Quantity* | 0.196 | 0.025 | <0.0001 |
| *Avg_Crossbuy* | 6.436 | 1.861 | 0.0005 |
| *Avg_Ret_Expense* | 0.857 | 0.405 | 0.0343 |
| *Avg_Ret_Expense_SQ* | −0.050 | 0.015 | 0.0010 |
| *Gender* | 0.589 | 1.024 | 0.5654 |
| *Married* | 1.079 | 1.146 | 0.3462 |
| *Income* | 1.780 | 0.360 | <0.0001 |
| *First_Purchase* | 0.052 | 0.013 | <0.0001 |
| *Loyalty* | 5.126 | 1.081 | <0.0001 |
| *Sigma* ($\sigma$) | 10.732 | 0.367 | <0.0001 |

We gain the following insights from these results. We find that all the variables with the exception of *Purchase_Rate*, *Gender*, and *Married* are statistically significant at $p < 0.05$. We find that *Avg_Order_Quantity* is positive, suggesting that the higher the average past order values of the customer, the higher the SOW. We find that *Avg_Crossbuy* is positive, suggesting that the more a customer has bought across multiple categories in the past, the higher the customer's SOW. We find that *Ret_Expense* is positive with a diminishing return, as noted by the positive coefficient on *Ret_Expense* and the negative coefficient on *Ret_Expense_SQ*. This means that marketing efforts to retain and build relationships with the customer do cause the customer to have a higher SOW. Then, after the threshold is reached, marketing efforts actually decrease the SOW on average. This is likely due to the fact that overly contacting customers can often strain the relationship between the customer and firm. We find that three of the customer characteristic variables are positive (*Income*, *First_Purchase*, and *Loyalty*) suggesting that customers with a higher income, higher first-purchase value, and who are members of the loyalty program are likely to have a higher SOW.

Our next step is to predict the value of SOW to see how well our model compares to the actual values. We do this by starting with the equation for expected SOW. Given that we have a two-way censored model, we obtain the following equation:

$$\mathrm{E}(SOW) = \Phi(a_i)L_i + (X'\beta + \lambda\sigma_i)(\Phi(b_i) - \Phi(a_i)) + (1 - \Phi(b_i))R_i$$

where

$$a_i = \left(\frac{L_i - X'\beta}{\sigma}\right), \quad b_i = \left(\frac{R_i - X'\beta}{\sigma}\right), \quad \lambda = \left(\frac{\phi(a_i) - \phi(b_i)}{\Phi(b_i) - \Phi(a_i)}\right).$$

In this case $\Phi$ is the normal CDF distribution, $\phi$ is the normal PDF distribution, $L_i$ is 1, $R_i$ is 100, $X$ is the matrix of independent variable values from the SOW equation, $\beta$ is the vector of parameter estimates from the SOW equation, $\lambda$ is the inverse Mills ratio, and $\sigma$ is the standard error of the SOW equation. Once we have predicted SOW for each of the customers we want to compare this to the actual value from the database. We do this by computing the MAD. The equation is as follows:

$$\text{MAD} = \text{Mean}\{\text{Absolute Value } [\mathrm{E}(SOW) - SOW]\}.$$

We find for the acquired customers that MAD = 7.36, or on average 7.36% from the actual SOW. If we were to instead use the mean value of SOW (52.98) across all customers as our prediction for all customers (this would be the benchmark model case), we would find that MAD = 25.84, or on average 25.84% from the actual SOW. As we can see, our model does a significantly better job of predicting the value of SOW than the benchmark case.

### 4.6.2 How do you implement it?

In this example we used a two-sided censored regression to understand the drivers of SOW and predict SOW. Given the limited nature of the dependent variable we used PROC QLIM in SAS to estimate the model with a lower bound of 1 and an upper bound of 100. While we did use SAS to implement this modeling framework, programs such as MATLAB, R, and GAUSS can also be used.

## 4.7 Profitability (CLV)

The ultimate goal of customer retention is to increase the profitability of companies. Reinartz, Thomas, and Kumar (2005) linked customer acquisition, relationship duration, and profitability using a probit two-stage lease squares model. In the simultaneous equation model, a probit model was used to capture the acquisition process and a Tobit model was used to estimate the relationship duration of acquired customers. In the profitability model, these authors included as explanatory variables the firm actions, customer actions, control variables, predicted acquisition probability, and predicted relationship duration. To account for the right-censoring, the authors used a standard right-censored Tobit model to estimate the profitability.

Besides stochastic methods, researchers have developed deterministic methods to capture the optimal retention spending to maximize profitability. Just as Blattberg and Deighton [16] determined the optimal level of acquisition spending, they asked managers similar questions about companies' past retention activities to draw the retention curve. The answers to these questions indicate the retention expenditure per prospect, $\$R$, the retention rate obtained as a result of that expenditure, $r$, and the ceiling rate on the retention curve. The curve is assumed to be exponential and captured by the equation

$$r = ceiling\,rate^{*}[1 - \exp(-k_2{}^{*}\$R)] \tag{4.40}$$

where $k_2$ is a parameter controlling the shape of the exponential curve. To draw the other part of the graph of customer equity against the retention budget, the authors assumed the same margin in each year, $\$m$. The value of the customer in any given year is then given by

$$year\ y\ contribution\,from\ retention = r^{y}\left(\$m - \frac{\$R}{r}\right). \tag{4.41}$$

The authors then sum up the contribution for each year of the customer's projected life, add the contribution of a first-year customer, discount to a present value, and so get the amount of customer equity for that customer. From the upper and lower parts of the graph, the authors can easily determine the optimal level of retention spending which maximizes customer equity. They further expressed customer equity generated from customer acquisition and retention efforts in the following form:

$$customer\ equity = a\ \$m - \$A + a\left(\$m - \frac{\$R}{r}\right)\left[\frac{r'}{1 - r'}\right] \tag{4.42}$$

where $r' = r/(1 + d)$.

### 4.7.1 Empirical example: Profitability (CLV)

The ultimate question for firms that are interested in customer retention is related to how profitable a current customer is likely to be in the future. In this example we will not focus on how a customer's expected future profitability, or CLV, is predicted. Instead, we want to focus on the drivers of CLV. Thus, we provide a prediction of CLV for each of the customers in the sample. We then want to use that prediction to understand which of the variables in our database help to explain the future value of a customer. If the drivers effectively explain the future value of a customer, we should be able to use the results of the estimation in the prediction of CLV for any customer not in the current sample. At the end of this example we should be able to do the following:

1. Determine the drivers of CLV.
2. Predict the expected CLV for each customer.
3. Determine the predictive accuracy of the model.

The information we need for this model includes the following list of variables:

| | |
|---|---|
| **Dependent variables** | |
| *CLV* | *The discounted value of all expected future profits, or customer lifetime value* |
| **Independent variables** | |
| *Purchase_Rate* | *The average value for purchases across all 12 quarters* |
| *Avg_Order_Quantity* | *The average dollar value of the purchases in all 12 quarters* |
| *Avg_Crossbuy* | *The average value for cross-buy across all 12 quarters* |
| *Avg_Ret_Expense* | *Average dollars spent on marketing efforts to try and retain that customer in all 12 quarters* |
| *Avg_Ret_Expense_SQ* | *Square of average dollars spent on marketing efforts to try and retain that customer in all 12 quarters* |
| *Gender* | *1 if the customer is male, 0 if the customer is female* |
| *Married* | *1 if the customer is married, 0 if the customer is not married* |
| *Income* | *1 if income < \$30 000*<br>*2 if \$30 001< income < \$45 000*<br>*3 if \$45 001 < income < \$60 000*<br>*4 if \$60 001 < income < \$75 000*<br>*5 if \$75 001 < income < \$90 000*<br>*6 if income > \$90 001* |
| *First_Purchase* | *The value of the first purchase made by the customer in quarter 1* |
| *Loyalty* | *1 if the customer is a member of the loyalty program, 0 if not* |
| *Share-of-Wallet (SOW)* | *The percentage of purchases the customer makes from the given firm, given the total amount of purchases across all firms in that category* |

In this case we have a dependent variable *CLV* which represents the expected discounted future profits from each customer. The *CLV* variable is continuous and not bound by any threshold. Thus, given the assumption that *CLV* is normally distributed around a specific mean, we merely want to run an OLS regression to uncover the drivers of *CLV*. We get an equation in the following format:

$$CLV_i = X'\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

where *CLV* is the customer lifetime value of a given customer, $X$ is the matrix of independent variables, $\beta$ is the vector of coefficients of the independent variables, and $\varepsilon$ is the random error term which is normally distributed with a mean of 0 and variance of $\sigma^2$. We get the following results when we estimate the model:

| Variable | Estimate | Standard error | *p*-value |
|---|---|---|---|
| *Intercept* | −226.150 | 16.274 | <0.0001 |
| *Purchase_Rate* | −36.919 | 24.411 | 0.1311 |
| *Avg_Order_Quantity* | 7.578 | 0.090 | <0.0001 |
| *Avg_Crossbuy* | 20.758 | 7.953 | 0.0093 |
| *Avg_Ret_Expense* | 4.373 | 1.790 | 0.0149 |
| *Avg_Ret_Expense_SQ* | −0.232 | 0.068 | 0.0006 |
| *Gender* | 20.284 | 4.462 | <0.0001 |
| *Married* | 4.741 | 5.029 | 0.3463 |
| *Income* | 6.319 | 1.647 | 0.0001 |
| *First_Purchase* | 1.894 | 0.056 | <0.0001 |
| *Loyalty* | 26.344 | 4.856 | <0.0001 |

We gain the following insights from these results. We find that all the variables with the exception of *Purchase_Rate* and *Married* are statistically significant at $p < 0.05$. We find that *Avg_Order_Quantity* is positive, suggesting that the higher the average past order values of the customer, the higher the CLV. We find that *Avg_Crossbuy* is positive, suggesting that the more a customer has bought across multiple categories in the past, the higher the customer's CLV. We find that *Ret_Expense* is positive with a diminishing return, as noted by the positive coefficient on *Ret_Expense* and the negative coefficient on *Ret_Expense_SQ*. This means that marketing efforts to retain and build relationships with the customer do cause the customer to have a higher CLV. Then after the threshold is reached, marketing efforts actually decrease the CLV on average. This is likely due to the fact that overly contacting customers can often strain the relationship between the customer and firm, and because when marketing efforts lose effectiveness the costs continue to increase without corresponding profit increases. We find that four of the customer characteristic variables are positive (*Gender*, *Income*, *First_Purchase*, and *Loyalty*) suggesting that customers who are male, with a higher income, higher first purchase value, and members of the loyalty program are likely to have a higher CLV.

Our next step is to predict the value of CLV to see how well our model compares to the actual values. We do this by starting with the equation for expected CLV. Given that we have an OLS regression model, we obtain the following equation:

$$\mathrm{E}(CLV) = X'\beta.$$

In this case $X$ is the matrix of independent variable values from the CLV equation and $\beta$ is the vector of parameter estimates from the CLV equation. Once we have predicted CLV for each of the customers we want to compare this to the actual value from the database. We do this by computing the MAD. The equation is as follows:

$$\mathrm{MAD} = \mathrm{Mean}\{\mathrm{Absolute\ Value}[\mathrm{E}(CLV) - CLV]\}.$$

We find for the acquired customers that MAD = 39.01, or on average \$39.01 from the actual CLV. If we were to instead use the mean value of CLV (\$1126.73) across all customers as our prediction for all customers (this would be the benchmark model case), we would find that MAD = 734.51, or on average \$734.51 from the actual CLV. As we can see, our model does a significantly better job of predicting the value of CLV than the benchmark case.

### 4.7.2 How do you implement it?

In this example we used PROC Reg from SAS to estimate the OLS regression model to explain the variance of CLV. While we did use SAS to implement this modeling framework, programs such as SPSS, MATLAB, R, and GAUSS can be used as well.

## 4.8 Chapter summary

The purpose of this chapter was to explore the current models for customer retention and provide some empirical examples as to how firms can apply this knowledge to their own customer databases. And we have shown that when firms are able to understand each of the aspects of customer retention, they can effectively manage their current customers for profit.

## Customer retention – SAS code

```
/* Import Data - Library: statcrm */

proc import out = statcrm.customer_trans
   datafile = "C:\_Your_Data_Location_\Statistics in CRM\Customer
Retention\Customer Retention.xls" dbms = excel replace;
   range = "Transactions$"; getnames = yes; mixed = no; scantext =
yes; usedate = yes; scantime = yes;
run; quit;

proc import out = statcrm.customer_demo
   datafile = "C:\_Your_Data_Location_\Statistics in CRM\Customer
Retention\Customer Retention.xls" dbms = excel replace;
   range = "Demographics$"; getnames = yes; mixed = no; scantext =
yes; usedate = yes; scantime = yes;
run; quit;

/* Combine Data */

proc sql;
create table statcrm.customer_retention as
select distinct a.*, b.*
from statcrm.customer_trans a left join statcrm.customer_demo b
on a.customer = b.customer order by customer; quit;

/* Repurchase Probability */

data statcrm.retention_repurchase;
set statcrm.customer_retention;
lcustomer = lag(customer);
if crossbuy >1 then cb = 1; else cb = 0;
if customer = lcustomer then do; lpurchase = lag(purchase);
lcrossbuy = lag(crossbuy); lcb = lag(cb); end;
else do; lpurchase = .; lcrossbuy = .; lcb = .; end;
if _n_ = 2 then do; lpurchase = 1; lcrossbuy = 1; lcb = 0; end;
run; quit;

data statcrm.retention_repurchase1;
set statcrm.retention_repurchase;
by customer;
if first.customer then do; quantity_sum = order_quantity; end;
else do; quantity_sum + order_quantity; end;
```

```
avg_order_quantity = quantity_sum / quarter;
run; quit;

proc logistic data = statcrm.retention_repurchase1 descending;
model purchase = lpurchase avg_order_quantity
ret_expense ret_expense_sq gender married income first_purchase
loyalty / rsquare;
output out = statcrm.retention_repurchase_pred p = pred;
run; quit;

/* Purchase Accuracy */

data statcrm.retention_repurchase_pred1;
set statcrm.retention_repurchase_pred;
if pred >= 0.5 then pred_pur = 1; else pred_pur = 0;
act_pur = purchase;
run; quit;

proc freq data = statcrm.retention_repurchase_pred1;
table pred_pur * act_pur; where quarter >1; run; quit;

/* P(Alive) */

proc sql;
create table statcrm.palive_x as
select customer, sum(purchase) as x
from statcrm.customer_retention
group by customer order by customer; quit;

proc sql;
create table statcrm.palive_tx as
select customer, max(quarter) as tx
from statcrm.customer_retention
where purchase = 1 group by customer order by customer; quit;

data statcrm.palive_T;
set statcrm.customer_retention;
T = 12; keep customer T;
run; quit;

proc sql;
create table statcrm.palive_xtx as
select a.*, b.tx
from statcrm.palive_x a left join statcrm.palive_tx b
on a.customer = b.customer; quit;
```

```
proc sql;
create table statcrm.palive_xtxT as
select distinct a.*, b.T
from statcrm.palive_xtx a left join statcrm.palive_T b
on a.customer = b.customer order by customer; quit;

PROC EXPORT DATA= STATCRM.PALIVE_XTXT OUTFILE=
"C:\_Your_Data_Location_\Statistics in CRM\Customer Retention
\palive.xls" DBMS = EXCEL REPLACE; NEWFILE = YES; run; quit;

/* Order Quantity */

proc logistic data = statcrm.retention_repurchase1 descending;
model purchase = lpurchase avg_order_quantity ret_expense
ret_expense_sq gender married income first_purchase loyalty / link
= probit;
output out = statcrm.retention_repurchase_probit xbeta =
xb_probit;
run; quit;

data statcrm.repurchase_imr;
set statcrm.retention_repurchase_probit;
imr_purchase = (pdf('Normal',xb_probit))/(probnorm(xb_probit));
run; quit;

proc reg data = statcrm.repurchase_imr;
model order_quantity = lpurchase avg_order_quantity ret_expense
ret_expense_sq gender married income first_purchase loyalty
imr_purchase;
output out = statcrm.repurchase_imr1 p = xbeta;
where order_quantity >0;
run; quit;

data statcrm.order_quantity1;
set statcrm.repurchase_imr;
xbeta = -68.63 + 50.77*lpurchase + 0.82+avg_order_quantity +
1.36*ret_expense - 0.04*ret_expense_sq + 11.38*gender
   + 11.04*income + 0.23*first_purchase + 6.00*loyalty +
35.17*imr_purchase;
pred_oq = probnorm(xb_probit)*(xbeta);
ad = abs(order_quantity - pred_oq);
ad1 = abs(order_quantity - 129.1);
if quarter >1 then output;
```

```
run; quit;

proc sql; select mean(order_quantity) as mean_oq, mean(ad) as mad,
mean(ad1) as random_mad from statcrm.order_quantity1 where
quarter >1; quit;

/* Crossbuy */

proc logistic data = statcrm.retention_repurchase1 descending;
model cb = lpurchase lcrossbuy avg_order_quantity ret_expense
ret_expense_sq gender married income first_purchase loyalty /
rsquare;
output out = statcrm.retention_cb_pred p = pred;
where purchase = 1;
run; quit;

data statcrm.retention_cb_pred1;
set statcrm.retention_cb_pred;
if pred >= 0.5 then pred_cb = 1; else pred_cb = 0;
act_cb = cb;
run; quit;

proc freq data = statcrm.retention_cb_pred1;
table pred_cb * act_cb; where quarter >1; run; quit;

/* SOW */

proc sql;
create table statcrm.sow as select customer, mean(purchase) as
purchase_rate, mean(order_quantity) as avg_order_quantity,
mean(crossbuy) as avg_crossbuy, mean(ret_expense) as avg_ret_exp,
(mean(ret_expense)*mean(ret_expense)) as avg_ret_exp_sq, gender,
married, income, first_purchase, sow, loyalty
from statcrm.retention_repurchase1 group by customer, gender,
married, income, first_purchase, sow, loyalty order by customer;
quit;

proc qlim data = statcrm.sow;
model sow = purchase_rate avg_order_quantity avg_crossbuy
avg_ret_exp avg_ret_exp_sq gender married income first_purchase
loyalty /
censored(lb = 1 ub = 100);
output out = statcrm.sow_pred predicted;
run; quit;
```

```
data statcrm.sow_pred1;
set statcrm.sow_pred;
ad = abs(sow - p_sow);
ape = ad/sow;
ad1 = abs(sow - 54.1);
ape1 = ad1/sow;
run; quit;

proc sql; select mean(sow) as mean_sow, mean(ad) as mad, mean(ape)
as mape, mean(ad1) as random_mad, mean(ape1) as random_mape from
statcrm.sow_pred1; quit;

/* Profitability */

proc sql;
create table statcrm.clv as select customer, mean(purchase) as
purchase_rate, mean(order_quantity) as avg_order_quantity,
mean(crossbuy) as avg_crossbuy, mean(ret_expense) as avg_ret_exp,
(mean(ret_expense)*mean(ret_expense)) as avg_ret_exp_sq, gender,
married, income, first_purchase, sow, clv, loyalty
from statcrm.retention_repurchase1 group by customer, gender,
married, income, first_purchase, sow, clv, loyalty order by
customer; quit;

proc reg data = statcrm.clv;
model clv = purchase_rate avg_order_quantity avg_crossbuy
avg_ret_exp avg_ret_exp_sq gender married income first_purchase
loyalty sow / rsquare;
output out = statcrm.clv_pred p = pred_clv;
run; quit;

data statcrm.clv_pred1;
set statcrm.clv_pred;
ad = abs(clv - pred_clv);
ape = abs(ad/clv);
ad1 = abs(clv - 1126.73);
ape1 = abs(ad1/clv);
run; quit;

proc sql; select mean(clv) as mean_clv, mean(ad) as mad, mean(ape)
as mape, mean(ad1) as random_mad, mean(ape1) as random_mape from
statcrm.clv_pred1; quit;
```

# Customer retention – SAS output

```
                  The SAS System
                The LOGISTIC Procedure
                  Model Information
Data Set                        STATCRM.RETENTION_REPURCHASE1
Response Variable               Purchase                   Purchase
Number of Response Levels       2
Model                           binary logit
Optimization Technique          Fisher's scoring

          Number of Observations Read          6000
          Number of Observations Used          5500

                       Response Profile
          Ordered                              Total
          Value             Purchase       Frequency

            1                   1               2678
            2                   0               2822

                Probability modeled is Purchase=1.

NOTE: 500 observations were deleted due to missing values for the response or
explanatory variables.

                Model Convergence Status

        Convergence criterion (GCONV=1E-8) satisfied.

                  Model Fit Statistics
                                        Intercept
                         Intercept            and
          Criterion           Only     Covariates

          AIC            7622.848        3192.342
          SC             7629.461        3258.467
          -2 Log L       7620.848        3172.342

R-Square    0.5546    Max-rescaled    R-Square    0.7397

                    The SAS System

                The LOGISTIC Procedure

          Testing Global Null Hypothesis: BETA=0

Test                  Chi-Square        DF          Pr > ChiSq
Likelihood Ratio      4448.5065         9              <.0001
Score                 3490.9130         9              <.0001
Wald                  1722.8722         9              <.0001
```

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| Intercept | 1 | -4.9881 | 0.2244 | 493.9003 | <.0001 |
| lpurchase | 1 | 2.1446 | 0.0996 | 463.6408 | <.0001 |
| avg_order_quantity | 1 | 0.0184 | 0.000660 | 775.1934 | <.0001 |
| Ret_Expense | 1 | 0.1071 | 0.00975 | 120.6179 | <.0001 |
| Ret_Expense_SQ | 1 | -0.00237 | 0.000213 | 123.9991 | <.0001 |
| Gender | 1 | -0.2023 | 0.0931 | 4.7182 | 0.0298 |
| Married | 1 | -0.0589 | 0.1084 | 0.2949 | 0.5871 |
| Income | 1 | 0.1317 | 0.0318 | 17.1545 | <.0001 |
| First_Purchase | 1 | 0.00122 | 0.00101 | 1.4693 | 0.2255 |
| Loyalty | 1 | 0.2679 | 0.1003 | 7.1384 | 0.0075 |

Odds Ratio Estimates

| Effect | Point Estimate | 95% Wald Confidence | Limits |
|---|---|---|---|
| lpurchase | 8.539 | 7.024 | 10.379 |
| avg_order_quantity | 1.019 | 1.017 | 1.020 |
| Ret_Expense | 1.113 | 1.092 | 1.134 |
| Ret_Expense_SQ | 0.998 | 0.997 | 0.998 |
| Gender | 0.817 | 0.681 | 0.980 |
| Married | 0.943 | 0.762 | 1.166 |
| Income | 1.141 | 1.072 | 1.214 |
| First_Purchase | 1.001 | 0.999 | 1.003 |
| Loyalty | 1.307 | 1.074 | 1.591 |

Association of Predicted Probabilities and Observed Responses

| | | | |
|---|---|---|---|
| Percent Concordant | 94.8 | Somers' D | 0.897 |
| Percent Discordant | 5.1 | Gamma | 0.898 |
| Percent Tied | 0.1 | Tau-a | 0.448 |
| Pairs | 7557316 | c | 0.949 |

```
          The SAS System
       The FREQ Procedure
    Table of pred_pur by act_pur
pred_pur       act_pur

Frequency,
Percent  ,
Row Pct  ,
Col Pct  ,       0,      1,      Total
ffffffff^ffffffff^fffffff^
       0 ,    2494 ,    330 ,      2824
         ,   45.35 ,   6.00 ,     51.35
         ,   88.31 ,  11.69 ,
         ,   88.38 ,  12.32 ,
```

```
ƒƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^
       1 ,     328 ,    2348 ,     2676
         ,    5.96 ,   42.69 ,    48.65
         ,   12.26 ,   87.74 ,
         ,   11.62 ,   87.68 ,
ƒƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^
Total         2822     2678       5500
             51.31    48.69     100.00

             The SAS System
           The LOGISTIC Procedure
             Model Information

Data Set                      STATCRM.RETENTION_REPURCHASE1
Response Variable             Purchase                        Purchase
Number of Response Levels     2
Model                         binary probit
Optimization Technique        Fisher's scoring

         Number of Observations Read        6000
         Number of Observations Used        5500

                   Response Profile
            Ordered                         Total
              Value       Purchase      Frequency

                  1              1           2678
                  2              0           2822

                   Probability modeled is Purchase=1.
NOTE: 500 observations were deleted due to missing values for the response or
explanatory variables.

                   Model Convergence Status
           Convergence criterion (GCONV=1E-8) satisfied.

                   Model Fit Statistics

                                          Intercept
                         Intercept              and
      Criterion               Only       Covariates

      AIC                 7622.848         3188.215
      SC                  7629.461         3254.340
      -2 Log L            7620.848         3168.215

          Testing Global Null Hypothesis: BETA=0

  Test                   Chi-Square        DF          Pr > ChiSq
  Likelihood Ratio        4452.6334         9          <.0001
  Score                   3490.9130         9          <.0001
  Wald                    2325.0095         9          <.0001

             The SAS System
```

```
                     The LOGISTIC Procedure
               Analysis of Maximum Likelihood Estimates

                                        Standard          Wald
Parameter            DF    Estimate        Error    Chi-Square   Pr > ChiSq

Intercept             1     -2.7828       0.1176      559.7470       <.0001
lpurchase             1      1.2314       0.0564      476.6149       <.0001
avg_order_quantity    1      0.0102     0.000350      842.7393       <.0001
Ret_Expense           1      0.0590      0.00522      127.8442       <.0001
Ret_Expense_SQ        1    -0.00131     0.000114      130.9564       <.0001
Gender                1     -0.1143       0.0502        5.1864       0.0228
Married               1     -0.0247       0.0581        0.1802       0.6712
Income                1      0.0724       0.0171       17.8861       <.0001
First_Purchase        1    0.000731     0.000543        1.8079       0.1788
Loyalty               1      0.1520       0.0542        7.8567       0.0051

        Association of Predicted Probabilities and Observed Responses

             Percent Concordant        94.8    Somers' D    0.897
             Percent Discordant         5.1    Gamma        0.898
             Percent Tied               0.1    Tau-a        0.448
             Pairs                  7557316    c            0.949

              The SAS System

                 The REG Procedure
                   Model:   MODEL1
          Dependent Variable: Order_Quantity Order_Quantity

         Number of Observations Read                        3178
         Number of Observations Used                        2678
         Number of Observations with Missing Values          500

                          Analysis of Variance

                                  Sum of          Mean
   Source               DF       Squares        Square    F Value    Pr > F

   Model                10      11201592       1120159     239.37    <.0001
   Error              2667      12480389    4679.56090
   Corrected Total    2677      23681981

             Root   MSE            68.40732    R-Square     0.4730
             Dependent Mean       265.14456    Adj R-Sq     0.4710
             Coeff Var             25.80001

                         Parameter Estimates

                                      Parameter     Standard
Variable         Label          DF     Estimate        Error   t Value    Pr > |t|

Intercept       Intercept        1   -68.63013     16.94526     -4.05      <.0001
lpurchase                        1    50.76880      6.79251      7.47      <.0001
avg_order_
```

```
quantity                         1    0.81572    0.03666    22.25    <.0001
Ret_Expense     Ret_Expense      1    1.36074    0.29566     4.60    <.0001
Ret_Expense_SQ  Ret_Expense_SQ   1   -0.03678    0.00664    -5.54    <.0001
Gender          Gender           1   11.37668    2.70669     4.20    <.0001
Married         Married          1   -2.46884    3.04825    -0.81    0.4181
Income          Income           1   11.03687    0.96485    11.44    <.0001
First_Purchase  First_Purchase   1    0.23232    0.02865     8.11    <.0001
Loyalty         Loyalty          1    6.00460    3.00479     2.00    0.0458
imr_purchase                     1   35.16844    8.27214     4.25    <.0001

                The SAS System

         mean_oq          mad       random_mad
         ffffffffffffffffffffffffffffff
         129.1013        54.76559    133.7096

                The SAS System

                        The LOGISTIC Procedure

                          Model Information

                Data Set                    STATCRM.RETENTION_REPURCHASE1
                Response Variable           cb
                Number of Response Levels   2
                Model                       binary logit
                Optimization Technique      Fisher's scoring

                Number of Observations Read      3178
                Number of Observations Used      2678

                           Response Profile

                      Ordered              Total
                      Value        cb      Frequency
                          1         1          1361
                          2         0          1317

                      Probability modeled is cb=1.

NOTE: 500 observations were deleted due to missing values for the response or
explanatory variables.
                        Model Convergence Status

              Convergence criterion (GCONV=1E-8) satisfied.

                         Model Fit Statistics
                                              Intercept
                            Intercept              and
              Criterion          Only       Covariates

              AIC            3713.773         2656.618
              SC             3719.666         2721.440
              -2 Log L       3711.773         2634.618
```

```
R-Square    0.3312    Max-rescaled  R-Square    0.4416

  The SAS System

            The LOGISTIC Procedure

     Testing Global Null Hypothesis: BETA=0

Test                    Chi-Square      DF      Pr > ChiSq

Likelihood Ratio        1077.1549       10          <.0001
Score                    896.2550       10          <.0001
Wald                     623.1437       10          <.0001

          Analysis of Maximum Likelihood Estimates

                                         Standard        Wald
Parameter            DF    Estimate         Error  Chi-Square     Pr > ChiSq

Intercept             1     -7.6414        0.3568    458.5606         <.0001
lpurchase             1      0.1373        0.1687      0.6624         0.4157
lcrossbuy             1      0.1616        0.0440     13.4830         0.0002
avg_order_quantity    1      0.0117      0.000831    199.1893         <.0001
Ret_Expense           1      0.0380       0.00969     15.3602         <.0001
Ret_Expense_SQ        1    -0.00077      0.000219     12.3786         0.0004
Gender                1      0.2603        0.0969      7.2224         0.0072
Married               1      0.1806        0.1107      2.6647         0.1026
Income                1      0.2330        0.0353     43.6735         <.0001
First_Purchase        1      0.0243       0.00123    389.0530         <.0001
Loyalty               1      0.2583        0.1077      5.7535         0.0165

                      Odds Ratio Estimates
                                    Point          95% Wald
        Effect                   Estimate      Confidence Limits

        lpurchase                   1.147         0.824       1.597
        lcrossbuy                   1.175         1.078       1.281
        avg_order_quantity          1.012         1.010       1.013
        Ret_Expense                 1.039         1.019       1.059
        Ret_Expense_SQ              0.999         0.999       1.000
        Gender                      1.297         1.073       1.569
        Married                     1.198         0.964       1.488
        Income                      1.262         1.178       1.353
        First_Purchase              1.025         1.022       1.027
        Loyalty                     1.295         1.048       1.599

        Association of Predicted Probabilities and Observed Responses

          Percent Concordant       84.1     Somers' D    0.684
          Percent Discordant       15.8     Gamma        0.685
          Percent Tied              0.1     Tau-a        0.342
          Pairs                 1792437     c            0.842

             The SAS System

          The FREQ Procedure

        Table of pred_cb by act_cb
```

```
pred_cb      act_cb

Frequency,
Percent ,
Row Pct ,
Col Pct ,      0,      1,      Total
ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^
0 ,      980 ,     308 ,    1288
  ,    36.59 ,   11.50 ,   48.10
  ,    76.09 ,   23.91 ,
  ,    74.41 ,   22.63 ,
ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^
1 ,      337 ,    1053 ,    1390
  ,    12.58 ,   39.32 ,   51.90
  ,    24.24 ,   75.76 ,
  ,    25.59 ,   77.37 ,

ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^ƒƒƒƒƒƒƒ^
Total    1317    1361    2678
        49.18   50.82  100.00

          The SAS System
```

The QLIM Procedure

Summary Statistics of Continuous Responses

| Variable | Mean | Standard Error | Type | Lower Bound | Upper Bound | N Obs Lower Bound | N Obs Upper Bound |
|---|---|---|---|---|---|---|---|
| sow | 52.984 | 29.680391 | Censored | 1 | 100 | 6 | 58 |

Model Fit Summary

| | |
|---|---|
| Number of Endogenous Variables | 1 |
| Endogenous Variable | sow |
| Number of Observations | 500 |
| Log Likelihood | -1695 |
| Maximum Absolute Gradient | 0.0002575 |
| Number of Iterations | 99 |
| Optimization Method | Newton-Raphson |
| AIC | 3413 |
| Schwarz Criterion | 3464 |

Algorithm converged.

Parameter Estimates

| Parameter | DF | Estimate | Standard Error | t Value | Approx Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | -1.737978 | 3.710723 | -0.47 | 0.6395 |
| purchase_rate | 1 | 7.846176 | 6.279358 | 1.25 | 0.2115 |
| avg_order_quantity | 1 | 0.195716 | 0.024778 | 7.90 | <.0001 |
| avg_crossbuy | 1 | 6.435820 | 1.861412 | 3.46 | 0.0005 |
| avg_ret_exp | 1 | 0.856724 | 0.404742 | 2.12 | 0.0343 |
| avg_ret_exp_sq | 1 | -0.049692 | 0.015098 | -3.29 | 0.0010 |
| Gender | 1 | 0.588789 | 1.024313 | 0.57 | 0.5654 |
| Married | 1 | 1.079447 | 1.146020 | 0.94 | 0.3462 |

```
Income              1     1.799558    0.360212      5.00       <.0001
First_Purchase      1     0.052324    0.012645      4.14       <.0001
Loyalty             1     5.126359    1.081014      4.74       <.0001
_Sigma              1    10.732490    0.367225     29.23       <.0001

                The SAS System

      mean_sow        mad          mape      random_mad    random_mape
      ffffffffffffffffffffffffffffffffffffffffffffffffffffffff
       52.984      7.364584    0.442466      25.8364       1.555554

                      The SAS System

                    The REG Procedure
                     Model: MODEL1
               Dependent Variable: CLV CLV

          Number of Observations Read        500
          Number of Observations Used        500

                      Analysis of Variance
                            Sum of          Mean
    Source            DF       Squares         Square      F Value    Pr > F

    Model             11     359881255       32716478     13944.3   <.0001
    Error            488       1144962     2346.23384
    Corrected Total  499     361026217

         Root MSE            48.43794     R-Square      0.9968
         Dependent Mean    1126.73178     Adj R-Sq      0.9968
         Coeff Var            4.29898

                        Parameter Estimates
                                  Parameter     Standard
Variable        Label        DF    Estimate         Error   t Value    Pr > |t|

Intercept       Intercept     1  -226.15047      16.27385    -13.90     <.0001
purchase_rate                 1   -36.91945      24.41084     -1.51     0.1311
avg_order                     1     7.57766       0.09018     84.03     <.0001
_quantity
avg_crossbuy                  1    20.75796       7.95307      2.61     0.0093
avg_ret_exp                   1     4.37313       1.78959      2.44     0.0149
avg_ret_exp_sq                1    -0.23209       0.06762     -3.43     0.0006
Gender          Gender        1    20.28437       4.46160      4.55     <.0001
Married         Married       1     4.74112       5.02879      0.94     0.3463
Income          Income        1     6.31902       1.64712      3.84     0.0001
First_Purchase  First_        1     1.89440       0.05644     33.56     <.0001
                Purchase 1
Loyalty         Loyalty       1    26.34402       4.85643      5.42     <.0001
SOW             SOW           1     0.50199       0.21363      2.35     0.0192

              The SAS System

     mean_clv     mad       mape       random_mad      random_mape
     fffffffffffffffffffffffffffffffffffffffffffffffffffff
     1126.732    39.01268    0.899111    734.5071     22.9432
```

## References

1. Meer, G.V. (2006) Customer development and retention on a Web-banking site. *Journal of Interactive Marketing*, **20**(1), 58–64.
2. Seo, D., Ranganathan, C., and Babad, Y. (2008) Two-level customer retention in the US mobile telecommunications service market. *Telecommunications Policy*, **32**, 182–196.
3. Lewis, M. (2006a) Customer acquisition promotions and customer asset value. *Journal of Marketing Research*, **43**(2), 195–203.
4. Verhoef, P.C. (2003) Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing*, **67**(4), 30–45.
5. Verhoef, P.C. and Donkers, B. (2005) The effect of acquisition channels on customer loyalty and cross-buying. *Journal of Interactive Marketing*, **19**(2), 31–43.
6. Bhattacharya, C.B. (1998) When customers are members: customer retention in paid membership contexts. *Journal of the Academy of Marketing Science*, **26**(1), 31–44.
7. Schweidel, D.A., Fader, P.S., and Bradlow, E.T. (2008) A bivariate timing model of customer acquisition and retention. *Marketing Science*, **27**(5), 829–843.
8. Reinartz, W.J. and Kumar, V. (2000) On the profitability of long-life customers in a non-contractual setting: an empirical investigation and implications for marketing. *Journal of Marketing*, **64**(4), 17–35.
9. Meyer-Waarden, L. (2007) The effects of loyalty programs on customer lifetime duration and share of wallet. *Journal of Retailing*, **83**(2), 223–236.
10. Bolton, R.N. (1998) A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction. *Marketing Science*, **17**(1), 45–65.
11. Fader, P.S. and Hardie, B.G.S. (2007) How to project customer retention. *Journal of Interactive Marketing*, **21**(1), 76–90.
12. Thomas, J.S. (2001) A methodology for linking customer acquisition to customer retention. *Journal of Marketing Research*, **38**(2), 262–268.
13. Bolton, R.N. and Lemon, K.N. (1999) A dynamic model of customers' usage of services: usage as an antecedent and consequence of satisfaction. *Journal of Marketing Research*, **36**(2), 171–186.
14. Anderson, E.T. and Simester, D.I. (2004) Long-run effects of promotion depth on new versus established customers: three field studies. *Marketing Science*, **23**(1), 4–20.
15. Lewis, M. (2006b) The effect of shipping fees on customer acquisition, customer retention, and purchase quantities. *Journal of Retailing*, **82**(1), 13–23.
16. Blattberg, R.C. and Deighton, J. (1996) Manage marketing by the customer equity test. *Harvard Business Review*, **74**(4), 136–144.
17. Berger, P.D. and Bechwati, N.N. (2001) The allocation of promotion budget to maximize customer equity. *Omega*, **29**(1), 49–61.
18. Lemon, K.N., White, T.B., & Winer, R.S. (2002) Dynamic customer relationship management: incorporating future considerations into the service retention decision. *Journal of Marketing*, **66**(1), 1–14.
19. Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., and Dedene, G. (2002) Bayesian neural network learning for repeat purchase modelling in direct marketing. *European Journal of Operational Research*, **138**(1), 191–211.

20. Bolton, R.N., Lemon, K.N., and Bramlett, M.D. (2006) The effect of service experiences over time on a supplier's retention of business customers. *Management Science*, **52**(12), 1811–1823.
21. Kivetz, R., Urminsky, O., and Zheng, Y. (2006) The goal-gradient hypothesis resurrected: purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research*, **43**(1), 39–58.
22. Bolton, R., Kannan, P., and Bramlett, M. (2000) Implications of loyalty program membership and service experiences for customer retention and value. *Journal of the Academy of Marketing Science*, **28**(1), 95–108.
23. Borle, S., Singh, S.S., and Jain, D.C. (2008) Customer lifetime value measurement. *Management Science*, **54**(1), 100–112.
24. Zhang, J.Q., Dixit, A., and Friedmann, R. (2010) Customer loyalty and lifetime value: an empirical investigation of consumer packaged goods. *Journal of Marketing Theory & Practice*, **18**(2), 127–139.
25. Reinartz, W., Thomas, J.S. and Kumar, V. (2005) Balancing acquisition and retention resources to maximize customer profitability. *Journal of Marketing*, **69**(1), 63–79.
26. Leenheer, J., van Heerde, H.J., Bijmolt, T.H.A., and Smidts, A. (2007) Do loyalty programs really enhance behavioral loyalty? An empirical analysis accounting for self-selecting members. *International Journal of Research in Marketing*, **24**(1), 31–47.
27. Greene, W.H. (2000) *Econometric Analysis*, 4th ed., Upper Saddle River, NJ: Prentice Hall.

# 5

# Balancing acquisition and retention

## 5.1 Introduction

The previous two chapters have discussed the integrated frameworks and key modeling issues with regard to customer acquisition and customer retention. In both cases the chapters looked at these two phenomena in the customer life cycle independently. Each chapter did this because it is critical to understand these phenomena in great depth to understand the drivers of the different aspects of customer acquisition and customer retention. However, it is myopic for a marketing manager to ignore the fact that these two processes are inherently linked together. All customers of a firm must first be acquired before effort can be placed on retention. And, eventually, all customers will churn from the firm at some point in time in the future. Ideally a firm wants to manage this process such that new customers are replacing those current customers who churn at an acquisition rate that is greater than or equal to the rate of churn. This way, the size of the customer base of the firm continues to grow (or at least does not shrink). In addition, once a customer is acquired the firm wants to build the relationship with the customer to extend the customer's life and profitability with the firm. Thus, a firm needs to 'balance' the marketing effort between acquisition and retention in order to maximize the profitability of the customer base over time.

There are often several key questions that need to be answered with regard to customer acquisition and retention (see Figure 5.1 for a graphical representation of the research issues in balancing customer acquisition and retention). Some of these questions are similar to those that were posed in the previous two chapters. However, we must still address these questions to be able to properly balance the

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

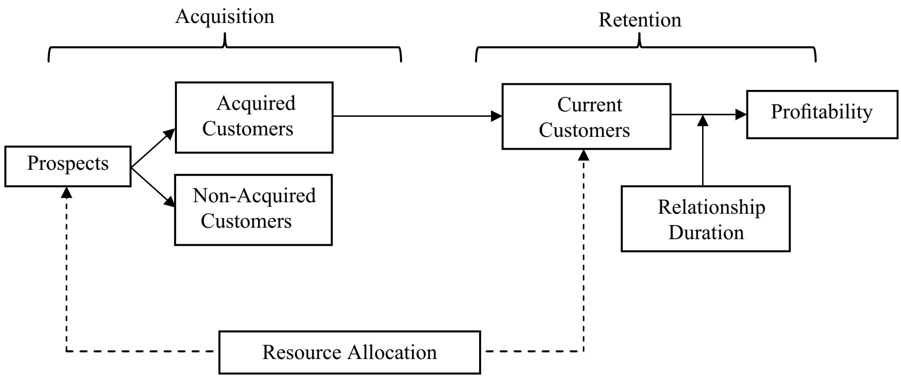



*Figure 5.1 Issues addressed in balancing acquisition and retention modeling.*

marketing resources between these two parts of the customer life cycle. These questions include:

- What are the drivers of customer acquisition?
- How much do I need to spend in order to acquire a customer?
- Once I acquire a customer, how long can I expect that customer to be a customer?
- How much do I need to invest in the relationship in order to keep the customer longer into the future?
- Given my resource constraints, how much should I spend on acquisition efforts versus retention efforts to maximize long-term profitability?

There have been several research studies that have been conducted with the goal of trying to answer these key questions. Table 5.1 contains a representative set of studies that have considered these issues and accounted for many of the problems that might occur in the model-building process. To provide a comprehensive understanding of how to balance customer acquisition and retention, we will review the issues in the studies along with the related modeling techniques. We will also provide empirical examples at the end of chapter to demonstrate how to apply this knowledge to a representative sample of customers from a B2B firm.

Similar to customer acquisition and customer retention models, the first question that needs to be answered when considering model selection is whether the firm's customers purchase in a contractual versus non-contractual manner. In most instances this will determine the type of statistical model that needs to be used in order to gain any insights from the data.

### 5.1.1 Data for empirical examples

In this chapter we will be providing a description of the key modeling frameworks that attempt to answer each key research question raised at the beginning of the

Table 5.1 Review of customer acquisition and retention models.

| Research interest | Specification | Estimation | Representative studies |
|---|---|---|---|
| Balancing acquisition and retention | Probit/Tobit | MLE | Thomas [1]<br>Berger and Nasr-Bechwati [2]<br>Reinartz, Thomas, and Kumar [3] |
| | Optimal resource allocation | Decision calculus | Blattberg and Deighton [4] |
| | | Simulation | Berger and Nasr-Bechwati<br>Reinartz, Thomas, and Kumar [3] |

chapter. We will also be providing one empirical example at the end of the chapter which will show how sample data can be used to answer these key research questions. For all the empirical examples in this chapter we provide a dataset titled 'Balancing Acquisition and Retention.' In this dataset you will find one table of data which include a representative sample of 500 customers from a typical B2B firm, where all the customers are from the same cohort. In this case, the cohort consists of a random sample of 500 customers who all made their first purchase with the firm at the same time. We provide the transaction and firmographic information for each customer. Thus, the data table here consists of 500 customers × 14 variables (15 total columns).

The first data table (labeled *Transactions*) includes the following variables, which will be used in some combination for each of the subsequent analyses:

| **Variable** | |
|---|---|
| *Customer* | *Customer number (from 1 to 500)* |
| *Acquisition* | *1 if the prospect was acquired, 0 otherwise* |
| *Duration* | *The number of days the customer was a customer of the firm, 0 if Acquisition = 0* |
| *Profit* | *The CLV of a given customer, –(Acq_Exp) if the customer is not acquired* |
| *Acq_Exp* | *The total dollars spent on trying to acquire this prospect* |
| *Ret_Exp* | *The total dollars spent on trying to retain this customer* |
| *Acq_Exp_SQ* | *The square of the total dollars spent on trying to acquire this prospect* |
| *Ret_Exp_SQ* | *The square of the total dollars spent on trying to retain this customer* |
| *Freq* | *The number of purchases the customer made during that customer's lifetime with the firm, 0 if Acquisition = 0* |
| *Freq_SQ* | *The square of the number of purchases the customer made during that customer's lifetime with the firm* |
| *Crossbuy* | *The number of product categories the customer purchased from during that customer's lifetime with the firm, 0 if Acquisition = 0* |
| *Share-of-Wallet (SOW)* | *The percentage of purchases the customer makes from the given firm, given the total amount of purchases across all firms in that category* |
| *Industry* | *1 if the prospect is in the B2B industry, 0 otherwise* |
| *Revenue* | *Annual sales revenue of the prospect's firm (in millions of dollars)* |
| *Employees* | *Number of employees in the prospect's firm* |

This data table will be used for the example presented at the end of this chapter. This example will cover models of customer acquisition, customer retention, and optimal resource allocation.

## 5.2 Acquisition and retention

Customer acquisition and customer retention are two processes that are essentially correlated. Researchers have modeled these two issues both separately and together. Lewis [5] investigated whether shipping fees differentially influence customer acquisition and retention. In a system of simultaneous equations, the author examined the effects shipping fees and other marketing variables on the number of new customers acquired the average order size for new customers, and the number of daily orders and the average order size for established customer. The models were estimated in a system because the dependent variables were possibly correlated. Furthermore, to account for the possible correlation between various equations and the possibility of endogenously determined explanatory variables, the author estimated the equations using three-stage least squares. Lagged measures of the endogenous variables, such as prices and average order size, were used as instruments in the estimation.

In another study, Lewis [6] investigated the influence of customer acquisition promotion depth on customer retention, including repeat purchasing and the time being a customer. In an online retailing setting, the author adopted a logistic regression to model whether the customer makes subsequent purchases within the next three quarters using acquisition discount as an explanatory variable. In a newspaper subscription setting, the author adopted accelerated failure time models to model the time as a subscriber using acquisition discount and its quadratic form as an explanatory variable. The author further examined the effects of acquisition discount on customer asset value.

Extending the proposed model by Blattberg and Deighton [4], Berger and Nasr-Bechwati [2] optimized the allocation of the promotion budget between acquisition spending and retention spending. The latter authors considered several different market situations, while here we only report the simplest situation. When companies are considering one market segment and using one promotion method, such as direct mailing, managers decide the allocation of existing budget between acquisition and retention to maximize customer equity. Following Blattberg and Deighton [4],

$$\textit{customer equity} = a\$m - \$A + a\left(\$m - \frac{\$R}{r}\right)\left[\frac{r'}{1-r'}\right] \quad \text{with } r' = r/(1+d),$$

while

1. $\$A + (a \cdot \$R) = \$B$
2. $A \geq 0$
3. $R \geq 0$

where $B$ is the average budget per prospect. These authors multiplied $R$ by the acquisition rate because a firm can only retain customers that have been acquired. The problem is to maximize the objective function which can also be expressed as a function of $A$ because

$$\$R = (\$B - \$A)/a. \tag{5.1}$$

Thus, managers just need to decide the acquisition spending to obtain the maximized customer equity by nonlinear programming.

Researchers have also considered methods to link customer acquisition and retention together instead of treating them as independent variables. Thomas [1] proposed a method known as a Tobit model with selection to account for the impact that the customer acquisition process has on the retention process. The standard right-censored Tobit model was specified as

$$\begin{aligned} y_i^* &= \beta_s' x_i + \varepsilon_{is} \\ y_i &= c_i \quad \text{if } y_i^* \geq c_i \\ y_i &= y_i^* \quad \text{if } y_i^* < c_i \end{aligned} \tag{5.2}$$

where $y_i^*$ is the index variable, $y_i$ the length of customer $i$'s lifetime, $c_i$ the censoring point for customer $i$ (the customer's maximum observable lifetime), $x_i$ the vector of covariates affecting the length of the customer's lifetime, $\varepsilon_{is} \sim N(0, \sigma_{\varepsilon s}^2)$, and $s$ is the segment $(1, 2, \ldots, S)$. The corresponding conditional likelihood for any individual in segment $s$ is as follows:

$$L_{i|s} = \left\{ \frac{1}{\sigma_{\varepsilon s}} \phi \left[ \frac{(y_i^* - \beta_s' x_i)}{\sigma_{\varepsilon s}} \right] \right\}^{1-\delta_i} \left\{ 1 - \Phi \left[ \frac{(c_i - \beta_s' x_i)}{\sigma_{\varepsilon s}} \right] \right\}^{\delta_i} \tag{5.3}$$

where $L_{i|s}$ is the likelihood function for individual $i$ given that he/she is in segment $s$, $\delta_i$ the right-censoring indicator (1 = observation $i$ is right-censored, 0 = not right-censored), $\Phi(\cdot)$ the standard normal cumulative distribution function, $\phi(\cdot)$ the standard normal density function, $c_i$ the value of the censoring variable for observation $i$, $\sigma_{\varepsilon s}$ the standard deviation of the Tobit error term, and $s = 1, 2, \ldots, S$.

The author added truncation to the specification and obtained what is known as a Tobit model with selection as

$$\begin{aligned} y_i^* &= \beta_s' x_i + \varepsilon_{is} \ \text{(main regression/retention equation)} \\ y_i &= c_i \quad \text{if } y_i^* \geq c_i \text{ and } z_i = 1 \\ y_i &= y_i^* \quad \text{if } y_i^* < c_i \text{ and } z_i = 1 \end{aligned} \tag{5.4}$$

and

$$\begin{aligned} z_i^* &= \alpha_s' v_i + \mu_{is} \ \text{(selection/acquisition equation)} \\ z_i &= 1 \quad \text{if } z_i^* > 0 \\ z_i &= 0 \quad \text{if } z_i^* \leq 0 \end{aligned} \tag{5.5}$$

where $y_i$ is the length of customer $i$'s lifetime, $y_i^*$ the index variable, $z_i^*$ customer $i$'s unobserved utility, $z_i$ the acquisition of customer $i$ (1 = acquired, 0 = not acquired), $x_i$ a vector of covariates affecting the retention of customer $i$, $v_i$ a vector of covariates affecting the acquisition of customer $i$, and $s$ the segment $(1, 2, \ldots, S)$. The proposed model successfully links customer acquisition to retention because the length of the customer's lifetime is observed conditional on the customer being acquired, and because the error term in the acquisition equation is possibly correlated with that in the retention equation. The author asserted that the proposed model is similar to the two-step procedure which combines a binary probit model with a standard Tobit model. But the two-step procedure is not applicable here since the information of non-acquired customers is not available. The author asserted that the proposed model could be consistently estimated with maximum likelihood estimation [7]. The estimation details are recorded in Appendix A.

Furthermore, Reinartz, Thomas, and Kumar [3] modeled customer acquisition, retention, and profitability all together in a system of equations, a probit two-stage least squares model. The probit model estimates the acquisition process and two standard right-censored Tobit models estimate the customer relationship duration and customer profitability. The mathematical specifications of the models are expressed as

$$y_{L_i}\begin{cases} = \beta_{Ls}' x_{Li} + \gamma_s' y_{Di} + \varepsilon_{Lis} & \text{if } z_i = 1 \\ = 0 & \text{otherwise} \end{cases} \quad \text{(cumulative profitability equation)} \tag{5.6}$$

$$y_{Di}\begin{cases} = \beta_{Ds}' x_{Di} + \varepsilon_{Dis} & \text{if } z_i = 1 \\ = 0 & \text{otherwise} \end{cases} \quad \text{(duration equation)} \tag{5.7}$$

$$\begin{aligned} z_i^* &= \alpha_s' v_i + \mu_{is} \quad \text{(acquisition equation)} \\ z_i &= \begin{cases} 1 & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \le 0 \end{cases} \end{aligned} \tag{5.8}$$

where $z_i^*$ is a latent variable indicating customer $i$'s utility to engage in a relationship with a firm, $z_i$ an indicator variable showing whether customer $i$ is acquired ($z_i = 1$) or not ($z_i = 0$), $v_i$ a vector of covariates affecting the acquisition of customer $i$, $y_{Di}$ the duration of customer $i$'s relationship with the firm, $x_{Di}$ a vector of covariates affecting the duration of customer $i$'s relationship with the firm, $y_{L_i}$ the cumulative profitability of customer $i$, $x_{Li}$ a vector of covariates affecting customer $i$'s lifetime value, $\alpha_s$, $\beta_{Ls}$, $\beta_{Ds}$ segment-specific parameters, and $\mu_{is}$, $\varepsilon_{Dis}$, and $\varepsilon_{Lis}$ error terms.

The three models are linked because the error terms in three equations are correlated. And these authors assume that the error terms ($\mu_{is}$, $\varepsilon_{Dis}$, and $\varepsilon_{Lis}$) are multivariate normal with a mean vector zero and a covariance matrix specified as

$$\Sigma = \begin{pmatrix} \Sigma_{LL} & \Sigma_{LD} & \Sigma_{L\mu} \\ & \Sigma_{DD} & \Sigma_{D\mu} \\ & & 1 \end{pmatrix}. \tag{5.9}$$

To estimate the system of equations, three-step procedures were adopted. In the first step, a probit model was applied to all prospects, acquired and non-acquired. The estimated probit model gave a selectivity variable, lambda ($\lambda_{is}$), for the acquired customers. The variable lambda ($\lambda_{is}$) was later included as an independent variable in the duration and profitability equations to correct for selection bias due to non-acquired customers. In the second step, a standard right-censored Tobit model was used to estimate duration with the estimated lambda ($\lambda_{is}$) and covariates affecting duration. In the third step, a standard right-censored Tobit model was used to estimate profitability with the estimated lambda ($\lambda_{is}$), the estimated duration, and covariates affecting the long-term profitability of customers.

Schweidel, Fader, and Bradlow [8] argued that previous research on customer acquisition and retention considered acquisition only as a binary variable but did not consider the time when customers were acquired. These authors considered that the two constructs, the time that elapses before a prospective customer acquires a particular service and the subsequent duration for which a customer retains service before dropping it, may be related. The authors adopted the Sarmanov family of multivariate distributions with the inclusion of survival functions to address the correlation between the two constructs and the censoring issues in acquisition and retention. According to these authors, the Sarmanov family works by defining

$$f(x,y) = f_x(x) \times f_y(y) \times \left[1 + \omega\varphi_x(x)\varphi_y(y)\right] \tag{5.10}$$

where $f_x(x)$ and $f_y(y)$ are univariate probability density functions, and $\varphi_x(x)$ and $\varphi_y(y)$ are bounded mixing functions such that

$$\int_{-\infty}^{\infty} f_z(z)\varphi_z(z)\,dz = 0$$

for $z = x, y$. In order for $f(x,y)$ to be a bivariate density function, $\omega$, $\varphi_x(x)$, and $\varphi_y(y)$ must satisfy the condition $1 + \omega\varphi_x(x)\varphi_y(y) \geq 0$ for all values of $x$ and $y$. Lee [9] showed that the mixing distribution for any univariate function $f(x)$ is given by

$$\varphi(x) = f(x) - \int_{-\infty}^{\infty} f^2(t)\,dt. \tag{5.11}$$

Then replacing $f(x)$ in Equation 5.11 with the probability mass functions given in Equations 11 and 24 yields the mixing functions for acquisition and retention processes, denoted $\varphi_A(t_A)$ and $\varphi_R(t_R)$, respectively

$$\begin{aligned}\varphi_A(t_A|\Theta) = {} & S_A(t_A - 1|\Theta) - S_A(t_A) \\ & + 2\sum_{t=1}^{\infty}[S_A(i|\Theta)(S_A(i-1|\Theta) - S_A(i|\Theta))] - 1\end{aligned} \tag{5.12}$$

and

$$\begin{aligned}\varphi_R(t_R\,|\,\beta,\Phi,X(t)) &= S_R(t_R-1\,|\,\beta,\Phi,X(t)) - S_R(t_R\,|\,\beta,\Phi,X(t))\\ &+2\sum_{i=1}^{\infty} S_R\{(i\,|\,\beta,\Phi,X(t))[S_R(i-1\,|\,\beta,\Phi,X(t)) - S_R(i\,|\,\beta,\Phi,X(t))]\} - 1.\end{aligned}\tag{5.13}$$

The joint distribution of acquisition time and duration of service at the customer level is then given by

$$\begin{aligned}j(t_A,t_R|\omega,\Theta,\beta,\Phi,X(t)) &= f(t_A|\Theta)g(t_R|\beta,\Phi,X(t))\\ &\times[1+\omega\varphi_A(t_A|\Theta)\varphi_R(t_R|\beta,\Phi,X(t))].\end{aligned}\tag{5.14}$$

In summary, the customer acquisition modeling is actually a probability prediction and the customer retention modeling essentially concerns the duration of customer lifetime. Acquisition probability can be estimated by a probit or logit model and hazard models can also be used if the timing of the incidence is concerned. Duration data are usually right-censored, so Tobit or hazard models are by nature suitable estimation techniques. Researchers link acquisition and retention modeling together either by specifying the correlation of the error terms in probability and duration models or by specifying the joint distribution of acquisition and retention for estimation.

### 5.2.1 Empirical example: Balancing acquisition and retention

In order to understand how to allocate resources across customer acquisition and customer retention, we first need to develop a set of models which describe the acquisition and retention process. This will involve three different models: the acquisition model, duration model, and profit model. Once we have the results from the three models, then we can determine how to optimally shift resources between acquisition and retention efforts to maximize profitability. Thus, the modeling framework will take the following format (similar to the modeling framework as described earlier in the chapter):

$$y_{L_i}\begin{cases}=\beta_L' x_{Li}+\gamma' y_{Di}+\varepsilon_{Li} & \text{if } z_i=1\\ =0 & \text{otherwise}\end{cases}\quad\text{(cumulative profitability equation)}\tag{5.15}$$

$$y_{Di}\begin{cases}=\beta_D' x_{Di}+\varepsilon_{Di} & \text{if } z_i=1\\ =0 & \text{otherwise}\end{cases}\quad\text{(duration equation)}\tag{5.16}$$

$$\begin{aligned}z_i^* &= \alpha' v_i+\mu_i \quad \text{(acquisition equation)}\\ z_i &= \begin{cases}1 & \text{if } z_i^*>0\\ 0 & \text{if } z_i^*\le 0\end{cases}\end{aligned}\tag{5.17}$$

where $z_i^*$ is a latent variable indicating customer $i$'s utility to engage in a relationship with a firm, $z_i$ an indicator variable showing whether customer $i$ is acquired ($z_i = 1$) or not ($z_i = 0$), $v_i$ a vector of covariates affecting the acquisition of customer $i$, $y_{Di}$ the duration of customer $i$'s relationship with the firm, $x_{Di}$ a vector of covariates affecting the duration of customer $i$'s relationship with the firm, $y_{L_i}$ the cumulative profitability of customer $i$, $x_{Li}$ a vector of covariates affecting customer $i$'s lifetime value, $\alpha$, $\beta_L$, $\beta_D$ a vector of parameters, and $\mu_i$, $\varepsilon_{Di}$, and $\varepsilon_{Li}$ error terms. Given that the modeling framework is recursive in nature, we can estimate it in a stepwise fashion. Thus, we will proceed in the following manner. There will be four subsections in this empirical example. The first subsection will describe and estimate the acquisition model, the second will describe and estimate the duration model, the third will describe and estimate the profit model, and the fourth will describe and show how to optimally allocate resources between acquisition and retention.

#### 5.2.1.1 Acquisition model

The key question we want to answer with regard to customer acquisition here is whether we can determine which future prospects have the highest likelihood of adoption. To do this we first need to know which past prospects were acquired and which were not. In the dataset provided for this chapter we have a binary variable which identifies whether or not a prospect was acquired by the firm (and hence became a customer) and a set of drivers which are likely to help explain a customer's decision to adopt. A random sample of 500 prospects (some of whom became customers) was taken from a B2B firm. The information we need for our acquisition model includes the following list of variables:

| | |
|---|---|
| **Dependent variable** | |
| *Acquisition* | *1 if the prospect was acquired, 0 otherwise* |
| **Independent variables** | |
| *Acq_Exp* | *The total dollars spent on trying to acquire this prospect* |
| *Acq_Exp_SQ* | *The square of the total dollars spent on trying to acquire this prospect* |
| *Industry* | *1 if the prospect is in the B2B industry, 0 otherwise* |
| *Revenue* | *Annual sales revenue of the prospect's firm (in millions of dollars)* |
| *Employees* | *Number of employees in the prospect's firm* |

In this case, we have a binary dependent variable (*Acquisition*) which tells us whether the prospect did adopt (= 1) or did not adopt (= 0). We also have five independent variables that we believe will be drivers of adoption. First, we have how many dollars the firm spent on each prospect (*Acq_Expense*) and the squared value of that variable (*Acq_Expense_SQ*). We want to use both the linear and squared terms since we expect that for each additional dollar spent on the acquisition effort for a given prospect, there will be a diminishing return to the value of that dollar. Second, since the focal firm of this example is a B2B firm, the other three variables are firmographic variables of the prospects. These include whether the prospect sells to B2B (= 1) or B2C (= 0) customers (*Industry*), how much (in millions) that the prospect firm makes in annual revenue (*Revenue*), and how many employees the prospect firm has (*Employees*).

First, we need to model the probability that a prospect will adopt. Since our dependent variable (*Acquisition*) is binary and we need an error structure that is similar to the duration and profit models (both normally distributed), we select a probit regression for this model. Choosing a logistic regression would require transforming the model output before integrating the results with the other two equations. In this case the $y$ variable is *Acquisition* and the $x$ variables represent the five independent variables in our database. When we run the probit regression we get the following result:

| Variable | Estimate | Standard error | $p$-value |
|---|---|---|---|
| *Intercept* | −8.255 | 0.811 | < 0.0001 |
| *Acq_Exp* | 0.017 | 0.002 | < 0.0001 |
| *Acq_Exp_SQ* | −0.000 02 | 0.000 002 | < 0.0001 |
| *Industry* | 1.217 | 0.168 | < 0.0001 |
| *Revenue* | 0.043 | 0.008 | < 0.0001 |
| *Employees* | 0.004 | 0.0004 | < 0.0001 |

As we can see from the results, all five independent variables are significant at a $p$-value of 5% or better. First, this means that acquisition expense has a positive, but diminishing, effect (*Acq_Exp* > 0 and *Acq_Exp_SQ* < 0) on acquisition likelihood. Second, prospects who are B2B (vs. B2C), have a higher *Revenue*, and have more *Employees* are more likely to be acquired.

Now that we have determined the drivers of customer acquisition we need to use the output of the model to determine our model's predictive accuracy. To do this we need to use the estimates we obtained from the acquisition model to help us determine the predicted probability that each customer will repurchase. We use the parameter estimates from the repurchase model and values for the $x$ variables for each customer in each time period to predict whether a customer is likely to purchase in that time period. For a probit regression we must apply the proper probability function

$$P(Acquisition = 1\,|\,X\beta) = \frac{1}{2}\left[1 + erf\left(\frac{X\beta - \mu}{\sqrt{2\sigma^2}}\right)\right]$$

where $X$ is the matrix of variables, $\beta$ is the vector of coefficients, $\mu$ is the mean of the error distribution (in this case 0), $\sigma$ is the standard deviation of the error distribution (in this case 1 since it is a standard normal distribution), and *erf* is the error function which is equal to

$$erf(x) = \frac{2}{\pi}\int_0^X \exp(-t^2)\,dt.$$

Once we compute the probability of repurchase, we need to create a cutoff value to determine at which point we are going to divide the customers into the two groups – predicted to adopt and predicted not to adopt. There is no rule that explicitly tells us what that cutoff number should be. Often by default we select 0.5 since it is equidistant from 0 and 1. However, it is also reasonable to check multiple cutoff values and choose the one that provides the best predictive accuracy for the dataset. By using 0.5 as the cutoff for our example, any customer whose predicted probability of acquisition is greater than or equal to 0.5 is classified as predicted to adopt and the rest are predicted not to adopt. To determine the predictive accuracy we compare the predicted to the actual acquisition values in a $2 \times 2$ table. For our sample of 500 customers we get the following table:

| | | Actual acquisition | | |
|---|---|---|---|---|
| | | 0 | 1 | *Total* |
| Predicted acquisition | 0 | *105* | *38* | ***143*** |
| | 1 | *57* | *300* | ***357*** |
| | *Total* | ***162*** | ***338*** | ***500*** |

As we can see from the table, our in-sample model accurately predicts 64.8% of the prospects who chose not to adopt (105/162) and 88.7% of the prospects who chose to adopt (300/338). For the prediction of prospects who did adopt this is a significant increase in the predictive capability of a random guess model[1] which would be only 67.6% accurate for this dataset. However, we see that our model slightly underperforms the random guess model with regard to the prospects that chose not to adopt. To determine overall model prediction performance we look at the diagonal and see that overall our prediction accuracy is 81.0% (405/500). Given that the model in general predicts better than the random guess model, we would determine that the model prediction is good.

As a result we now know how changes in acquisition expense and prospect characteristics are likely to either increase or decrease the likelihood of adoption. And we also know that these drivers do a good job in helping us predict whether a prospect is going to adopt or not.

#### 5.2.1.2 Duration model

The second step of this process is to estimate the duration model. The purpose of this model is to understand the drivers that describe the length of time a customer is likely to be a customer conditional on the fact that adoption occurred. Thus the

[1] A random guess model would do the following. First, it would determine which bucket acquire or not acquire has more customers in it. In this case 338 customers adopted versus 162 who did not. Then it would predict that all customers would adopt and it would be accurate 67.6% of the time (338/500).

equation takes the following format:

$$\mathrm{E}(Duration) = \mathrm{P}(Acquisition = 1) * \mathrm{E}(Duration|Acquisition = 1).$$

This equation shows us that the expected duration is a function of the probability that the prospect is acquired multiplied by the expected value of duration given that the prospect was acquired. If we were to merely run a regression with *Duration* as the dependent variable and ignore the probability that the customer will make a purchase, we would get biased estimates due to a potential sample selection bias.

Sample selection bias is a problem that is common in many marketing problems and has to be statistically accounted for in many modeling frameworks. In this case the customer has a choice as to whether or not to be adopted before deciding how long the relationship will last. If we were to ignore this choice we would bias the estimates from the model and we would have less precise predictions for the value of *Duration*. To account for this issue we need to be able to predict the value for both the probability of *Acquisition* (what we did in the first step of this example) and the expected value of *Duration* given that the prospect is expected to adopt. To account for this issue we use a two-stage modeling framework similar to that described earlier in this chapter and found in Reinartz *et al.* [3].

We will use the output and predictions of the probit model from the first step of this example to create a new variable, $\lambda$, which will represent the correlation in the error structure across the two equations. This variable, also known as the sample selection correction variable, will then be used as an independent variable in the *Duration* model to remove the sample selection bias in the estimates. To compute $\lambda$ we use the following equation, also known as the inverse Mills ratio:

$$\lambda = \frac{\phi(X'\beta)}{\Phi(X'\beta)}.$$

In this equation $\phi$ represents the normal probability density function, $\Phi$ represents the normal cumulative density function, $X$ represents the value of the variables in the acquisition model, and $\beta$ represents the coefficients derived from the estimation of the acquisition model.

Finally, we want to estimate a regression model for *Duration* and include the variable $\lambda$ as an additional independent variable. This is done in a straightforward manner using the following equation:

$$Duration = \gamma'\alpha + \mu\lambda + \varepsilon.$$

In this case *Duration* is the value of the duration, $\gamma$ is the matrix of variables used to help explain the value of *Duration*, $\alpha$ are the coefficients for the independent variables, $\mu$ is the coefficient on the inverse Mills ratio, $\lambda$ is the inverse Mills ratio, and $\varepsilon$ is the error term. Thus, for this example we will use the following list of variables:

| | |
|---|---|
| **Dependent variable** | |
| *Duration* | *The number of days the customer was a customer of the firm, 0 if Acquisition = 0* |
| **Independent variables** | |
| *Ret_Exp* | *The total dollars spent on trying to retain this customer* |
| *Ret_Exp_SQ* | *The square of the total dollars spent on trying to retain this customer* |
| *Freq* | *The number of purchases the customer made during that customer's lifetime with the firm, 0 if Acquisition = 0* |
| *Freq_SQ* | *The square of the number of purchases the customer made during that customer's lifetime with the firm* |
| *Crossbuy* | *The number of product categories the customer purchased from during that customer's lifetime with the firm, 0 if Acquisition = 0* |
| *SOW* | *The percentage of purchases the customer makes from the given firm given the total amount of purchases across all firms in that category* |
| *Lambda(λ)* | *The computed inverse Mills ratio from the acquisition model* |

When we estimate the second-stage of the model, we get the following parameter estimates from the second of the two equations (the parameter estimates for the acquisition model are detailed in the first part of this example):

| Variable | Estimate | Standard error | $p$-value |
|---|---|---|---|
| *Intercept* | 91.008 | 9.756 | < 0.0001 |
| *Ret_Exp* | 2.528 | 0.029 | < 0.0001 |
| *Ret_Exp_SQ* | −0.001 | 0.000 02 | < 0.0001 |
| *Freq* | 7.072 | 0.806 | < 0.0001 |
| *Freq_SQ* | −0.842 | 0.040 | < 0.0001 |
| *Crossbuy* | 3.196 | 0.479 | < 0.0001 |
| *SOW* | 0.353 | 0.045 | < 0.0001 |
| *Lambda(λ)* | 29.520 | 2.557 | < 0.0001 |

We gain the following insights from the results. We see that λ is positive and significant. We can interpret this to mean that there is a potential selection bias problem since the error term of our selection equation is correlated positively with the error term of our regression equation. We also see that all other variables of the duration model are significant, meaning that we have likely uncovered many of the drivers of duration.

We find that *Ret_Exp* is positive with a diminishing return, as noted by the positive coefficient on *Ret_Exp* and the negative coefficient on *Ret_Exp_SQ*. This means that marketing efforts to retain and build relationships with the customer do cause the customer to stay longer in the relationship, to a point. Then, after the threshold is reached, marketing efforts actually decrease the length of duration of the relationship on average. This is likely due to the fact that overly contacting customers can often strain the relationship between the customer and firm. We find that

*Freq* is also positive with a diminishing return, as noted by the positive coefficient on *Freq* and a negative coefficient on *Freq_SQ*. This means that customers who purchase a moderate number of times are likely to have the longest relationships with the firm. And customers who purchase less frequently (or very frequently) are more likely to leave the relationship earlier. We find the coefficient on *Crossbuy* is positive, suggesting that customers who purchase across more categories are more likely to stay in the relationship longer. Finally, we find the coefficient on *SOW* to be positive, suggesting that customers who purchase a larger percentage of their budget for a given set of items at the focal store are more likely to have a longer relationship.

Our next step is to predict the value of *Duration* to see how well our model compares to the actual values. We do this by starting with the equation for expected duration at the beginning of this example:

$$\begin{aligned}\mathrm{E}(Duration) &= \mathrm{P}(Acquisition = 1) * \mathrm{E}(Duration|Acquisition = 1)\\ &= \Phi(X'\beta) * (\gamma'\alpha + \mu\lambda).\end{aligned}$$

In this case $\Phi$ is the normal CDF distribution, $X$ is the matrix of independent variable values from the *Acquisition* equation, $\beta$ is the vector of parameter estimates from the *Acquisition* equation, $\gamma$ is the matrix of independent variables from the *Duration* equation, $\alpha$ is the vector of parameter estimates from the *Duration* equation, $\mu$ is the parameter estimate for the inverse Mills ratio, and $\lambda$ is the inverse Mills ratio. Once we have predicted the *Duration* value for each of the customers we want to compare this to the actual value from the database. We do this by computing the MAD. The equation is as follows:

$$\mathrm{MAD} = \mathrm{Mean}\{\mathrm{Absolute\ Value}[\mathrm{E}(Duration) - Duration]\}.$$

We find for all customers that MAD = 144.02. This means that on average each of our predictions of *Duration* deviates from the actual value by about 144 days. If we were to instead use the mean value of *Duration* (742.45) across all customers as our prediction for all prospects (this would be the benchmark model case), we would find that MAD = 484.09, or about 484 days. As we can see, our model does a significantly better job of predicting the length of the customer relationship than the benchmark case.

#### 5.2.1.3 Profit model

The third step of this process is to estimate the profit model. The purpose of this model is to understand the drivers that describe the expected value of the CLV. Thus the equation takes the following format:

$$\mathrm{E}(Profit) = \mathrm{P}(Acquisition = 1) * \mathrm{E}(Profit|Acquisition = 1, \mathrm{E}(Duration)).$$

This equation shows us that the expected duration is a function of the probability that the prospect is acquired multiplied by the expected value of profit given that the prospect was acquired and the estimated duration of the customer's relationship with the firm. Again, if we were to merely run a regression with *Profit* as the dependent variable and ignore the probability that the customer will make a purchase and the estimated duration, we would get biased estimates due to a potential sample selection bias.

Thus, we will use the λ variable as an additional variable in the model which is computed using the following equation:

$$\lambda = \frac{\phi(X'\beta)}{\Phi(X'\beta)}.$$

In this equation $\phi$ represents the normal probability density function, $\Phi$ represents the normal cumulative density function, $X$ represents the value of the variables in the acquisition model, and $\beta$ represents the coefficients derived from the estimation of the acquisition model.

We will also use the expected value of *Duration* from the second step of this example in our profit model. The expected value of *Duration* is merely computed as

$$\mathrm{E}(Duration) = \mathrm{P}(Acquisition = 1) * \mathrm{E}(Duration|Acquisition = 1).$$

Finally, we want to estimate a regression model for *Profit* and include the variables λ and E(*Duration*) as additional independent variables. This is done in a straightforward manner using the following equation:

$$Profit = \gamma'\alpha + \mu\lambda + \rho Dur\hat{a}tion + \varepsilon.$$

In this case *Profit* is the value of the profit, $\gamma$ is the matrix of variables used to help explain the value of *Profit*, $\alpha$ are the coefficients for the independent variables, $\mu$ is the coefficient on the inverse Mills ratio, λ is the inverse Mills ratio, $\rho$ is the coefficient on the expected duration, *Durâtion* is the expected duration, and ε is the error term. Thus, for this example we will use the following list of variables:

| | |
|---|---|
| **Dependent variable** | |
| *Profit* | *The CLV of a given customer, –(Acq_Exp) if the customer is not acquired* |
| **Independent variables** | |
| *Acq_Exp* | *The total dollars spent on trying to acquire this prospect* |
| *Acq_Exp_SQ* | *The square of the total dollars spent on trying to acquire this prospect* |
| *Ret_Exp* | *The total dollars spent on trying to retain this customer* |
| *Ret_Exp_SQ* | *The square of the total dollars spent on trying to retain this customer* |
| *Freq* | *The number of purchases the customer made during that customer's lifetime with the firm, 0 if Acquisition = 0* |
| *Freq_SQ* | *The square of the number of purchases the customer made during that customer's lifetime with the firm* |

| | |
|---|---|
| *Crossbuy* | *The number of product categories the customer purchased from during that customer's lifetime with the firm, 0 if Acquisition = 0* |
| *SOW* | *The percentage of purchases the customer makes from the given firm, given the total amount of purchases across all firms in that category* |
| *Lambda(λ)* | *The computed inverse Mills ratio from the acquisition model* |
| *Durâtion* | *The number of days the customer was a customer of the firm, 0 if Acquisition = 0* |

When we estimate the third stage of the model, we get the following parameter estimates (the parameter estimates for the acquisition model are detailed in the first part of this example and the parameter estimates of the duration model are detailed in the second part):

| Variable | Estimate | Standard error | $p$-value |
|---|---|---|---|
| *Intercept* | 7.038 | 48.899 | 0.8856 |
| *Acq_Exp* | 3.344 | 0.132 | < 0.0001 |
| *Acq_Exp_SQ* | −0.001 | 0.0001 | < 0.0001 |
| *Ret_Exp* | 3.846 | 0.185 | < 0.0001 |
| *Ret_Exp_SQ* | −0.001 | 0.0001 | < 0.0001 |
| *Freq* | 17.820 | 2.751 | < 0.0001 |
| *Freq_SQ* | −1.349 | 0.144 | < 0.0001 |
| *Crossbuy* | 18.189 | 1.636 | < 0.0001 |
| *SOW* | 2.075 | 0.158 | < 0.0001 |
| *Lambda(λ)* | 107.586 | 44.839 | 0.0170 |
| *Durâtion* | 0.496 | 0.066 | < 0.0001 |

We gain the following insights from the results. We see that λ is positive and significant. We can interpret this to mean that there is a potential selection bias problem since the error term of our selection equation is correlated positively with the error term of our regression equation. We also see that all other variables of the profit model are significant, meaning that we have likely uncovered many of the drivers of customer lifetime value.

We find that *Acq_Exp* is positive with a diminishing return, as noted by the positive coefficient on *Acq_Exp* and the negative coefficient on *Acq_Exp_SQ*. This means that marketing efforts to acquire prospects do cause the customer to be more profitable, to a threshold. Then, after the threshold is reached, marketing efforts on acquisition actually decrease the profitability of the customer on average. This is because, after a certain amount of spending on acquisition, there is no additional benefit (i.e., it costs more for no benefit in spending). We find that *Ret_Exp* is positive with a diminishing return, as noted by the positive coefficient on *Ret_Exp* and the negative coefficient on *Ret_Exp_SQ*. This means that marketing efforts to retain and build relationships with the customer do cause the customer to be more profitable, to a point. Then, after the threshold is reached, marketing efforts on retention actually decrease the profitability on average. This is likely due to the fact that overly contacting customers can often strain the relationship between the customer and firm and drive the customer not to purchase. We find that *Freq* is also positive with a diminishing return, as noted by the positive coefficient on *Freq* and negative coefficient on *Freq_SQ*. This means that customers who purchase a moderate number of times are likely to have the highest profitability with the firm. And customers who purchase less frequently (or very frequently) are more likely to less profitable. We find the coefficient on *Crossbuy* is positive, suggesting that customers who purchase across more categories are more likely to be profitable. We find the coefficient on *SOW* to be positive, suggesting that customers who purchase a larger percentage of their budget for a given set of items at the focal store are more likely to be profitable. Finally, we find the coefficient on expected duration to be positive, suggesting that customers who are in the relationship longer are more likely to profitable.

Our next step is to predict the value of *Profit* to see how well our model compares to the actual values. We do this by starting with the equation for expected profit at the beginning of this example:

$$\begin{aligned}\mathrm{E}(\mathit{Profit}) &= \mathrm{P}(\mathit{Acquisition} = 1) * \mathrm{E}(\mathit{Profit}\,|\,\mathit{Acquisition} = 1, \mathrm{E}(\mathit{Duration}))\\ &= \Phi(X'\beta) * (\gamma'\alpha + \mu\lambda + \rho \mathit{Dur\hat{a}tion} + \varepsilon).\end{aligned}$$

In this case $\Phi$ is the normal CDF distribution, $X$ is the matrix of independent variable values from the *Acquisition* equation, $\beta$ is the vector of parameter estimates from the *Acquisition* equation, $\gamma$ is the matrix of independent variables form the *Profit* equation, $\alpha$ is the vector of parameter estimates from the *Profit* equation, $\mu$ is the parameter estimate for the inverse Mills ratio, $\lambda$ is the inverse Mills ratio, $\rho$ is the coefficient on the expected duration, and *Durâtion* is the expected duration. Once we have predicted the *Profit* value for each of the customers we want to compare this to the actual value from the database. We do this by computing the MAD. The equation is as follows:

$$\mathrm{MAD} = \mathrm{Mean}\{\mathrm{Absolute\ Value}[\mathrm{E}(\mathit{Profit}) - \mathit{Profit}]\}.$$

We find for all the customers that MAD $= 752.26$. This means that on average each of our predictions of *Profit* deviates from the actual value by about \$752.26. If we were to instead use the mean value of *Profit* (\$2403.84) across all customers as our prediction (this would be the benchmark model case), we would find that MAD $= 1881.40$, or \$1881.40. As we can see, our model does a significantly better job of predicting the expected profit of customers than the benchmark case.

## 5.3 Optimal resource allocation

Our final task is to see what our potential profitability would be if we were to allow the resources spent on acquisition (*acq_exp*) and retention (*ret_exp*) to vary. To do this we need to take the results of the first three steps of this example and simulate different potential outcomes based on a set of constraints. To start we set up the objective function we wish to maximize. For the purpose of this example there will be have four scenarios that we will run. We will change the average amount we spend on acquisition and retention based on these desired outcomes:

1. Maximize the acquisition rate across all customers given the current marketing budget.
2. Maximize the average duration across all customers given the current marketing budget.
3. Maximize the total profit across all customers given the current marketing budget.
4. Maximize the total profit across all customers given no budget limitation.

For simplicity, in each case we will keep the average values of the other variables (outside of acquisition expense and retention expense) the same throughout the exercise. Then we want to compare the results of the optimization exercises to the current scenario to see where improvements are made and how the allocations change to get those improvements. The current level of the potential outcome variables and the acquisition and retention expenses are the following:

| | E(*Profit*) | E(*Duration*) | E(*Acq. %*) | *Acq_Exp* | *Ret_Exp* | Total expense |
|---|---|---|---|---|---|---|
| Current | \$1.166 million | 1098 d | 67.60% | \$493.35 | \$497.43 | \$415 000 |

In this case E(*Profit*) is the expected total profit across all prospects and customers, E(*Duration*) is the average expected duration given that the prospect becomes a customer, E(*Acq. %*) is the expected acquisition rate, or E(*# Acquired*)/500, *Acq_Exp* is the average acquisition expense across all prospects, *Ret_Exp* is the expected retention expense across all acquired customers, and Total expense is the total spent across prospects and customers.

For the first three scenarios we plan to simulate, we will try to maximize the three dependent variables (*Profit*, *Duration*, and *Acquisition*) across all prospects and customers given the following constraints. First, the amount of spending on acquisition and retention efforts needs to be positive (i.e., we cannot spend a negative amount of money on either activity). Second, the total amount spent on acquisition and retention cannot exceed the current budget of \$415 000. Third, the duration of any customer relationship cannot be negative (this is conceptually straightforward, but since we are using expected duration, the prediction of duration can become negative if the acquisition and retention variables change too dramatically). In addition for the first scenario where we want to maximize acquisition rate, since acquisition rate does not impact duration directly, we will assume that the firm allocates all remaining resources not desired for acquisition efforts to retention efforts (so the sum of all expenditures is still \$415 000). For the last scenario the only change to our constraints is to relax the budget constraint and allow the firm to spend an unlimited amount of money on acquisition and retention efforts. When we run the simulation we get the following results:

| | E(*Profit*) | E(*Duration*) | E(*Acq. %*) | *Acq_Exp* | *Ret_Exp* | Total expense |
|---|---|---|---|---|---|---|
| Current | \$1.17 million | 1098 d | 67.60% | \$493.35 | \$497.43 | \$415 000 |
| E(*Acq. %*) | \$1.44 million | 900 d | 77.18% | \$432.50 | \$514.52 | \$415 000 |
| E(*Duration*) | \$1.45 million | 1000 d | 69.13% | \$321.75 | \$734.65 | \$415 000 |
| E(*Profit*) | \$1.47 million | 976 d | 74.78% | \$370.38 | \$614.11 | \$415 000 |
| E(*Profit*) with/without limits | \$2.71 million | 470 d | 76.95% | \$452.04 | \$2273.73 | \$1.10 million |

We get the following results for each of the scenarios:

- **Maximize E(Acq. %).** For the scenario where we want to maximize acquisition rate, the expected total profit is higher by \$270 000, the expected duration is shorter by 198 days, the acquisition rate is higher by 9.58%, the average acquisition expense is lower by \$60.85, the average retention expense is higher by \$17.09, and the total marketing budget is the same. This shows that maximizing acquisition rate is better than the current scenario since the total profit is higher when we try to acquire the most customers. This happens because even though the average acquisition spending is lower, the spending on acquisition is targeted at the prospects who are more likely to join and not targeted at the prospects who are unlikely to join. This is why the expected acquisition rate is so much higher than in the current scenario. Now in this case the duration of the customers is less most likely due to the fact that the average spending on each customer had to decrease, given that so many more customers were acquired in this scenario and the budget was limited to be the same as before.
- **Maximize E(duration).** For the scenario where we want to maximize the average duration of the acquired customers, the expected profit is higher by \$280 000, the expected duration is shorter by 98 days, the acquisition rate is higher by 1.53%, the average acquisition expense is lower by \$171.60, the average retention expense is higher by \$237.22, and the total marketing budget is the same when compared to the current scenario. This shows that maximizing the duration of the acquired customers is better than the current scenario or even the scenario where we maximized acquisition rate, since the total profit is higher when we try to increase the duration of each customer relationship. However, we notice that the expected duration is less than the current scenario even when we are trying to maximize customer duration. This is happening because we are maximizing the customer acquisition of the average customer who is acquired. Since we are not bringing on the ‘best’ customers we are only able to maximize the duration of the customers we

were actually able to bring on. This means the firm was already acquiring better than 'average' customers in its current scenario.

- **Maximize E(profit).** For the scenario where we want to maximize the expected profit across all prospects and customers, the expected profit is higher by \$300 000, the expected duration is shorter by 122 days, the acquisition rate is higher by 7.18%, the average acquisition expense is lower by \$122.97, the average retention expense is higher by \$116.68, and the total marketing budget is the same. This shows that maximizing the profit does lead to a much higher profit level when compared to the current scenario and also a little higher when compared to both of the previous two scenarios where we wanted to maximize customer acquisition rate and customer duration in isolation. These results also show that balance is critical to maximizing profit since the expected acquisition rate and the expected duration are between the scenarios where we maximize acquisition rate and maximize duration. Again we see that the entire budget is used, suggesting that our simulation would potentially find it more lucrative to spend even more money on acquisition and retention efforts.
- **Maximize E(profit) with unlimited budget.** For the scenario where we want to maximize the expected profit across all prospects and customers regardless of budget level, the expected profit is higher by \$1.54 million, the expected duration is shorter by 628 days, the acquisition rate is higher by 9.35%, the average acquisition expense is lower by \$41.31, the average retention expense is higher by \$1776.30, and the total marketing budget is higher by \$685 000 when compared to the current scenario. We see that when we take away the budget constraint the desired spending level is significantly higher than the current spending level of the firm. While it may not be the case that the firm can afford to spend this much on marketing efforts for this group of prospects/customers, it does suggest that if the firm is holding back on the current budget level, there is some justification to spend more, especially on retention efforts.

The end result of this optimal resource allocation exercise is that the firm can be more profitable if it balances its acquisition and retention efforts. By focusing too much on acquisition or on retention (duration), the firm is not maximizing its profit. It only maximizes profit when it balances its spending across both activities. In addition, we see that the firm is dramatically underspending on its marketing efforts if it desires to maximize long-term profitability.

### 5.3.1 How do you implement it?

For this empirical exercise several different methods were used. First, to estimate the probit regression for the acquisition model we used PROC Logistic in SAS with the probit link function. Second, to estimate the censored regression for both the duration model and the profit model in the second and third steps we used PROC

Reg in SAS. Finally, to run the optimization routine we used Solver in Excel. There are numerous other programs such as MATLAB, GAUSS, and R which could be used to estimate these models and run the optimizations.

## 5.4 Chapter summary

The purpose of this chapter was to explore the current models for balancing customer acquisition and retention and to provide an empirical example as to how firms can apply this knowledge to their own customer databases. We have shown that when firms are able to first understand the drivers of customer acquisition, customer retention, and customer profitability and then run some simulation exercises with these results, the firms can achieve a much higher level of profitability.

## Balancing acquisition and retention – SAS code

```
/* Import Data - Library: statcrm */

proc import out = statcrm.acquisition_retention
    datafile = "C:\_Your_Data_Location_\Statistics in CRM\Acquisition and
Retention\Acquisition-Retention.xls" dbms = excel replace;
    range = "Summary$"; getnames = yes; mixed = no; scantext = yes;
usedate = yes; scantime = yes;
run; quit;

/* Acquisition Probability */

proc logistic data = statcrm.acquisition_retention descending; model
acquisition = acq_exp acq_exp_sq Industry Revenue Employees / link =
probit;
output out = statcrm.acq_ret_probit xbeta = xb_probit p = pred;
run; quit;

/* Purchase Accuracy */

data statcrm.acq_ret_accuracy;
set statcrm.acq_ret_probit;
if pred >= 0.5 then pred_acq = 1; else pred_acq = 0;
act_acq = acquisition;
run; quit;

proc freq data = statcrm.acq_ret_accuracy;
table pred_acq * act_acq; run; quit;
/* Inverse Mills Ratio */

data statcrm.acq_ret_imr;
set statcrm.acq_ret_probit;
imr_acq = (pdf('Normal',xb_probit))/(probnorm(xb_probit));
run; quit;

/* Duration */

proc reg data = statcrm.acq_ret_imr;
model duration = ret_exp ret_exp_sq freq freq_sq crossbuy sow imr_acq /
rsquare;
where acquisition = 1;
run; quit;

data statcrm.acq_ret_duration1;
set statcrm.acq_ret_imr;
xbeta_duration = 91.00828 + 2.52829*ret_exp - 0.00102*ret_exp_sq +
7.07247*freq - 0.84187*freq_sq + 3.19638*crossbuy + 0.35287*sow +
29.51954*imr_acq;
pred_dur = probnorm(xb_probit)*(xbeta_duration);
ad = abs(duration - pred_dur);
ad1 = abs(duration - 742.45);
run; quit;
```

```
proc sql; select mean(duration) as mean_dur, mean(ad) as mad, mean(ad1) as
random_mad from statcrm.acq_ret_duration1; quit;

/* Profit */

proc reg data = statcrm.acq_ret_duration1;
model profit = acq_exp acq_exp_sq ret_exp ret_exp_sq freq freq_sq crossbuy
sow imr_acq pred_dur / rsquare;
where acquisition = 1;
run; quit;

data statcrm.acq_ret_profit1;
set statcrm.acq_ret_duration1;
xbeta_profit = 3.34409*acq_exp - 0.00126*acq_exp_sq + 3.84649*ret_exp -
0.00066891*ret_exp_sq + 17.81953*freq - 1.34911*freq_sq +
      18.18908*crossbuy + 2.07463*sow + 107.58570*imr_acq +
0.49555*pred_dur; pred_profit = probnorm(xb_probit)*(xbeta_profit);
ad = abs(profit - pred_profit);
ad1 = abs(profit - 2403.843);
run; quit;

proc sql; select mean(profit) as mean_profit, mean(ad) as mad, mean(ad1) as
random_mad from statcrm.acq_ret_profit1; quit;
```

## Balancing acquisition and retention – SAS output

```
                    The SAS System

                 The LOGISTIC Procedure

                   Model Information

Data Set                    STATCRM.ACQUISITION_RETENTION
Response Variable               Acquisition              Acquisition
Number of Response Levels       2
Model                           binary probit
Optimization Technique          Fisher's scoring

                 Number of Observations Read          500
                 Number of Observations Used          500

                          Response Profile

                  Ordered                             Total
                  Value         Acquisition       Frequency

                      1              1                  338
                      2              0                  162

             Probability modeled is Acquisition = 1.

                    Model Convergence Status

       Convergence criterion (GCONV = 1E-8) satisfied.

                     Model Fit Statistics
                                                  Intercept
                                  Intercept             and
                  Criterion            Only      Covariates

                  AIC               631.848         363.277
                  SC                636.062         388.564
                  -2 Log L          629.848         351.277

              Testing Global Null Hypothesis: BETA = 0

            Test                    Chi-Square       DF      Pr > ChiSq

            Likelihood Ratio          278.5713        5        <.0001
            Score                     212.7075        5        <.0001
            Wald                      143.8489        5        <.0001
```

The SAS System

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| Intercept | 1 | -8.2554 | 0.8106 | 103.7102 | <.0001 |
| Acq_Exp | 1 | 0.0173 | 0.00244 | 50.2920 | <.0001 |
| Acq_Exp_SQ | 1 | -0.00002 | 2.425E-6 | 51.5224 | <.0001 |
| Industry | 1 | 1.2167 | 0.1682 | 52.3471 | <.0001 |
| Revenue | 1 | 0.0425 | 0.00817 | 27.0653 | <.0001 |
| Employees | 1 | 0.00432 | 0.000434 | 99.0474 | <.0001 |

Association of Predicted Probabilities and Observed Responses

| | | | |
|---|---|---|---|
| Percent Concordant | 90.7 | Somers' D | 0.814 |
| Percent Discordant | 9.2 | Gamma | 0.815 |
| Percent Tied | 0.1 | Tau-a | 0.357 |
| Pairs | 54756 | c | 0.907 |

The SAS System

The FREQ Procedure

Table of pred_acq by act_acq

```
pred_acq     act_acq

Frequency,
Percent ,
Row Pct ,
Col Pct ,        0,         1,     Total

   ffffffffff^ffffffff^ffffffff^
0 ,           105 ,       38 ,       143
 ,          21.00 ,     7.60 ,     28.60
 ,          73.43 ,    26.57 ,
 ,          64.81 ,    11.24 ,
 fffffffff^ffffffff^ffffffff^
1 ,            57 ,      300 ,       357
 ,          11.40 ,    60.00 ,     71.40
 ,          15.97 ,    84.03 ,
 ,          35.19 ,    88.76 ,
 ffffffff^ffffffff^ffffffff^
Total          162        338        500
             32.40      67.60     100.00
```

```
                           The SAS System

                          The REG Procedure
                           Model: MODEL1
                  Dependent Variable: Duration Duration

                Number of Observations Read         338
                Number of Observations Used         338

                        Analysis of Variance

                                           Sum of         Mean
Source                  DF      Squares     Square      F Value      Pr > F

Model                    7     15730414    2247202      7673.94      <.0001
Error                  330        96636  292.83567
Corrected Total        337     15827050

        Root MSE               17.11244     R-Square        0.9939
        Dependent Mean       1098.30473     Adj R-Sq        0.9938
        Coeff Var               1.55808

                        Parameter Estimates

                                      Parameter     Standard
Variable      Label          DF        Estimate        Error   t Value   Pr > |t|

Intercept     Intercept       1        91.00827      9.75567      9.33   <.0001
Ret_Exp       Ret_Exp         1         2.52829      0.02929     86.31   <.0001
Ret_Exp_SQ    Ret_Exp_SQ      1        -0.00102   0.00002424    -42.02   <.0001
Freq          Freq            1         7.07247      0.80596      8.78   <.0001
Freq_SQ       Freq_SQ         1        -0.84187      0.04031    -20.89   <.0001
Crossbuy      Crossbuy        1         3.19638      0.47933      6.67   <.0001
SOW           SOW             1         0.35287      0.04545      7.76   <.0001
imr_acq                       1        29.51954      2.55664     11.55   <.0001

                           The SAS System

                 mean_dur        mad        random_mad
                  ffffffffffffffffffffffffffffff
               742.454      144.0177    484.0866

                           The SAS System
                          The REG Procedure
                           Model: MODEL1
                    Dependent Variable: Profit Profit

             Number of Observations Read  338
             Number of Observations Used  338
```

Analysis of Variance

| Source | DF | Squares | Sum of Square | Mean F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 10 | 144748053 | 14474805 | 4314.17 | <.0001 |
| Error | 327 | 1097143 | 3355.17707 | | |
| Corrected Total | 337 | 145845196 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 57.92389 | R-Square | 0.9925 |
| Dependent Mean | 3792.64059 | Adj R-Sq | 0.9922 |
| Coeff Var | 1.52727 | | |

Parameter Estimates

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| Intercept | Intercept | 1 | 7.03798 | 48.89864 | 0.14 | 0.8856 |
| Acq_Exp | Acq_Exp | 1 | 3.34409 | 0.13222 | 25.29 | <.0001 |
| Acq_Exp_SQ | Acq_Exp_SQ | 1 | -0.00126 | 0.00013165 | -9.57 | <.0001 |
| Ret_Exp | Ret_Exp | 1 | 3.84649 | 0.18521 | 20.77 | <.0001 |
| Ret_Exp_SQ | Ret_Exp_SQ | 1 | -0.00066891 | 0.00010968 | -6.10 | <.0001 |
| Freq | Freq | 1 | 17.81953 | 2.75056 | 6.48 | <.0001 |
| Freq_SQ | Freq_SQ | 1 | -1.34911 | 0.14377 | -9.38 | <.0001 |
| Crossbuy | Crossbuy | 1 | 18.18908 | 1.63568 | 11.12 | <.0001 |
| SOW | SOW | 1 | 2.07463 | 0.15796 | 13.13 | <.0001 |
| imr_acq | | 1 | 107.58570 | 44.83913 | 2.40 | 0.0170 |
| pred_dur | | 1 | 0.49555 | 0.06584 | 7.53 | <.0001 |

The SAS System

```
mean_profit      mad       random_mad
ffffffffffffffffffffffffffffffff
2403.843      752.2664     1881.398
```

# References


1. Thomas, J.S. (2001) A methodology for linking customer acquisition to customer retention. *Journal of Marketing Research*, **38**(2), 262–268.
2. Berger, P.D. and Bechwati, N.N. (2001) The allocation of promotion budget to maximize customer equity. *Omega*, **29**(1), 49–61.
3. Reinartz, W., Thomas, J.S. and Kumar, V. (2005) Balancing acquisition and retention resources to maximize customer profitability. *Journal of Marketing*, **69**(1), 63–79.
4. Blattberg, R.C. and Deighton, J. (1996) Manage marketing by the customer equity test. *Harvard Business Review*, **74**(4), 136–144.
5. Lewis, M. (2006b) The effect of shipping fees on customer acquisition, customer retention, and purchase quantities. *Journal of Retailing*, **82**(1), 13–23.

6. Lewis, M. (2006a) Customer acquisition promotions and customer asset value. *Journal of Marketing Research*, **43**(2), 195–203.
7. Bloom, E.D. and Killingsworth, R.M. (1985) Correction for truncation bias caused by a latent truncation variable. *Journal of Econometrics*, **27**(1), 131–135.
8. Schweidel, D.A., Fader, P.S., and Bradlow, E.T. (2008) A bivariate timing model of customer acquisition and retention. *Marketing Science*, **27**(5), 829–843.
9. Lee, M.L.T. (1996) Properties and applications of the Sarmanov family of bivariate distributions. *Communications in Statistics: Theory Methods*, **25**(6), 1207–1222.

# 6

# Customer churn

## 6.1 Introduction

In the previous chapters the focus has been on models that help with acquiring the most profitable customers, retaining the most profitable customers, and finally balancing marketing resources across acquisition and retention efforts to maximize profitability. While the ultimate goal of customer retention is to prevent customers from switching merchandisers and stay with the current service provider, there is always going to be customer attrition. And this customer churn can be extremely costly for firms that do not understand when customers are likely to quit the relationship or what warning signals exist that help explain why a customer is likely to quit. For instance, a firm will have to spend significantly more to acquire a new customer than to continue to retain most customers on average. Thus, it becomes financially important to try and get a clear picture of the customer churn process and manage it appropriately.

The most effective way to manage customer churn is to understand the causes or determinants of customer churning behavior, predict which customers are most likely to leave, and conduct promotions or other strategies to encourage them to stay (given that they are likely to be profitable to bring back). For this chapter of the book the focus will be on the following two questions:

- What are the drivers of customer churn?
- Given that a customer has not yet left the firm, when is the customer likely to end the relationship?

In the next chapter we will address the strategies to win the customers back by using promotions or other strategies to encourage the customers to stay. There have been many studies which have been conducted to try and answer the two key

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

Table 6.1 Review of customer churn models.

| Research interest | Specification | Estimation | Representative studies |
|---|---|---|---|
| Churn or not | Binomial logit | MLE | Kim and Yoon [1] |
| | Hierarchical logistic regression | | Capraro, Broniarczyk, and Srivastava [2] |
| | Logistic regression | | Buckinx and Van den Poel [3] |
| | | | Ahn, Han, and Lee [4] |
| | | | Brockett *et al.* [5] |
| | ARD neural network | Bayesian | Buckinx and Van den Poel [3] |
| | Random forests | — | Buckinx and Van den Poel [3] |
| | Bagging and boosting classification trees | — | Lemmens and Croux [6] |
| | Cost-sensitive classifier | — | Glady, Baesens, and Croux [7] |
| | Time series regression | MLE | Danaher [8] |
| | Hazard | | Dekimpe and Degraeve [9] |
| | | | Jamal and Bucklin [10] |
| | Proportional hazard | | Van den Poel and Larivière [11] |
| | | | Brockett *et al.* [5] |

questions listed above. Table 6.1 lists the representative studies on customer churn modeling. To provide a comprehensive understanding of how to model customer churn, we will review the issues in the studies one by one along with the related modeling techniques. We will also provide an empirical example at the end of the chapter which will demonstrate how to apply this knowledge to a representative sample of customers from a B2C firm.

Similar to customer acquisition and customer retention models, the first question that needs to be answered is whether the firm's customers purchase in a contractual versus non-contractual manner. In most instances this will determine the type of statistical model that needs to be used in order to gain any insights from the data.

### 6.1.1 Data for empirical examples

In this chapter we will be providing a description of the key modeling frameworks that attempt to answer each key research question raised at the beginning of the chapter. We will also be providing an empirical example at the end of the chapter which will show how sample data can be used to answer these key research questions. For the empirical example we provide a dataset titled 'Customer Churn.' In this dataset you will find a representative sample of 500 customers and former customers from a typical B2B firm. The data include the following variables:

| Variable | |
|---|---|
| *Customer* | *Customer number (from 1 to 500)* |
| *Duration* | *The time in days that the acquired prospect has been or was a customer, right-censored at 730 d* |
| *Censor* | *1 if the customer was still a customer at the end of the observation window, 0 otherwise* |
| *Avg_Ret_Exp* | *Average number of dollars spent on marketing efforts to try and retain that customer per month* |
| *Avg_Ret_Exp_SQ* | *Square of the average number of dollars spent on marketing efforts to try and retain that customer per month* |
| *Total_Crossbuy* | *The total number of categories the customer has purchased from during the customer's lifetime* |
| *Total_Freq* | *The total number of purchase occasions the customer had with the firm in the customer's lifetime* |
| *Total_Freq_SQ* | *The square of the total number of purchase occasions the customer had with the firm in the customer's lifetime* |
| *Industry* | *1 if the prospect is in the B2B industry, 0 otherwise* |
| *Revenue* | *Annual sales revenue of the prospect's firm (in millions of dollars)* |
| *Employees* | *Number of employees in the prospect's firm* |

These data will be used for the empirical example at the end of the chapter, which will help us to understand the drivers and likelihood of customer churn.

## 6.2 Customer churn

Researchers have a strong interest in the causes of customer churn or switching behaviors. In the mobile telephony industry, Kim and Yoon [1] adopted a binomial logit model to investigate the causes for customers switching carriers. The dependent variable of the model recorded the churning behavior (1 for switching, 0 for staying). These authors assumed that customers switch carriers because the utility of churning is greater than no churning. The utility can be expressed as

$$U_{jn} = U\left(z_{jn}, s_n\right), \quad j \in \{churn, no\ churn\} \tag{6.1}$$

where $z_{jn}$ denotes service attributes and $s_n$ denotes individual-specific characteristics, and

$$U_{jn} = V_{jn} + e_{jn}. \tag{6.2}$$

The probability of the $n$th subscriber to churn can be expressed as

$$\Pr(churn|j) = \Pr\left(U_{churn,n} > U_{no\ churn,n}\right). \tag{6.3}$$

Since the unobserved part, $e_{jn}$, could be identically and independently distributed (i.i.d.) following a logistic distribution, a binomial logit model is suitable to model the utility and the probability of churning. The model is expressed as follows:

$$P_{jn} = F\left(x'_{jn}\beta\right), \tag{6.4}$$

where $P_{jn}$ is the probability that the $n$th subscriber will switch from the $j$th carrier to another, $x'_{jn}$ is a vector of explanatory variables, and $F$ denotes the cumulative logistic distribution function. For the explanatory variables, the authors included service attributes, such as call quality and price level, and demographic and phone usage characteristics, such as income, age, and subscription duration.

Since whether to churn or not is a binary decision, logistic regression remains the popular method used. [2] adopted a hierarchical logistic regression to investigate the relationship between satisfaction, switching risk, objective and subjective knowledge, and likelihood of defection. Buckinx and Van den Poel [3] argued that, unlike in contractual settings where companies can detect when customers totally defect, in non-contractual settings customers may only display partial defection, such as decreasing some of their purchase from current shops. In a retailing industry, these authors used logistic regression to classify partial defectors and non-partial defectors. The authors included as independent variables the observed purchase behavior and customer information, including interpurchase time, frequency of purchase, monetary indicators, shopping behavior across product categories, brand purchase and promotional behavior, length of relationship, timing of shopping, mode of payment, and customer demographics. In the insurance industry, Brockett *et al.* [5] used logistic regression to estimate the probability of policy holders' total simultaneous cancellation. In the telecommunications industry, [4] adopted two binary logistic regressions to explore churn determinants, including customer dissatisfaction, switching costs and service usage, and the mediating role of customer status (active, non-use, and suspended).

Besides logistic regression, the binary classification problem of churning could be modeled by techniques from data mining and machine learning. For churn classification purposes, Buckinx and Van den Poel [3] adopted MacKay's Bayesian automatic relevance determination (ARD) neural network and random forests as proposed in Breiman [12]. These authors selected Mackay's Bayesian ARD neural network framework because it has the appealing property of providing a Bayesian hyperparameter per input variable, representing the importance of the variable. Lemmens and Croux [6] adopted bagging and boosting classification trees to predict churn in the wireless telecommunications industry. These authors evaluated the predictive accuracy of their churn model not only on the misclassification rate but also on the Gini coefficient and the top-decile lift. [7] used two cost-sensitive classifiers, the AdaCost boosting and the cost-sensitive decision tree, to predict churn in the financial services industry. These authors also compared the prediction accuracy of the proposed techniques to other classification methods, such as logistic regression, decision trees, and neural networks.

In the telecommunications industry, service usage duration is possibly correlated with customer attrition behavior and ignoring this condition will lead to biased estimation. In a market experiment, Danaher [8] developed time series analysis to model two phenomena, attrition and usage, conditional on retention. Following Hausman and Wise [13] and extending their models to multiple time periods, the author used linear regression to explain the cell phone airtime usage. The model was expressed as

$$y_{it} = X_{it}\beta + \varepsilon_{it}, \qquad i = 1, 2, \ldots, n \text{ and } t = 1, 2, \ldots, T \tag{6.5}$$

where the independent variables $X_{it}$ change with time, and the error term $\varepsilon_{it}$ is i.i.d. with an individual component $\mu_i$ and an uncorrelated time effect $\eta_{it}$. The properties of the error term are

$$\varepsilon_{it} = \mu_i = \eta_{it}, \qquad E[\varepsilon_{it}] = 0, \tag{6.6}$$

$$var(\varepsilon_{it}) = \sigma_\mu^2 + \sigma_\eta^2 = \sigma^2 \quad \text{and} \quad \varepsilon_{it} \sim N(0, \sigma^2). \tag{6.7}$$

The author considered that a decreasing trend of service usage may influence customers' intention to continue with the contract and induce faster attrition. Thus, following the approach by Hausman and Wise [13] and Winer [14], the author defined an indicator variable $a_{it}$ (1, if person $i$ remains in the trial at time $t$; 0, if the person drops out) to incorporate attrition effects that depend on usage. Then, $y_{it}$ is defined to be observed if

$$A_{it} = \alpha y_{it} + X_{it}\theta + W_{it}\gamma + \omega_{it} \geq 0, \tag{6.8}$$

where $W_{it}$ is a matrix containing variables that do not affect $y_{it}$ but do influence the probability of observing $y_{it}$. Substituting for $y_{it}$ in Equation 6.8 gives

$$A_{it} = \alpha(X_{it}\beta + \varepsilon_{it}) + X_{it}\theta + W_{it}\gamma + \omega_{it} = R_{it}\delta + \tau_{it} \tag{6.9}$$

where $R_{it} = (X_{it}, W_{it})$, $\delta = (\alpha\beta + \theta, \gamma)'$, and $\tau_{it} = \alpha\varepsilon_{it} + \omega_{it}$. The author assumed that $\omega_{it}$ is normally distributed with mean 0 and variance $\sigma_\omega^2$ and $\tau_{it} \sim N(0, \sigma_\tau^2)$. As was done by Hausman and Wise [13], the author normalized the variance of $\tau_{it}$ by setting it to 1, so that the models for the retention and attrition probabilities can be given by the probit model as

$$\begin{aligned} \Pr(a_{it} = 1) &= \Pr(R_{it}\delta + \tau_{it} \geq 0) = \Phi(R_{it}\delta) \\ \Pr(a_{it} = 0) &= 1 - \Phi(R_{it}\delta) \end{aligned} \tag{6.10}$$

where $\Phi(\cdot)$ is the standard normal distribution function. For the usage model, the author included independent variables, such as service access price and usage price, demographic variables, and dummy variables for three of the year's four quarters. For the attrition model, the author included the same independent variables as the usage model for the identifiability purpose of the Hausman and Wise model.

Since the purpose of customer churn models is to understand the occurrence and time for customer attrition, the family of survival analysis includes popular techniques for such modeling. Parametric survival models usually assume a baseline distribution of certain forms, such as the exponential or Weibull distribution. Dekimpe and Degraeve [9] applied a hazard-rate model to analyze the attrition rate of volunteers with the Belgian Red Cross. These authors assumed an exponential form for the lifetime distribution which assumes a homogeneous population and does not assume any time dependence. Following Dekimpe and Degraeve's baseline model, let $T$ denote the random duration of a volunteer with probability density function $f(t)$, cumulative distribution function $F(t)$, and hazard function $h(t)$:

$$f(t;\lambda) = \lambda\exp(-\lambda t), \tag{6.11}$$

$$F(t;\lambda) = 1 - \exp(-\lambda t); \tag{6.12}$$

$$h(t;\lambda) = \frac{f(t;\lambda)}{1 - F(t;\lambda)} = \lambda. \tag{6.13}$$

The authors stated that they knew during what months the volunteers joined and left but did not know the timing of these events within a given month. To account for the discrete nature of the data, the authors defined monthly grouping intervals $[t_{k-1}, t_k)$, $k = 1, 2, \ldots, m+1$, $t_0 = 0$, and $t_{m+1} = \infty$, and recorded quitting in duration interval $[t_{k-1}, t_k)$ as $t_k$. They addressed that the likelihood contribution of any volunteer $i$ can be expressed as

$$L_i\left(t_{i,1}, t_{i,2}\right) = \left[\frac{S\left(t_{i,1} - 1\right) - S\left(t_{i,1}\right)}{S\left(t_{i,2}\right)}\right]^{1-d_i} \left[\frac{S\left(t_{i,1} - 1\right)}{S\left(t_{i,2}\right)}\right]^{d_i}, \tag{6.14}$$

where $t_{i,1}$ equals the total number of months with the Red Cross, and $t_{i,2}$ is the number of months the volunteer had been with the Red Cross before a certain time. $S(t_{i,j}) = 1 - F(t_{i,j})$, $j = 1, 2$, is the survival function and gives the probability that volunteer $i$ stays for at least $t_{i,j}$ periods; and $d_i$ is a censoring dummy variable which equals 0 if the volunteer has left and 1 if the volunteer is still active. After substituting the expression for the survival function of the exponential distribution, the log-likelihood for a set of $N$ volunteers (who are all assumed to have the same mean quitting rate $\lambda$) is then equal to

$$LL = \sum_{i=1}^{N} (1 - d_i)\ln\left[\frac{\exp\left(-\lambda(t_{i,1} - 1)\right) - \exp\left(-\lambda t_{i,1}\right)}{\exp\left(-\lambda t_{i,2}\right)}\right] - d_i\lambda\left(t_{i,1} - 1\right) + d_i\lambda t_{i,2}, \tag{6.15}$$

which is maximized to get an estimate of λ. The covariates examined are cohort, gender, age when joining as an active volunteer, education level when joining, and seniority.

Jamal and Bucklin [10] adopted a Weibull hazard model with time-varying covariates to predict customer churn. The Weibull distribution gives a smoothly increasing function of duration time, and churn rate is shown to be an increasing function of duration time. These authors stated that the Weibull hazard probability (conditional probability of churn) of customer $i$ at duration time $t$ conditional on $i$ belonging to segment $j$ is given by

$$h_j(t|x_{it}, i \in j) = h_{0j}(t)\exp(x_{it}\beta_j) \tag{6.16}$$

where

$$h_{0j}(t) = \left(\frac{1}{\sigma_j}\right)t^{(1/\sigma_j)-1}\exp(\beta_{0j})$$

is the baseline hazard function for segment $j$ with Weibull distribution parameters $\sigma_j$ and $\beta_{0j}$, $x_{it}$ are the customer and time-specific covariates, and $\beta_j$ are the response parameters for segment $j$. The authors gave the likelihood contribution of observation $it$ for customer $i$ at time period $t$ conditional on $i$ belonging to segment $j$ as

$$L(it|i \in j) = S_j(t|x_{it}, i \in j)^{1-d_{it}} f_j(t|x_{it}, i \in j)^{d_{it}} \tag{6.17}$$

where $S$ is the survival probability, $f$ the probability density function, and $d_{it}$ an indicator variable for censoring. The unconditional likelihood contribution of customer $i$ over all the $t$ observations is expressed as

$$L(i) = \sum_{j-1}^{J} L(i|i \in j)\Pr(j), \quad j = 1, 2, \ldots, J \tag{6.18}$$

where $J$ is the number of latent segments and $\Pr(j)$ is the prior probability of segment $j$, parameterized as

$$\Pr(j) = \frac{\exp(\theta_j)}{\sum_J^{j=1}}\exp(\theta_j),$$

where $\theta_j = 0$. The model parameters were estimated using maximum likelihood estimation. The covariates examined included customer service experience, failure recovery, and payment equity.

The proportional hazards model [15] incorporates the effects of individual customer covariates but assumes a common hazard function of which each individual hazard function is a multiple. Van den Poel and Larivière [11] adopted a proportional hazards model to analyze customer attrition in the context of the European financial services industry. These authors combined three categories of predictors of churn behavior, namely, customer behavior, customer demographics, and macroenvironment, into the independent variables and allowed these variables to take different values over time. Brockett *et al.* [5] used proportional hazards models to understand sequential policy cancellations. These authors modeled the time between first cancellation notification and the final complete withdrawal by assuming that there is a baseline distribution for the time a customer will take for defection, and the relative risk of an individual customer defecting completely changes from this baseline according to his/her particular set of individual household covariates. The authors adopted the empirically based Breslow estimator [16] to estimate the parameters.

Hadden *et al.* [17] and Neslin *et al.* [18] provided an overview of the techniques used in customer churn prediction. Hadden *et al.* [17] proposed a five-stage churn management framework, including identification of the best data, data semantics, feature selection, development of predictive model, and validation of results. To develop a predictive model for customer churn, these authors reviewed traditional methods, including decision trees, neural networks, regression analysis, Bayesian network classifiers, semi-Markov processes, support vector machine, $K$-nearest neighbor (KNN), and soft computing, including fuzzy logic, evolutionary computation, artificial neural networks, probabilistic computing, and their combinations. Based on a tournament, Neslin *et al.* [18] tried to identify the most suitable methodological approaches for predicting customer churn. These authors considered methods such as logistic regression, decision trees, neural nets, discriminant analysis, cluster analysis, and Bayesian analysis.

### 6.2.1 Empirical example: Customer churn

One of the many critical questions a firm needs to answer is related to whether that firm can predict why and when a customer is likely to churn. In this example we will focus on an example from a typical B2B firm with a contractual basis with its customers. In this case the firm is actually able to observe when the customer churns from the database. However, we do observe the entire lifetime of every customer in the database. Instead we only observe whether or not the customer defected from the firm in the first two years of the relationship (or 730 days). For all customers who have not defected from the firm, we only observe a censored lifetime of 730 days. Thus, in this example we want to model the drivers of customer churn to try and understand if there is a difference between the customers who have already left the firm and the customers who have yet to leave the firm. At the end of this example we should be able to do the following:

1. Determine the drivers of customer churn.

2. Predict the expected duration of the customers who have yet to churn.

3. Determine the predictive accuracy of the model.

The information we need for this model includes the following list of variables:

| Variable | |
|---|---|
| *Customer* | *Customer number (from 1 to 500)* |
| *Duration* | *The time in days that the acquired prospect has been or was a customer, right-censored at 730 d* |
| *Censor* | *1 if the customer was still a customer at the end of the observation window, 0 otherwise* |
| *Avg_Ret_Exp* | *Average number of dollars spent on marketing efforts to try and retain that customer per month* |
| *Avg_Ret_Exp_SQ* | *Square of the average number of dollars spent on marketing efforts to try and retain that customer per month* |
| *Total_Crossbuy* | *The total number of categories the customer has purchased from during the customer's lifetime* |
| *Total_Freq* | *The total number of purchase occasions the customer had with the firm in the customer's lifetime* |
| *Total_Freq_SQ* | *The square of the total number of purchase occasions the customer had with the firm in the customer's lifetime* |
| *Industry* | *1 if the prospect is in the B2B industry, 0 otherwise* |
| *Revenue* | *Annual sales revenue of the prospect's firm (in millions of dollars)* |
| *Employees* | *Number of employees in the prospect's firm* |

In this case we will be using both *Duration* and *Censor* as our dependent variables and the remaining variables as our independent variables. Since we actually observe a customer's defection from the firm, we can choose a modeling framework with an observed dependent variable. In the case where the relationship between the customer and firm was non-contractual we would have to model the probability of customer churn as a stochastic process (see the empirical example on lifetime duration in Chapter 4 for a description of modeling this process). For this case we will choose an accelerated failure time (AFT) model. AFT models are parametric models that provide an alternative to modeling failure time data using a proportional hazards model (PHM). As a result we can model them as a linear model with the time until failure (*Duration*) transformed by the logarithmic function. We get a model of the following format:

$$\ln(Duration_i) = X_i'\beta + \sigma\varepsilon_i$$

where $\ln(Duration_i)$ is the natural logarithm of the duration of customer $i$, $X_i$ is a matrix of the time-invariant independent variables for each customer $i$, $\beta$ is a vector of parameter estimates, $\sigma$ is the estimated scale parameter, and $\varepsilon_i$ is the random disturbance term. In the case we will be estimating the values of $\beta$ and $\sigma$. We note here that if the value of $\sigma$ is 1 and there are no censored values (i.e., every customer churns during the time window) we merely have an OLS regression where the dependent variable is transformed by the natural logarithm function and the independent variables are linear. However, given the censored nature of the data, we cannot use an OLS regression. For the purpose of this example we choose to estimate the model with a log-normal distribution in *Duration*, where the estimated distribution of *Duration* can take many different forms (e.g., Weibull, log-normal, exponential, etc.). Also note here that we are not choosing the distribution of ln (*Distribution*) or $\varepsilon$. The selection of the distribution for all AFT models is always in the original time variable and not in the transformed variable or the random disturbance term. When we estimate the model we get the following results:

| Variable | Estimate | Standard error | $p >$ ChiSq |
|---|---|---|---|
| *Intercept* | 5.770 | 0.052 | < 0.0001 |
| *Avg_Ret_Exp* | 0.009 | 0.001 | < 0.0001 |
| *Avg_Ret_Exp_SQ* | −0.0002 | 0.000 01 | < 0.0001 |
| *Total_Crossbuy* | 0.098 | 0.007 | < 0.0001 |
| *Total_Freq* | 0.028 | 0.007 | < 0.0001 |
| *Total_Freq_SQ* | −0.001 | 0.000 3 | 0.0050 |
| *Industry* | −0.028 | 0.019 | 0.1372 |
| *Revenue* | 0.004 | 0.001 | < 0.0001 |
| *Employees* | 0.0004 | 0.000 01 | < 0.0001 |
| *Scale* ($\sigma$) | 0.158 | 0.007 | |

As we can see from the results, all of the independent variables with the exception of *Industry* are significant at $p < 0.05$. We find that *Avg_Ret_Exp* is positive with a diminishing return, suggesting that the higher the average monthly spending on retention efforts (to a threshold), the longer the duration before the customer is likely to churn. We find that *Total_Crossbuy* is positive, suggesting that the more categories a customer purchases, the longer the expected lifetime. We find that *Total_Freq* is positive with a diminishing return, suggesting (similar to average retention expense) that the more frequently a customer purchases (to a threshold), the longer the customer's expected lifetime. This means that customers who do not purchase very frequently are not likely to stay long and customers who purchase very frequently are likely to exhaust their need to purchase quickly and leave earlier as well. It is the customers who purchase at a moderate frequency that are likely to have the longest lifetime with the firm. We find that *Revenue* and *Employees* are positively related, the expected duration of the customer lifetime meaning that customers who have higher revenue and more employees are more likely to have a longer duration with the firm. Finally, we see a scale value $\sigma$ of 0.158. The scale value is merely an estimated value that helps to scale the random disturbance term. While for some distributions it can affect the shape of the hazard function, in this case given a log-normal distribution of *Duration*, changes in the scale parameter only serve to compress or stretch the hazard function.

It is also important to understand exactly how changes in the drivers of customer duration are likely to lead to either increases or decreases in that customer's expected duration. To do this we need to understand how to interpret the parameter coefficients from the AFT model. We see that we have a model format with a logged dependent variable of time and linear independent variables (log-linear format). As a result, we see that the ratio of survival times between the baseline and current case is the following:

$$\frac{T(X_i + \delta)}{T(X_i)} = \exp((X_i + \delta) - X_i)\beta = \exp(\beta\delta)$$

where $T(\cdot)$ is the hazard model, $X_i$ is the value of the focal independent variable for customer $i$, $\delta$ is the change in the value of the independent variable, and $\exp(\cdot)$ is the exponential function. When the change in the value of the independent variable is only 1, we see that the above function simplifies to

$$\frac{T(X_i + 1)}{T(X_i)} = \exp(\beta).$$

When we compute the ratio for each of the statistically significant variables we get the following results for an increase in 1 unit of the independent variable:

| Variable | Duration ratio |
|---|---|
| *Avg_Ret_Exp* | (0.0088−0.0004*$Avg\_Ret\_Exp$) |
| *Total_Crossbuy* | 1.103 |
| *Total_Freq* | (0.027−0.002*$Total\_Freq$) |
| *Revenue* | 1.004 |
| *Employees* | 1.0004 |

We gain the following insights from the ratios. With regard to *Avg_Ret_Exp*, we see that the ratio is dependent on the level of *Avg_Ret_Exp*. This is due to the fact that we include both the level and squared terms for *Avg_Ret_Exp*. For example, if we usually spend \$15 on a given customer per month, by spending \$16 we should see an increase in the expected duration by $\exp(0.0088-0.0004^*15) = \exp(0.0028) = 1.003$. This means that by increasing our spending from \$15 to \$16, we should see an increase in expected duration by 0.3%. Also, it is important to note that this will vary depending on the initial level of *Avg_Ret_Exp*. With regard to *Total_Crossbuy*, we see that for each increase in cross-buy by one category, the expected duration should increase by 10.3%. With regard to *Total_Freq*, we see that the ratio is dependent on the level of *Total_Freq*. This is due to the fact that we include both the level and squared terms (similar to the case for *Avg_Ret_Exp*) for *Total_Freq*. For example, if we see a customer who has purchased five times in his/her lifetime, by purchasing a sixth time we should see an increase in the expected duration by $\exp(0.027-0.002^*5) = \exp(0.017) = 1.017$. This means that when we see the customer increase total purchases from five to six, we should see an increase in expected duration by 1.7%. Again, it is important to note that this will vary depending on the initial level of *Total_Freq*. With regard to *Revenue*, we see that for each increase in *Revenue* by \$1 million the expected duration should increase by 0.4%. Finally, with regard to *Employees*, we see that for each increase in *Employees* by one person the expected duration should increase by 0.04%.

Our next step is to determine how well the model does at explaining the expected duration of a customer. Since we have not observed the churn of 268 customers, we only predict the expected duration for the 232 customers we actually observe churning from the customer database and compare the actual durations we observe to the predicted durations. We make the prediction of duration using the following equation:

$$\mathrm{E}(Duration) = \exp(X'_i\beta)$$

where $\exp(\cdot)$ is the exponential function, $X_i$ is the matrix of independent variables, and $\beta$ is the vector of estimated coefficients from the modeling exercise. Once we have the predicted duration value for each of the customers who already churned, we compute the MAD using the following formula:

$$\text{MAD} = \text{Mean}\{\text{Absolute Value}[\mathrm{E}(Duration) - Duration]\}.$$

We find that MAD = 76.7. This means that on average each of our predictions of *Duration* deviates from the actual value by about 77 days. If we were to instead use the mean value of *Duration* (502.68) as our prediction for all the customers who churned (this would be the benchmark model case), we would get MAD = 142.4, or about 142 days. As we can see, our model does a significantly better job of predicting the length of time until customer churn than the benchmark case.

While we cannot determine the predictive accuracy of our model against the actual durations of the customers who have yet to churn without waiting for them to churn, we can see if our model does a good job of classifying customers as churning within the observed time window or being censored at the end of the time window. In this case we give a value of 1 to actual churn (*Act_Ch*) when a customer has a duration of less than 730, and 0 when the customer has a censored duration of 730. We then give a value of 1 to predicted churn (*Pred_Ch*) when the prediction of the customer duration is less than 730, and 0 when the prediction of customer duration is greater than or equal to 730. We can then compare the hit rate between the predicted and actual values of churn as in the following table:

| | | Actual churn | | |
|---|---|---|---|---|
| | | 0 | 1 | *Total* |
| Predicted churn | 0 | *231* | *38* | ***269*** |
| | 1 | *37* | *194* | *231* |
| | *Total* | ***268*** | ***232*** | ***500*** |

As we can see from the table, our in-sample model accurately predicts 86.2% of the customers who have not churned (231/268) and 83.6% of the customers who already churned (194/232). This is a significant increase in the predictive capability of a random guess model[1] which would be only 53.6% accurate for this dataset. Since our model is significantly better than the best alternative, in this case a random guess model, we determine that the predictive accuracy of the model is good. If there are other benchmark models available for comparison, the 'best' model would be the one which provides the highest accuracy of both the churn and not churn, or in other words the prediction would provide the highest sum of the diagonal. In this case the sum of the diagonal is 425 and it is accurate 85.0% of the time (425/500).

### 6.2.2 How do you implement it?

In this example we used PROC LIFEREG from SAS to estimate the AFT model to explain the drivers of customer churn. While we did use SAS to implement the modeling framework, programs such as GAUSS, MATLAB, and R can be used as well.

## 6.3 Chapter summary

Customer churn can be considered a negative outcome of the customer retention process. The purpose of retention modeling is to understand the effects of marketing variables on the duration of customer lifetime, while the churn modeling is to estimate the possible causes which induce customer defection or the ending of customer lifetime duration. The modeling of churn is as simple as a probability modeling, whether customers will churn or not, and it can be estimated by a logit model. As churn modeling can also be viewed as a binary classification problem, techniques from the machine learning field have been adopted, such as neural networks, bagging and boosting classification trees, and cost-sensitive classifiers. When the churn data are in panel form, time series techniques can be adopted to correct the possible correlation of lifetime duration and customer defection behaviors. Hazard models (as used in the empirical example for this chapter) are also suitable for churn modeling and baseline hazard distribution, such as the exponential, Weibull, and log-normal, among others, have been selected in the model specification. The PHM is often used and it is able to estimate the effect of time-variant covariates on the hazard rate.

---

[1] A random guess model would do the following. First, it would determine which bucket churn or not churn has more customers (or former customers) in it. In this case 268 customers have not churned versus 232 who have churned. Then it would predict that all customers would not churn and it would be accurate 53.6% of the time (268/500).

# Customer churn – SAS code

```
/* Import Data - Library: statcrm */
proc import out = statcrm.customer_churn
   datafile = "C:\_Your_Data_Location_\Statistics in CRM\Churn
\Customer Churn.xls" dbms = excel replace;
   range = " 'Customer Churn Data$' "; getnames = yes; mixed = no;
scantext =
yes; usedate = yes; scantime = yes;
run; quit;

/* Probability of Churn */

proc lifereg data = statcrm.customer_churn;
model duration*censor(1) = avg_ret_exp avg_ret_exp_sq industry
revenue employees total_crossbuy total_freq total_freq_sq /
dist = lognormal;
output out = statcrm.churn_pred p = pred_churn xbeta =
xbeta_churn;
run; quit;

data statcrm.churn_pred1;
set statcrm.churn_pred;
ad = abs(duration - pred_churn);
ad1 = abs(502.68 - duration);
if pred_churn < 730 then pred_ch = 1; else pred_ch = 0;
if duration < 730 then act_ch = 1; else act_ch = 0;
run; quit;

/* Predictive Accuracy */

proc sql;
select mean(duration) as mean_dur, mean(ad) as mad_dur, mean
(ad1) as
mad_dur_rand
from statcrm.churn_pred1 where censor = 0; quit;

proc freq data = statcrm.churn_pred1;
table pred_ch*act_ch; run; quit;
```

## Customer churn – SAS output

```
                    The SAS System
               The LIFEREG Procedure
                  Model Information
Data Set                       STATCRM.CUSTOMER_CHURN
Dependent Variable                      Log(Duration)     Duration
Censoring Variable                             Censor     Censor
Censoring Value(s)                                  1
Number of Observations                            500
Noncensored Values                                232
Right Censored Values                             268
Left Censored Values                                0
Interval Censored Values                            0
Name of Distribution                        Lognormal
Log Likelihood                           22.045240067

                  Number of Observations Read     500
                  Number of Observations Used     500

                              Fit Statistics

         -2 Log Likelihood                            -44.090
         AIC (smaller is better)                      -24.090
         AICC (smaller is better)                     -23.641
         BIC (smaller is better)                       18.056

  Algorithm converged.

                     Type III Analysis of Effects

                                      Wald
       Effect              DF     Chi-Square      Pr > ChiSq

       Avg_Ret_Exp         1        110.2181        <.0001
       Avg_Ret_Exp_SQ      1        734.4065        <.0001
       Industry            1          2.2088        0.1372
       Revenue             1         38.5267        <.0001
       Employees           1        295.8348        <.0001
       Total_Crossbuy      1        217.4082        <.0001
       Total_Freq          1         16.0197        <.0001
       Total_Freq_SQ       1          7.8647        0.0050

             Analysis of Maximum Likelihood Parameter Estimates

                                Standard    95% Confidence    Chi-
Parameter      DF   Estimate     Error       Limits       Square Pr >   ChiSq

Intercept       1     5.7700    0.0522    5.6677    5.8723   12213.0   <.0001
Avg_Ret_Exp     1     0.0089    0.0008    0.0072    0.0106    110.22   <.0001
```

```
                    The SAS System
                  The LIFEREG Procedure
         Analysis of Maximum Likelihood Parameter Estimates

                              Standard  95% Confidence   Chi-
Parameter       DF   Estimate    Error     Limits      Square Pr >   ChiSq

Avg_Ret_Exp_SQ  1    -0.0002   0.0000  -0.0002  -0.0002  734.41  <.0001
Industry        1    -0.0283   0.0191  -0.0657   0.0090    2.21  0.1372
Revenue         1     0.0035   0.0006   0.0024   0.0046   38.53  <.0001
Employees       1     0.0004   0.0000   0.0004   0.0005  295.83  <.0001
Total_Crossbuy  1     0.0977   0.0066   0.0847   0.1107  217.41  <.0001
Total_Freq      1     0.0275   0.0069   0.0140   0.0409   16.02  <.0001
Total_Freq_SQ   1    -0.0009   0.0003  -0.0015  -0.0003    7.86  0.0050
Scale           1     0.1582   0.0074   0.1444   0.1733

                    The SAS System

                mean_dur   mad_dur   mad_dur_rand

                ffffffffffffffffffffffffffffffff
                502.681  76.6954   142.3593

                    The SAS System
                   The FREQ Procedure
                 Table of pred_ch by act_ch

               pred_ch   act_ch

               Frequency,
               Percent ,
               Row Pct ,
               Col Pct ,    0,    1, Total
               ffffffff^fffffff^fffffff^
                  0 ,    231 ,      38 ,  269
                    ,  46.20 ,    7.60 , 53.80
                    ,  85.87 ,   14.13 ,
                    ,  86.19 ,   16.38 ,
               ffffffff^fffffff^fffffff^
                  1 ,     37 ,    194 ,     231
                    ,   7.40 ,  38.80 ,   46.20
                    ,  16.02 ,  83.98 ,
                    ,  13.81 ,  83.62 ,
               ffffffff^fffffff^fffffff^
               Total      268      232      500
                        53.60    46.40   100.00
```

# References


1. Kim, H.-S. and Yoon, C.-H. (2004) Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy*, **28**(9–10), 751–765.
2. Capraro, A., Broniarczyk, S., and Srivastava, R. (2003) Factors influencing the likelihood of customer defection: the role of consumer knowledge. *Journal of the Academy of Marketing Science*, **31**(2), 164–175.

3. Buckinx, W. and Van den Poel, D. (2005) Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, **164**(1), 252–268.
4. Ahn, J.-H., Han, S.-P., and Lee, Y.-S. (2006) Customer churn analysis: churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy*, **30**(10–11), 552–568.
5. Brockett, P.L., Golden, L.L., Guillen, M. *et al.* (2008) Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection? *Journal of Risk and Insurance*, **75**(3), 713–737.
6. Lemmens, A. and Croux, C. (2006) Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, **43**(2), 276–286.
7. Glady, N., Baesens, B., and Croux, C. (2009) A modified Pareto/NBD approach for predicting customer lifetime value. *Expert Systems with Applications*, **36**(2), 2062–2071.
8. Danaher, P.J. (2002) Optimal pricing of new subscription services: analysis of a market experiment. *Marketing Science*, **21**(2), 119–138.
9. Dekimpe, M.G. and Degraeve, Z. (1997) The attrition of volunteers. *European Journal of Operational Research*, **98**(1), 37–51.
10. Jamal, Z. and Bucklin, R.E. (2006) Improving the diagnosis and prediction of customer churn: a heterogeneous hazard modeling approach. *Journal of Interactive Marketing*, **20**(3–4), 16–29.
11. Van den Poel, D. and Larivière, B. (2004) Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, **157**(1), 196–217.
12. Breiman, L. (2001) Random forests. *Machine Learning*, **45**(1), 5–32.
13. Hausman, A.J. and Wise, A.D. (1979) Attrition bias in experimental and panel data: the Gary income maintenance experiment. *Econometrica*, **47**(2), 455–473.
14. Winer, S.R. (1983) Attrition bias in econometric models estimated with panel data. *Journal of Marketing Research*, **20**(2), 177–186.
15. Cox, R.D. (1972) Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological*), **34**(2), 187–220.
16. Breslow, N. (1974) Covariance analysis of censored survival data. *Biometrics*, **30**(1), 89–99.
17. Hadden, J., Tiwari, A., Roy, R., and Ruta, D. (2005) Computer assisted customer churn management: state-of-the-art and future trends. *Computers & Operations Research*, **34**(10), 2902–2917.
18. Neslin, S.A., Gupta, S., Kamakura, W. *et al.* (2006) Defection detection: measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, **43**(2), 204–211.

# 7

# Customer win-back

## 7.1 Introduction

Managers usually think customer churn is the end of a customer's life cycle. But it does not have to be the end of the life cycle. Companies can still win the lost customers back and give them a second life. Unlike new customers, lost customers have certain knowledge about the products and services of the company and have their own judgment on the attributes and functions of the products and services. On the one hand, it is easier to approach lost customers since they are familiar with the company. On the other hand, lost customers often switch merchandisers because they are not satisfied with the product, so it is not necessarily easy to change their attitude and persuade them to come back. We also have to consider whether it is worth trying to bring the customers back to the firm – not all customers are worth chasing after they leave the firm.

Once we determine when a customer is likely to churn (see Chapter 6 for a detailed analysis on determining the probability of churn), we have to decide whether it is in the firm's best interest to attempt to win the customer back. So, there are several key questions that should be answered. These include:

- Should we intervene with the customer churn?
- How should we approach the customers to win them back?

The answers to each of these questions will drive the optimal customer win-back strategy. There have been a few research studies in the area, although research in customer win-back is still scarce. We review the representative studies of customer win-back here in Table 7.1. To provide a comprehensive understanding of how to model customer win-back, we will review the issues in the studies one by one along with the related modeling techniques. We will also provide an empirical

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

Table 7.1 Review of customer win-back models.

| Research interest | Specification | Estimation | Representative studies |
|---|---|---|---|
| Win-back management | Conceptual model and case study | N/A | Strauss and Friege [1] |
| | Split hazard | Bayesian MCMC | Thomas, Blattberg, and Fox [2] |
| | Quasi-experimental design | ANOVA | Tokman, Davis, and Lemon [3] |

example at the end of the chapter which will demonstrate how to apply this knowledge to a representative sample of customers from a B2C firm.

Similar to customer acquisition and customer retention models, the first question that needs to be answered is whether the firm's customers purchase in a contractual versus non-contractual manner. In most instances this will determine the type of statistical model that needs to be used in order to gain any insights from the data.

### 7.1.1 Data for empirical examples

In this chapter we will be providing a description of the key modeling frameworks that attempt to answer each key research question raised at the beginning of the chapter. We will also be providing an empirical example at the end of the chapter which will show how sample data can be used to answer these key research questions. For the empirical example we provide a dataset titled 'Customer Win-back.' In this dataset you will find a representative sample of 500 customers and former customers from a typical B2C firm. The data include the following variables:

| Variable | |
|---|---|
| *Customer* | *Customer number (from 1 to 500)* |
| *Reacquire* | *1 if the customer is reacquired, 0 if not* |
| *Duration_2* | *Time in days of the customer's second life cycle with the company, 0 if not reacquired* |
| *SCLV* | *The CLV of the customer in the second life cycle* |
| *Duration_1* | *Time in days of the customer's first life cycle with the company* |
| *Offer* | *The value of the offer provided to the customer for reacquisition* |
| *Duration_lapse* | *Time in days since the customer was lost to when the offer to reacquire was given* |
| *Price_Change* | *The increase (or decrease) in price of the subscription the customer received between the first life cycle and the second life cycle, 0 if not reacquired* |
| *Gender* | *1 if male, 0 if female* |
| *Age* | *Age in years of the customer at the time of the attempt to reacquire* |

These data will be used for the empirical example at the end of the chapter, which will help us to understand the drivers and likelihood of customer win-back.

## 7.2 Customer win-back

Strauss and Friege [1] provided a conceptual framework for lost customers' regain management which consists of analysis, actions, and controlling. To determine which customers are worthy of regaining, these authors proposed the second lifetime value (SLTV). In the controlling process, they proposed to maximize the regain profit (RP) function

$$RP = \sum_{i=1}^{n} SLTV_i * L_i * rr_i(RC_i) - \sum_{i=1}^{n} cc_i * L_i - \sum_{i=1}^{n} uc_i * Lp_i - \sum_{i=1}^{n} sc_i * Lu_i \quad (7.1)$$

where $L$ is the number of lost customers attempted to regain for segment $i$, $rr(RC_i)$ is the regain ratio which is a function influenced by the regain cost, and $\sum_{i=1}^{n} cc_i * L_i$, $\sum_{i=1}^{n} uc_i * Lp_i$, and $\sum_{i=1}^{n} sc_i * Lu_i$ are various costs occurred for regaining lapsed customers.

Thomas *et al.* [2] investigated the best price strategy for reacquisition of lapsed customers. Different from the proportional hazards model, the proposed model used in Thomas *et al.* [2] is called 'split hazard model.' These authors focused on the second life of lapsed customers and linked the reacquisition and duration of second life together. The authors specified a probit model for the probability of reacquisition as

$$z_i = \begin{cases} 1 & \text{if } z_i^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (7.2)$$

where the latent variable $z_i^*$ is modeled with a linear specification as

$$z_i^* = w_i'\gamma + \eta_i. \quad (7.3)$$

To examine the effects of price changes on the probability of relationship termination, the authors adopted the continuous Markov process assumption that the probability of the customer terminating the relationship at any point in time is independent of the current duration. They separated a customer's second lifetime into many 'subspells' and assumed that the duration of a subspell does not depend on the length of a prior subspell. The authors modeled the duration for customer $i$ during subspell $s_i$ ($s_i = 1, \ldots, S_i$) with a conditional regression

$$y_{is_i} = \begin{cases} y_{is_i}^* & \text{if } y_{is_i} < c_{is_i} \\ c_{is_i} & \text{otherwise} \end{cases} \quad (7.4)$$

where $y^*_{is_i}$ is the latent duration of relationship and $c_{is_i}$ is the censoring value. They specified the latent duration of subspell $s_i$ of customer $i$ as

$$\ln\left(y^*_{is_i}\right) = x'_{is_i}\beta + \varepsilon_{is_i}. \tag{7.5}$$

To link reacquisition and duration of second lifetime together, the authors specified the errors of the two models, respectively, as

$$\eta_i = \iota_i + \Psi_i \tag{7.6}$$

$$\varepsilon_{is_i} = \alpha_i + \xi_{is_i} \tag{7.7}$$

where $\Psi_i \sim N(0, \sigma^2_\Psi)$, $\xi_{is_i} \sim N(0,\ \sigma^2_\xi)$, and $\iota_i$ and $\alpha_i$ represent customer-specific preferences. The authors also allowed the distributions of customer-specific preferences to be correlated so that $\theta_i \sim BVN(\bar{\theta},\ \Sigma_\theta)$, where

$$\theta_i = \begin{pmatrix} \iota_i \\ \alpha_i \end{pmatrix}, \qquad \bar{\theta} = \begin{pmatrix} \bar{\iota} \\ \bar{\alpha} \end{pmatrix}, \quad \text{and} \quad \Sigma_\theta = \begin{pmatrix} \sigma^2_i & \sigma_{\iota\alpha} \\ \sigma_{\iota\alpha} & \sigma^2_\alpha \end{pmatrix}. \tag{7.8}$$

In this way, customers' preference to be reacquired is correlated with customers' preference for the duration of the relationship. In summary, the authors specified the error variances of Equations 7.3 and 7.5 as:

$$Var(\eta_i) = \sigma^2_\Psi + \sigma^2_i + \sigma_{\iota\alpha} \tag{7.9}$$

$$Var(\varepsilon_{is}) = \sigma^2_\xi + \sigma^2_\alpha + \sigma_{\iota\alpha}. \tag{7.10}$$

The authors estimated the variance component specification in a Bayesian framework by using Markov chain Monte Carlo methods.

Customer win-back is still a new field and needs further research. Intuitively, customer win-back concerns the reacquisition probability and the duration of customer second lifetime. Researchers should consider developing suitable models to link the two processes together. In addition, most companies consider lapsed customers dead and do not store any past information of these customers, which presents a big challenge for customer win-back modeling.

### 7.2.1 Empirical example: Customer win-back

In order to understand whether we should try to win back a lost customer, we need to develop a set of models which describe the reacquisition and second customer life cycle process. This will involve three different models: the reacquisition model, second duration model, and second CLV model. Once we have the results from the three models, we will have a better understanding of how to reacquire customers and which lost customers are worth reacquisition. The three models we will estimate are the following:

$$y_{L_i} = \begin{cases} \beta_L' x_{Li} + \gamma' y_{Di} + \varepsilon_{Li} & \text{if } z_i = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{(second CLV equation)} \tag{7.11}$$

$$y_{Di} = \begin{cases} \beta_D' x_{Di} + \varepsilon_{Di} & \text{if } z_i = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{(second duration equation)} \tag{7.12}$$

$$\begin{aligned} z_i^* &= \alpha' v_i + \mu_i \quad \text{(reacquisition equation)} \\ z_i &= \begin{cases} 1 & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \leq 0 \end{cases} \end{aligned} \tag{7.13}$$

where $z_i^*$ is a latent variable indicating customer $i$'s utility to engage in a second relationship with a firm, $z_i$ an indicator variable showing whether customer $i$ is reacquired ($z_i = 1$) or not ($z_i = 0$), $v_i$ a vector of covariates affecting the reacquisition of customer $i$, $y_{Di}$ the second duration of customer $i$'s relationship with the firm, $x_{Di}$ a vector of covariates affecting the second duration of customer $i$'s relationship with the firm, $y_{L_i}$ the second CLV of customer $i$, $x_{Li}$ a vector of covariates affecting customer $i$'s second CLV, $\alpha$, $\beta_L$, $\beta_D$ a vector of parameters, and $\mu_i$, $\varepsilon_{Di}$, and $\varepsilon_{Li}$ error terms. Given that the modeling framework is recursive in nature, we can estimate it in a stepwise fashion. Thus, we will proceed in the following manner. There will be three subsections (similar to the estimation in Chapter 5) in this empirical example. The first subsection will describe and estimate the reacquisition model, the second will describe and estimate the second duration model, and the third will describe and estimate the second CLV model.

#### 7.2.1.1 Reacquisition model

The key question we want to answer with regard to customer acquisition here is whether we can determine which future prospects have the highest likelihood of reacquisition. To do this we first need to know which previously lost customers were reacquired and which were not. In the dataset provided for this chapter we have a binary variable which identifies whether or not a previously lost customer was reacquired by the firm (and hence became a customer again) and a set of drivers which are likely to help explain a customer's decision to rejoin the firm. A random sample of 500 previously lost customers (some of whom became customers again) was taken from a B2C firm. The information we need for our reacquisition model includes the following list of variables:

| | |
|---|---|
| **Dependent variable** | |
| *Reacquire* | *1 if the customer is reacquired, 0 if not* |
| **Independent variables** | |
| *Duration_1* | *Time in days of the customer's first life cycle with the company* |
| *Offer* | *The value of the offer provided to the customer for reacquisition* |
| *Duration_lapse* | *Time in days since the customer was lost to when the offer to reacquire was given* |
| *Price_Change* | *The increase (or decrease) in price of the subscription the customer received between the first life cycle and the second life cycle, 0 if not reacquired* |
| *Gender* | *1 if male, 0 if female* |
| *Age* | *Age in years of the customer at the time of the attempt to reacquire* |

In this case, we have a binary dependent variable (*Reacquire*) which tells us whether the prospect did join again (= 1) or did not (= 0). We also have six independent variables that we believe will be drivers of reacquisition. First, we have how long the customer lasted in the relationship the first time around (*Duration_1*). Second, we have the value of the offer that the firm provided the customer to entice the lapsed customer to readopt (*Offer*). Third, we have the time since the customer left the relationship with the firm to the time of the offer (*Duration_lapse*). Fourth, we have the change in subscription price of the new subscription in the second lifetime from the subscription price in the first lifetime (*Price_Change*). Finally, we have two variables which describe the lapsed customer's demographics. These include both the *Gender* and *Age* of the customer.

First, we need to model the probability that a prospect will be reacquired. Since our dependent variable (*Reacquire*) is binary and we need an error structure that is similar to the second duration and second CLV models (both normally distributed), we select a probit regression for this model. Choosing a logistic regression would require us to transform the model output before integrating the results with the other two equations. In this case the $y$ variable is *Reacquire* and the $x$ variables represent the nine independent variables in our database. When we run the probit regression we get the following result:

| Variable | Estimate | Standard error | $p$-value |
|---|---|---|---|
| *Intercept* | −1.499 | 0.663 | 0.0237 |
| *Duration_1* | 0.009 | 0.001 | < 0.0001 |
| *Offer* | 0.157 | 0.026 | < 0.0001 |
| *Duration_lapse* | −0.033 | 0.004 | < 0.0001 |
| *Gender* | −0.206 | 0.195 | 0.2900 |
| *Age* | −0.039 | 0.007 | < 0.0001 |

As we can see from the results, five of the six independent variables are significant at a $p$-value of 5% or better – with the only non-significant variable being

*Gender*. First, this means that the longer the customer's initial relationship with the firm (*Duration_1*), the higher the likelihood of reacquisition. Second, the results show that the higher the *Offer* made for the former customer to rejoin, the more likely the reacquisition. Third, the results show that the longer the time since the first customer relationship lapsed (*Duration_Lapse*), the less likely the customer will be reacquired. Finally, the results show that the older the customer (*Age*), the less likely the customer will be reacquired.

Now that we have determined the drivers of customer reacquisition we need to use the output of the model to determine our model's predictive accuracy. To do this we need to use the estimates we obtained from the reacquisition model to help us determine the predicted probability that each customer will be reacquired. We use the parameter estimates from the reacquisition model and values for the $x$ variables for each customer to predict whether a customer is likely to be reacquired. For a probit regression we must apply the proper probability function

$$P(Acquisition = 1|X\beta) = \frac{1}{2}\left[1 + erf\left(\frac{X\beta - \mu}{\sqrt{2\sigma^2}}\right)\right]$$

where $X$ is the matrix of variables, $\beta$ is the vector of coefficients, $\mu$ is the mean of the error distribution (in this case 0), $\sigma$ is the standard deviation of the error distribution (in this case 1 since it is a standard normal distribution), and *erf* is the error function which is equal to

$$erf(x) = \frac{2}{\pi}\int_0^X \exp(-t^2)\, dt.$$

Once we compute the probability of reacquisition, we need to create a cutoff value to determine at which point we are going to divide the customers into the two groups – predicted to reacquire and predicted not to reacquire. There is no rule that explicitly tells us what that cutoff number should be. Often by default we select 0.5 since it is equidistant from 0 and 1. However, it is also reasonable to check multiple cutoff values and choose the one that provides the best predictive accuracy for the dataset. By using 0.5 as the cutoff for our example, any customer whose predicted probability of reacquisition is greater than or equal to 0.5 is classified as predicted to be reacquired and the rest are predicted not to be reacquired. To determine the predictive accuracy we compare the predicted to the actual reacquisition values in a $2 \times 2$ table. For our sample of 500 customers we get the following table:

| | | Actual reacquisition | | |
|---|---|---|---|---|
| | | 0 | 1 | *Total* |
| Predicted | 0 | *172* | *35* | ***207*** |
| reacquisition | 1 | *33* | *260* | ***293*** |
| | *Total* | ***205*** | ***295*** | ***500*** |

As we can see from the table, our in-sample model accurately predicts 83.9% of the customers who chose not to readopt (172/205) and 88.1% of the customers who chose to readopt (260/295). For the prediction of customers who did chose to reacquire the product and for the prediction of customers who chose not to reacquire the product, this is a significant increase in the predictive capability of a random guess model[1] which would be only 59% accurate for this dataset. To determine overall model prediction performance we look at the diagonal and see that overall our prediction accuracy is 86.4% (432/500). Given that the model in general predicts better than the random guess model, we would determine that the model prediction is good.

As a result we now know how analyzing a customer's past lifetime duration, the time since that customer disadopted, the level of the offer we provide the customer to incentivize readoption, and customer characteristics are likely to either increase or decrease the likelihood of readoption. And we also know that these drivers do a good job in helping us predict whether a customer is going to readopt or not.

#### 7.2.1.2 Second duration model

The second step of this process is to estimate the second duration model. The purpose of this model is to understand the drivers that describe the length of time a customer is likely to be a customer for the second time, conditional on the fact that readoption occurred. Thus the equation takes the following format:

$$\mathrm{E}(\mathit{Duration\_2}) = \mathrm{P}(\mathit{Reacquisition} = 1) * \mathrm{E}(\mathit{Duration\_2} | \mathit{Reacquisition} = 1).$$

This equation shows us that the expected second duration is a function of the probability that the customer is reacquired multiplied by the expected value of second duration given that the customer was reacquired. If we were to merely run a regression with *Duration_2* as the dependent variable and ignore the probability that the customer will readopt, we would get biased estimates due to a potential sample selection bias.

Sample selection bias is a problem that is common in many marketing problems and has to be statistically accounted for in many modeling frameworks. In this case the customer has a choice as to whether or not to reacquire the product before deciding how long the second relationship will last. If we were to ignore this choice we would bias the estimates from the model and we would have less precise predictions for the value of *Duration_2*. To account for this issue we need to be able to predict the value of both the probability of *Reacquisition* (what we did in the first step of this example) and the expected value of *Duration_2* given that the customer is expected to readopt. To account for this issue we use a two-stage modeling

[1] A random guess model would do the following. First, it would determine which bucket reacquire or not reacquire has more customers in it. In this case 295 customers readopted versus 205 who did not. Then it would predict that all customers would readopt and it would be accurate 59% of the time (295/500).

framework similar to that described earlier in this chapter and found in Reinartz *et al.* [4].

We will use the output and predictions of the probit model from the first step of this example to create a new variable, $\lambda$, which will represent the correlation in the error structure across the two equations. This variable, also known as the sample selection correction variable, will then be used as an independent variable in the *Duration_2* model to remove the sample selection bias in the estimates. To compute $\lambda$ we use the following equation, also known as the inverse Mills ratio:

$$\lambda = \frac{\phi(X'\beta)}{\Phi(X'\beta)}.$$

In this equation $\phi$ represents the normal probability density function, $\Phi$ represents the normal cumulative density function, $X$ represents the value of the variables in the reacquisition model, and $\beta$ represents the coefficients derived from the estimation of the reacquisition model.

Finally, we want to estimate a regression model for *Duration_2* and include the variable $\lambda$ as an additional independent variable. This is done in a straightforward manner using the following equation:

$$Duration\_2 = \gamma'\alpha + \mu\lambda + \varepsilon.$$

In this case *Duration_2* is the value of the second duration, $\gamma$ is the matrix of variables used to help explain the value of *Duration_2*, $\alpha$ are the coefficients for the independent variables, $\mu$ is the coefficient on the inverse Mills ratio, $\lambda$ is the inverse Mills ratio, and $\varepsilon$ is the error term. Thus, for this example we will use the following list of variables:

| **Dependent variable** | |
|---|---|
| *Duration_2* | *Time in days of the customer's second life cycle with the company, 0 if not reacquired* |
| **Independent variables** | |
| *Duration_1* | *Time in days of the customer's first life cycle with the company* |
| *Offer* | *The value of the offer provided to the customer for reacquisition* |
| *Duration_lapse* | *Time in days since the customer was lost to when the offer to reacquire was given* |
| *Price_Change* | *The increase (decrease) in price of the subscription the customer received between the first life cycle and the second life cycle, 0 if not reacquired* |
| *Gender* | *1 if male, 0 if female* |
| *Age* | *Age in years of the customer at the time of the attempt to reacquire* |
| *Lambda* ($\lambda$) | *The computed inverse Mills ratio from the reacquisition model* |

When we estimate the second-stage of the model, we get the following parameter estimates from the second of the two equations (the parameter estimates for the reacquisition model are detailed in the first part of this example):

| Variable | Estimate | Standard error | *p*-value |
|---|---|---|---|
| *Intercept* | 152.638 | 13.376 | $< 0.0001$ |
| *Duration_1* | 1.212 | 0.011 | $< 0.0001$ |
| *Offer* | 2.351 | 0.452 | $< 0.0001$ |
| *Duration_lapse* | −0.972 | 0.054 | $< 0.0001$ |
| *Price_Change* | −7.838 | 0.116 | $< 0.0001$ |
| *Gender* | 10.335 | 3.662 | 0.0051 |
| *Age* | −2.506 | 0.127 | $< 0.0001$ |
| *Lambda* ($\lambda$) | 19.138 | 6.823 | 0.0054 |

We gain the following insights from the results. We see that $\lambda$ is positive and significant. We can interpret this to mean that there is a potential selection bias problem since the error term of our selection equation is correlated positively with the error term of our regression equation. We also see that all other variables of the *Duration_2* model are significant, meaning that we have likely uncovered many of the drivers of second duration.

We find that *Duration_1* is positive, suggesting that the longer the customer's first lifetime with the company, the longer the second lifetime the customer will have with the company. We find that *Offer* is positive, suggesting that the higher the offer amount (i.e., the greater the incentive), the longer the second lifetime duration. We find that the longer the time since the customer disadopted from the first relationship with the company (*Duration_lapse*), the shorter the second lifetime with the company. We find that when the price the customer pays for the product in the second lifetime is lower ($Price\_Change < 0$), the second lifetime duration of the customer is longer. Finally, we find that male customers or younger customers are more likely to have a longer second lifetime than female customers or older customers.

Our next step is to predict the value of *Duration_2* to see how well our model compares to the actual values. We do this by starting with the equation for expected duration at the beginning of this example:

$$\begin{aligned} \mathrm{E}(Duration\_2) &= \mathrm{P}(Reacquisition = 1) * \mathrm{E}(Duration\_2 | Reacquisition = 1) \\ &= \Phi(X'\beta) * (\gamma'\alpha + \mu\lambda). \end{aligned}$$

In this case $\Phi$ is the normal CDF distribution, $X$ is the matrix of independent variable values from the *Reacquisition* equation, $\beta$ is the vector of parameter estimates from the *Reacquisition* equation, $\gamma$ is the matrix of independent variables from the *Duration_2* equation, $\alpha$ is the vector of parameter estimates from the *Duration_2* equation, $\mu$ is the parameter estimate for the inverse Mills ratio, and $\lambda$ is the inverse

Mills ratio. Once we have predicted the *Duration_2* value for each of the customers (both those we reacquired and those we did not reacquire) we want to compare this to the actual value from the database. We do this by computing the MAD. The equation is as follows:

$$\text{MAD} = \text{Mean}\{\text{Absolute Value}[\text{E}(\textit{Duration\_2}) - \textit{Duration\_2}]\}.$$

We find for all customers that MAD = 67.88. This means that on average each of our predictions of *Duration_2* deviates from the actual value by about 68 days. If we were to instead use the mean value of *Duration_2* (394.61) across all customers as our prediction for all second lifetimes (this would be the benchmark model case), we would find that MAD = 353.01, or about 353 days. As we can see, our model does a significantly better job of predicting the length of the customer relationship than the benchmark case.

### 7.2.1.3 SCLV Model

The third step of this process is to estimate the SCLV model. The purpose of this model is to understand the drivers that describe the expected value of the customer's second lifetime value. Thus the equation takes the following format:

$$\text{E}(\textit{SCLV}) = \text{P}(\textit{Reacquisition} = 1) * \text{E}(\textit{SCLV} \mid \textit{Reacquisition} = 1, E(\textit{Duration\_2})).$$

This equation shows us that the expected *SCLV* is a function of the probability that the customer is reacquired multiplied by the expected value of SCLV given that the customer was reacquired and the estimated second duration of the customer's relationship with the firm. Again, if we were to merely run a regression with *SCLV* as the dependent variable and ignore the probability that the customer will be reacquired and the estimated second duration, we would get biased estimates due to a potential sample selection bias.

Thus, we will use the $\lambda$ variable as an additional variable in the model, which is computed using the following equation:

$$\lambda = \frac{\phi(X'\beta)}{\Phi(X'\beta)}.$$

In this equation $\phi$ represents the normal PDF, $\Phi$ represents the normal CDF, $X$ represents the value of the variables in the reacquisition model, and $\beta$ represents the coefficients derived from the estimation of the reacquisition model.

We will also use the expected value of *Duration_2* from the second step of this example in our SCLV model. The expected value of *Duration_2* is merely computed as

$$\text{E}(\textit{Duration\_2}) = \text{P}(\textit{Reacquisition} = 1) * \text{E}(\textit{Duration\_2} \mid \textit{Reacquisition} = 1).$$

Finally, we want to estimate a regression model for *SCLV* and include the variables $\lambda$ and E(*Duration_2*) as additional independent variables. This is done in a straightforward manner using the following equation:

$$SCLV = \gamma'\alpha + \mu\lambda + \rho Dur\hat{a}tion\_2 + \varepsilon.$$

In this case *SCLV* is the value of the second lifetime, $\gamma$ is the matrix of variables used to help explain the value of *SCLV*, $\alpha$ are the coefficients for the independent variables, $\mu$ is the coefficient on the inverse Mills ratio, $\lambda$ is the inverse Mills ratio, $\rho$ is the coefficient on the expected second duration, *Durâtion_2* is the expected second duration, and $\varepsilon$ is the error term. Thus, for this example we will use the following list of variables:

| **Dependent variable** | |
|---|---|
| *SCLV* | *The CLV of the customer in the second life cycle* |
| **Independent variables** | |
| *Duration_1* | *Time in days of the customer's first life cycle with the company* |
| *Offer* | *The value of the offer provided to the customer for reacquisition* |
| *Price_Change* | *The increase (or decrease) in price of the subscription the customer received between the first life cycle and the second life cycle, 0 if not reacquired* |
| *Gender* | *1 if male, 0 if female* |
| *Age* | *Age in years of the customer at the time of the attempt to reacquire* |
| *Lambda* ($\lambda$) | *The computed inverse Mills ratio from the reacquisition model* |
| *Durâtion_2* | *The expected number of days the customer will be with the firm for the second lifetime* |

When we estimate the third stage of the model, we get the following parameter estimates (the parameter estimates for the reacquisition model are detailed in the first part of this example and the parameter estimates of the second duration model are detailed in the second part):

| Variable | Estimate | Standard error | *p*-value |
|---|---|---|---|
| *Intercept* | 337.821 | 27.503 | < 0.0001 |
| *Duration_1* | 0.932 | 0.060 | < 0.0001 |
| *Offer* | 1.625 | 0.809 | 0.0454 |
| *Price_Change* | −7.730 | 0.442 | < 0.0001 |
| *Gender* | 14.159 | 7.014 | 0.0445 |
| *Age* | −2.155 | 0.254 | < 0.0001 |
| *Lambda* ($\lambda$) | 124.053 | 18.404 | < 0.0001 |
| *Durâtion_2* | 0.701 | 0.055 | < 0.0001 |

We gain the following insights from the results. We see that $\lambda$ is positive and significant. We can interpret this to mean that there is a potential selection bias problem since the error term of our selection equation is correlated positively with the error term of our regression equation. We also see that all other variables of the *SCLV* model are significant, meaning that we have likely uncovered many of the drivers of second customer lifetime value.

We find that *Duration_1* is positive, suggesting that the longer the duration of the first relationship, the higher the expected *SCLV*. We find that the higher the incentive provided to the customer for reacquisition (*Offer*), the higher the expected *SCLV*. We find that when the price the customer pays for the product in the second lifetime is lower (*Price_Change* $< 0$), *SCLV* is higher. We find that customers who are male or who are younger are more likely to have a higher *SCLV*. Finally, we find the coefficient on expected second duration to be positive, suggesting that customers who are in the relationship longer the second time around are more likely to profitable.

Our next step is to predict the value of *SCLV* to see how well our model compares to the actual values. We do this by starting with the equation for expected *SCLV* at the beginning of this example:

$$\begin{aligned}\mathrm{E}(SCLV) &= \mathrm{P}(Reacquisition = 1) * \mathrm{E}(SCLV \mid Reacquisition = 1, E(Duration\_2)) \\ &= \Phi(X'\beta) * (\gamma'\alpha + \mu\lambda + \rho Dur\hat{a}tion\_2 + \varepsilon).\end{aligned}$$

In this case $\Phi$ is the normal CDF distribution, $X$ is the matrix of independent variable values from the *Reacquisition* equation, $\beta$ is the vector of parameter estimates from the *Reacquisition* equation, $\gamma$ is the matrix of independent variables from the *SCLV* equation, $\alpha$ is the vector of parameter estimates from the *SCLV* equation, $\mu$ is the parameter estimate for the inverse Mills ratio, $\lambda$ is the inverse Mills ratio, $\rho$ is the coefficient on the expected second duration, and $Dur\hat{a}tion\_2$ is the expected second duration. Once we have predicted the *SCLV* value for each of the customers we want to compare this to the actual value from the database. We do this by computing the MAD. The equation is as follows:

$$\mathrm{MAD} = \mathrm{Mean}\{\text{Absolute Value}[\mathrm{E}(SCLV) - SCLV]\}.$$

We find for all the customers that MAD $= 140.59$. This means that on average each of our predictions of *SCLV* deviates from the actual value by about \$140.59. If we were to instead use the mean value of *SCLV* (\$730.15) across all customers as our prediction (this would be the benchmark model case), we would find that MAD $= 630.23$, or \$630.23. As we can see, our model does a significantly better job of predicting the expected profit of customers than the benchmark case.

### 7.2.2 How do you implement it?

For this empirical exercise several different methods were used. First, to estimate the probit regression for the reacquisition model we used PROC Logistic in SAS with the probit link function. Second, to estimate the censored regression for both the second duration model and the SCLV model in the second and third steps we used PROC Reg in SAS. There are numerous other programs such as MATLAB, GAUSS, and R which could be used to estimate these models.

## 7.3 Chapter summary

The purpose of this chapter was to explore the current models for customer win-back and provide an empirical example as to how firms can apply this knowledge to their own customer databases. We have shown that when firms are able to first understand the drivers of customer reacquisition, customer duration in a second lifetime, and second lifetime customer profitability, the firms can make more strategic decisions about which customers to try and win back and what level of offers to provide those customers to maximize customer profitability.

# Customer win-back – SAS code

```
/* Import Data - Library: statcrm */
proc import out = statcrm.winback
    datafile = "C:\_Your_Data_Location_\Statistics in CRM\Customer Win-
back\Customer Win-back.xls" dbms = excel replace;
  range = "Data$"; getnames = yes; mixed = no; scantext = yes; usedate =
yes; scantime = yes;
run; quit;

/* Reacquisition Probability */
proc logistic data = statcrm.winback descending;
model reacquire = Duration_1 Offer Duration_lapse Gender Age / link =
probit;
output out = statcrm.winback_probit xbeta = xb_probit p = pred;
run; quit;

/* Purchase Accuracy */
data statcrm.winback_accuracy;
set statcrm.winback_probit;
if pred >= 0.5 then pred_win = 1; else pred_win = 0;
act_win = reacquire;
run; quit;
proc freq data = statcrm.winback_accuracy;
table pred_win * act_win; run; quit;

/* Inverse Mills Ratio */
data statcrm.winback_imr;
set statcrm.winback_probit;
imr_win = (pdf('Normal',xb_probit))/(probnorm(xb_probit));
run; quit;
/* Duration */
proc reg data = statcrm.winback_imr;
model duration_2 = Duration_1 Offer Duration_lapse price_change Gender
Age imr_win / rsquare;
where reacquire = 1;
run; quit;
data statcrm.winback_duration1;
set statcrm.winback_imr;
xbeta_duration2 = 152.6382 + 1.21243*Duration_1 + 2.35055*Offer -
0.97247*Duration_Lapse - 7.83839*Price_Change +
  10.33485*Gender - 2.50636*Age + 19.13765*imr_win;
pred_dur2 = probnorm(xb_probit)*(xbeta_duration2);
ad = abs(duration_2 - pred_dur2);
ad1 = abs(duration_2 - 394.61);
run; quit;
proc sql; select mean(duration_2) as mean_dur, mean(ad) as mad, mean
(ad1) as random_mad from statcrm.winback_duration1; quit;
```

```
/* Profit */
proc reg data = statcrm.winback_duration1;
model sclv = duration_1 Offer price_change Gender Age imr_win pred_dur2 /
rsquare vif;
where reacquire = 1;
run; quit;
data statcrm.winback_sclv1;
set statcrm.winback_duration1;
xbeta_sclv = 337.8208 + 0.93246*Duration_1 + 1.62464*Offer -
7.73031*Price_Change + 14.15901*Gender - 2.15518*Age +
124.05317*imr_win + 0.70135*pred_dur2;
pred_sclv = probnorm(xb_probit)*(xbeta_sclv);
ad = abs(sclv - pred_sclv);
ad1 = abs(sclv - 730.1482);
run; quit;
proc sql; select mean(sclv) as mean_sclv, mean(ad) as mad, mean(ad1) as
random_mad from statcrm.winback_sclv1; quit;
```

# Customer win-back – SAS output

```
                      The SAS System

                   The LOGISTIC Procedure

                     Model Information

     Data Set                     STATCRM.WINBACK
     Response Variable            Reacquire       Reacquire
     Number of Response Levels    2
     Model                        binary probit
     Optimization Technique       Fisher's scoring

            Number of Observations Read     500
            Number of Observations Used     500

                      Response Profile

              Ordered                              Total
                Value          Reacquire       Frequency

                    1                  1             295
                    2                  0             205

          Probability modeled is Reacquire=1.
                  Model Convergence Status

       Convergence criterion (GCONV=1E-8) satisfied.

                   Model Fit Statistics
                                               Intercept
                           Intercept                 and
            Criterion           Only          Covariates
            AIC              678.859             244.351
            SC               683.073             269.638
            -2 Log L         676.859             232.351
            Testing Global Null Hypothesis: BETA=0

   Test                Chi-Square        DF          Pr > ChiSq
   Likelihood Ratio      444.5080         5              <.0001
   Score                 280.6332         5              <.0001
   Wald                  118.4638         5              <.0001

                      The SAS System

                   The LOGISTIC Procedure

          Analysis of Maximum Likelihood Estimates

                                       Standard          Wald
 Parameter      DF     Estimate          Error    Chi-Square    Pr > ChiSq

 Intercept       1      -1.4987         0.6628        5.1131        0.0237
 Duration_1      1      0.00910       0.000864      111.0338        <.0001
```

```
 Offer            1       0.1571     0.0261       36.3549      <.0001
 Duration_lapse   1      -0.0333    0.00358       86.8678      <.0001
Gender            1      -0.2064     0.1950        1.1198      0.2900
Age               1      -0.0388    0.00708       29.9645      <.0001

   Association of Predicted Probabilities and Observed Responses

       Percent Concordant  96.3   Somers' D        0.927
       Percent Discordant   3.6   Gamma            0.927
       Percent Tied         0.0   Tau-a            0.449
       Pairs              60475   c                0.963

                 The SAS System

              The FREQ Procedure

                Table of pred_win by act_win
        pred_win   act_win

        Frequency,
         Percent ,
         Row Pct ,
         Col Pct ,      0,      1,   Total
         ffffffff^fffffff^fffffff^
              0 ,   172 ,    35 ,     207
                , 34.40 ,  7.00 ,   41.40
                , 83.09 , 16.91 ,
                , 83.90 , 11.86 ,
         ffffffff^fffffff^fffffff^
              1 ,    33 ,   260 ,     293
                ,  6.60 , 52.00 ,   58.60
                , 11.26 , 88.74 ,
                , 16.10 , 88.14 ,
         ffffffff^fffffff^fffffff^
           Total    205     295          500
                   41.00  59.00       100.00

                 The SAS System

               The REG Procedure

                 Model: MODEL1

       Dependent Variable: Duration_2 Duration_2

       Number of Observations Read     295
       Number of Observations Used     295

                 Analysis of Variance

                           Sum of        Mean
 Source          DF       Squares      Square         F Value       Pr > F

 Model            7      26119513     3731359         4265.01       <.0001
 Error          287        251090   874.87753
```

| | | |
|---|---|---|
| Corrected Total | 294 | 26370603 |

| | | | |
|---|---|---|---|
| Root MSE | 29.57833 | R-Square | 0.9905 |
| Dependent Mean | 668.83390 | Adj R-Sq | 0.9902 |
| Coeff Var | 4.42237 | | |

Parameter Estimates

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| Intercept | Intercept | 1 | 152.63820 | 13.37648 | 11.41 | <.0001 |
| Duration_1 | Duration_1 | 1 | 1.21243 | 0.01111 | 109.17 | <.0001 |
| Offer | Offer | 1 | 2.35055 | 0.45184 | 5.20 | <.0001 |
| Duration_lapse | Duration_lapse | 1 | -0.97247 | 0.05380 | -18.08 | <.0001 |
| Price_Change | Price_Change | 1 | -7.83839 | 0.11605 | -67.54 | <.0001 |
| Gender | Gender | 1 | 10.33485 | 3.66216 | 2.82 | 0.0051 |
| Age | Age | 1 | -2.50636 | 0.12703 | -19.73 | <.0001 |
| imr_win | | 1 | 19.13765 | 6.82310 | 2.80 | 0.0054 |

The SAS System

```
mean_dur         mad          random_mad
ffffffffffffffffffffffffffff
394.612        67.87737        353.006
```

The SAS System 12:34 Tuesday, December 13, 2011 141

The REG Procedure

Model: MODEL1
Dependent Variable: SCLV SCLV

| | |
|---|---|
| Number of Observations Read | 295 |
| Number of Observations Used | 295 |

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 7 | 60711747 | 8673107 | 2714.93 | <.0001 |
| Error | 287 | 916848 | 3194.59261 | | |
| Corrected Total | 294 | 61628595 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 56.52073 | R-Square | 0.9851 |
| Dependent Mean | 1241.01390 | Adj R-Sq | 0.9848 |
| Coeff Var | 4.55440 | | |

Parameter Estimates

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr>\|t\| | Variance Inflation |
|---|---|---|---|---|---|---|---|
| Intercept | Intercept | 1 | 337.82080 | 27.50279 | 12.28 | <.0001 | 0 |
| Duration_1 | Duration_1 | 1 | 0.93246 | 0.06007 | 15.52 | <.0001 | 18.10038 |
| Offer | Offer | 1 | 1.62464 | 0.80850 | 2.01 | 0.0454 | 1.13313 |

```
Price_Change  Price_Change  1   -7.73031   0.44232  -17.48  <.0001  3.98627
Gender        Gender        1   14.15901   7.01448    2.02  0.0445  1.11176
Age           Age           1   -2.15518   0.25375   -8.49  <.0001  1.23249
imr_win                     1  124.05317  18.40432    6.74  <.0001  4.21886
pred_dur2                   1    0.70135   0.05459   12.85  <.0001 31.51271

                          The SAS System

              mean_sclv     mad      random_mad
              ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
               730.1482   140.5866     630.2259
```

# References


1. Strauss, B. and Friege, C. (1999) Regaining service customers: costs and benefits of regaining management. *Journal of Service Research*, **1**(4), 347–361.
2. Thomas, J., Blattberg, R.C., and Fox, E.J. (2004) Recapturing lost customers. *Journal of Marketing Research*, **41**(1), 31–45.
3. Tokman, M., Davis, L.M., and Lemon, K.N. (2007) The WOW factor: creating value through win-back offers to reacquire lost customers. *Journal of Retailing*, **83**(1), 47–64.
4. Reinartz, W.J., Thomas, J., and Kumar, V. (2005) Balancing acquisition and retention resources to maximize customer profitability. *Journal of Marketing*, **69**(1), 63–79.

# 8

# Implementing CRM models

## 8.1 Introduction

Up until now, we have explored and understood the four important steps of the CRM process: customer acquisition, customer retention, customer churn, and customer win-back. For each of these topics, we reviewed the key concepts, modeling issues, and key drivers that determine the acquisition, retention, churn, and win-back rates of a firm. A keen understanding of these CRM processes would enable firms to select the right strategies to maximize CLV, and thereby enhance firm profitability.

However, when it comes to implementing one or more of these four CRM processes, they are rarely based on any singular step, but rather based on an integrated outlook that is cognizant of the interlinkage between all them. For instance, depending on the desired outcome of the CRM initiatives, a manager might have to consider how to optimally allocate different resources across multiple steps of the process to increase the customer value (see Chapter 5 on balancing acquisition and retention as an example). This would make the manager speed acquisition, increase revenue through retention, delay attrition/churn, or make win-back more effective, or involve a combination of all of them. Figure 8.1 illustrates the typical customer life cycle and the strategic impact of implementing the CRM processes on the traditional CLV curve.

Now we will step back and take a macroscopic view of the entire CRM process to see how all the CRM processes fall into place, when attempting to maximize firm value. In this chapter, we will discuss the application of the CLV metric. Specifically, we will focus on methodology to measure the individual customer's CLV and to implement a CLV-based marketing framework through two real-life case studies, one in a B2B and the other in a B2C setting.

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

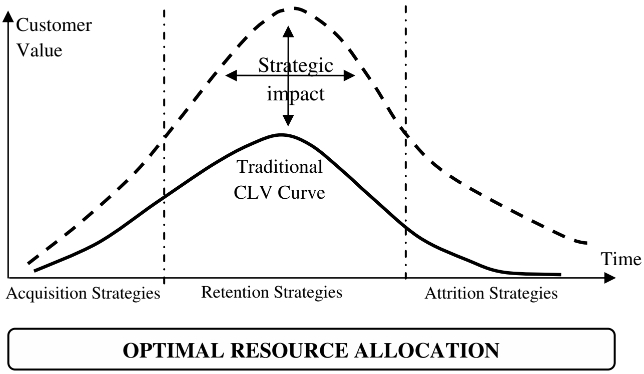



*Figure 8.1 A typical customer life cycle.*

The chapter is organized as follows. First, the approach to measure CLV and a numerical example are presented. Then, we illustrate the implementation of the CLV management framework through two case studies conducted in IBM (B2B setting) and a fashion retailer (B2C setting). Finally, the challenges in the implementation of the CLV model in these firms will also be discussed.

## 8.2 CLV measurement approach

CLV is defined as the 'sum of cumulated future cash flows – discounted using the weighted average cost of capital (WACC) – of a customer over their entire lifetime with the company' [1]. In this regard, CLV can also be referred to as the net present value of future cash flows from a customer. Increasingly, adoption of CLV and CLV-based strategies are being well accepted by marketing practitioners, due to its forward-looking component, unlike the traditional measures of profitability.

When computing CLV, managers have to consider the setting in which the customer purchases are being made, that is, contractual and non-contractual. A contractual setting is one where the customers are bound by a contract such as a mobile phone subscription. On the contrary, in a non-contractual setting the customers are not bound by a contract such as grocery store purchases. That is, in a contractual setting, the firm gets fixed monthly revenue through the subscription and observes when churn occurs (i.e., when the subscription is cancelled). But the observation of the churn would be missing in a non-contractual setting. Therefore, these differences will have to be included while computing CLV. To cover both situations, CLV can be expressed in the following form:

$$CLV_i = \sum_{t=1}^{T} \frac{Base\ GC}{(1+r)^t} + \sum_{t=1}^{T} \frac{\hat{p}(Buy_{it}=1)^{*}\hat{G}C_{it}}{(1+r)^t} - \frac{\hat{M}C_{it}}{(1+r)^t} \qquad (8.1)$$

where:

$CLV_i$ = lifetime value for customer $i$

$\hat{p}(Buy_{it})$ = predicted probability that customer $i$ will purchase additional product(s)/service(s) in time period $t$

$\hat{G}C_{it}$ = predicted gross contribution margin provided by customer $i$ in time period $t$

$\hat{M}C_{it}$ = predicted marketing costs directed toward customer $i$ in time period $t$

$t$ = index for time periods, such as months, quarters, years, and so on

$T$ = the end of the calibration or observation time frame

$r$ = monthly discount factor

Base GC = predicted base monthly gross contribution margin

Equation 8.11 can also be written as follows:

$$CLV_i = Baseline\ CLV + Augmented\ CLV.$$

As is evident from Equation 8.1, this formula can be applied in both the contractual and non-contractual settings. Now, let us look how it works in each of the settings.

In a contractual setting, the first term in Equation 8.1 or the baseline CLV corresponds to the constant gross contribution that the customer is going to give to the firm. This could be on a monthly, quarterly, or annual basis depending on the time frame of the subscription. This term includes the marketing cost to retain the customer at the current level of *Base GC*. The second term in the equation corresponds to the predicted net present value of future purchases by the customer at a particular time period. The third term corresponds to the additional marketing costs incurred to sell more product(s)/service(s) to the customer. The final value of these three components would yield the CLV in a contractual setting.

In a non-contractual setting, the first term in Equation 8.1 would not be valid as there is no constant flow of base income from a customer on a periodic basis. In short, there is no assured income due to subscriptions. Therefore, there would be no baseline CLV for the firm. As with the previous case, the second term in the equation corresponds to the predicted net present value of future purchases by the customer at a particular time period. Then, the third term corresponds to the additional marketing costs incurred to sell more product(s)/service(s) to the customer. The final value of these three components would yield the CLV in a non-contractual setting.

Now let us now consider a numerical example in a contractual setting to see how CLV is computed. Consider the case of Amy, a customer of a mobile phone

Table 8.1 Transaction details of Amy.

| | June | July | August | September |
|---|---|---|---|---|
| $\hat{p}(Buy)$ | 0.55 | 0.50 | 0.40 | 0.20 |
| Monthly purchase amount ($) | 20 | 10 | 10 | 15 |
| Profit margin (%) | 20 | 20 | 20 | 20 |
| Marketing cost ($) | 5 | 7 | 7 | 10 |

company. The monthly subscription or the *Base GC* provided by Amy is $40. Standing at the end of May, the company wants to know the value Amy is likely to provide to the company in the next four months (June, July, August, and September). Table 8.1 provides Amy's probability of buying additional services or $\hat{p}(Buy)$ (such as downloads, ringtones, text messaging, etc.) for the next four months, her monthly purchase amount, the percentage of margin for each purchase, and the marketing cost incurred by the company in contacting Amy.

Assuming an annual discount rate ($r$) of 12% (or 1% monthly rate), we can now compute the CLV of Amy for the next four months. First, let us compute the CLV of Amy at the end of July using Equation 8.1.

$$\text{Amy's lifetime value at the end of June} = \frac{40}{(1.01)^1} + \frac{(0.55)^*(20^*20\%)}{(1.01)^1} - \frac{5}{(1.01)^1}.$$

Therefore, $CLV_{Amy,\ June} = \$36.9$.

Similarly, we can compute the value Amy would give to the company at the end of each subsequent month as follows: July, $33.2; August, $32.8; and September, $29.6. A summation of all the four months' CLVs would yield a value of $132.5. In other words, over the next four months Amy would provide $132.5 in value to the company through her subscription and additional purchases.

Table 8.1 can also be used to explain the case of non-contractual purchases. Assume that the table indicates Amy's monthly purchases from her nearby café. In this case, $\hat{p}(Buy)$ would denote the probability of Amy buying products from the café. However, this case will not have the monthly subscription or baseline CLV. Now, let us compute the CLV of Amy at the end of July using Equation 8.1:

$$\text{Amy's lifetime value at the end of July} = \frac{(0.55)^*(20^*20\%)}{(1.01)^1} - \frac{5}{(1.01)^1}.$$

Therefore, $CLV_{Amy,\ June} = -\ \$2.7$.

In other words, Amy will be costing the café $2.7 in June by receiving communications from the café and by being a part of its customer base. Similarly, we can compute the value Amy would give to the café at the end of each subsequent month as follows: July, −$6.0; August, −$6.0; and September, −$9.0. A summation of all the four months' CLVs would yield a value of −\$23.7. That is, over the next four months Amy would cost the café \$23.7 in value by being its customer. In other words, the amount spent by the café on marketing to Amy will be more than the profit before marketing costs contributed by her to the café.

The above illustration is a simplistic scenario for computing CLV. Given the changes and challenges in data availability and business needs, several other approaches to model and compute CLV have been developed. The two case studies that will be presented in the following section highlight the following important CLV maximization strategies:

- Efficient customer selection by targeting customers with high profit potential.
- Managing existing sets of customers and rewarding them based on their profit potential.
- Investing in high-profit customers to prevent attrition and ensure future profitability.

## 8.3 CRM implementation at IBM

In this section, we will present the CRM implementation at IBM (a B2B company), from initial model development to actual field study implementation. The whole idea of the case study is to showcase how CLV was used as a useful indicator of customer profitability for IBM's marketing decisions toward maximizing the firm's value [2]. This study shows that CLV performs better in measuring customer value as compared to IBM's customer selection metric.

By implementing CLV measurement and maximization on about 35 000 customers, IBM was able to increase its revenue 10-fold without any changes in the level of marketing investment. This increase in revenue was made possible just by reallocating the marketing communication resources to the right customer groups.

### 8.3.1 IBM background

IBM, a multinational technology and consulting service provider to B2B customers, aimed to maximize its overall profitability through customer-centric CRM strategies. Specifically, IBM sought to maximize its profitability by prioritizing its resource allocation to the most profitable customers. To measure customer profitability, it used Customer Spending Score (CSS), which was defined as the total revenue that can be expected from a customer in the next year. Following the scoring, IBM then ranked customers into 10 deciles according to their corresponding CSS values. The top one or two deciles were then targeted and allocated more marketing resources.

In 2004, IBM felt the need to move to implementing a better indicator for customer value measurement than CSS, because CSS focused primarily on customer

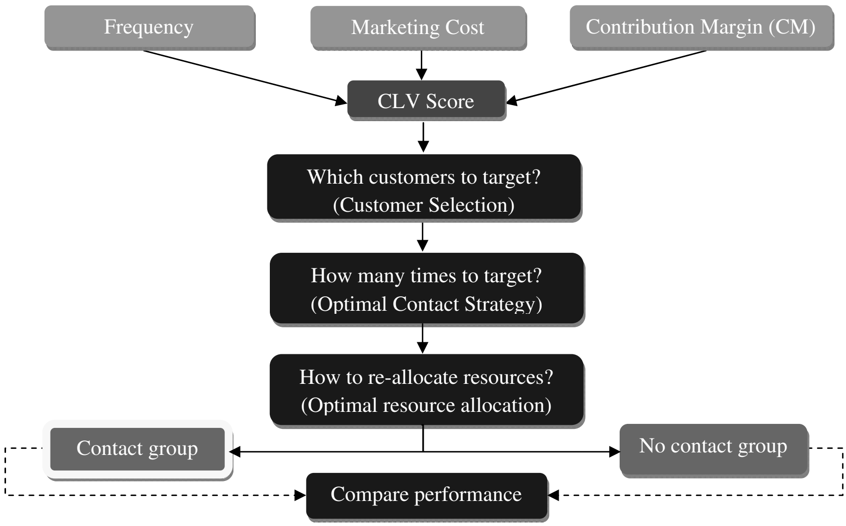



*Figure 8.2 CLV-based management framework at IBM.*

revenues and largely ignored the variable cost of servicing customers. CLV was then proposed as an alternative indicator, and a CLV-based management framework was suggested for implementation for IBM's customers to improve their profitability.

Specifically, IBM wanted to test the following belief: 'When all other drivers remain unchanged, can an increase in the level of contacts with the right customers create high value from the low-value customers?' To accomplish this objective, a CLV-based management framework was adopted to design customer management initiatives, as illustrated in Figure 8.2.

## 8.3.2 Implementing a CLV management framework at IBM

A two-stage process was used to develop and implement the CLV management framework at IBM. In the first stage, several models were developed to generate inputs for implementation. In the second stage, a field study was conducted based on recommendations from the models developed in the first stage. The aim of the field study was to test IBM's main objective – *whether an increase in contacts to the right customer creates high value from low-value customers when all other drivers are similar*. Figure 8.3 illustrates the two-stage approach involved in implementing a CLV management framework at IBM.

Data on 20 000 mid-market companies of IBM were used in this study and the subsequent implementation. Mid-market companies are defined as companies that have the number of employees within the range of 100 and 999 at the enterprise level, and with total enterprise revenues to IBM from 2001 to 2003 that were over

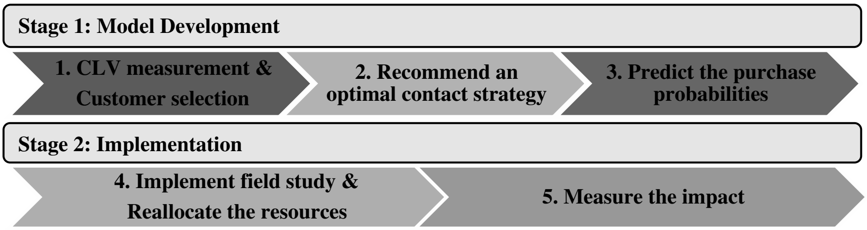



*Figure 8.3 Process assessing the impact of the CLV management framework at IBM.*

$25 000. The data were collected at the monthly transaction level, and per establishment. Each enterprise can have more than one establishment (e.g., different locations making independent decisions). In total, 2.5 million observations were collected (72 months each for the 35 131 establishments).

#### 8.3.2.1 Stage 1: Model development

The model development stage involves developing a framework to measure individual customer's CLV and selecting the right customers, propose an optimized resource allocation and contact strategy, and build propensity models. Specifically, this stage involves the following three phases:

- *Phase 1.* Measure CLV score for individual customer, and select the right customers to target based on the CLV scores.
- *Phase 2.* Optimize resource allocation to the marketing contacts with different groups of targeted customers.
- *Phase 3.* Develop propensity models for each product category to identify the product to feature.

**Phase 1: CLV measurement and customer selection**


In this phase, the *always-a-share* approach for measuring CLV was adopted because of its relevance to the non-contractual setting of IBM. This approach assumes that customers never 'quit' their relationship with a firm; rather they demonstrate only dormancy in their relationship with the firm while transacting with other firms and always have a probability (no matter how small) of purchasing in a given time period [3]. When they return to the relationship, they retain the memory about their prior relationship with the firm. Accordingly, the CLV was measured using the following formula:

$$CLV_i = \sum_{j=T+1}^{T+36} \frac{\left(p(Buy_{ij} = 1) \times \hat{C}M_{ij}\right)}{(1+r)^{j-T}} - \hat{M}T_{ij} \times \frac{\overline{MC}}{(1+r)^{j-T}} \tag{8.2}$$

where:

| | | |
|---|---|---|
| $CLV_i$ | $=$ | lifetime value for customer $i$ |
| $p(Buy_{ij} = 1)$ | $=$ | predicted probability that customer $i$ will purchase in time period $j$ |
| $\hat{C}M_{i,t}$ | $=$ | predicted contribution margin provided by customer $i$ in time period $j$ |
| $\overline{MC}$ | $=$ | average cost of a single marketing contact per customer, assumed to be \$7 in the study |
| $\hat{M}T_{i,t}$ | $=$ | number of contacts directed to customer $i$ in time period $j$ |
| $j$ | $=$ | index for time periods, months in this case |
| $T$ | $=$ | the end of the calibration or observation time frame |
| $r$ | $=$ | monthly discount factor, 0.0125 in this case (amounting to a 15% annual rate) |

In Equation 8.2, the computation of CLV involved predictions of three aspects: (a) the level of marketing contacts directed toward customer $i$ in time period $j$ ($MT$), (b) the probability that a customer would purchase in each time period ($p(Buy)$), and (c) the contribution (in \$) provided by the customer in each time period ($CM$). A description of the prediction of these three aspects is provided first here.

*(a) Predicting the marketing contacts (MT)* The marketing contacts allocated by the firm toward customer $i$ are determined by the following model:

$$\log\left(MT_{ij}\right) = \alpha_{1i} + x_{1ij}^{T}\beta_1 + u_{1ij} \tag{8.3}$$

where $x_{1ij}$, $\beta_1$, $\alpha_{1i}$, $u_{1ij}$ are, respectively, a vector of predictor variables, a vector of corresponding coefficients, an individual-level intercept, and an error term.

*(b) Predicting the purchase propensity p(Buy)* It was assumed that customer $i$ purchases from the firm only when the customer's latent utility for purchasing from the firm ($Buy_{ij}^{*}$) exceeds a certain threshold, set to zero in this case. In this study, only the binary outcome variable regarding whether or not the customer purchased in time period $j$ was observed. Consequently, the latent utility of the customer was not

observed. The latent utility is mapped to the binary outcome variable ($Buy_{ij}$) as follows:

$$(Buy^*_{ij}) > 0, \text{ if } Buy_{ij} = 1; \quad (Buy^*_{ij}) < 0, \text{ if } Buy_{ij} = 0. \tag{8.4}$$

The latent utility, ($Buy^*_{ij}$), for customer $i$ to purchase from the firm in time period $j$ was then modeled as a function of predictor variables in a linear model:

$$Buy^*_{ij} = \alpha_{2i} + x^T_{2ij}\beta_2 + u_{2ij} \tag{8.5}$$

where, similar to Equation 8.3, $x_{2ij}$, $\beta_2$, $\alpha_{2i}$, $u_{2ij}$ are, respectively, a vector of predictor variables, a vector of corresponding coefficients, an individual-level intercept, and an error term.

*(c) Predicting the contribution margin (CM)* We assume that a latent variable, $CM^*_{ij}$, represents the amount spent by customer $i$ in time period $j$, irrespective of whether it is with the firm, as a function of predictor variables with a linear structure

$$CM^*_{ij} = \alpha_{3i} + x^T_{3ij}\beta_3 + u_{3ij} \tag{8.6}$$

where, similar to Equations 8.3 and 8.5, $x_{3ij}$, $\beta_3$, $\alpha_{3i}$, $u_{3ij}$ are, respectively, a vector of predictor variables, a vector of corresponding coefficients, an individual-level intercept and an error term.

If the customer purchased from the firm in time period $j$, then the firm observed the contribution margin provided by the customer as follows:

$$CM_{ij} = CM^*_{ij}, \text{ if } Buy_{ij} = 1; \quad CM_{ij} = unobserved, \text{ if } Buy_{ij} = 0. \tag{8.7}$$

*(d) Model likelihood* The three aspects involved in the computation of CLV are inherently correlated. The level of marketing contacts directed toward a customer depends on customer characteristics, past customer behavior, and the past level of marketing resources allocated to the customer. The probability that a customer would purchase is likely to be dependent on the level of marketing resources directed toward the customer, and, finally, the customer provides profits to the firm only if a purchase is made. In the modeling framework used in this study, these aspects of firm and customer behavior were allowed to be correlated with each other. To jointly model the marketing contacts, the probability of purchase, and the contribution margin, a model structure based on a 'seemingly unrelated regression' (SUR) was used. The likelihood that summarizes our model structure is provided

below:

$$L(MT, Buy, CM) = \alpha \prod_{i=1}^{N} \prod_{j=1}^{T} \Pr\left(B\dot{u}y_{ij} \leqslant 0, MT_{ij}\right)^{1-Buy_{ij}} \cdot \Pr\left(C\dot{M}_{ij} = CM_{ij}, B\dot{u}y_{ij} > 0, MT_{ij}\right)^{Buy_{ij}} \quad (8.8)$$

where $B\dot{u}y_{ij}$ is customer $i$'s latent utility for purchasing in time period $j$.

The covariance structure of the errors in Equations 8.3, 8.5, and 8.6 was modeled as

$$\begin{pmatrix} u_{1ij} \\ u_{2ij} \\ u_{3ij} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}\right) = N_3\left(0, \Sigma_y\right). \quad (8.9)$$

Such a method of modeling provided the possibility of correlations among the three residuals. Further, $\sigma_{11}$ was fixed to be equal to 1 to ensure model identification. The covariance structure of the errors accounted for any unobserved dependence between a firm's decision to contact a customer (*MT*), a customer's decision to purchase from the firm (*Buy*), and the amount of money the customer spends with the firm (*CM*). By allowing $\beta = [\beta_1, \beta_2, \beta_3]$ and $\alpha = [\alpha_1, \alpha_2, \alpha_3]$, the simultaneous equation model gives rise to the likelihood specified in Equation 8.8 – the model likelihood equation. The customer-specific intercept terms were obtained from a multivariate normal distribution as follows:

$$\alpha_i \sim MVN(\Delta Z_i, \Sigma_\alpha)$$

where:

$Z_i$ = a $p \times 1$ vector of customer characteristics

$\Delta$ = a $3 \times p$ matrix of coefficients for the customer characteristics

$\Sigma_\alpha$ = a $3 \times 3$ variance–covariance matrix

p = number of customer characteristics that are used to capture heterogeneity

Using the modeling approach described above, the first 54 months of data were used to estimate model parameters, and the CLV score was computed for each customer for 36 months (2002 through 2004).

*(e) Drivers of CLV* In calculating the CLV score, the drivers of the respective parameters, such as marketing contacts, probability of purchase, and contribution margin at IBM, were identified. The interactions between the drivers were also evaluated so that the CLV modeling was conducted along four critical dimensions: (1) modeling firm decisions, (2) providing a forward-looking cost allocation

strategy, (3) imputing missing contribution margin, and (4) accommodating unobserved dependence among levels of marketing contacts, purchase incidence, and contribution margin.

The drivers of marketing contacts were past customer spending, past levels of marketing contacts, cross-buying and recency, and past purchase activity. The drivers of probability of purchase and contribution margin were categorized into customer relationship characteristics and customer firmographics. The customer relationship characteristics included drivers such as past customer spending level, cross-buying behavior, purchase frequency, recency of purchase, and past purchase activity, and the marketing contacts by the firm. The customer firmographics included drivers such as sales of an establishment (a measure of the size of the establishment), and the installed level of PCs in the establishment (a measure of the level of demand for IT products in the establishment).

*(f) CLV measurement results* The coefficient estimates of the drivers of CLV are reported in Table 8.2. The values are the posterior means and variances. A parameter is considered not significant if a zero exists within the 2.5th percentile and 97.5th percentile values of the posterior distribution for that parameter.

According to Table 8.2, the following key observations were made:

- Customers who have spent more, have made a recent purchase, and have purchased across a wider range of product categories are more likely to purchase in the current month.
- Marketing contacts have a positive influence on customer purchase incidence, and for customers who have made a recent purchase, the influence of marketing contacts is enhanced.
- Customers who have a greater cross-buying, and customers who have purchased frequently in the past, provide a higher contribution margin.
- While IBM allocated more marketing contacts for customers who have higher sales, the purchase incidence and contribution margin are lower for these customers. This is possible because customers who have higher sales in general split their purchases across several vendors.
- Customers who have a large installed base of PCs have a higher purchase incidence and contribution margin and are contacted more by IBM.

*(g) Customer selection* The CLV score for each customer was computed using information from the predictions of marketing contacts, purchase incidence, and contribution margin, as well as the unit marketing costs for each channel. In all, 72 months of historical data were available for model development. Traditional metrics were also computed using the first 54 months of data. Customers were then rank ordered based on the CLV measure as well as on the traditionally used metrics. The comparative performance of the customers (i.e., the observed profits provided

Table 8.2 Estimation results.

| Estimation results<br>Dependent variables | | |
|---|---|---|
| CLV | *The discounted value of all expected future profits, or customer lifetime value*<br>Coefficients | |
| Independent variables | Mean* | Variance* |
| *Level of marketing contacts* | | |
| Lagged level of contacts | 0.7366 | 0.0316 |
| Two-period lagged level of contacts | 0.3239 | 0.0333 |
| Lagged average number of purchases | 0.5836 | 0.2158 |
| Two-period indicator of purchase | 4.7114 | 1.4023 |
| Interaction of cross-buying and recency | −0.0149 | 0.0035 |
| Lagged contribution margin | 0.6016 | 0.1128 |
| *Purchase incidence* | | |
| Lagged indicator of purchase | 0.6573 | 0.0902 |
| Two-period lagged indicator of purchase | 0.2172 | 0.0891 |
| Lagged average level of contribution margin | 0.0056 | 0.0026 |
| Log of lagged level of contacts | 0.0041 | 0.0012 |
| Interaction of cross-buying and recency | −0.0047 | 0.0074 |
| Interaction of log of lagged level of contacts and lagged indicator of purchase | 0.0004 | 0.0002 |
| *Contribution margin* | | |
| Lagged contribution margin | 0.8612 | 0.0247 |
| Lagged average contribution margin | 0.7442 | 0.0325 |
| Cross-buying | 0.2858 | 0.1075 |
| Frequency of purchases | 7.3692 | 1.887 |
| Log of lagged level of contacts | 0.079 | 0.0105 |
| Interaction of cross-buying and recency | −0.038 | 0.0565 |

*Mean and variance are computed using the 5th through 95th percentiles of the posterior sample.

by the customers in the last 18 months) in the top 15% of each list of metrics clearly shows the power of CLV to identify the best customers for future targeting (see Table 8.3). Contrary to prior findings, this study found that in non-contractual settings, at least with regard to selecting high-potential customers for future targeting, current profit performs worse than estimates of future profitability.

Table 8.3 Profits generated using customer selection strategy by different metrics.

| Percentage of cohort (selected from top) | | Using the first 54 months of data to predict the next 18 months of purchase behavior | | | |
|---|---|---|---|---|---|
| | | CLV | CSS | RFM | PCV* |
| *15* | Average revenue | 30 427 | 21 789 | 22 622 | 23 542 |
| | Gross value | 9184 | 6659 | 6966 | 7185 |
| | Variable costs | 107 | 114 | 110 | 104 |
| | Net value | 9077 | 6544 | 6856 | 7081 |

*Past Customer Value.

Notes: The reported values are in dollars (expressed as a multiple of the actual numbers) per customer and are cell medians. The net result was to identify the top customers who provided the best customer value.

### Phase 2: Recommend optimal contact strategy


Using the CLV measurement and the subsequent customer selection approach, a forward-looking cost allocation strategy was implemented on IBM's customers. The cost allocation strategy involved used an optimization algorithm to maximize the sum of expected CLVs for all the customers, as the frequency of marketing contacts for each customer were allowed to be varied. This method facilitated identification of the optimal level of marketing contact for each customer that would maximize the sum of expected CLVs of all customers.

The optimization algorithm was conducted on 5000 customers, with the following parameters: population size = 200, probability of crossover = 0.8, probability of mutation = 0.25, and convergence criteria = difference in optimal solution over the last 10 000 iterations is less than 0.1%. The output from this optimal resource allocation model allowed for the determination of the optimal number of contacts for each customer and the corresponding marketing resources to be allocated to the customer.

The output from the optimal resource allocation model produced the input to the decision-making process about the number of contacts in each channel for each customer in various customer segments. The differences in suggested optimal contact frequencies across various customer segments are shown in Figure 8.4.

This figure shows the contact strategies recommended by IBM's CSS indicator and recommended by the CLV management framework respectively. It is important to note that the CLV framework suggests a rather different strategy. For example, customer segment D, which was suggested to receive the lowest level of contact by CSS indicator, should actually receive the highest level of contact by CLV model.

Based on the CLV contact approach, an optimal contact strategy for the customer segments involved in this study was developed. This strategy involved

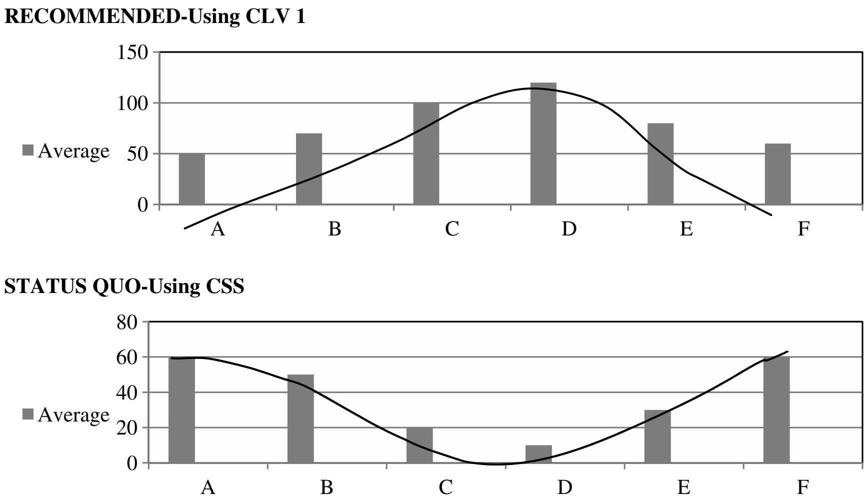



*Figure 8.4 A comparison of contact strategies.*

classifying the contact strategies into four buckets along the CSS and CLV metrics. Table 8.4 illustrates the optimization strategies.

As shown in Table 8.4, it was recommended that the contact interval through direct mail/telesales/catalog/e-mail to the Low CSS–High CLV and High CSS–High CLV customer segments be increased from 4.82 to 1.9 days and from 6.3 to 5.3 days, respectively. In other words, these customers will have to be contacted more often than the current practice. This increase in contact level would provide an increase in gross value of around 63 and 69% respectively. The biggest lift in

Table 8.4 Optimization strategies.

| CLV | CSS | |
|---|---|---|
| | Low | High |
| High | Direct mail/telesales/catalog/e-mail:<br>Current interval: 4.82 d<br>Optimal interval: 1.9 d<br>Gross value:<br>Current value: \$10 936<br>Optimal value: \$17 809 | Direct mail/telesales/catalog/e-mail:<br>Current interval: 6.3 d<br>Optimal interval: 5.3 d<br>Gross value:<br>Current value: \$53 488<br>Optimal value: \$90 522 |
| Low | Direct mail/telesales/catalog/e-mail:<br>Current interval: 9.7 d<br>Optimal interval: 12.6 d<br>Gross value:<br>Current value: \$743<br>Optimal value: \$1203 | Direct mail/telesales/catalog/e-mail:<br>Current interval: 8.4 d<br>Optimal interval: 8.3 d<br>Gross value:<br>Current value: \$1091<br>Optimal value: \$2835 |

gross value is observed in the High CSS–Low CLV group segment, approximately 160%, if IBM were to reduce the current contact interval from 8.4 to 8.3 days. Finally, the Low CSS–Low CLV segment customers should be contacted less frequently, thus less marketing resources should be allocated to this group, which will result in a 62% increase in gross value.

Following this, to optimally reallocate the marketing resources, and to test/assess whether the different resource allocation strategy suggested by the CLV model will translate into higher profits for the firm, a field experiment was conducted. Before conducting the field experiment, the study sought to find out what products to pitch to which customers.

**Phase 3: Predict purchase probabilities**


In this phase, the predicted purchase probability for a category, a product for each customer was identified. This allowed IBM to determine which products to pitch to the targeted customers. The rationale behind this phase was the assumption that the customer's need for a certain product type and familiarity with the focal categories are the key drivers of the customer's product category choice. The drivers of product category choice that have significant influence include: (1) the proportion of the same category purchases, that is, the *dominance* of a category over others; (2) the depth of same-category purchases measured as the number of products purchased within the focal category, that is, *knowledge* of the focal category; and (3) the breadth of same-category purchases measured as the number of different product types purchased within the focal category.

To determine a customer's propensity to purchase each of the product categories, propensity models were built to predict the propensity of buying products within the following categories: hardware, software, and services. A product message was used in a marketing contact to a customer if the predicted purchase propensity for the category is greater than 0.5 for that customer. The results of the propensity model, together with the CLV model, allowed IBM to convey the right product/service message to the right customers.

### 8.3.2.2 Stage 2: Model implementation

The model implementation stage involved implementing the customer selection and resource allocation strategies proposed in Stage 1 in a field study on selected IBM customers. The specific objectives of the field study were to:

- reallocate the optimal resources and align marketing resources to contacting customers who have the highest potential to bring profitability to the firm;
- compare the performance of the contact group and no-contact group to evaluate the impact of the implementation of the CLV framework on IBM's customers.

Table 8.5 Resource reallocation based on CLV.

| Decile | Not Contacted until 2004 | Contacted by 2004 | Customer Segment |
|---|---|---|---|
| **1** | 350,471 | 2,124,483 | Super High CLV |
| **2** | 993 | 125,460 | High CLV |
| **3** | 669 | 43,681 | |
| **4** | 638 | 23,624 | |
| **5** | 623 | 17,499 | Medium CLV |
| **6** | 611 | 13,675 | |
| **7** | 534 | 10,513 | |
| **8** | 444 | 8,051 | |
| **9** | 369 | 5,023 | Low CLV |
| **10** | 80 | (35) | |

## Phase 4: Implement field study and reallocate resources


To determine the level of resources to be allocated to the respective groups of customers, 35 131 customers were selected and divided into two groups, the Contacted by 2004 (customers who have been contacted previously by IBM until 2004) and the Not Contacted until 2004 (customers who have not been contacted until 2004). In each group, the customers were then rank ordered into deciles, according to their respective CLV scores, as shown in Table 8.5.

As can be seen from this table, customers in decile 10 in the Contact group were not profitable. It was recommended that resources from this group be reallocated to customers in the Not Contacted until 2004 group in deciles 1, 2, 3. Customers in deciles 1–3 of the Not Contacted until 2004 group were identified as those having high purchase propensity for at least one of the three product categories. Such a reallocation implied that customers with higher CLV will be given higher priority to be allocated resources first. The level of resources to be allocated was determined based on the optimum contact strategy described in Phase 2 (Table 8.4). This reallocation process resulted in some customers in decile 1, all customers in deciles 2 and 3 in the Not Contacted group being allocated marketing resources for 2005.

## Phase 5: Implement field study and measure the impact


As a result of an improved targeting strategy, the revenue of the Not Contacted until 2004 group increased 10 times in 2005 compared to revenues in 2004. The lift in revenues for the Not Contacted until 2004 but Contacted in 2005 group of customers was about \$19.2 million. The incremental revenue due to resource reallocation (after adjusting for the annual growth in customer revenue) among the Not Contacted until 2004 but Contacted in 2005 group of customers was about \$19.1 million.

This incremental value was derived from two sources: (a) \$7.6 million (nearly 40%) was obtained from the increase in purchase amount from customers who were active in 2004, and (b) \$11.4 million (nearly 60 %) was obtained from the reactivated customers (about 273 customers) who were dormant in 2004. Therefore, the average increase in revenue from reactivating dormant customers was about \$41 758, and the average increase in revenue from existing customers was about \$4160. The effectiveness of this CLV model was reflected in the superior performance of the sales revenue metric. The improved profitability for IBM was made possible by the successful implementation of the CLV-based strategies.

## 8.4 CRM implementation at a B2C firm

A similar CLV management framework was also be implemented at a fashion retailer in a B2C setting. Like the IBM case study, this case study also highlights the power of the CLV metric and its related strategies in maximizing the retailer's profitability. This also showcases that CLV scores can be measured and effectively applied by the retailer to maximize both customer and store profitability. For instance, by performing segmentation, profile, and impact analysis in conjunction with the CLV score of the customers, retailers can uncover valuable customer-level insights thereby enabling them to deploy various customer management strategies. Similarly, CLV computation of customers by different stores can enable retailers to uncover some interesting and often counter-intuitive store-level insights leading to store management strategies.

### 8.4.1 The focal firm background

The focal firm mentioned in this study is a fashion retailer, whose name is not revealed for confidentiality reasons. The retailer, which sells apparel, shoes, and accessories for both men and women, has a chain of 30 stores across the USA, with a relatively larger concentration of stores on the east and west coasts. Each of the stores was more or less of the same size and located in regions having similar demographics, thus there is no need to normalize the store's profit potential for these factors.

### 8.4.2 Implementing the CLV management framework at a fashion retailer

In essence, the key objective of the retailer was to maximize its profitability. Specifically, the study sought to answer the following research questions for the fashion retailer in particular and retailers in general [4]:

1. What is the right metric to manage customer programs, for example, customer loyalty programs? Can CLV outperform traditional metrics?

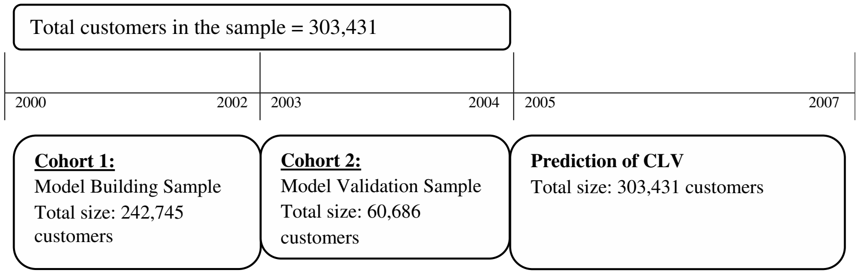



*Figure 8.5 Data analysis timeline.*

2. How can the CLV concept be applied to *measure* and *manage* customer value?

3. How can the CLV concept be applied to *manage* store performance?

The customer data from the fashion retailer were utilized for this study. The total dataset comprised 303 431 customers after excluding 4% outliers from the original dataset that were consistently very high- or very low-value purchasers and probably would skew the sample's distribution. These 303 431 customers were then divided into two cohorts, as shown in Figure 8.5. Cohort 1, called 'Model Building Sample,' consisted of customers who made at least one purchase prior to December 31, 2003. Cohort 2, called 'Model Validation Sample,' consisted of customers who made at least one purchase prior to December 31, 2004 and who are not included in Cohort 1. Cohort 1 was used to calibrate the model and Cohort 2 was used to validate the calibrated model. There were 242 745 customers in Cohort 1 and 60 686 customers in Cohort 2.

The process adopted to develop the CLV-based framework for maximizing the retailer profits is illustrated in Figure 8.6. The idea was to use CLV as a basis for comparing the profit each customer brings to the firm. The aggregate profits all customers bring to a firm constitute its overall profit level. Therefore, to maximize the firm's overall profitability, the task was to implement marketing strategies that can maximize the profitability of each individual customer, or maximize the CLV score of each customer.

The framework provided in Figure 8.6 can be classified into three main stages: (1) model development, (2) model implementation, and (3) tactics and strategy recommendations. The following subsection reviews all these three stages and explains the process adopted in implementing this approach.

### 8.4.3 Process to implement the CLV management framework at a fashion retailer

A clear picture of the process to implement the CLV management framework at the fashion retailer is provided in Figure 8.7. There are three main stages. In the first stage, just like the IBM case study, a CLV model was proposed and demonstrated

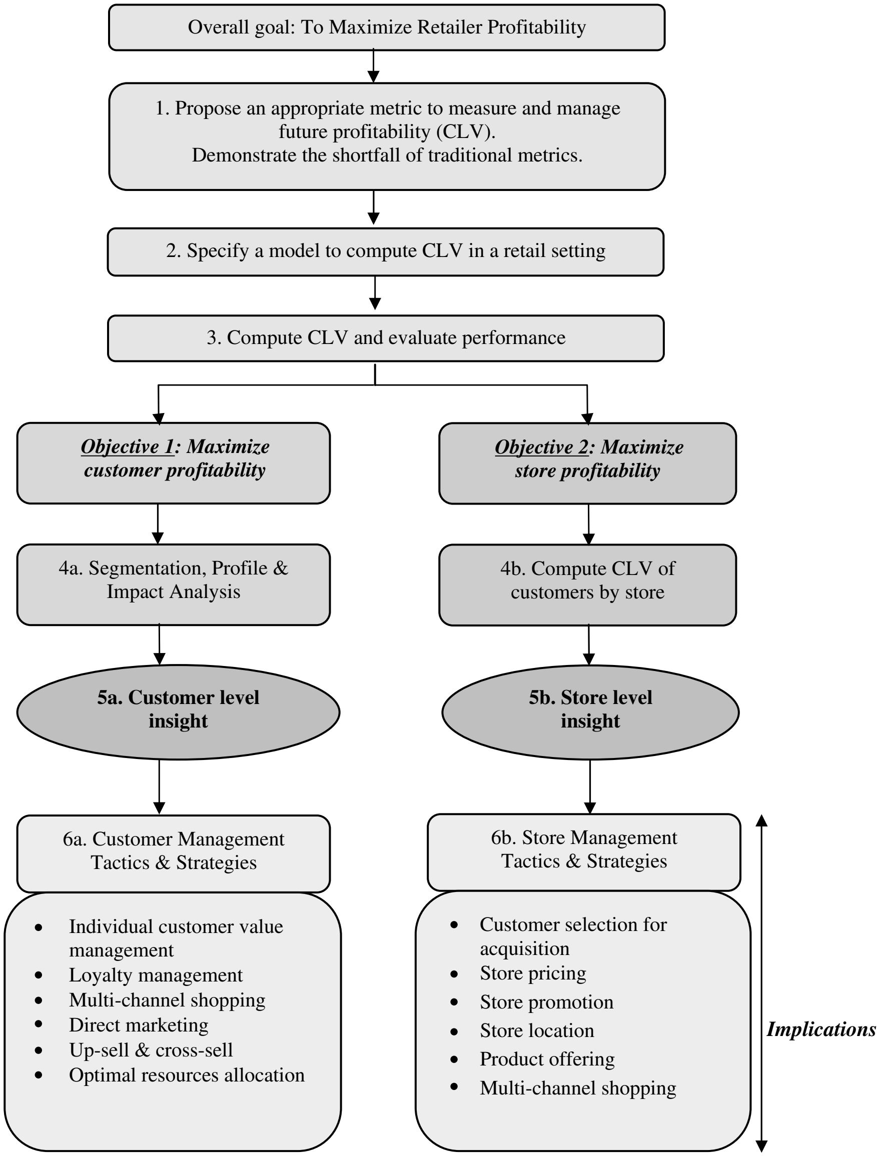



*Figure 8.6 A CLV-based framework for maximizing retail profits.*

as the most suitable model to measure future profitability for the retailer. In the second stage, based on the proposed CLV model, the CLVs of individual customers were measured, based on which the analyses of customers and stores were conducted. In the last stage, tactics and strategic recommendations were made based on the result of findings in the previous stages.

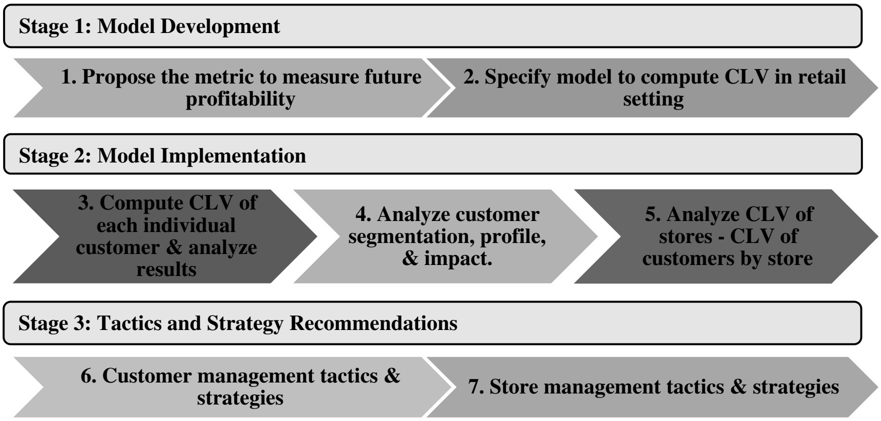


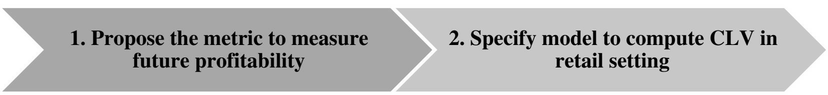

*Figure 8.7 CLV management framework implementation process used at the fashion retailer.*

### 8.4.3.1 Stage 1: Model development

The model development stage involves two phases:

- *Phase 1.* Propose the metric to measure future profitability.
- *Phase 2.* Specify model to compute CLV in retail setting.

**Phase 1: Propose the metric to measure future profitability**




In this phase, the specific goals were to: (1) demonstrate the shortfall of traditional metrics, and (2) propose the CLV model as the more suitable model to measure the future profitability of a firm. The backward-looking characteristics of the traditional metrics were addressed, with specific reference to poor loyalty and profitability relationships shown by past research. Also, the forward-looking characteristic of the CLV metric was highlighted because it incorporated the future values/profitability that a customer can bring to a firm. Finally, the CLV metric was suggested as the superior metric to measure customer value and firm profitability.

**Phase 2: Specify model to compute CLV in retail setting**

In this phase, the desired output was the development of the CLV computation model. The *always-a-share* approach was adopted due to the non-contractual nature of the retailing sector. Based on this approach, a CLV model was formulated using discounted future value methodologies, which incorporated predictions of three key components, namely, purchase frequency, contribution margin, and marketing cost, as follows:

$$CLV_i = \sum_{t=1}^{T_i} \frac{GC_{i,t}}{(1+r)^{t/frequency_i}} - \sum_{l=1}^{n} \frac{\sum_m c_{i,m,l} \times x_{i,m,l}}{(1+r)^l} \tag{8.10}$$

where:

$GC_{i,t}$ = the gross contribution from customer $i$ on purchase occasion $t$

$c_{i,m,l} \times x_{i,m,l}$ = the marketing cost (unit marketing cost × number of contacts) to customer $i$, in channel $m$, in time period $l$

$frequency_i$ = the purchase frequency, computed as $12/expint_i$ because months are used as the units of analysis ($expint_i$ is the expected interpurchase time for customer $i$)

$r$ = discounted rate for money

$n$ = number of years to forecast

$T_i$ = total number of purchases made by customer $i$

$i$ = index for customer

$l$ = index for time period

$t$ = index for time

$m$ = index for channel

#### 8.4.3.2 Stage 2: Model implementation

In this stage, the suggested CLV model was used to compute the actual CLV score of each individual customer, and conduct analyzes of the customers and stores, with the goal of maximizing the profitability of the retailer. Specifically, it involves the three following phases:

- *Phase 3.* Compute CLV of each individual customer and analyze results.
- *Phase 4.* Analyze customer segmentation, customer profile, and customer impact.
- *Phase 5.* Analyze stores – compute CLV of customers by store.

**Phase 3: Compute CLV of each individual customer and analyze results**


This phase involved the prediction of the three main components of CLV model – purchase frequency, contribution margin, and marketing cost – by using the generalized gamma distribution, panel-data regression, and discounted future value methodologies respectively. Below is the detailed explanation of each step of this phase.

*(a) Predicting the purchase frequency for each customer* To get the desired output, prediction of the purchase frequency for each customer, the purchase frequency was modeled using the likelihood function, generalized gamma distribution, and inverse generalized gamma distribution.

*The likelihood function for purchase frequency model* This model is given by

$$L = \prod_{i=1}^{n} \prod_{j=1}^{J_i} \prod_{k=1}^{K} \Phi_{ijk} f_k (t_{ij}|\alpha_k, \lambda_{ik}, \gamma_k)^{c_{ij}} S_k \times (t_{ij}|\alpha_k, \lambda_{ik}, \gamma_k)^{(1-c_{ij})} \tag{8.11}$$

where:

| | | |
|---|---|---|
| $f(t_{ij} \mid \alpha, \lambda_i, \gamma)$ | = | the density function for the generalized gamma distribution, in other words the probability of the *j*th purchase for customer *i* occurring at time period *t*, given $\alpha$, $\lambda_i$, $\gamma$. |
| $S(t_{ij} \mid \alpha, \lambda_i, \gamma)$ | = | the survival function for the generalized gamma distribution, in other words the probability of the *j*th purchase for customer *i* occurring at a time period greater than *t*, given $\alpha$, $\lambda_i$, $\gamma$. |
| $c_{ij}$ | = | the censoring indicator, where |
| $c_{ij}$ | = | 1: if the *j*th interpurchase time for the *i*th customer is *not* right-censored |
| $c_{ij}$ | = | 0: if the *j*th interpurchase time for the *i*th customer is right-censored |
| $\Phi_{ijk}$ | = | the probability of observation *j* for the *i*th customer belonging to subgroup *k* |
| $\alpha$, $\lambda_i$, $\gamma$ | = | the parameters of the generalized gamma distribution |

*The expected time until the next purchase model* According to Equation 8.10, purchase frequency was computed as $12/expint_i$ because months were used as the units of analysis, where $expint_i$ was expected interpurchase time for customer $i$, or the expected time until the next purchase. Given that the generalized gamma distribution was used to model the interpurchase time and the likelihood function in Equation 8.11, this expected time until the next purchase would be given by

$$\sum_k \Phi_{ik} \left\{ \frac{[\Gamma(\alpha_k + (1/\gamma_k))]}{\Gamma(\alpha_k)} \right\} \lambda_k \tag{8.12}$$

where:

$\alpha, \gamma$ = the parameters that establish the shape of the interpurchase time distribution

$\lambda_i$ = the individual specific purchase rate parameter, which in the generalized gamma distribution is assumed to be a random draw from an inverse generalized gamma distribution (IGG), specifically

$$\lambda_i \sim IGG(\upsilon, \theta, \gamma) = \frac{\gamma}{\Gamma(\upsilon)\theta^{\upsilon\gamma}} \gamma_i^{-\upsilon\gamma-1} (\exp(-1/\theta\gamma_i))\gamma$$

where:

$\upsilon, \gamma$ = the shape and scale parameters of the IGG distribution respectively

$\gamma$ = common parameters to both the generalized gamma distribution and the IGG distribution

$k$ = the number of subgroups the population was assumed to consist of

$\Phi_{ik}$ = the probability of a customer belonging to each subgroup, and also provided the mass point (i.e., weight) for each subgroup. $\Phi_{ik}$ was modeled as a probit function of the antecedents and covariates of the purchase frequency. Specifically, the link function can be represented as $\Phi_{ik} = f(x_{ij}\beta_i)$, where $x_{ij}$ are the antecedents and covariates of purchase frequency for customer $i$ on purchase occasion $j$, and $\beta_i$ are the customer-specific response coefficients.

The model framework presented in Equation 8.12 resembles a hierarchical Bayes formulation of the concomitant continuous mixture model. In order for the issue of endogeneity to be addressed, the one-period lagged value for all antecedents and covariates was used in the analysis [5,6]. In addition, the performance of the proposed model framework was also evaluated with several alternative model formulations such as: (a) the simple generalized gamma distribution model, (b) the continuous mixture model, and (c) the latent class mixture model. Upon evaluation, the proposed model framework outperformed all the alternative model

Table 8.6 Correlation of observed future profitability and different measures of loyalty.

| | Consistency of purchase frequency | RFM | Relationship duration |
|---|---|---|---|
| CLV | $R = -0.10$ | $R = 0.25$ | $R = 0.37$ |

specifications. Finally, to account for any extraneous factors not accounted for by the antecedent and covariate set, the logarithm of the lagged interpurchase time was used. The specification of the model allowed the individual customer-level coefficients to be estimated for the influence of the various covariates on the probability of a customer belonging to a particular subgroup, hence the interpurchase times.

*Results and discussion of findings* The finding showed a weak correlation between loyalty and observed future profitability. As given in Table 8.6, the correlation coefficients between observed future profitability and consistency of purchase frequency and RFM were both less than 0.3, while the correlation coefficient between observed future profitability (OFP) and relationship duration was the highest, 0.37. This result reinforced the argument that the retailer cannot afford to use traditional loyalty metrics to manage the customer relationship, which were backward looking. To manage both loyalty and profitability simultaneously, the retailer needed to adopt a forward-looking metric such as the CLV metric.

*(b) Predicting the contribution margin for each customer*

*Gross contribution model* The gross contribution for each customer was modeled using the panel-data regression method. Gross contribution margin was defined as the revenue that the customer provides to the firm, whenever a purchase is made, minus the cost of goods sold. While the cost of goods sold does not change very much over time, the revenue a customer provides is likely to change over time. As a result, the gross contribution model was given by

$$GC_{i,j} = \beta_0 + \sum_{k=1}^{n} \beta_k x_{i,j-1} + e_{i,j} \tag{8.13}$$

where:

$GC_{ij}$ = the gross contribution for customer $i$ on purchase occasion $j$, measured in dollars

$X_{i,j-1}$ = the independent variable relevant to customer $i$ on purchase occasion $j-1$

$e_{i,j}$ = the error term

$n$ = the number of independent variables

$k$ = the index for independent variables

$i$ = the index for the customer

$\beta_0, \beta_j$ = the coefficients

The performance of the proposed model framework, Equation 8.13, was evaluated with various alternative functional forms for gross contribution margin model, including (a) using log-margin as a dependent variable, and (b) a polynomial specification for the lagged dependent variables to allow for nonlinearity. The specification of Equation 8.13 provided the best in-sample fit and predictive accuracy. Any issues related to endogeneity were addressed by using lagged endogenous variables wherever necessary. The results were tested for multicollinearity using the variance inflation factor and eigenvalues of the principal component analysis of the predictors.

*(c) Results and discussion of findings* The result of the gross contribution model is provided in Table 8.7. This model provided a good model fit ($R^2 = 0.71$) and the mean absolute deviation (MAD) was 65 for the estimation sample and 70 for the holdout sample. The existence of a unit root in the gross contribution provided by a customer was rejected by the Dickey–Fuller test. On comparison, the proposed model outperformed the log-linear model. In addition, no significant nonlinear effects were observed for any of the parameters in the dataset.

Table 8.8 summarizes the results for the frequency model. The model had a log marginal likelihood (LMD) of $3.4 \times 10^4$ and MAD = 2.4.

Table 8.9 provides a comparison of the performance of the above-proposed frequency model and alternative competing models described in earlier sections. As illustrated in Table 8.9, the proposed frequency model provided a better LMD and forecasting performance than the competing models. In essence, the concomitant mixture framework appears to be better suited to model abrupt changes in the interpurchase times, which was observed in the dataset.

Table 8.7 Results of the gross contribution model.

| Variables | Standardized coefficients |
|---|---|
| Lagged contribution margins | $0.83^a$ |
| Lagged cumulative revenue from other channels | $0.76^a$ |
| Amount ($) spent on product category A on previous purchase occasion | $0.61^b$ |
| Time elapsed since last purchase | $0.11^a$ |
| Square of time elapsed since last purchase | $-0.88^b$ |

[a]Significant at $\alpha = 0.01$ level.
[b]Significant at $\alpha = 0.05$ level.

Table 8.8 Generalized gamma purchase frequency model.

| Variables | Parameter estimates |
|---|---|
| Cross-purchase | 6.9[a] |
| Number of distinct channels | 6.1[b] |
| Number of returns | 3.1[a] |
| Number of returns square | −1.7[b] |
| Log of lagged interpurchase time | −5.2[a] |

[a]Posterior sample values between the 2.5th and 97.5th percentile do not contain zero.
[b]Posterior sample values between the 0.5th and 99.5th percentile do not contain zero.

Table 8.9 Comparison of proposed model with competing purchase frequency models.

| Variables | The proposed model | Model A (simple model) | Model B (continuous mixture model) | Model C (latent class mixture model) |
|---|---|---|---|---|
| LMD | $-3.4 \times 10^4$ | $-4.7 \times 10^4$ | $-4.1 \times 10^4$ | $-3.9 \times 10^4$ |
| MAD | 2.4 | 4.2 | 3.5 | 3.1 |

*(d) Predicting the marketing cost for each customer* There are several ways to compute the marketing cost for each customer. For example, it may be calculated as an aggregate measure by dividing the total marketing budget by the number of customers. A more sophisticated approach would entail calculating the total marketing cost separately for each customer based on the various marketing channels expected to be used to interact with that customer. This approach to computing the marketing costs is given by

$$MC_i = \sum_{l=1}^{n} \frac{\sum_m c_{i,m,l} x_{i,m,l}}{(1+r)^l} \tag{8.14}$$

where:

$MC_i$ = the total marketing cost for customer $i$

$c_{i,m,l}$ = the unit marketing cost for customer $i$, in channel $m$, in time period $l$

$m$ = the marketing channel (for this retailer it was the Web and catalog)

$r$ = the discount rate

$l$ = the index for time

$n$ = the number of years to forecast

Equation 8.14 gives retailers the means to calculate the marketing cost not only on an individual customer basis, but also on the basis of individual marketing channels adopted for the same customer. This is very useful given the fact that direct marketing costs could vary widely across customers and across communication channels. For example, digital communications would cost much less than direct mailing. Furthermore, with the increasing trend of multi-channel marketing to the same customer, computing marketing costs as shown in Equation 8.14 can help retailers accurately arrive at a fair estimate of the expected marketing cost per customer. Accordingly, the expected marketing cost was discounted by $l$ years to arrive at the present value of market cost.

For the purpose of computation, it was assumed that the direct marketing cost for each customer will be the same for the next three years. This assumption was applied because (1) the customer historical database shows that the historical direct marketing cost per customer remained more or less constant for the retailer, and (2) the retailer's directives were followed strictly. According to the retailer, the best prediction of future marketing cost of each customer over the next three years could be taken as three times the direct marketing cost in the most recent year (i.e., 2004). In this particular case study, the available marketing cost in the retailer's database was the total marketing cost (across all channels) for each customer.

After the three main components of the CLV model were predicted, Equation 8.10 was applied to calculate the CLV score of each individual customer.

### Phase 4: Analyze customer segmentation, customer profile, and customer impact


The desired output of this phase was to answer the following questions:

- Which of the demographics, life style, and shopping behavior vary significantly across the high and low CLV segments?
- What is the impact of contribution margin and purchase frequency on high CLV segments of customers?

*(a) Analyzing customer segmentation* Here, the CLV scores computed in Phase 3 were ranked in descending order and grouped into deciles so that each decile represents 10% of the customer base. Figure 8.8 shows the distribution of CLV scores across 10 deciles.

The result provided interesting insights. Most companies believe in the Pareto principle, or the 80:20 rule, which indicates that the top 20% (or 30%) of the customers typically generate 80% (or 70%) of the revenue or profits [7]. However, contrary to this belief, as shown in Figure 8.8 the top 20% (decile 1) were actually accounting for 95% of profits. This is because the bottom 30% had negative CLV.

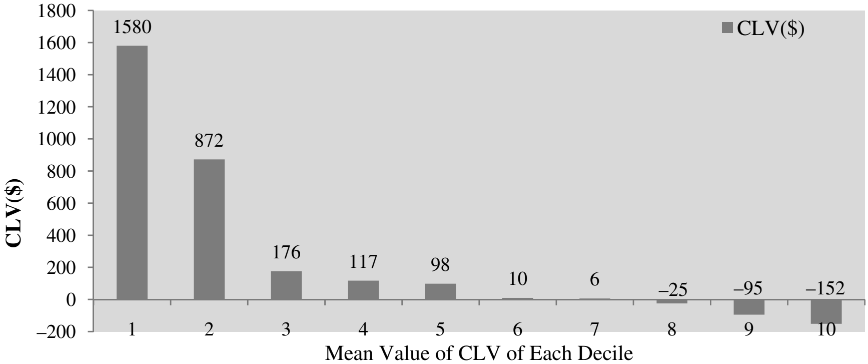



*Figure 8.8 Distribution of CLV scores across deciles (*N = *303 431).*

Based on the above distribution of CLV scores, the entire customer group was divided into three segments: high CLV (comprising deciles 1 and 2), medium CLV (comprising deciles 3, 4, and 5) and low CLV segments (comprising deciles 6, 7, 8, 9, and 10). It should be noted that this form of segmentation is sensitive to the CLV scores of individual customers, thus the relative size of the segments will be subject to change over time if there are any changes to the CLV distribution.

*(b) Analyzing impacts of CLV drivers on CLV scores of customers* In this section, the impact of CLV drivers on the three segments of high, medium, and low CLV was measured in the three following steps.

First, the key drivers of CLV (as outlined in Table 8.10) were selected to assess their impacts on CLV for the high CLV segment of customers. Second, the logistic regression method was used to determine which of these variables in the dataset (variables related to demographic, life style, and shopping behavior (DLSB)) varied significantly across the high and low CLV segments. In this case study, the medium CLV segment was excluded from the analysis to highlight the main difference between a high CLV and a low CLV customer. However, a similar procedure can be extended for medium CLV customers also if need be.

The logistic regression method employed here is

$$Prob_i = \frac{1}{1 + \exp(-Y_i)}, \quad \text{where } Y_i = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ik} \tag{8.15}$$

where:

$Prob_i$ = the probability of customer $i$ to be in high or low CLV segment

$Y_i$ = the outcome of regressing $X_{ik}$ on high/low CLV segment indicator

$X_{ik}$ = the '$K$' DLSB (predictor) variables measured for customer $i$

$\beta_0$, $\beta_k$ = the coefficients estimated from the data

Table 8.10 Variables included for predicting purchase frequency and/or gross contribution.

| Variables | Operationalization | Expected effect on | |
|---|---|---|---|
| | | Purchase frequency | Gross contribution |
| Cross-buying | Number of different product categories a customer has purchased | $+$ | N/A |
| Returns | Number of products the customer returns on average between two observed purchases | $\cap$ | N/A |
| Purchase of specific product category | Indicator variables to indicate purchase of specific product category of item | N/A | $\pm$ |
| Multi-channel shopping behavior | Cumulative revenue from other channels (Web and catalog purchase) | N/A | $+$ |
| | Number of distinct channels used for transactions | $+$ | N.A. |
| Time elapsed between successive purchases | Time duration between the current purchase and the most recent purchase | N/A | $\cap$ |
| | Lagged interpurchase time | $-$ | N/A |

This logistic regression method offered a simple procedure to estimate the coefficients by employing maximum likelihood.

Third, this model was run separately for the high and low CLV segments. In the model for the high CLV segment, the dependent variable was set as 1 if the customer belonged to the high CLV segment, or else 0. Similarly, in the model for the low CLV segment, the dependent variable was set as 1 if the customer belonged to the low CLV segment, or else 0. The findings from this analysis were then used to draw corresponding implications for customer profile analysis.

Finally, to assist the retailer's managers and develop strategies to increase the CLV of each customer, the impact of changing the values of some of the drivers of CLV in Table 8.10 was measured. High CLV customer segment and certain customer behavior-related variables were chosen for this measurement. Specifically, the magnitude of each driver was increased by 15% and the corresponding increase in CLV in each customer was evaluated. In practice, managers can make a similar impact on the value of each driver through appropriate marketing interventions, that is, promoting cross-buying behavior by offering customized incentives to customers to purchase across different product categories. The results presented in Figure 8.9 show that a 15% change in these key drivers significantly impacts the CLV of

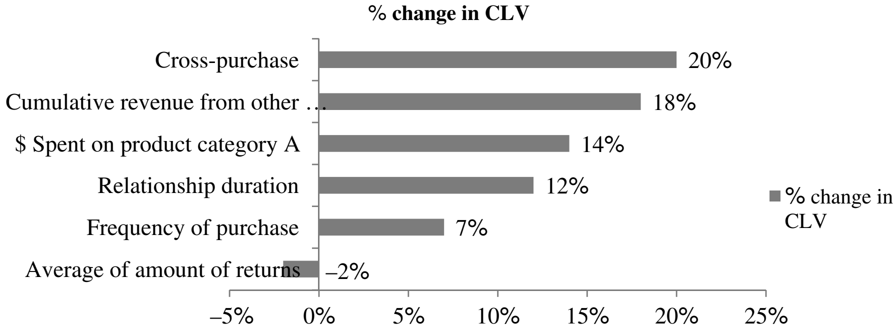



*Figure 8.9 Impact on CLV corresponding to change of select customer behavior-related variables.*

customers. These findings provide useful implications for customer and store management strategies.

*(c) Analyzing customer profile* Based on the findings from the impact analysis, several group-level differences were identified. For instance, most high CLV customers were professionally employed and married females in the age range of 30–49 years old with at least one child, having high household income and the store's loyalty card. They also shopped in many channels and lived relatively close to the store. Low CLV customers, on the other hand, were primarily single men, in the age range of between 24 and 44 years old, having relatively low income and not necessarily the store's loyalty card. They shopped in one channel and lived further away from the retailer.

One advantage of this profile analysis is that it could help the retailer put a 'face' to the CLV score of the customer, which would give useful insights in acquiring new customers. This method also overcomes the shortcomings of most CLV studies, which do not include demographics and product usage variables [8].

**Phase 5: Analyze CLV of stores**


The desired output of this phase was the CLV of a store. Customers of the retailer shop at several of the 30 store chains nationwide. Thus, it was possible to calculate the CLV of a store based on the CLVs of customers who have ever purchased from that store. This phase was divided into two smaller steps:

- To find the CLV score of customers by store.
- To find the CLV of a store.

The CLV score of customers by store was measured as the weighted average CLV of all customers shopping at that store. For example, suppose that customer 1 made 80% of his/her purchases from store A, 10% from store B, and 10% from store C. This information was used to assign customer value weights to each store. That is, store A can be assigned 80% of CLV from customer 1, 20% from customer 2, 0% from customer 3 (who did not shop at store A at all), and so on. The same procedure was then applied for all the stores.

The CLV of a store was measured as the weighted sum of the net present value of the CLVs of all customers that shop at that store minus the present value of the rent for the store. It should be noted that this method was based on the current set of customers, and did not take into account new customers that might be acquired by the store in the future.

After all values of store CLVs were computed, all the stores were ranked according to both their past three-year profitability and future three-year profitability (store CLVs) (see Table 8.11). There was no need to normalize the stores' profit potential for the difference in size, location, and demographics, as they were more or less similar in terms of these factors. Finally, Spearman's correlation coefficients between the past and future profitability ranks were computed to draw corresponding implications for store management tactics and strategies.

Table 8.11 Comparison of store performance based on revenue and profitability.

| Store number | Store revenue ($ million) | Revenue rank | Store profitability based on CLV of the store's customers ($ million) | Profitability rank |
|---|---|---|---|---|
| 1 | 13.45 | 1 | 2.36 | 4 |
| 2 | 7.90 | 2 | 1.23 | 12 |
| 3 | 6.50 | 3 | −3.21 | 29 |
| 4 | 5.79 | 4 | 5.46 | 1 |
| 5 | 4.32 | 5 | 1.84 | 6 |
| 6 | 4.20 | 6 | −4.18 | 30 |
| 7 | 3.38 | 7 | 3.33 | 2 |
| 8 | 2.86 | 8 | 2.55 | 3 |
| 9 | 2.78 | 9 | 0.12 | 20 |
| 10 | 2.58 | 10 | 0.73 | 17 |
| 11 | 2.56 | 11 | 0.79 | 16 |
| 12 | 2.51 | 12 | 2.18 | 5 |
| 13 | 2.50 | 13 | −0.26 | 22 |
| 14 | 1.90 | 14 | 1.45 | 10 |
| 15 | 1.88 | 15 | −2.33 | 28 |
| 16 | 1.85 | 16 | 1.72 | 7 |
| 17 | 1.70 | 17 | −0.27 | 23 |
| 18 | 1.69 | 18 | 1.62 | 8 |

| | | | | |
|---|---|---|---|---|
| 19 | 1.52 | 19 | 1.00 | 15 |
| 20 | 1.33 | 20 | 1.29 | 11 |
| 21 | 1.24 | 21 | 1.58 | 9 |
| 22 | 0.99 | 22 | 1.05 | 13 |
| 23 | 0.98 | 23 | 0.56 | 18 |
| 24 | 0.98 | 24 | 1.01 | 14 |
| 25 | 0.92 | 25 | −0.35 | 24 |
| 26 | 0.86 | 26 | −1.10 | 27 |
| 27 | 0.72 | 27 | 0.31 | 19 |
| 28 | 0.64 | 28 | −0.20 | 21 |
| 29 | 0.46 | 29 | −0.80 | 25 |
| 30 | 0.27 | 30 | −0.84 | 26 |

#### 8.4.3.3 Stage 3: Tactics and strategy recommendations

In this final stage, the result findings were used to suggest suitable tactics and strategy recommendations. The stage includes two phases:

- *Phase 6.* Customer management tactics and strategies.
- *Phase 7.* Store management tactics and strategies.

**Phase 6: Customer management tactics and strategies**


The following managerial implications were suggested for the management of the retailer:

- *Paradigm shift in management.* The case study suggested that the CLV metric facilitated a paradigm shift in doing business by shifting the emphasis from *managing customer relationship* to managing *customer value.* This is because, while customer relationships are critical for all firms, they are just the means for the firms to achieve the overall goal of maximizing customer profitability. Managing customer value will directly have an impact on the firms' profitability.
- *Customer value management.* The result findings showed that as much as 30% of the retailer's customers had a negative CLV. Thus, it is important to

obtain customer-level insights to design marketing strategies unique to each customer in order to maximize the value of the customer.

- *Loyalty management.* Given the weak relationship between the profitability and loyalty of customers, firms should adopt forward-looking metrics, such as the CLV metric, rather than backward-looking metrics in the management of both loyalty and profitability.

- *Multi-channel shopping.* The customer impact analysis revealed that a 15% increase in spending from other channels (e.g., through the Web and catalogs) by customers in the top two deciles leads to an 18% increase in their CLV scores. Therefore, it is advisable for the retailer to promote multi-channel shopping options to customers.

- *Direct marketing.* The customer profile analysis provided a clear snapshot of typical high and low CLV customer profiles. It also allowed the retailer to prioritize and target its sales and promotion campaigns effectively. For example, since the probability of finding less profitable customers, bargain hunters, is higher in the low CLV segment, the retailer should target its sales and promotional campaign to low CLV segments to avoid cannibalization of its business.

- *Up-sell and cross-sell.* Cross-purchase was a significant driver of the customer CLV that the retailer's managers should consider. In customer impact analysis, a 15% increase in cross-purchase of customers in the top two deciles resulted in the highest increase in their CLV, namely, 20%. Thus it was recommended that the retailer promote up-selling and cross-selling to customers that have low CLVs but have profiles similar to high CLV customers.

- *Optimal resource allocation.* Based on customer-level CLV score, the retailer was advised to set a ceiling on the dollar value for investing marketing resources in each customer to avoid overspending.

**Phase 7: Store management tactics and strategies**


The retailer was advised to incorporate CLV metrics to store performance in order to increase its overall profitability. Two main management strategies were recommended:

- *Identify the right customers for each store to target* Based on customer profile analysis, managers should target or compare prospective customers to the high and low CLV customer profiles to make a decision on the right level of resources to acquire those customers.

- *Determine the right marketing mix (4* Ps) to maximize the retailer's profitability. Recommendations were made for the following four strategies: pricing, promotion, location, and product, as follows. For the pricing strategy, the retailer should align price setting with the retailer's CLV structures, distribution, and characteristics, such as the price sensitivity of the high, medium and low CLV customers, the possibility of market cannibalization if the price for products typically purchased by high CLV customers was lowered, and so on. For the promotion strategy, the retailer was advised to align promotion with the retailer's profiles of the high, medium, and low CLV segments (such as responsiveness to in-store display and direct marketing, price sensitivities, special promotions, clearance sales, etc.). For the location strategy, since prime location might not translate into high profitability for a store, the retailer should evaluate a typical profile of high CLV customers and find a location that has a relatively high density of the targeted customers. For the product strategy, the retailer was advised to study the types of products that high, medium, and low CLV customer will buy, and target the right customer group accordingly. Importantly, it should ensure that a wide range of shopping options are available to customers.

## 8.5 Challenges in implementing the CLV management framework

We have looked at the development and implementation process of the CLV management framework at two firms, in two business settings, B2B and B2C. However, when it comes to practical aspects of the implementation, project managers have to face many other challenges. Most importantly, more challenges will be faced by the management of the focal firm in implementing the recommended strategic implications. In this last section of the chapter, we will address those challenges and suggest viable approaches that firms can apply to overcome such issues.

### 8.5.1 Challenges in data collection and internal collaboration

As technology advances, firms have access to a much wider variety of customer-level data necessary for CLV computing. Some of such key information includes demographics, firmographic data, purchase value, products purchased on each occasion, the number, time, and types of marketing contacts, prospects, and competitors' customers. However, not all firms have access to all of the information they need, especially information regarding prospects and competitors' customers. In the meantime, given the complexity of business settings nowadays, many firms find it challenging to identify the right informational needs, to collect the huge amount of data, and to utilize these data. Sometimes, in order to get the right information, managers will need to resort to collaborative efforts from multiple departments. Failure to properly address these issues might lead to unreliable result findings or delays in implementation, which are costly to the firm.

To overcome these challenges, before starting to collect information, firms should be clear in terms of their informational needs. Firms should also be resourceful in making use of the available secondary information sources and data vendors. Making use of customer data intermediaries (CDIs) – firms specializing in collecting customer behavioral and demographic data – is also a viable option. A more long-term approach would be to invest in an effective data management system internally. To effectively coordinate collaborations across the firm, management must ensure that the CRM initiatives are well communicated to the relevant parties. Any delays or discrepancies should be promptly communicated so that timely rectification can be made.

## 8.5.2 Challenges in implementing the customer-centric approach

The findings in both case studies suggest several valuable implications for maximizing the firms' profitability. In implementing those recommendations, the management of each firm have to make certain systematic changes. The most critical change would be to shift the firm's approach from product-centric to customer-centric [9].

Traditionally, firms focus on the product-centric approach. That is, selling products to whoever is willing to buy them. However, given the increasing competition among firms, customers who are supposedly satisfied have many reasons to defect and move on. The key for firms to maximize their profitability is to maximize their individual customers' profitability, which can only be done by focusing on the customers' needs. This requires firms to shift their focus from the product level to the customer level. Several firms such as Wells Fargo, Apple, and so on, have moved away from a product-centric to a customer-centric approach. This fundamental shift requires important changes in the operations element and changes to the workforce. Table 8.12 highlights the fundamental differences between the product-centric and customer-centric approaches.

To ensure a customer-centric approach to CRM initiatives, changes pertaining to two crucial areas of the organization will have to be addressed: operational elements and workforce elements [10].

### 8.5.2.1 Changes to operational elements

To effectively implement the customer-centric approach, firms have to focus on interaction orientation [11]. When a firm adopts interaction orientation, its marketing activities are conducted with the customers. Each customer is viewed both as a source of business and as a potential business resource to the firm. The strategic importance of customer-to-customer linkages is recognized and included in the customer empowerment component. Examples of such incorporation of customer empowerment in firms' marketing activities are as follows:

- Making decisions on a per customer point of view.
- Providing rapid responses to customer needs, creating a rich customer experience.

Table 8.12 Differences between the product-centric approach and the customer-centric approach

| Product-centric approach | Customer-centric approach |
|---|---|
| Ideology is to sell products from a portfolio of products; and business is based on transactions | Ideology is to serve a portfolio of customers; and business is based on relationship |
| How many customers can we sell this product to? | How many products can we sell to this customer? |
| Firm is internally focused on product development and market share | Firm is externally focused on customer relationships and customer profitability |
| Firm highlights product features and advantages | Firm highlights product benefits satisfying customer needs |
| Firm is structured based on profit centers and sales performance | Firm is structured based on customer segments and customer relationship |
| Firm performance is measured by profitability per product and market share | Firm performance is measured by customer satisfaction, CLV, and customer equity |

- Allowing customers to exchange information and reviews about product and experiences with other customers.
- Encouraging customers to connect with the firm and design the nature of transactions.

With regard to internal operations, firms are required to realign their structures to successfully adopt the customer-centric approach. The management of these firms need to revisit their business dimension and shift all their performance criteria from product based to customer based. Following the changes in structures, and internal performance metrics, firms should tailor the offerings that focus on the customers, rather than on the products. Product positioning should highlight the benefits in terms of meeting individual customer needs, rather than simply good product features. Specifically, firms should implement the following changes:

- Customer segment centers, customer relationship managers, and customer segment sales teams will replace product profit centers, product managers, and product sales teams.
- Performance metrics should be customer share-of-wallet, customer satisfaction, CLV, and customer equity, instead of profitability per product, market share per product, or number of new products.
- Firms should ask ‘How many products can we sell to this customer?’, rather than ‘How many customers can we sell this product to?’’

#### 8.5.2.2 Changes to workforce elements

The workforce of a firm is the most valuable asset, as the success of any CLV implementation depends on how well it is executed by the staff of the firm. The last step in ensuring the implementation process can be effectively conducted is to make key changes to workforce elements.

First of all, top management must support the integration of all corporate functions to focus on customer value. The management should understand the importance of the fundamental changes to the customer-centric approach, and make sure it is communicated to the staff. Finally, the management have to ensure employees' interest in participating and supporting the change by communicating the initiative's effectiveness and potential benefits. Prioritization of customer needs should be made and any barriers should be removed timely through constant monitoring and evaluation efforts.

## References


1. Kumar, V. (2008) *Managing Customers for Profit: Strategies to Increase Profits and Build Loyalty*, Prentice Hall, Upper Saddle River, NJ.
2. Kumar, V., Venkatesan, R., Bohling, T., and Beckmann, D. (2008) The power of CLV: managing customer lifetime value at IBM. *Marketing Science*, **27**(4), 585–599.
3. Dwyer, F.R. (1997) Customer lifetime valuation to support marketing decision making. *Journal of Direct Marketing*, **11**(4), 6–13.
4. Kumar, V., Shah, D., and Venkatesan, R. (2006) Managing retailer profitability – one customer at a time. *Journal of Retailing*, **82**(4), 277–294.
5. Venkatesan, R. and Kumar, V. (2004) A customer lifetime value framework for customer selections and resource allocation strategy. *Journal of Marketing*, **68**(4), 106–125.
6. Villas-Boas, M.J. and Winer, R.S. (1999) Endogeneity in brand choice models. *Management Science*, **45**(10), 1324–1338.
7. Zeithaml, 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 Journal*, **1**(28), 67–85.
8. Jain, D. and Singh, S. (2002) Customer lifetime value research in marketing: a review and future directions. *Journal of Interactive Marketing*, **16**(2), 34–46.
9. Kumar, V., Venkatesan, R., and Rajan, B. (2009) Implementing profitability through a customer lifetime value management framework. *GfK Marketing Intelligence Review*, **2**(4), 32–43.
10. Kumar, V. (2007) *Customer Lifetime Value – The Path to Profitability*, Now Publisher, Hanover, MA.
11. Ramani, G. and Kumar, V. (2008) Interaction orientation and firm performance. *Journal of Marketing*, **72**(1), 27–45.

# 9

# The future of CRM

## 9.1 Introduction

In the previous chapters of this book we have examined how CRM can be implemented within an organization. The concepts, models, and strategies discussed in these chapters will provide firms with a good understanding of how to acquire high-value customers, how to retain them, how to prevent customers from churning, and how to increase customer win-back rate, all in an effort to enhance firm profitability. Next, firms need to apply this understanding to the actual resource allocation across customer segments, marketing activities, and channels within the organization. As technology develops, organizations have to constantly update themselves with the latest changes and adapt their marketing strategies to remain competitive in the market. Therefore, it is important for CRM managers to be aware of these changes and take them into account when formulating and implementing the marketing strategies [1].

In this chapter, we present three marketing methods that are receiving increased attention from marketers – social media, mobile marketing, and customized campaigns. The chapter is organized as follows. We first discuss the emergence of social media. Then we talk about the influence of mobile channels in implementing CRM initiatives. Finally, we discuss how firms can implement customized marketing campaigns by focusing on developing relationships with individual customers. Popular commercial uses and expected future development trends are also identified for each of the three marketing methods.

## 9.2 Social media

Social media typically cover all interactive media outlets in which consumers can comment on or contribute to the medium's content. As the Internet becomes one of the most important distribution channels, and a major source of customer

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

information and empowerment, social media, also referred to as Web 2.0, are increasingly influencing how customers interact with each other and how businesses use social medial tools to manager customer relationships. Internet users are turning to one another as trusted sources for their purchase decision making. In riding this social media wave, marketers have significantly increased their marketing budgets on social media channels. For instance, Internet advertising spending in the USA increased by 23.1% in the first half of 2011, to $14.9 billion, from $12.1 billion for the same period in 2010, according to Interactive Advertising Bureau. Some of the popular social media channels that marketers are increasingly adopting include blogs, social networks, content communities, and social coupons:

*Blogs* are online journals, and are often a virtual storehouse of information. A corporate blog is a powerful tool for firms to create consumer involvement with their marketing activities and to keep customers informed about the company and its product offerings [2]. A recent addition to the world of blogs is micro-blogging. Micro-blogging sites such as Twitter are also popular social media channels that can reach out to millions of users. Many companies such as General Motors, H&R Block, and Kodak are using Twitter as the primary communication portal for customer service, and the feedback has helped to turn around their company image and improve CRM initiatives. Blogs are often combined with podcasts and video casts, digital audio or video that can be streamed or downloaded to portable devices.

*Social networking platforms* such as Facebook, Myspace, and LinkedIn allow users to create their profile and build a personal network that connects them to other users. The market research firm eMarketer estimates that advertising revenues on social networking platforms would reach nearly $7.72 billion worldwide in 2012, a 71.6% increase compared to 2010. Identifying and targeting micro-segments of consumers, starting dialogs with consumers, and building brand awareness are all benefits that firms can profit from by establishing and managing their presence on one of these platforms. Companies such as Papa John's, RedBull, and Target have used Facebook as a means of creating a new portal into their 'world.' While Papa John's engages in free give-aways for 'liking,' RedBull has essentially transposed its web site onto its Facebook page, and Target has used its page as a platform for promoting its CSR (Corporate Social Responsibility) efforts. All three companies have used distinctly unique approaches, with the same common goal of raising and improving brand awareness.

Due to the proliferation of social media into everyday marketing, researchers have started investigating ways to measure/quantify a user's influence in the social media realm in a manner that would be of use to marketers. For instance, Kumar *et al.* [3] provide a research approach to measure social media ROI and customer's word-of-mouth value by creating the Customer Influence Effect (CIE) metric. This metric measures the net influence wielded by a user in a social network and helps in predicting the user's ability to generate viral information spread. The study also develops another metric –

Customer Influence Value (CIV) – to link word of mouth to the actual sales generated. When this method was implemented at HokeyPokey, a chain of ice-cream retailers in western India, the research findings show that social media can be used to generate sales, increase ROI, induce positive word of mouth, and spread brand knowledge.

*Content communities* are web sites that organize and share particular types of Web-based content such as videos (e.g., YouTube, and Google Videos), photos (e.g., Flickr and Picasa), social bookmarks (e.g., Digg.com and Del.icio.us), publicly edited encyclopedias (e.g., Wikipedia), and public forums (e.g., epinions.com) among others. For instance, Kohl's uses YouTube as a channel to introduce its new product offerings instantly to millions of users. Similarly, H&M has been successful in collaborating with well-known designers to bring the latest chic fashion to its target customers and feature its new collections through YouTube.

*Social coupons* are web sites that offer online daily deals and discounts, such as Groupon and LivingSocial. Such sites have attracted substantial attention from customers and businesses alike. Social coupons provide customers with significant discounts (ranging from 50 to 90% off) while providing small-business owners guaranteed revenue and a flood of new customers. According to Groupon, the firm received a huge demand for its services: about 13 million people have registered to receive its e-mails, and 35 000 businesses are waiting to be featured on its site.

Looking ahead, these forms of social media are expected to continue growing at a rapid rate. Given the existing technological development pace, social media are also expected to enable further collaboration and social connection among customers. Three major changes are expected to take place. First, there will be an increasing interaction between customers and firms. This phenomenon will lead to more direct involvement of customers in many aspects of the firm's business, such as product design, advertising, and branding. To CRM implementation managers, it also means opportunities to acquire customer feedback on firms' products, improved customer relationships, and increased customer loyalty. Second, there will be increased interaction among customers, which will lead to increased prominence of online communities, which in turn will gradually become a trusted source of information for consumers. Finally, social networks will play a key role in becoming the intermediary medium between consumers and brands. Social networking firms such as Facebook and Twitter have begun aggregating information regarding user profiles, information, and connectivity that can serve as an important source of customer-level data for businesses. It is therefore important for CRM implementation managers to be cognizant of such developments in the marketplace and make use of such avenues for their information needs. In light of the influence of such new technologies on marketing, the following issues warrant attention from academics and practitioners:

- What will be the right customer metrics to measure customers' value given the prevalence of online communities? What changes are required to the customer valuation models in order to accurately incorporate customers' values in terms of their adoption and usage of social media activities?

- Do older consumers with less developed social graphs have the same needs as the younger generation? How can marketers address this difference?
- How can firms make use of the vast information provided by social network providers to develop and implement CRM strategies to satisfy specific customer needs?
- Are social coupons really helping small and medium-size businesses ensure long-term profitability? What type(s) of customer segments (new or existing, local or walk-in, etc.) should businesses target in order to achieve maximized long-term profitability? [4]
- How can social networking be applied and evaluated in B2B and B2C settings, and to what extent will the user characteristics pertaining to CLV affect/influence online customer-to-customer (C2C) exchanges? [3]

## 9.3 Mobile marketing

In recent years, mobile marketing has emerged as a new marketing channel for firms to reach out to customers on an individual basis. For marketing managers, this is an effective channel to convey their message to the customers that they want to target. To devise the most effective CRM strategy, it is important for managers to have a good understanding of the key benefits and make the best use of this new communication channel.

Mobile marketing is conducted and communicated through hand-held mobile devices, such as mobile phones, PDAs, smart phones, and portable tablet computers using 3G technologies, such as iPads. Mobile marketing has become a new marketing paradigm because of the superior benefits it offers compared to other communication channels. First, the most important benefit of mobile marketing is that it works even when the intended recipients are on the move. This is not the case with traditional marketing media such as billboards, TV, print, and radio that can only be accessed through certain networks and only from a certain location. Marketers thus can reach out to their customers anytime, anywhere, almost instantly. Mobile platforms therefore provide a fundamentally different type of consumer experience. Second, another interesting feature of mobile marketing is the opt-in feature. The opt-in feature allows customers to indicate their willingness to participate in the call-to-action when they see the firms' advertisement. This feature also enables companies to gather customer information in exchange for special discounts, or privileges in the future. Such a source of customer data provides firms with a quantifiable measure of reach and impact of a marketing program [5]. Finally, through interaction with customers, mobile marketing can be used to collect more contextual and location-based types of data. For example, marketers can determine not only the exact location of a consumer at a given time, but also the context of why that individual might be there. Such pieces of information will be very useful to marketing managers for the purpose of designing and implementing a particular strategy.

Mobile marketing is being implemented by many firms. For example, commuters in Japan can scan their bus schedule with their phones and receive coupons from stores along their routes [6]. Banks such as Standard Chartered and HSBC send out promotional offers to customers through SMS to mobile phones. Also, cellular service providers such as AT&T and T-Mobile are constantly keeping their customers updated with the latest promotions by sending messages to the mobile devices.

Looking ahead, the future of mobile marketing looks promising. Mobile marketing is expected to be used as a complementary, and not an alternative, channel of communication besides the traditional channels such as TV and print. At present, most mobile marketing campaigns involve sending marketing messages to a large group of customers. To increase the effectiveness of mobile marketing in the future, managers can apply the findings provided in the previous chapters in various combinations to target the right individual customers. Adopting mobile marketing in such a manner, in conjunction with other marketing strategies, will multiply the overall effects on firm value. In effect, it will provide significant opportunities for firms to explore ways to enrich the customer–firm relationship. In an effort to bring mobile marketing into the mainstream marketing approach, the following issues will be of interest to marketers and researchers:

- How to measure, manage, and maximize the consumer response rates for a mobile marketing campaign?
- How to integrate mobile marketing methods into the overall brand management strategies of firms?
- How to address the privacy concerns of consumers regarding mobile marketing?

## 9.4 Customized marketing campaigns

Customized marketing campaigns, or one-to-one marketing campaigns, refer to campaigns containing marketing messages that are tailored to a selected individual customer based on his/her specific profile, preference, and needs. In the past, to reach the maximum number of clients, and to achieve economies of scale, firms employed mass marketing and targeted the entire market. As marketing progresses from being product-centric to becoming customer-centric, marketers have started to employ campaign-based marketing and target certain segmented groups of customers. Customized marketing is expected to be the next prevalent marketing trend that promises to enable firms to gain competitiveness through differentiation in an increasingly competitive market. It is important for marketing managers to be aware of this emerging trend as they formulate CRM strategies for the firm.

One-to-one marketing has become a widely used marketing strategy due to the number of benefits it offers. First, it is an effective way to ensure information is conveyed to the targeted customers by overcoming media clutter. Given the increasing number of media outlets, one common problem faced by mass marketers is that they are unable to convey their messages to consumers who are inundated with too much information. By targeting a specific individual customer, marketers build a customized marketing campaign and choose a more effective way to communicate to the customer, and ensure the message is conveyed to that particular customer. Second, implementing customized marketing campaigns is a perfect way for firms to build a long-term relationship with customers. Satisfied and loyal customers can eventually provide excellent referrals for firms. Customized marketing ultimately facilitates the paradigm shift from selling the product to selling the benefit to customers. Finally, implementing customized marketing campaigns is a great way for firms to differentiate themselves from their competitors.

Many firms are currently implementing or on their way toward implementing a one-to-one marketing strategy. Amazon.com uses information acquired from individual customer transactions and profiles to filter and determine the type of products it will recommend to each customer. Portola Plaza Hotel in California gathers information about guest preferences and, through detail analysis, send mails to individual guests to offer customized discounts and promotional deals. ICICI-Lombard, an insurance company, uses customer surveys to prepare personalized insurance plans for its customers.

Perhaps the biggest deterrents for customized marketing campaigns to be a preferred marketing approach are the amounts of customer-level information required to develop campaigns, and the high cost involved in obtaining such customer-level data. As customers are more involved in social network activities, this information will be more readily available to firms through many social network providers. Furthermore, the success of a customized campaign lies in identifying the right customer to target. As firms move from a mass marketing approach to a segmented marketing approach, the availability of appropriate customer-level information and the ability to identify the right customer to target will help companies design and implement successful customized campaigns. The quantitative models discussed in the previous chapters can serve as good starting points for firms to identify the right customer and launch effective customized marketing campaigns.

Looking ahead, customized campaigns can become a mainstream marketing approach when marketers address some important issues. First, firms must know their customers well, identify the most valuable customers, and maintain interaction with each of these customers. All of these require a thorough analysis of their customers, and the maintenance of a comprehensive customer database. Second, firms need to integrate different departments within an organization (e.g., marketing, sales, production, distribution, finance, and so on) to devise a customized marketing campaign for each of the targeted customers. Other important issues that require careful attention from marketers and researchers include the following:

- What are the relevant factors that impact firms' customized marketing strategies? How important are market-specific factors (e.g., market growth rate, competitive landscape, etc.) in the formulation of customized marketing strategies?

- Currently, firms usually have very limited information about competitors' customers. Under what conditions can sharing customer information with competitors be profitable? [7]
- What is the role of customized marketing in reducing information overload and aiding customer decisions? [8]
- What pricing structure should firms adopt for customized marketing?

## 9.5 Conclusion

The three marketing methods discussed here are only a selection of the emerging trends in marketing, and not a conclusive list of trends. Each of these methods works in different scenarios and presents a varied set of alternatives. Apart from their novelty aspect, social media open up a new source of information for marketers that can be used as input for developing CRM models. Mobile marketing offers a new, better, and more effective channel of communication to reach out to individual customers. Customized campaigns show promise in providing firms with the much needed competitive advantage in engaging and managing customers. Adopting one or all of these methods, as the company situation demands, will ensure firms' future success in ensuring a profitable customer relationship.

## References

1. Kumar, V. (2008) *Managing Customers for Profit: Strategies to Increase Profits and Build Loyalty*, Prentice Hall, Upper Saddle River, NJ.
2. Ahuja, V. and Medury, Y. (2010) Corporate blogs as e-CRM tools – building consumer engagement through content management. *Journal of Database Marketing & Customer Strategy Management*, **17**(2), 91–105.
3. Kumar, V., Bhaskaran, V., Mirchandani, R., and Shah, M. (forthcoming) Creating a measurable social media marketing strategy for hokey pokey: increasing the value and ROI of intangibles and tangibles. *Marketing Science*.
4. Kumar, V. and Rajan, B. (2011) Social coupons as a marketing strategy: a multifaceted perspective. *Journal of the Academy of Marketing Science*, **40**(1), 120–136.
5. Aaker, D., Kumar, V., Day, G., and Leone, R.P. (2010) *Marketing Research*, John Wiley & Sons, Inc., Hoboken, NJ.
6. Sultan, F. and Rohm, A. (2005) The coming era of brand-in-the-hand marketing. *Sloan Management Review*, **47**(Fall), 83–90.
7. Chen, Y., Narasimhan, C., and Zhang, J. (2001) Individual marketing with imperfect targetability. *Marketing Science*, **20**(1), 23–24.
8. Arora, N., Dreze, X., Ghose, A. *et al.* (2008) Putting one-to-one marketing to work: personalization, customization, and choice. *Marketing Letters*, **19**, 305–321.

# Appendix A

# Maximum likelihood estimation

## (Source: Franses and Paap [1])

In the estimation of maximum likelihood, the likelihood function is defined as

$$L(\beta) = p(y \,|\, X; \beta) \tag{A.1}$$

where $p(\cdot)$ is the joint density distribution for the observed variables $y$ given $X$, and $\beta$ summarizes the model parameters $\beta_0$ to $\beta_k$. The logarithmic likelihood function is

$$l(\beta) = \log(L(\beta)). \tag{A.2}$$

Since it is not possible to find an analytical solution for the value of $\beta$ that maximizes the log-likelihood function, the maximization has to be done using a numerical optimization algorithm. Here, Franses and Paap [1] prefer the Newton–Raphson method, which introduces the gradient $G(\beta)$ and the Hessian matrix $H(\beta)$ as

$$G(\beta) = \frac{\partial l(\beta)}{\partial \beta}, \tag{A.3}$$

$$H(\beta) = \frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'}. \tag{A.4}$$

Around a given value $\beta$ the first-order condition for the optimization problem can be linearized, resulting in $G(\beta_h) + H(\beta_h)(\beta - \beta_h) = 0$. Solving this gives the

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

sequence of estimates

$$\beta_{h+1} = \beta_h - H(\beta_h)^{-1} G(\beta_h) \tag{A.5}$$

where $G(\beta_h)$ and $H(\beta_h)$ are the gradient and Hessian matrix evaluated in $\beta_h$.

# Reference

1. Franses, P.H. and Paap, R. (2001) *Quantitative Models in Marketing Research*, Cambridge University Press, Cambridge.

# Appendix B

# Log-linear model—an introduction

## (Source: Knoke and Burke [1] and Course Material by Angela Jeansonne)

Log-linear analysis is an extension of the two-way contingency table where the conditional relationship between two or more discrete, categorical variables is analyzed by taking the natural logarithm of the cell frequencies within a contingency table. The log-linear model states the expected cell frequencies of a cross-tabulation (the $F_{ij}$) as functions of parameters representing characteristics of the categorical variables and their relationships with each other. The general log-linear model does not distinguish between independent and dependent variables. All variables are treated alike as 'response variables' whose mutual associations are explored.

The usual data suitable for log-linear analysis are contingency tables. Angela Jeansonne provided an example in the class materials. Suppose we are interested in the relationship between sex, heart disease, and body weight. We could take a sample of 200 subjects and determine the sex, approximate body weight, and who does and does not have heart disease. The continuous variable, body weight, is broken down into two discrete categories: not overweight and overweight. The contingency table containing the data may look like this:

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

| | | Heart disease | | Total |
|---|---|---|---|---|
| Body weight | Sex | Yes | No | |
| Not overweight | Male | 15 | 5 | 20 |
| | Female | 40 | 60 | 100 |
| Total | | 55 | 65 | 120 |
| Overweight | Male | 20 | 10 | 30 |
| | Female | 10 | 40 | 50 |
| Total | | 30 | 50 | 80 |

Angela Jeansonne in the class material states that the basic strategy in log-linear modeling involves fitting models to the observed frequencies in the cross-tabulation of categorical variables. The models can then be represented by a set of expected frequencies that may or may not resemble the observed frequencies. Models will vary in terms of the marginal they fit, and can be described in terms of the constraints they place on the associations or interactions that are present in the data. The pattern of association among variables can be described by a set of odds and by one or more odds ratios derived from them. Once expected frequencies are obtained, we then compare models that are hierarchical to one another and choose a preferred model, which is the most parsimonious model that fits the data. It is important to note that a model is not chosen if it bears no resemblance to the observed data. The choice of a preferred model is typically based on a formal comparison of goodness-of-fit statistics associated with models that are related hierarchically (models containing higher order terms also implicitly include all lower order terms). Ultimately, the preferred model should distinguish between the pattern of the variables in the data and sampling variability, thus providing a defensible interpretation.

Angela Jeansonne further wrote the log-linear model. The following model refers to the traditional chi-square test where two variables, each with two levels ($2 \times 2$ table), are evaluated to see if an association exists between the variables.=:

$$\ln\left(F_{ij}\right) = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \tag{B.1}$$

where $\ln(F_{ij})$ is the logarithm of the expected cell frequency of the cases for cell *ij* in the contingency table, $\mu$ is the overall mean of the natural logarithm of the expected frequencies, $\lambda$ terms each represent ‘effects’ which the variables have on the cell frequencies, *A* and *B* are the variables, and *i* and *j* the categories within the variables.

Angela Jeansonne also stated that the above model is considered a saturated model because it includes all possible one-way and two-way effects. Given that the saturated model has the same number of cells in the contingency table as it does effects, the expected cell frequencies will always exactly match the observed frequencies, with no degrees of freedom remaining (Knoke and Burke, [1]). For example, in a $2 \times 2$ table there are four cells and in a saturated model involving two variables there are four effects, $\mu$, $\lambda_i^A$, $\lambda_j^B$, $\lambda_{ij}^{AB}$, therefore the expected cell frequencies will exactly match the observed frequencies. Thus, in order to find a more parsimonious model that will isolate the effects best demonstrating the data patterns, a non-saturated model must be sought. This can be achieved by setting some of the effect parameters to zero. For instance, if we set the effects parameter $\lambda_{ij}^{AB}$ to zero (i.e., we assume that variable $A$ has no effect on variable $B$, or vice versa) we are left with the unsaturated model.

$$\ln\left(F_{ij}\right) = \mu + \lambda_i^A + \lambda_j^B. \tag{B.2}$$

## Reference

1. Knoke, D. and Burke, P.J. (1980) *Log-linear Models*, Sage, Beverly Hills, CA.

# Appendix C

# Vector autoregression modeling

## (Source: Dekimpe and Hanssens [1])

Dekimpe and Hanssens [1] introduced vector autoregression modeling under the framework of persistence modeling. Persistence modeling addresses the problem of long-run market-response quantification by combining into one measure of 'net long-run impact' the chain reaction of consumer response, firm feedback, and competitor response that emerges following the initial marketing action. Persistence modeling is a multi-step process. In the first step, unit-root tests are used to determine whether or not the different variables are stable or evolving. In case several of the variables are found to have a unit root, one subsequently tests for cointegration. Depending on the outcome of these two preliminary steps, one estimates a vector autoregression (VAR) model in the levels, in the differences, or in error-correction format. Finally, the parameter estimates from this VAR model are used to derive impulse-response functions (IRFs), from which various summary statistics on the short- and-long-run dynamics of the system can be derived. Each of these steps is elaborated briefly as below.

## C.1 Unit-root testing: Are performance and marketing variables stable or evolving?

The distinction between stability and evolution is formalized through the unit-root concept. Here, a simple case where the behavior over time of the variable

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

of interest (e.g., a brand's sales $S_t$) is described by a first-order autoregressive process:

$$(1 - \phi L)S_t = c + u_t \tag{C.1}$$

where $\phi$ is an autoregressive parameter, $L$ the lag operator (i.e., $L^k S_t = S_{t-k}$), $u_t$ a residual series of zero-mean, constant-variance $\sigma_u^2$, and uncorrelated random shocks, and $c$ a constant. Note that Equation C.1 may also be written in the more familiar form

$$S_t = c + \phi S_{t-1} + u_t \tag{C.2}$$

which corresponds to a simple regression model of $S_t$ on its own past, with $u_t$ the usual i.i.d. residuals. Applying successive backward substitutions allows us to write Equation C.1 as

$$S_t = [c/(1 - \phi)] + u_t + \phi u_{t-1} + \phi^2 u_{t-2} + \cdots, \tag{C.3}$$

in which the present value of $S_t$ is explained as a weighted sum of random shocks. Depending on the value of $\phi$, two scenarios can be distinguished. When $|\phi| < 1$, the impact of past shocks diminishes and eventually becomes negligible. Hence, each shock has only a temporary impact. In that case, the series has a fixed mean $c/(1 - \phi)$ and a finite variance $\sigma_u^2/(1 - \phi^2)$. Such a series is called stable. When $|\phi| = 1$, however, Equation C.3 becomes

$$S_t = (c + c + \cdots) + u_t + u_{t-1} + \cdots, \tag{C.4}$$

implying that each random shock has a permanent effect on the subsequent values of $S$. In that case, no fixed mean is observed, and the variance increases with time. Sales do not revert to a historical level, but instead wander freely in one direction or another, that is, they evolve. Distinguishing between both situations involves checking whether the parameter $\phi$ in Equation C.1 is less than or equal to 1.

Numerous tests have been developed to distinguish stable from evolving patterns. One popular test, due to Dickey and Fuller [2], is based on the following test equation:

$$(1 - L)S_t = \Delta S_t = \alpha_0 + bS_{t-1} + a_1 \Delta S_{t-1} + \cdots + a_m \Delta S_{t-m} + u_t. \tag{C.5}$$

The $t$-statistics of $b$ are compared to critical values and the unit-root null hypothesis is rejected if the obtained value is larger in absolute value than the critical value. The $m\Delta S_{t-j}$ terms reflect temporary sales fluctuations, and are added to make $u_t$ white noise. Because of these additional terms, one often refers to this test as the 'augmented' Dickey–Fuller (ADF) test. Villanueva *et al.* [3] used the ADF test with the null hypothesis of unit root. For the key decisions to be made when implementing ADF-like unit-root tests, readers may refer to Dekimpe and Hanssens [1] for details.

## C.2 Cointegration tests: Does a long-run equilibrium exist between evolving series?

Evolving variables are said to be cointegrated when a linear combination exists between them that results in stable residuals. Consider, without loss of generality, a three-variable example where a brand's sales (*S*), marketing support (*M*), and its competitors' marketing support (*CM*) are all evolving (i.e., they all have a unit root). The existence of a perfect equilibrium relationship between these three variables would imply (see Powers *et al.* [4] for a more in-depth discussion)

$$S_t = \beta_0 + \beta_1 M_t + \beta_2 CM_t. \tag{C.6}$$

In practice, however, we are unlikely to observe a perfect equilibrium in every single period. A more realistic requirement is that its deviations are mean reverting (stable) around zero, that is, $e_{S,t}$ in Equation C.7 below should no longer be evolving, even though each of the other variables in the equation is:

$$S_t = \beta_0 + \beta_1 M_t + \beta_2 CM_t + e_{S,t}. \tag{C.7}$$

For a detailed review of the test procedure for cointegration, readers may refer to Dekimpe and Hanssens [1].

## C.3 Var models: How to capture the dynamics in a system of variables?

The third step in persistence modeling is to specify a VAR model to link the (short-run) movements of the different variables under consideration. Depending on the outcomes of the proceeding unit-root and cointegration tests, these VAR models are specified in the levels (no unit roots), in the differences (unit roots without cointegration), or in error-correction format (cointegration).

For expository purposes, we first consider a model in levels, and focus on a simple three-equation model linking own sales performance (*S*), own marketing spending (*M*), and competitive marketing spending (*CM*). The corresponding VAR model (in which, for ease of notation, all deterministic components are omitted) becomes

$$\begin{pmatrix} S_t \\ M_t \\ CM_t \end{pmatrix} = \begin{pmatrix} \pi_{11}^1 & \pi_{12}^1 & \pi_{13}^1 \\ \pi_{21}^1 & \pi_{22}^1 & \pi_{23}^1 \\ \pi_{31}^1 & \pi_{32}^1 & \pi_{33}^1 \end{pmatrix} \begin{pmatrix} S_{t-1} \\ M_{t-1} \\ CM_{t-1} \end{pmatrix} + \cdots$$

$$+ \begin{pmatrix} \pi_{11}^J & \pi_{12}^J & \pi_{13}^J \\ \pi_{21}^J & \pi_{22}^J & \pi_{23}^J \\ \pi_{31}^J & \pi_{32}^J & \pi_{33}^J \end{pmatrix} \begin{pmatrix} S_{t-J} \\ M_{t-J} \\ CM_{t-J} \end{pmatrix} + \begin{pmatrix} u_{S,t} \\ u_{M,t} \\ u_{CM,t} \end{pmatrix}, \tag{C.8}$$

where $J$ is the order of the model, and where vector $u$ is $u = (\, u_{S,t} \quad u_{M,t} \quad u_{CM,t}\,)' \sim N(0, \Sigma)$.

This specification is very flexible, and reflects the forces or channels of influence discussed earlier: delayed response ($\pi^j_{12}, j = 1, \ldots J$), purchase reinforcement ($\pi^j_{11}$), performance feedback ($\pi^j_{21}$), inertia in decision making ($\pi^j_{22}$), and competitive reactions ($\pi^j_{32}$). Only instantaneous effects are not included directly, but these are reflected in the variance–covariance matrix of the residuals ($\Sigma$). Estimation of these models is straightforward: (a) all explanatory variables are predetermined, so there is no concern over the identification issues that are often encountered when specifying structural multiple-equation models; and (b) all equations in the system have the same explanatory variables so that OLS estimation can be applied without loss of efficiency.

However, this flexibility comes at a certain cost. First, the number of parameters may become exuberant. For $J = 8$, for example, the VAR model in Equation C.8 will estimate $9 \times 8 = 72$ autoregressive parameters. If, however, one considers a system with five endogenous variables, this number increases to $25 \times 8 = 200$. As a consequence, VAR modelers typically do not interpret the individual parameters themselves, but rather focus on the impulse-response functions (IRFs) derived from these parameters. IRFs trace, over time, the incremental performance and spending implications of an initial one-period change in one of the support variables. In so doing, they provide a concise summary of the information contained in this multitude of parameters, a summary that lends itself well to a graphical and easy-to-interpret representation.

Second, no direct estimate is provided of the instantaneous effects. The residual correlation matrix can be used to establish the presence of such an effect, but not its direction. Various procedures have been used in the marketing literature to deal with this issue, and readers may refer to Dekimpe and Hanssens [1] for a brief review.

If some of the variable have a unit root, the VAR in Equation C.8 is specified in the differences; for example, $S_t$, $S_{t-1}$, ... are replaced by $\Delta S_t$, $\Delta S_{t-1}$, .... If the variables are cointegrated as well, this model in differences is augmented with the lagged residuals of the respective long-run equilibrium relationships, resulting in the following specification:

$$\begin{pmatrix} \Delta S_t \\ \Delta M_t \\ \Delta CM_t \end{pmatrix} = \begin{pmatrix} \alpha_S & 0 & 0 \\ 0 & \alpha_M & 0 \\ 0 & 0 & \alpha_{CM} \end{pmatrix} \begin{pmatrix} e_{S,t-1} \\ e_{M,t-1} \\ e_{CM,t-1} \end{pmatrix} + \sum_{j=1}^{j} \begin{pmatrix} \pi^j_{11} & \pi^j_{12} & \pi^j_{13} \\ \pi^j_{21} & \pi^j_{22} & \pi^j_{23} \\ \pi^j_{31} & \pi^j_{32} & \pi^j_{33} \end{pmatrix} \begin{pmatrix} \Delta S_{t-j} \\ \Delta M_{t-j} \\ \Delta CM_{t-j} \end{pmatrix} + \begin{pmatrix} u_{S,t} \\ u_{M,t} \\ u_{CM,t} \end{pmatrix}. \tag{C.9}$$

The addition of the error-correction terms $[\alpha_S e_{S,t-1} \quad \alpha_M e_{M,t-1} \quad \alpha_{CM} e_{CM,t-1}]'$ implies that in every period there is a partial adjustment toward restoring the underlying, temporarily distributed, long-run equilibrium. Put differently, the system partially corrects for the previously observed deviations $[e_{S,t-1} \quad e_{M,t-1} \quad e_{CM,t-1}]'$, and the respective $\alpha$ coefficients reflect the speed of adjustment of the corresponding dependent variable toward the equilibrium. A good review of the implementation issues involved can be found in Franses and Paap [5].

## C.4 Impulse-response function derivation

An IRF traces the incremental effect of a 1 unit (or one standard deviation) shock in one of the variables on the future values of the other endogenous variables. The first steps of this process are depicted later (where for expository purposes, a VAR model of order 1 is considered). IRFs can also been seen as the difference between two forecasts: a first extrapolation based on an information set that does not take the marketing shock into account, and another prediction based on an extended information set that takes this action into account. As such, IRFs trace the incremental effect of the marketing action reflected in the shock. Note that marketing actions (e.g., a price promotion) are operationalized as deviations from a benchmark, which is derived as the expected value of the marketing mix variable (e.g., the price) as predicted through the dynamic structure of the VAR model. For more details about IRFs, interested readers may refer to Dekimpe and Hanssens [1].

## C.5 Impulse-response functions: Mathematical derivations

$$\begin{pmatrix} S_t \\ M_t \\ CM_t \end{pmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} \begin{pmatrix} S_{t-1} \\ M_{t-1} \\ CM_{t-1} \end{pmatrix} + \begin{pmatrix} u_{S,t} \\ u_{M,t} \\ u_{CM,t} \end{pmatrix},$$

One sets

$$\begin{aligned} \begin{pmatrix} u_S & u_M & u_{CM} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix} & \text{ prior to } t \\ \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} & \text{ at time } t \\ \begin{pmatrix} 0 & 0 & 0 \end{pmatrix} & \text{ after } t \end{aligned}$$

and computes (simulates) the future values for the various endogenous variable, that is,

$$\begin{pmatrix} S_t \\ M_t \\ CM_t \end{pmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} S_{t+1} \\ M_{t+1} \\ CM_{t+1} \end{pmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \pi_{12} \\ \pi_{22} \\ \pi_{32} \end{pmatrix},$$

$$\begin{pmatrix} S_{t+2} \\ M_{t+2} \\ CM_{t+2} \end{pmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} \begin{pmatrix} \pi_{12} \\ \pi_{22} \\ \pi_{32} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \pi_{11}\pi_{12} & \pi_{12}\pi_{22} & \pi_{13}\pi_{32} \\ \pi_{21}\pi_{12} & \pi_{22}\pi_{22} & \pi_{23}\pi_{32} \\ \pi_{31}\pi_{12} & \pi_{32}\pi_{22} & \pi_{33}\pi_{32} \end{pmatrix}, \ \ldots .$$

# References


1. Dekimpe, M.G. and Hanssens, D.M. (2004) *Persistence Modeling for Assessing Marketing Strategy Performance*, Marketing Science Institute, Cambridge, MA.
2. Dickey, D.A. and Fuller, W.A. (1979) *Distribution of the estimators for autoregressive time series with a unit root*. *Journal of the American Statistical Association*, **74**(366), 427–431.
3. Villanueva, J., Yoo, S., and Hanssens, D.M. (2008) *The impact of marketing-induced versus word-of-mouth customer acquisition on customer equity growth*. *Journal of Marketing Research*, **45**(1), 48–59.
4. Powers, K., Hanssens, D.M., Hser, Y.-I., and Anglin, M.D. (1991) *Measuring the long term effects of public policy: the case of narcotics use and property crime*. *Management Science*, **37**(6), 627–644.
5. Franses, P.H. and Paap, R. (2001) *Quantitative Models in Marketing Research*, Cambridge University Press, Cambridge.

# Appendix D

# Accelerated lifetime model

## (Source: Franses and Paap [1])

In the estimation of duration models, the log-likelihood function is

$$L(\theta) = \prod_{i=1}^{N} f(t_i \mid X_i)^{d_i} Sf(t_i|X_i)^{1-d_i} \tag{D.1}$$

where $d_i$ is defined as a 0/1 dummy that is 1 if the observation is not censored and 0 if the observation is censored, and $\theta$ is a vector of the model parameters consisting of $\beta$ and the distribution-specific parameters. The log-likelihood function is given by

$$l(\theta) = 1 \sum_{i=1}^{N} d_i \log f(t_i|X_i) + (1-d_i)\log S(t_i|X_i). \tag{D.2}$$

The maximum likelihood (ML) estimator $\hat{\theta}$ is the solution of the equation

$$\frac{\partial l(\theta)}{\partial \theta} = 0. \tag{D.3}$$

In general, there are no closed-form expressions for this estimator and we have to use numerical optimization algorithms such as Newton–Raphson to maximize the log-likelihood function. The ML estimates can be found by iterating over

$$\theta_h = \theta_{h-1} - H(\theta_{h-1})^{-1} G(\theta_{h-1}) \tag{D.4}$$

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

until convergence, where $G(\theta)$ and $H(\theta)$ denote the first- and second-order derivatives of the log-likelihood function. The analytical form of the first- and second-order derivatives of the log-likelihood depends on the form of the baseline hazard.

The hazard function of an accelerated lifetime model with a Weibull specification reads as

$$\lambda(t_i|X_i) = \exp(X_i\beta)\lambda_0(\exp(X_i\beta)t_i) = \exp(X_i\beta)\alpha(\exp(X_i\beta)t_i)^{\alpha-1} \tag{D.5}$$

where we put $\gamma = 1$ for identification. The survival function is then given by

$$S(t_i|X_i) = \exp(-(\exp(X_i\beta)t_i)^{\alpha}). \tag{D.6}$$

To facilitate the differentiation of the likelihood function in an accelerated lifetime model, it is convenient to define

$$z_i = \alpha \ln(\exp(X_i\beta)t_i) = \alpha(\ln t_i + X_i\beta). \tag{D.7}$$

Straightforward substitution in Equation D.6 results in the survival function and the density function of $t_i$ expressed in terms of $z_i$, that is,

$$S(t_i|X_i) = \exp(-\exp(z_i)) \tag{D.8}$$

$$f(t_i|X_i) = \alpha \exp(z_i - \exp(z_i)). \tag{D.9}$$

The log-likelihood function can be written as

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{N}[d_i \log f(t_i|X_i) + (1-d_i)\log S(t_i|X_i)] \\ &= \sum_{i=1}^{N}[d_i(z_i + \log(\alpha)) - \exp(z_i)] \end{aligned} \tag{D.10}$$

where $\theta = (\beta, \alpha)$.

The first-order derivative of the log-likelihood equals $G(\theta) = (\partial l(\theta)/\partial\beta', \partial l(\theta)/\partial\alpha)'$ with

$$\frac{\partial l(\theta)}{\partial\beta} = \sum_{i=1}^{N}\alpha(d_i - \exp(Z_i))X'_i \tag{D.11}$$

$$\frac{\partial l(\theta)}{\partial\alpha} = \sum_{i=1}^{N}\frac{d_i(z_i+) - \exp(z_i)z_i}{\alpha} \tag{D.12}$$

where we use $\partial z_i/\partial\alpha = z_i/\alpha$ and $\partial z_i/\partial\beta = \alpha X_i'$. The Hessian is

$$H(\theta)\begin{pmatrix} \dfrac{\partial^2 l(\theta)}{\partial\beta\partial\beta'} & \dfrac{\partial^2 l(\theta)}{\partial\alpha\partial\beta} \\ \dfrac{\partial^2 l(\theta)}{\partial\alpha\partial\beta'} & \dfrac{\partial^2 l(\theta)}{\partial\alpha\partial\alpha} \end{pmatrix} \tag{D.13}$$

where

$$\frac{\partial^2 l(\theta)}{\partial\beta\partial\beta'} = -\sum_{i=1}^{N} \alpha^2 \exp(z_i) X'_i X_i \tag{D.14}$$

$$\frac{\partial^2 l(\theta)}{\partial\alpha\partial\beta} = -\sum_{i=1}^{N} (d_i - \exp(z_i)) X'_i \tag{D.15}$$

$$\frac{\partial^2 l(\theta)}{\partial\alpha\partial\alpha} = -\sum_{i=1}^{N} \frac{d_i + \exp(z_i) z_i^2}{\alpha^2}. \tag{D.16}$$

The ML estimates are found by iterating over Equation D.4 for properly chosen starting values for $\beta$ and $\alpha$.

## Reference

1. Franses, P.H. and Paap, R. (2001) *Quantitative Models in Marketing Research*, Cambridge University Press, Cambridge.

# Appendix E

# Type-1 Tobit model

## (Source: Franses and Paap [1])

For a type-1 Tobit model, the censored variable $Y_i$ is 0 if the unobserved latent variable $y_i^*$ is less than or equal to 0 and $Y_i = y_i^*$ if $y_i^*$ is positive,

$$Y_i = X_i\beta + \varepsilon_i \quad \text{if } y_i^* = X_i\beta + \varepsilon_i > 0 \tag{E.1}$$

$$Y_i = 0 \quad \text{if } y_i^* = X_i\beta + \varepsilon_i \leq 0 \tag{E.2}$$

with $\varepsilon_i \sim N(0, \sigma^2)$. For observations $y_i$ that are 0, we know only that

$$\Pr(Y_i = 0|X_i) = \Pr(X_i\beta + \varepsilon_i \leq 0|X_i) = \Pr(\varepsilon_i \leq -X_i\beta|X_i) = \Phi\left(-\frac{X_i\beta}{\sigma}\right). \tag{E.3}$$

Maximum likelihood estimation is used to estimate the Tobit model. The likelihood function consists of two parts: the probability that an observation is censored is given by Equation E.3; and the density of the non-censored observations is a standard normal density. The likelihood function is

$$L(\theta) = \prod_{i=1}^{N} \Phi\left(-\frac{X_i\beta}{\sigma}\right)^{I[y_i=0]} \left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(y_i - X_i\beta)^2\right)\right]^{I[y_i>0]}, \tag{E.4}$$

where $\theta = (\beta, \sigma)$. It is more convenient to reparameterize the model according to $\gamma = \beta/\sigma$ and $\xi = 1/\sigma$. The log-likelihood function in terms of $\theta^* = (\gamma, \xi)$ reads

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

$$l(\theta^*) = \sum_{i=1}^{N} \left\{ I[y_i = 0] \log \Phi(-X_i\gamma) + I[y_i > 0] \left[ \log\xi - \frac{1}{2}\log(2\pi) - \frac{1}{2}(\xi y_i - X_i\gamma)^2 \right] \right\}. \tag{E.5}$$

The first-order derivatives of the log-likelihood function with respect to $\gamma$ and $\xi$ are

$$\frac{\partial l(\theta^*)}{\partial \gamma} = \sum_{i=1}^{N} \left\{ -I[y_i = 0]\lambda(X_i\gamma)X'{}_i + I[y_i > 0](\xi y_i - X_i\gamma)X_i' \right\} \tag{E.6}$$

$$\frac{\partial l(\theta^*)}{\partial \xi} = \sum_{i=1}^{N} I[y_i > 0][1/\xi - (\xi y_i - X_i\gamma)y_i] \tag{E.7}$$

and the second-order derivatives are

$$\frac{\partial l(\theta^*)}{\partial \gamma \partial \gamma'} = \sum_{i=1}^{N} \left\{ I[y_i = 0]\left( -\lambda(X_i\gamma)^2 + X_i\gamma\lambda(X_i\gamma)X_i'X_i - I[y_i > 0]X_i'X_i \right) \right\} \tag{E.8}$$

$$\frac{\partial l(\theta^*)}{\partial \gamma \partial \xi} = \sum_{i=1}^{N} I[y_i > 0]y_iX_i' \tag{E.9}$$

$$\frac{\partial l(\theta^*)}{\partial \xi \partial \xi} = \sum_{i=1}^{N} I[y_i > 0]\left( -\frac{1}{\xi^2} - y_i^2 \right). \tag{E.10}$$

# Reference

1. Franses, P.H. and Paap, R. (2001) *Quantitative Models in Marketing Research*, Cambridge University Press, Cambridge.

# Appendix F

# Multinomial logit model

## (Sources: Greene [1] and Borooah [2])

The unordered-choice model can be motivated by a random utility model. For the $i$th customer faced with $J$ choices, suppose that the utility of choice $J$ is

$$U_{ij} = \beta' Z_{ij} + \varepsilon_{ij}. \tag{F.1}$$

If the consumer makes choice $j$, then we assume that $U_{ij}$ is the maximum among the $J$ utilities. Hence, the statistical model is driven by the probability that choice $j$ is made, which is

$$\begin{aligned} &Prob\left(U_{ij} > U_{ik}\right) \quad \text{for all other}\, k \neq j \\ &\Rightarrow \Pr\left(Z_{ij} + \varepsilon_{ij} > Z_{ik} + \varepsilon_{ik}\right) \\ &\Rightarrow \Pr\left(\varepsilon_{ik} - \varepsilon_{ij} < Z_{ij} - Z_{ik}\right). \end{aligned} \tag{F.2}$$

McFadden [3] has shown that if and only if the $J$ disturbances are independent and identically distributed with Weibull distribution

$$F\left(\varepsilon_{ij}\right) = \exp\left(-\exp(-\varepsilon i_j)\right) \tag{F.3}$$

then

$$Prob(Y_i = j) = \frac{\exp(\beta' Z_{ij})}{\sum_J^{j=1} \exp(\beta' Z_{ij})} \tag{F.4}$$

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

where $Y_i$ is defined as a random variable that indicates the choice made. This is called the conditional logit model. The explanatory variables include two groups of factors, individual- and choice-specific attributes. Let $Z_{ij} = [X_{ij}, W_i]$. The components of $X_{ij}$ are typically called the attributes of the choices and $W_i$ contain the characteristics of the individual. The multinomial model is the one that incorporates only individual characteristic effects, so that for such models all the $X_{ij} = 0$.

In the multinomial logit model, the probabilities are

$$Prob(Y = j) = \frac{\exp(\beta_j' x_i)}{1 + \sum_{k=1}^{J} \exp(\beta_k' x_i)} \quad \text{for } j = 1, 2, \ldots, J, \tag{F.5}$$

$$Prob(Y = 0) = \frac{1}{1 + \sum_{k=1}^{J} \exp(\beta_k' x_i)}. \tag{F.6}$$

The model implies that we can compute $J$ log-odds ratios

$$\ln\left[\frac{P_{ij}}{P_{i0}}\right] = \beta_j' x_i \tag{F.7}$$

$$\ln\left[\frac{P_{ij}}{P_{ik}}\right] = x_i'(\beta_j - \beta_k). \tag{F.8}$$

The multinomial logit model is estimated by the maximum likelihood estimation method. The log-likelihood can be derived by defining, for each individual, $d_{ij} = 1$ if alternative $j$ is chosen by individual $i$, and 0 if not, for the $J + 1$ possible outcomes. Then, for each $i$, one and only one of the $d_{ij}$ is 1. The log-likelihood is a generation of that for the binomial probit or logit model:

$$\ln L = \sum_{i=1}^{n} \sum_{j=0}^{J} d_{ij} \ln Prob(Y_i = j)'. \tag{F.9}$$

The derivatives have the characteristically simple form

$$\frac{\partial \ln L}{\partial \beta_j} = \sum_i \left[d_{ij} - P_{ij}\right] x_i \quad \text{for } j = 1, \ldots, J. \tag{F.10}$$

The exact second derivatives matrix has $J^2 K \times K$ blocks,

$$\frac{\partial^2 \ln L}{\partial \beta_j \partial \beta_l'} = -\sum_{i=1}^{n} P_{ij}[1(j = l) - P_{il}] x_i x', \tag{F.11}$$

where $1(j = 1)$ equals 1 if $j$ equals $l$, and 0 if not.

# References

1. Greene, W.H. (2000) *Econometric Analysis*, 4th ed., Prentice Hall, Upper Saddle River, NJ.
2. Borooah, V.K. (2001) *Logit and Probit: Ordered and Multinomial Models*, Sage, Thousand Oaks, CA.
3. McFadden, D. (1973) Conditional logit analysis of qualitative choice behaviour, in *Frontiers in Econometrics* (ed. P. Zarembka), Academic Press, New York.

# Appendix G

# Survival analysis – an introduction

## (Sources: Greene [1], Cameron and Trivedi [2], and Le [3])

Suppose that the random variable $T$ has a continuous probability distribution $f(t)$, where $t$ is a realization of $T$. The cumulative probability is

$$F(t) = \int_0^t f(s)\,ds = Prob(T \leq t). \tag{G.1}$$

We will usually be more interested in the probability that the spell is of length at least $t$, which is given by the survival function

$$S(t) = 1 - F(t) = Prob(T \geq t). \tag{G.2}$$

Consider the question: 'Given that the spell has lasted until time $t$, what is the probability that it will end in the next short interval of time, say $\Delta$?' This is

$$l(t, \Delta) = Prob(t \leq T \leq t + \Delta | T \geq t). \tag{G.3}$$

A useful function for characterizing this aspect of the distribution is the hazard rate,

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

$$\begin{aligned}\lambda(t) &= \lim\nolimits_{\Delta\to 0}\frac{Prob(t\le T\le t+\Delta|T\ge t)}{\Delta}\\ &= \lim\nolimits_{\Delta\to 0}\frac{F(t+\mathrm{D})-F(t)}{\Delta S(t)}\\ &= \frac{f(t)}{S(t)}.\end{aligned}\tag{G.4}$$

The hazard rate is the rate at which spells are completed after duration $t$, given that they last at least until $t$. The hazard function can also be expressed in the form of a survival function, the density or distribution function:

$$\lambda(t)=\frac{-d\ln S(t)}{dt}\tag{G.5}$$

and

$$f(t)=S(t)\lambda(t).\tag{G.6}$$

The hazard $\lambda(t)$ specifies the distribution of $T$. In particular, integrating $\lambda(t)$ and using $S(0)=1$ shows that

$$S(t)=\exp\left(-\int_0^t\lambda(u)\,du\right).\tag{G.7}$$

A final related function is the cumulative hazard function or integrated hazard function

$$\begin{aligned}\Lambda(t) &= \int_0^t\lambda(s)\,ds\\ &= -\ln S(t).\end{aligned}\tag{G.8}$$

Estimation of the survival function can be done by maximum likelihood, the procedure provided in Appendix D. Here, we provide an example of the estimation of the exponential distribution from Le [3].

Suppose that we may be able to assume a parametric model for survival times, for example, an exponential model with density $f(t;\rho)$. Given a random sample of survival times

$$\{t_i\}_{i=1}^n\tag{G.9}$$

and a density function, denoted $f(t_i;\theta)$, the likelihood function of $\theta$ is

$$\begin{aligned}L(\theta) &= \prod_{i=1}^n f(t_i;\theta)\\ &= \prod_{i=1}^n\lambda\exp(-\lambda t_i)\\ &= \lambda^n\exp(-\lambda\textstyle\sum t_i)\\ &= \lambda^n\exp(-\lambda T)\quad\text{with } T=\textstyle\sum t_i \text{ being total time.}\end{aligned}\tag{G.10}$$

This leads to

$$\ln L = n \ln \lambda - \lambda T. \tag{G.11}$$

The first-order derivative gives

$$\hat{\lambda} = \frac{n}{T}. \tag{G.12}$$

From the second-order derivative, we get

$$-\frac{d^2}{d\lambda^2} \ln L = \frac{n}{\lambda^2} \tag{G.13}$$

and the standard error (SE) of $\hat{\lambda}$ is given by

$$SE(\hat{\lambda}) = \frac{\hat{\lambda}}{\sqrt{n}}. \tag{G.14}$$

# References

1. Greene, W.H. (2000) *Econometric Analysis*, 4th ed., Prentice Hall, Upper Saddle River, NJ.
2. Cameron, A.C. and Trivedi, P.K. (2005) *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.
3. Le, C.T. (1997) *Applied Survival Analysis*, John Wiley & Sons, Inc., New York.

# Appendix H

# Discrete-time hazard

## (Source: Cameron and Trivedi [1])

The starting point is to define the discrete-time hazard function as the probability of transaction at discrete time $t_j$, $j = 1, 2, \ldots$, given survival to time $t_j$:

$$\begin{aligned}\lambda_j &= \Pr\left[T = t_j | T \geq t_j\right] \\ &= f^d(t_j)/S^d(t_j),\end{aligned} \tag{H.1}$$

where the superscript $d$ denotes discrete, and where $S^d(a_-) = \lim_{t \to a_-} S^d(t_j)$, an adjustment made because formally $S^d(t)$ equals $\Pr[T > t]$ rather than $\Pr[T \geq t]$.

The discrete-time survivor function is obtained recursively from the hazard function as

$$\begin{aligned}S^d(t) &= \Pr[T \geq t] \\ &= \prod_{j \mid t_j \leq t} (1 - \lambda_j).\end{aligned} \tag{H.2}$$

For example, $\Pr[T > t_2]$ equals the probability of no transition at time $t_1$ times the probability of no transition at time $t_2$ conditional on surviving to just before $t_2$, so that $\Pr[T > t_2] = (1 - \lambda_1) \times (1 - \lambda_2)$. The function $S^d(t)$ is a decreasing step function with steps at $t_j$, $j = 1, 2, \ldots$.

The discrete-time cumulative hazard function is

$$\Lambda^d(t) = \sum_{j \mid t_j \leq t} \lambda_j. \tag{H.3}$$

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

Using Equation H.1, we have that the discrete probability that the spell ends at $t_j$ is $\lambda_j S^d(t_j)$.

## Reference

1. Cameron, A.C. and Trivedi, P.K. (2005) *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.

# Appendix I

# Proportional hazards model

## (Sources: Kalbfleisch and Prentice [1], Franses and Paap [2])

A second way to include explanatory variables in a duration model is to scale the hazard function by the function $\psi(\cdot)$, that is,

$$\lambda(t_i \,|\, x_i) = \psi(x_i)\lambda_0(t_i), \tag{I.1}$$

where $\lambda_0(t_i)$ denotes an arbitrary, unspecified, baseline hazard function for continuous $T$. Again, because the hazard function has to be non-negative, one usually specifies $\psi(\cdot)$ as

$$\psi(x_i) = \exp(\beta_0 + \beta_1 x_i). \tag{I.2}$$

If the intercept $\beta_0$ is unequal to 0, the baseline hazard is identified upon a scalar. Here, if one opts for a Weibull or an exponential baseline hazard one again has to restrict $\gamma$ to 1 to identify the parameters.

The conditional density function of $T$ given covariates $z$ is

$$f(t; z) = \lambda_0(t)\exp(z\beta)\exp\left[-\exp(z\beta)\int_0^t \lambda_0(u)\,du\right]. \tag{I.3}$$

The conditional survivor function for $T$ given covariates $z$ is

$$F(t; z) = [F_0(t)]^{\exp(z\beta)} \tag{I.4}$$

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

where

$$F_0(t) = \exp\left[-\int_0^t \lambda_0(u)\,du\right]. \tag{I.5}$$

The log-likelihood function for the proportional hazards model

$$\lambda(t_i|X_i) = \exp(X_i\beta)\lambda_0(t_i)$$

is given by

$$l(\theta) = \sum_{i=1}^{N} [d_i X_i\beta + d_i \log\lambda_0(t_i) - \exp(X_i\beta)\Lambda_0(t_i)], \tag{I.6}$$

which allows for various specifications of the baseline hazard. If we assume that the parameters of the baseline hazard are summarized in $\alpha$, the first-order derivatives of the log-likelihood are given by

$$\frac{\partial l(\theta)}{\partial\beta} = \sum_{i=1}^{N} [d_i - \exp(X_i\beta)\Lambda_0(t_i)]X_i' \tag{I.7}$$

$$\frac{\partial l(\theta)}{\partial\alpha} = \sum_{i=1}^{N} \left(\frac{d_i}{\lambda_0(t_i)}\frac{\partial\lambda_0(t_i)}{\partial\alpha} - \exp(X_i\beta)\frac{\partial\Lambda_0(t_i)}{\partial\alpha}\right). \tag{I.8}$$

The second-order derivatives are given by

$$\frac{\partial^2 l(\theta)}{\partial\beta\partial\beta'} = -\sum_{i=1}^{N} \exp(X_i\beta)\Lambda_0(t_i)X_i'X_i \tag{I.9}$$

$$\frac{\partial^2 l(\theta)}{\partial\alpha\partial\beta} = \sum_{i=1}^{N} \exp(X_i\beta)\frac{\partial \Lambda_0(t_i)}{\partial\alpha'}X' \tag{I.10}$$

$$\frac{\partial^2 l(\theta)}{\partial\alpha\partial\alpha'} = -\sum_{i=1}^{N} \left(\frac{d_i}{\lambda_0(t_i)}\frac{\partial^2\lambda_0(t_i)}{\partial\alpha\partial\alpha'} - \frac{d_i}{\lambda_0(t_i)^2}\frac{\partial\lambda_0(t_i)}{\partial\alpha}\frac{\partial\lambda_0(t_i)}{\partial\alpha'} - \exp(X_i\beta)\frac{\partial^2\Lambda_0(t_i)}{\partial\alpha\partial\alpha'}\right), \tag{I.11}$$

which shows that we need the first- and second-order derivatives of the baseline hazard and the integrated baseline hazard. If we assume a Weibull baseline hazard with $\gamma = 1$, the integrated baseline hazard is $\Lambda_0(t) = t^\alpha$. Straightforward differentiation gives

$$\frac{\partial \lambda_0(t_i)}{\partial \alpha} = (1 + \alpha \log(t))t^\alpha \quad \text{and} \quad \frac{\partial^2 \lambda_0(t_i)}{\partial \alpha^2} = (2\log(t) + \alpha \log(t))^2 t^{\alpha - 1} \tag{I.12}$$

$$\frac{\partial \Lambda_0(t_i)}{\partial \alpha} = t^\alpha \log(t) \quad \text{and} \quad \frac{\partial^2 \Lambda_0(t_i)}{\partial \alpha^2} = t^\alpha (\log(t))^2. \tag{I.13}$$

The maximum likelihood estimates are found by iterating over

$$\theta_h = \theta_{h-1} - H(\theta_{h-1})^{-1} G(\theta_{h-1}) \tag{I.14}$$

for properly chosen starting values for $\beta$ and $\alpha$.

# References

1. Kalbfleisch, J.D. and Prentice, R.L. (1980) *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, Inc., New York.
2. Franses, P.H. and Paap, R. (2001) *Quantitative Models in Marketing Research*, Cambridge University Press, Cambridge.

# Appendix J

# Random intercept model

## (Sources: Cameron and Trivedi [1], Hsiao [2])

The random effects model can be written as

$$y_{it} = \mu + x_{it}'\beta + \alpha_i + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{J.1}$$

or

$$y_i = \tilde{X}_i\delta + v_i, \quad i = 1, 2, \ldots N, \tag{J.2}$$

where $\tilde{X}_i = (e, X_i)$, $\delta' = (\mu, \beta')$, $e' = (1, 1, \ldots, 1)$, $v_i = (v_{i1}, \ldots, v_{iT})$. The variance–covariance matrix of $v_i$ is

$$Ev_iv_i' = \sigma_u^2 I_T + \sigma_\alpha^2 ee' = V. \tag{J.3}$$

Its inverse is

$$V^{-1} = \frac{1}{\sigma_u^2}\left(I_T - \frac{\sigma_\alpha^2}{\sigma_u^2 + T\sigma_\alpha^2} ee'\right). \tag{J.4}$$

The individual-specific effects $\alpha_i$ are assumed to be realizations of i.i.d. random variables with distribution $[0, \sigma_\alpha^2]$ and the error $u_{it}$ is i.i.d. $[0, \sigma_u^2]$. The model can be re-expressed as $y_{it} = \mu + x_{it}'\beta + v_{it}$, where the error term $v_{it}$ has two components $v_{it} = \alpha_i + u_{it}$. For this reason the random effects model is also called the error

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

components model or random intercept model. The random intercept model can be estimated by the maximum likelihood method.

When $\alpha_i$ and $u_{it}$ are random and normally distributed, the logarithm of the likelihood function is

$$\begin{aligned}\log L &= -\frac{NT}{2}\log 2\pi - \frac{N}{2}\log|V| - \frac{1}{2}\sum_{i=1}^{N}(y_i - e\mu - X_i\beta)'V^{-1}(y_i - e\mu - X_i\beta) \\ &= -\frac{NT}{2}\log 2\pi - \frac{N(T-1)}{2}\log\sigma_u^2 - \frac{N}{2}\log(\sigma_u^2 + T\sigma_\alpha^2) \\ &\quad -\frac{1}{2\sigma_u^2}\sum_{i=1}^{N}(y_i - e\mu - X_i\beta)'Q(y_i - e\mu - X_i\beta) \\ &\quad -\frac{T}{2(\sigma_u^2 + T\sigma_\alpha^2)}\sum_{i=1}^{N}(\bar{y}_i - \mu - \beta'\bar{x}_i)^2\end{aligned} \tag{J.5}$$

where

$$|V| = \sigma_u^{2(T-1)}(\sigma_u^2 + T\sigma_\alpha^2). \tag{J.6}$$

The maximum likelihood estimator (MLE) of $(\mu, \beta', \sigma_u^2, \sigma_\alpha^2) = \tilde{\delta}'$ is obtained by solving the following first-order conditions simultaneously:

$$\frac{\partial \log L}{\partial \mu} = \frac{T}{\sigma_u^2 + T\sigma_\alpha^2}\sum_{i=1}^{N}(\bar{y}_i - \mu - \beta'\bar{x}_i) = 0, \tag{J.7}$$

$$\frac{\partial \log L}{\partial \beta} = \frac{1}{\sigma_u^2}\left[\sum_{i=1}^{N}(y_i - e\mu - X_i\beta)'QX_i - \frac{T\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}\sum_{i=1}^{N}(\bar{y}_i - \mu - \bar{x}_i'\beta)\bar{x}_i'\right] = 0, \tag{J.8}$$

$$\begin{aligned}\frac{\partial \log L}{\partial \sigma_u^2} &= -\frac{N(T-1)}{2\sigma_u^2} - \frac{N}{2\left(\sigma_u^2 + T\sigma_\alpha^2\right)} \\ &\quad + \frac{1}{2\sigma_u^4}\sum_{i=1}^{N}(y_i - e\mu - X_i\beta)'Q(y_i - e\mu - X_i\beta) \\ &\quad + \frac{T}{2(\sigma_u^2 + T\sigma_\alpha^2)}\sum_{i=1}^{N}(\bar{y}_i - \mu - \bar{x}_i'\beta)^2 = 0,\end{aligned} \tag{J.9}$$

$$\frac{\partial \log L}{\partial \sigma_\alpha^2} = -\frac{NT}{2(\sigma_u^2 + T\sigma_\alpha^2)} + \frac{T^2}{2(\sigma_u^2 + T\sigma_\alpha^2)^2}\sum_{i=1}^{N}(\bar{y}_i - \mu - \bar{x}_i'\beta)^2 = 0. \tag{J.10}$$

Simultaneous solution of Equations J.7–J.10 is complicated. The Newton–Raphson iterative procedure can be used to solve for the MLE. The procedure uses an initial trial value $\hat{\delta}^{(1)}$ of $\tilde{\delta}$ to start the iteration by substituting it into the formula

$$\tilde{\delta}^{(j)} = \hat{\delta}^{(j-1)} - \left[\frac{\partial^2 \log L}{\partial \tilde{\delta}^{(j)} \partial \delta'}\right]^{-1}_{\tilde{\delta}^{(j)} = \hat{\delta}^{(j-1)}} \frac{\partial \log L}{\partial \tilde{\delta}}_{\,|\tilde{\delta} = \hat{\delta}^{(j-1)}} \tag{J.11}$$

to obtain a revised estimate of $\tilde{\delta}$, $\hat{\delta}^{(2)}$. The process is repeated until the *j*th iterative solution $\hat{\delta}^{(j)}$ is close to the $(j-1)$th iterative solution $\hat{\delta}^{(j-1)}$.

# References


1. Cameron, A.C. and Trivedi, P.K. (1998) *Regression Analysis of Count Data*, Cambridge University Press, New York.
2. Hsiao, C. (2003) *Analysis of Panel Data*, 2nd ed., Cambridge University Press, Cambridge.

# Appendix K

# Poisson regression model

## (Source: Greene [1])

The Poisson regression model specifies that each $y_i$ is drawn from a Possion distribution with parameter $\lambda_i$, which is related to the regressors $x_i$. The primary equation of the model is

$$Prob(Y_i = y_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \ldots . \tag{K.1}$$

The most common formulation for $\lambda_i$ is the log-linear model,

$$\ln \lambda_i = \beta' x_i. \tag{K.2}$$

It is easily shown that the expected number of events *per period* is given by

$$\mathrm{E}[y_i|x_i] = Var[y_i|x_i] = \lambda_i = \exp(\beta' x_i), \tag{K.3}$$

so

$$\frac{\partial \mathrm{E}[y_i|x_i]}{\partial x_i} = \lambda_i \beta. \tag{K.4}$$

With the parameter estimates in hand, this vector can be computed using any data vector desired.

In practice, the Poisson model is simply a nonlinear regression. But it is far easier to estimate the parameters with maximum likelihood techniques. The

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

log-likelihood function is

$$\ln L = \sum_{i=1}^{n} [-\lambda_i + y_i \beta' x_i - \ln y_i!]. \tag{K.5}$$

The likelihood equations are

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n} (y_i - \lambda_i) x_i = 0. \tag{K.6}$$

The Hessian is

$$\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} \lambda_i x_i x_i'. \tag{K.7}$$

The Hessian is negative definite for all $x$ and $\beta$. Newton's method is a simple algorithm for this model and will usually converge rapidly. At convergence,

$$\left[ \sum_{i=1}^{n} \hat{\lambda}_i x_i x_i' \right]^{-1}$$

provides an estimator of the asymptotic covariance matrix for the parameter estimates. Given the estimates, the prediction for observation $i$ is $\hat{\lambda}_i = \exp(\bar{\beta}' x)$. A standard error for the prediction interval can be formed by using a linear Taylor series approximation. The estimated variance of the prediction will be $\hat{\lambda}_i^2 x_i' V x_i$, where $V$ is the estimated asymptotic covariance matrix for $\hat{\beta}$.

# Reference

1. Greene, W.H. (2000) *Econometric Analysis*, 4th ed., Prentice Hall, Upper Saddle River, NJ.

# Appendix L

# Negative binomial regression

## (Source: Cameron and Trivedi [1])

In the Poisson regression model $y_i$ has mean $\lambda_i = \exp(X_i'\beta)$ and variance $\lambda_i$. We now relax the variance assumption, because data almost always reject the restriction that the variance equals the mean, and we maintain the assumption that the mean is $\exp(X_i'\beta)$.

We use the general notation

$$\omega_i = V[y_i|x_i] \tag{L.1}$$

to denote the conditional variance of $y_i$. It is natural to continue to model the variance as a function of the mean, with

$$\omega_i = \omega(\lambda_i, \alpha) \tag{L.2}$$

for some specified function $\omega(\cdot)$ and where $\alpha$ is a scalar parameter. Most models specialize this to the general variance function

$$\omega_i = \lambda_i + \alpha\lambda_i^p \tag{L.3}$$

where the constant $p$ is specified. Analysis is usually restricted to two special cases, in addition to the Poisson case of $\alpha = 0$.

First, the NB1 variance function sets $p = 1$. Then the variance

$$\omega_i = (1 + \alpha)\lambda_i \tag{L.4}$$

is a multiple of the mean.

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

Second, the NB2 variance function sets $p = 2$. Then the variance is quadratic in the mean:

$$\omega_i = \lambda_i + \alpha\lambda_i^2. \tag{L.5}$$

In both cases the dispersion parameter $\alpha$ is to be estimated.

The most common implementation of the negative binomial is the NB2 model, with NB2 variance function $\lambda_i + \alpha\lambda_i^2$. It has density

$$f(y|\lambda,\alpha) = \frac{\Gamma(y+\alpha^{-1})}{\Gamma(y+1)\Gamma(\alpha^{-1})}\left(\tfrac{\alpha^{-1}}{\alpha^{-1}+\lambda}\right)^{\alpha^{-1}}\left(\tfrac{\lambda}{\alpha^{-1}+\lambda}\right)^{y}, \quad \alpha \geq 0, \quad y = 0,1,2,\ldots. \tag{L.6}$$

The function $\Gamma(\cdot)$ is a gamma function, defined as

$$\Gamma(\alpha) = \int_0^\infty \exp(-t)t^{a-1}\,dt, \quad a > 0 \tag{L.7}$$

It is shown that $\Gamma(y+\alpha)/\Gamma(\alpha) = \prod_{j=0}^{y-1}(j+a)$, if $y$ is an integer. Thus,

$$\ln\left(\frac{\Gamma(y+\alpha^{-1})}{\Gamma(\alpha^{-1})}\right) = \sum_{j=0}^{y-1}\ln\big(j+\alpha^{-1}\big). \tag{L.8}$$

Substituting Equation L.8 into L.7, the log-likelihood function for exponential mean $\lambda_i = \exp(X_i'\beta)$ is therefore

$$\ln L(\alpha,\beta) = \sum_{i=1}^{n}\left\{\left(\sum_{j=0}^{y_i-1}\ln\big(j+\alpha^{-1}\big)\right) - \ln y_i! - \big(y_i+\alpha^{-1}\big)\ln\big(1+\alpha\exp\big(X_i'\beta\big)\big) + y_i\ln\alpha + y_iX_i'\beta\right\}. \tag{L.9}$$

The NB2 MLE $\big(\hat{\beta}_{NB2},\hat{\alpha}_{NB2}\big)$ is the solution to the first-order conditions

$$\sum_{i=1}^{n}\frac{y_i-\lambda_i}{1+\alpha\lambda_i}x_i = 0 \tag{L.10}$$

$$\sum_{i=1}^{n}\left[\frac{1}{\alpha^2}\left(\ln(1+\alpha\lambda_i) - \sum_{j=0}^{y_i-1}\frac{1}{(j+\alpha^{-1})}\right) + \frac{y_i-\lambda_i}{\alpha(1+\alpha\lambda_i)}\right] = 0. \tag{L.11}$$

The NB1 log-likelihood function is

$$\ln L(\alpha,\beta) = \sum_{i=1}^{n}\left\{\left[\sum_{j=0}^{y_i-1}\ln\big(j+\alpha^{-1}\exp(X_i'\beta)\big)\right] - \ln y_i! - \big(y_i+\alpha^{-1}\exp(X_i'\beta)\big)\ln(1+\alpha) + y_i\ln\alpha\right\}. \tag{L.12}$$

The NB1 MLE solves the first-order conditions

$$\sum_{i=1}^{n} \left[ \left( \sum_{j=0}^{y_i-1} \frac{\alpha^{-1}\lambda_i}{j+\alpha^{-1}\lambda_i} \right) x_i + \alpha^{-1}\lambda_i x_i \right] = 0 \tag{L.13}$$

$$\sum_{i=1}^{n} \frac{1}{\alpha^2} \left[ -\left( \sum_{j=0}^{y_i-1} \frac{\lambda_i}{j+\alpha^{-1}} \right) - \alpha^2\lambda_i \ln(1+\alpha) - \frac{\alpha}{1+\alpha} + y_i\alpha \right] = 0. \tag{L.14}$$

# Reference

1. Cameron, A.C. and Trivedi, P.K. (1998) *Regression Analysis of Count Data*, Cambridge University Press, New York.

# Appendix M

# Estimation of Tobit model with selection

## (Source: Thomas [1])

The basic assumptions are that the error terms in Equations 5.4 and 5.5 are bivariate normal – that is, $\varepsilon_{is} \sim N(0, \sigma^2_{\varepsilon s})$, $\mu_{is} \sim N(0, 1)$, $(\varepsilon_{is}, \mu_{is}) \sim BVN(0, 0, \varepsilon_{is}, 1, \rho_s)$– and are not correlated with the regressors. Therefore, a truncated bivariate normal distribution is used to derive the likelihood function for the Tobit model with selection. Specifically, an individual's conditional likelihood function is as follows:

$$L_{i|s} = \left[\frac{Prob\left(z_i^* > 0|y_i\right)Prob\left(y_i^*\right)}{Prob\left(z_i^* > 0\right)}\right]^{1-\delta_i} \times \left[\frac{Prob\left(y_i^* \geq c_i, z_i^* > 0\right)}{Prob\left(z_i^* > 0\right)}\right]^{\delta_i}. \qquad \text{(M.1)}$$

Given the assumption of normality of the error terms, Equation M.1 can be respecified as follows:

$$L_{i|s} = \left\{\frac{[1 - \Phi(\varpi_{is})] \times \phi(\varepsilon_{is}/\sigma_{\varepsilon s})/\sigma_{\varepsilon s}}{1 - \Phi(-\alpha_s v_i)}\right\}^{1-\delta_i}$$
$$\times \left[\frac{\int_{(c_i - \beta_s' x_i)/\sigma_{\varepsilon s}}^{\infty} \int_{-\alpha_s v_i}^{\infty} SBVN(\varepsilon_{is}/\sigma_{\varepsilon s}, \mu_{is})\, d\mu\, d\varepsilon}{1 - \Phi(-\alpha_s v_i)}\right]^{\delta_i} \qquad \text{(M.2)}$$

where

$$\varpi_{is} = \frac{-\alpha_s v_i - (\sigma_{\mu\varepsilon_s}\varepsilon_{is}/\sigma^2_{\varepsilon s})}{[1 - (\sigma^2_{\mu\varepsilon_s}/\sigma^2_{\varepsilon s})]^{1/2}} \qquad \text{(M.3)}$$

*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.

and $\sigma_{\varepsilon s}$ is the standard deviation of $\varepsilon$ for segment $s$, $\sigma_{\mu\varepsilon_s}$ is the covariance between $\mu$ and $\varepsilon$ for segment $s$, and SBVN is the standard bivariate normal distribution.

Given the conditional likelihood function in Equation M.2, Kamakura and Russell's [2] approach asserts that the unconditional likelihood can be determined by the following:

$$L_i = \sum_{s=1}^{S} f_s \times L_{i|s} \tag{M.4}$$

such that

$$f_s = \frac{\exp(\lambda_s)}{\sum_{s=1}^{S} \exp(\lambda_s)}, \tag{M.5}$$

where $f_s$ is the probability that a consumer is in segment $s$, and $\lambda_s$ is a segment-specific parameter. Based on this specification, the sample likelihood is

$$L = \prod_{i=1}^{N} L_i \tag{M.6}$$

where $N$ is the total number of observations.

# References

1. Thomas, J.S. (2001) A methodology for linking customer acquisition to customer retention. *Journal of Marketing Research*, **38**(2), 262–268.
2. Kamakura, W.A. and Russell, G.J. (1989) A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, **26**(4), 379–390.

# Index



*Statistical Methods in Customer Relationship Management*, First Edition. V. Kumar and J. Andrew Petersen.