

**THE PERCEIVED CREDIBILITY OF
ELECTRONIC WORD-OF-MOUTH COMMUNICATION
ON E-COMMERCE PLATFORMS**

by
DIRK JOHANNES BOSMAN

**THESIS PRESENTED IN FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF COMMERCE**



**IN THE FACULTY OF ECONOMIC AND MANAGEMENT SCIENCES
UNIVERSITY OF STELLENBOSCH**

**SUPERVISORS:
PROF C. BOSHOFF AND PROF G.J. VAN ROOYEN**

December 2012

DECLARATION REGARDING PLAGIARISM

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature

Name in full

Date

ABSTRACT

Enterprises and more specifically, marketing departments, function in a complex global market, while trying to deliver products and services to satisfy the needs of consumers. It is estimated that by 2013, enterprises will be spending \$4.75 trillion and consumers \$330 billion by means of commercial transactions over the Internet, and that by 2050 most transactions – if not all transactions – will be e-commerce based (Laudon and Traver, 2010:1-7).

The 24-hour access to a global network of markets has brought about two major challenges for most enterprises. Firstly, the Internet as a publishing platform has exponentially increased the creation and sharing of information, which has significantly increased consumers' search cost; and secondly, as more electronic word-of-mouth (EWOM) is being generated online, a significant amount of power and influence over enterprises has shifted to consumers (Chen, Wu and Yoon, 2004:716-722; Tapscott and Williams, 2008:52-53). Ultimately, enterprises are challenged to harness the power of EWOM for more successful e-commerce strategies and increased market share.

Given previous studies, it was possible to extend the theoretical framework of EWOM communication in the fields of Internet marketing and online consumer behaviour. The purpose of this study was to create two models that could measure, over time, the impact of EWOM review communication on an e-commerce platform, specifically with regard to review credibility and sales levels. In using a non-probability judgement sampling procedure, it emerged that EWOM reviews do indeed influence the sales levels of e-commerce platform *Amazon.com*, and that certain review factors (platform, text length, time and star ratings) significantly influenced the credibility of *Amazon.com* and *Barnesandnoble.com* reviews. Furthermore, it was concluded that the overall credibility of reviews increases over time as more and more online users have the ability to scrutinise it. When *Amazon.com* and *Barnesandnoble.com*'s reviews were compared to each other, the results indicated that *Amazon.com* had more reviews than *Barnesandnoble.com*, and that the reviews posted at *Amazon.com* had on average longer text lengths and were found to be more helpful than the reviews at *Barnesandnoble.com*.

The study's major contribution is that it provides wide-ranging guidelines for usability and user experience design, sales and inventory forecasting, as well as benchmark statistics for marketing campaigns.

OPSOMMING

Ondernemings, en in die besonder bemarkingsafdelings, funksioneer in 'n komplekse globale mark, in hulle strewe om voortdurend produkte en dienste te lewer wat voldoen aan verbruikersbehoefes. Na raming sal ondernemings teen die jaar 2013 \$4.75 triljoen en verbruikers \$330 biljoen spandeer aan kommersiële transaksies oor die Internet, terwyl die meeste, indien nie alle transaksies, teen die jaar 2050 gebaseer gaan wees op e-handel (Laudon en Traver, 2010:1-7).

Die 24-uur toegang tot 'n globale netwerk van markte het twee hoofuitdagings vir die meeste ondernemings tot gevolg. In die eerste plek het die Internet as uitgewersplatform die skep en verspreiding van inligting eksponensieel laat toeneem, wat verbruikers se soekoste noemenswaardig verhoog het; en tweedens, namate elektroniese hoorsê aanlyn gegeneer word, het 'n beduidende hoeveelheid mag en invloed van ondernemings na die verbruiker verskuif (Chen, Wu en Yoon, 2004:716-722; Tapscott en Williams, 2008:52-53). Ondernemings word dus uitgedaag om die impak van elektroniese hoorsê (*electronic word-of-mouth*) te ontgin om meer suksesvolle e-handelstrategieë en verhoogde winste te verkry.

In die lig van vorige studies was dit moontlik om die tradisionele teoretiese raamwerk van hoorsêkommunikasie uit te brei na die veld van Internet-bemarking en verbruikersgedrag. Die doel van hierdie studie was om twee modelle te ontwikkel wat, met verloop van tyd, die impak van elektroniese hoorsêkommunikasie op 'n e-handelplatform kon meet, met spesifieke verwysing na resensiegeloofwaardigheid en verkoopsvlakke van boeke.

Deur die gebruik van 'n nie-waarskynlikheidsoordeel steekproefprosedure, het die studie bevind dat elektroniese hoorsêresensies inderdaad 'n invloed het op die verkoopsvlakke van elektroniese e-handelplatforms en dat sekere resensiefaktore (platform, teks, lengte, tyd en stergradering) die geloofwaardigheid van elektroniese e-handelplatforms se resensies beduidend beïnvloed het. Verder is die gevolgtrekking gemaak dat die oorhoofse geloofwaardigheid van resensies toeneem namate al hoe meer aanlynverbruikers die vermoë het om dit onder oë te kry. In 'n vergelyking tussen die resensies van twee e-handelplatforms *Amazon.com* en *Barnesandnoble.com*, is bevind dat *Amazon.com* meer resensies as *Barnesandnoble.com* het, dat die resensies op *Amazon.com* gemiddeld langer tekste het, en ook meer behulpsaam uit 'n verbruikersoogpunt is as die resensies op *Barnesandnoble.com*.

Die hoofbydrae van hierdie studie is dat dit riglyne bied aan e-handelplatforms om hul kliënte se gebruikerservaring beter te verstaan, en om sodoende beter verkope- en voorraadvooruitskattings te kan maak. Verder bied die studie ook riglyne t.o.v. doeltreffende bemarkingsveldtogte vir e-handelplatforms.

APPRECIATION

Jesus, for blessing me with the opportunity to have worked with an amazing group of people over the past 2 years of my life. Thank-you for the daily abundance of hope, joy and wisdom. Your grace and love is enough.

Professor Gert-Jan van Rooyen, for your immense wisdom and guidance in the field of technology. If there was a programme that one could download on how to approach the technology industry, I would want you to code it. For the long hours we have spent talking about *Buscape.com.br*, *Brandwoo.com*, *Amazon.com*, *Barnesandnoble.com* and *Kalahari.com* - I thank you for every minute! I also want to acknowledge you, Dr. Engelbrecht, Dr. McElory and the late Antonie Roux for all the time you invest into the MIH Media Lab - my home away from home. May the fruits of your hard labour be reflected in our work.

Professor Christo Boshoff, for guiding my broad thoughts on Internet marketing and consumer behaviour into refined research objectives. I now know why you are world-renown and highly respected. I am also immensely thankful for our coffee-breaks and your general wisdom on how I need to approach the next stage of my career. I also wanted to thank you for allowing me to pursue an international exchange opportunity, while writing the final part of my thesis. I am humbled by the experience.

Doctor Tom Berning, for your guidance and insights in statistical analysis.

Family and Friends, for the way you have supported me with my academic career. I thank you for allowing me to pursue this dream and standing by me all the way. For the late night tea sessions, the good food and love - I thank you all so dearly from the bottom of my heart! A special thanks to my close entourage: Louis, Narien, Carin, Cobus, Ouma "Joe", Derek, Casper, Elmi, Emma, Reenen, Nico and Marc.

Labrats, for all the people I had the privilege to study, work and play with, thank you! It was an honour. "*May the force be with you*", always!

The Department of Business Management, MIH Media Lab and the **Ackerman family** for all the support through the years.

“May the words of my mouth and the meditation of my heart be pleasing in your sight, O Lord, my Rock and my Redeemer” (Psalm 19:14)

**THE PERCEIVED CREDIBILITY OF
ELECTRONIC WORD-OF-MOUTH COMMUNICATION
ON E-COMMERCE PLATFORMS**

Table of Contents

1. INTRODUCTION	1
1.1 INTRODUCTION	1
1.2 BUSINESS AND MARKETING MANAGEMENT.....	2
1.3 WORD-OF-MOUTH COMMUNICATION	3
1.4 ELECTRONIC WORD-OF-MOUTH COMMUNICATION	3
1.4.1 The Internet and the implication it has on business management	3
1.4.2 Digital channels and information technology	4
1.4.3 Electronic word-of-mouth communication.....	4
1.5 CONSUMER WORD-OF-MOUTH COMMUNICATION.....	5
1.6 STATEMENT OF THE PROBLEM.....	5
1.7 THE DEPENDENT AND INDEPENDENT VARIABLES	6
1.8 CONTRIBUTION OF THE PRESENT STUDY	7
1.9 RESEARCH OBJECTIVES.....	7
1.10 HYPOTHESES	8
1.11 METHODOLOGY	10
1.11.1 Secondary research	10
1.11.2 Primary research.....	10
1.11.3 Sampling procedure	11
1.11.4 Data collection and analysis	11
1.12 ORIENTATION OF THE STUDY	12
2. BUSINESS MANAGEMENT	13
2.1 INTRODUCTION	13

2.2 THE ROLE OF BUSINESS IN SOCIETY.....	13
2.3 NEEDS AND NEED SATISFACTION.....	14
2.3.1 Hierarchy of human needs.....	14
2.3.2 Needs satisfaction cycle.....	15
2.4 THE NATURE OF BUSINESS MANAGEMENT.....	16
2.5 ENTERPRISE MANAGEMENT FUNCTIONS.....	17
2.5.1 General and strategic management function.....	17
2.5.2 Production function.....	18
2.5.3 Financial function.....	18
2.5.4 Purchasing function.....	19
2.5.5 Human resources function.....	20
2.5.6 Information function.....	20
2.5.7 Public relations function.....	20
2.5.8 Marketing function.....	21
2.6 CONCLUSION.....	26
3. CONSUMER WORD-OF-MOUTH COMMUNICATION.....	27
3.1 INTRODUCTION.....	27
3.2 WORD-OF-MOUTH COMMUNICATION.....	27
3.2.1 Defining word-of-mouth.....	27
3.2.2 Motives for participating in word-of-mouth communication.....	28
3.2.3 The communication process.....	29
3.2.4 The word-of-mouth communication model.....	31
3.3 ELECTRONIC AND CONSUMER WORD-OF-MOUTH COMMUNICATION.....	34

3.3.1 The power shift in business, marketing, and online consumer behaviour	35
3.3.2 A comparison between word-of-mouth and electronic word-of-mouth communication.....	37
3.4 TECHNOLOGIES ENABLING ELECTRONIC WORD-OF-MOUTH COMMUNICATION..	39
3.5 THE SEARCH COST AND CREDIBILITY OF INFORMATION ON REVIEW WEB PAGES OF E-COMMERCE PLATFORMS	41
3.6 CONCLUSION.....	42
4. RESEARCH METHODOLOGY	44
4.1 INTRODUCTION	44
4.2 THE MARKETING RESEARCH PROCESS	44
4.2.1 Step 1: Formulate the research problem	45
4.2.2 Step 2: Formulate the research objectives and hypotheses	46
4.2.3 Step 3: Conduct the research design	50
4.2.4 Step 4: Select a primary research method	52
4.2.5 Step 5: Conduct the sampling procedure	53
4.2.6 Step 6: Collect, process and analyse the data.....	55
4.3 CONCLUSION.....	71
5. RESEARCH RESULTS.....	73
5.1 INTRODUCTION	73
5.2 RESULTS.....	73
5.2.1 Descriptive statistics.....	73
5.2.2 Inferential statistics.....	83
5.3 CONCLUSION.....	98

6. CONCLUSION AND RECOMMENDATIONS	99
6.1 INTRODUCTION	99
6.2 CONCLUSIONS FROM THE STUDY	99
6.2.1 The target population and realised samples	99
6.2.2 Individual electronic word-of-mouth review factors and review credibility (model 1).....	100
6.2.3 Spotlight electronic word-of-mouth and sales rank levels (model 2)	106
6.3 IMPLICATIONS AND RECOMMENDATIONS	107
6.3.1 Usability and user experience design	108
6.3.2 Forecasting, review insights and benchmark statistics	109
6.4 LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH	110
6.4.1 Future samples	110
6.4.2 Future measurement units	111
6.5 CONCLUSION.....	111
LIST OF REFERENCES	113
ANNEXURE A: BOOK LIST	123
ANNEXURE B: WEB PAGE DATA	126
ANNEXURE C: SOFTWARE PROGRAMMES.....	147
ANNEXURE D: EXTERNAL SOURCES	157

LIST OF FIGURES

FIGURE 2.1: MASLOW'S HIERARCHY OF NEEDS.....	15
FIGURE 2.2: THE ENTERPRISE FUNCTIONS.....	17
FIGURE 2.3: THE MARKETING PROCESS.....	23
FIGURE 2.4: TARGET MARKETING STRATEGY.....	25
FIGURE 3.1: THE MARKETING COMMUNICATION PROCESS.....	30
FIGURE 3.2: THE WOM COMMUNICATION MODEL.....	31
FIGURE 4.1: THE MARKETING RESEARCH PROCESS.....	45
FIGURE 4.2: EWOM REVIEW FACTORS AND REVIEW CREDIBILITY (MODEL 1).....	49
FIGURE 4.3: SPOTLIGHT EWOM REVIEWS AND SALES RANK LEVELS (MODEL 2).....	50
FIGURE 4.4: TYPES OF RESEARCH DESIGNS.....	51
FIGURE 4.5: THE TOP-HALF OF THE SPOTLIGHT AMAZON.COM BOOK REVIEW PAGE (FREEDOM - AN OPRAH NOVEL).....	127
FIGURE 4.6: THE BOTTOM-HALF OF THE SPOTLIGHT AMAZON.COM BOOK REVIEW PAGE (FREEDOM - AN OPRAH NOVEL).....	128
FIGURE 4.7: THE SPOTLIGH BARNESANDNOBLE.COM BOOK REVIEW PAGE (FREEDOM - AN OPRAH NOVEL).....	129
FIGURE 4.8: BACKLOG AMAZON.COM RATINGS AND REVIEWS (FREEDOM - AN OPRAH NOVEL).....	130
FIGURE 4.9: BACKLOG BARNESANDNOBLE.COM RATINGS AND REVIEWS (FREEDOM - AN OPRAH NOVEL).....	131
FIGURE 4.10: SCREENSHOT OF THE SCRAPED WEB PAGES ON THE PERSONAL COMPUTER.....	132

FIGURE 4.11: PROCESSING OF BACKLOG AMAZON.COM REVIEW DATA.....	133
FIGURE 4.12: PROCESSING OF BACKLOG BARNESANDNOBLE.COM REVIEW DATA.....	134
FIGURE 4.13: EXCEL ALGORITHM FOR IDENTIFYING FOUND HELPFUL (MODEL 1 – STEP 1).....	135
FIGURE 4.14: EXCEL ALGORITHM FOR IDENTIFYING STARS (MODEL 1 – STEP 1).....	135
FIGURE 4.15: EXCEL ALGORITHM FOR IDENTIFYING WHERE A REVIEW BEGINS (MODEL 1 – STEP 1).....	135
FIGURE 4.16: EXCEL ALGORITHM FOR IDENTIFYING WHERE A REVIEW ENDS (MODEL 1 – STEP 1).....	136
FIGURE 4.17: EXCEL ALGORITHM FOR IDENTIFYING A REVIEW’S ENTIRE TEXT STRING (MODEL 1 – STEP 1).....	136
FIGURE 4.18: EXCEL ALGORITHM FOR CATEGORISING THE REVIEW’S TEXT STRING (MODEL 1 – STEP 1).....	136
FIGURE 4.19: EXCEL’S SORT AND FILTER FUNCTION (MODEL 1 – STEP 1).....	137
FIGURE 4.20: EXCEL ALGORITHM FOR WORD COUNT (MODEL 1 – STEP 2).....	138
FIGURE 4.21: PROCESSING OF AMAZON.COM MAIN LANDING WEB PAGE DATA - TOP HALF OF THE WEB PAGE (MODEL 2).....	141
FIGURE 4.22: PROCESSING OF AMAZON.COM MAIN LANDING WEB PAGE DATA - LEFT HAND SIDE ON THE BOTTOM HALF OF THE WEPAGE (MODEL 2).....	142
FIGURE 4.23: PROCESSING OF AMAZON.COM MAIN LANDING WEB PAGE DATA - RIGHT HAND SIDE ON THE BOTTOM HALF OF THE WEB PAGE (MODEL 2).....	143

FIGURE 5.1: EWOM REVIEW DATA ACCORDING TO E-COMMERCE PLATFORM.....	74
FIGURE 5.2: THE DISTRIBUTION OF EWOM REVIEW DATA OVER TIME.....	75
FIGURE 5.3: BI-MODAL DISTRIBUTION OF REVIEW FOUND HELPFUL DEGREE AT AMAZON.COM.....	75
FIGURE 5.4: BI-MODAL DISTRIBUTION OF REVIEW FOUND HELPFUL DEGREE AT BARNESANDNOBLE.COM.....	76
FIGURE 5.5: BI-MODAL DISTRIBUTION OF REVIEW STAR RATINGS AT AMAZON.COM.....	77
FIGURE 5.6: BI-MODAL DISTRIBUTION OF REVIEW STAR RATINGS AT BARNESANDNOBLE.COM.....	78
FIGURE 5.7: BI-MODAL DISTRIBUTION OF REVIEW TEXT SENTIMENT AT AMAZON.COM.....	79
FIGURE 5.8: BI-MODAL DISTRIBUTION OF REVIEW TEXT SENTIMENT AT BARNESANDNOBLE.COM.....	79
FIGURE 5.9: BI-MODAL DISTRIBUTION OF REVIEW TEXT LENGTH AT AMAZON.COM.....	80
FIGURE 5.10: BI-MODAL DISTRIBUTION OF REVIEW TEXT LENGTH AT BARNESANDNOBLE.COM.....	81
FIGURE 5.11: BOXPLOT OF AMAZON.COM REVIEW TEXT LENGTH (MODEL 1).....	145
FIGURE 5.12: BOXPLOT OF BARNESANDNOBLE.COM REVIEW TEXT LENGTH (MODEL 1).....	145
FIGURE 5.13: AVERAGE DAILY LOG OF SALES RANK VALUES OVER TIME.....	82
FIGURE 5.14: FITTED AND RESIDUAL VALUES (MODEL 1).....	87
FIGURE 5.15: THE NORMAL AND RESIDUAL DISTRIBUTION (MODEL 1).....	88

FIGURE 6.1: THE RELATIONSHIP BETWEEN EWOM REVIEW FACTORS AND REVIEW CREDIBILITY.....	104
FIGURE 6.2: THE RELATIONSHIP BETWEEN SPOTLIGHT EWOM AND SALES RANK.....	106

LIST OF TABLES

TABLE 4.1: PRIMARY RESEARCH APPROACHES.....	53
TABLE 4.2: TARGET POPULATION.....	124
TABLE 4.3: AMAZON.COM WEB SCRAPING PROGRAMMES.....	148
TABLE 4.4: BARNESANDNOBLE.COM WEB SCRAPING PROGRAMMES.....	148
TABLE 4.5: THE RELATIONSHIP BETWEEN REVIEW CREDIBILITY AND EWOM REVIEW FACTORS (MODEL 1).....	59
TABLE 4.6: AMAZON.COM FOUND HELPFUL, STAR RATING, PUBLISH DATE AND REVIEW TEXT (MODEL 1).....	60
TABLE 4.7: BARNESANDNOBLE.COM DATA (MODEL 1 – STEP 1).....	137
TABLE 4.8: AMAZON.COM PUBLISH DATE, FOUND HELPFUL, STAR RATING, REVIEW TEXT AND TEXT LENGTH (MODEL 1).....	61
TABLE 4.9: BARNESANDNOBLE.COM DATA (MODEL 1 – STEP 2).....	138
TABLE 4.10: THE PYTHON PROGRAMME THAT CONNECTS TO UCLASSIFY’S API FOR SENTIMENT CLASSIFICATION (MODEL 1).....	149
TABLE 4.11: AMAZON.COM PUBLISH DATE, FOUND HELPFUL, STAR RATING, REVIEW TEXT, TEXT LENGTH, POSITIVE AND NEGATIVE SENTIMENT (MODEL 1).....	61
TABLE 4.12: BARNESANDNOBLE.COM DATA (MODEL 1 – STEP 3).....	139
TABLE 4.13: AMAZON.COM AND BARNESANDNOBLE.COM BOOK TITLE, PUBLISH DATE, COLLECTION DATE, FOUND HELPFUL, STAR RATING, TEXT LENGTH AND SENTIMENT (MODEL 1).....	62
TABLE 4.14: AMAZON.COM AND BARNESANDNOBLE.COM INDEX (MODEL 1 – STEP 5).....	139
TABLE 4.15: ELAPSED TIME CALCULATION (MODEL 1 – STEP 5).....	140

TABLE 4.16: EX-SENT AND EX-STARS CALCULATION (MODEL 1 – STEP 5).....	140
TABLE 4.17: THE RELATIONSHIP BETWEEN AMAZON.COM SALES RANK AND EWOM STAR RATING FACTORS (MODEL 2).....	63
TABLE 4.18: ESTIMATED SALES RANK UPDATE FREQUENCY.....	64
TABLE 4.19: THE PYTHON PROGRAMME THAT EXTRACTS SALES RANK DATA FROM AMAZON.COM WEB PAGES (MODEL 2).....	150
TABLE 4.20: AMAZON.COM SALES RANK DATA (MODEL 2).....	64
TABLE 4.21: MATLAB PROGRAMMES USED TO CALCULATE SALES RANK VALUES (MODEL 2).....	151
TABLE 4.22: BOOK, DATE, PREVIOUS AND FOLLOWING WEEK SALES RANK DATA (MODEL 2).....	65
TABLE 4.23: THE PYTHON PROGRAMME THAT EXTRACTS SPOTLIGHT FOUND HELPFUL STAR RATINGS FROM AMAZON.COM WEB PAGES (MODEL 2).....	153
TABLE 4.24: AMAZON.COM SPOTLIGHT MOST HELPFUL STAR RATINGS (MODEL 2).....	66
TABLE 4.25: AMAZON.COM DAILY AVERAGE NUMBER OF STARS FOR THE THREE MOST HELPFUL REVIEWS (MODEL 2).....	66
TABLE 4.26: MATLAB PROGRAMMES USED TO CALCULATE AVERAGE STAR RATINGS AND MOST RECENT REVIEWS' STAR RATINGS (MODEL 2).....	154
TABLE 4.27: AMAZON.COM AVERAGE STAR RATING, NUMBER OF REVIEWS, MOST HELPFUL REVIEW'S STARS AND MOST RECENT REVIEW'S STARS (MODEL 2).....	144
TABLE 4.28: AMAZON.COM AVERAGE STAR RATING DATA (MODEL 2).....	68
TABLE 4.29: AMAZON.COM STAR RATING DATA (ADAPTED FROM MODEL 1).....	68

TABLE 4.30: AMAZON.COM AVERAGE STAR RATING CALCULATION (MODEL 2).....	68
TABLE 4.31: AMAZON.COM MOST RECENT REVIEWS' STAR RATINGS (MODEL 2).....	69
TABLE 5.1: AVERAGE REVIEW TEXT LENGTH PER STAR RATING (MODEL 1).....	146
TABLE 5.2: THE MULTIPLE STEPWISE REGRESSION RESULTS (MODEL 1).....	87
TABLE 5.3: CHI-SQUARED TEST OF NORMALITY (MODEL 1).....	89
TABLE 5.4: ORDINARY REGRESSION RESULTS (MODEL 1).....	90
TABLE 5.5: CORRELATION MATRIX OF THE INDEPENDENT VARIABLES (MODEL 1).....	91
TABLE 5.6: AMAZON.COM SALES RANK AND EWOM DATA (MODEL 2).....	93
TABLE 5.7: CHANGES IN EWOM COMMUNICATION (SAME, WORSE AND BETTER).....	94
TABLE 5.8: OBSERVED SALES RANK AND EWOM CHANGES.....	95
TABLE 5.9: PROBABILITIES OF SALES RANK AND EWOM CHANGES.....	95
TABLE 5.10: MARGINAL PROBABILITIES OF SALES RANK AND EWOM CHANGES.....	95
TABLE 5.11: JOINT PROBABILITIES OF SALES RANK AND EWOM CHANGES.....	96
TABLE 5.12: EXPECTED SALES RANK AND EWOM CHANGES.....	96
TABLE 5.13: CHI-SQUARE TEST RESULTS.....	97
TABLE 5.14: WEIGHTS, SALES RANK CHANGE FACTORS AND P-VALUES.....	98
TABLE 6.1: RETAIL E-COMMERCE (NON-TRAVEL) GROWTH RATES FOR 2011, 1 ST QUARTER.....	158

CHAPTER 1 INTRODUCTION

1.1 INTRODUCTION

Since the first website was launched in 1991, the Internet has continually transformed the way people and enterprises conduct marketing and business (Chaffey, Ellis-Chadwick, Mayer and Johnston, 2009:3). The web, according to Tapscott and Williams (2008:45), is “no longer about idly surfing and passively reading, listening, or watching. It’s about peering, sharing, socializing, collaborating, and, most of all, creating within loosely connected communities.” As more and more user content is generated, alone and in groups, online users exert more power on the Internet. Tapscott and Williams (2008:52-53) describe consumers who exert this power and influence as “prosumers”. Ultimately, enterprises are challenged to harness the power of word-of-mouth (WOM) for more successful marketing strategies to increase their online sales levels.

This study has analysed the characteristics and the use of electronic word-of-mouth (EWOM) communication as a marketing channel concerning a social object in an online community. A “social object” is defined by Kotler (1972:49) as any product, person, idea, place, service or organisation from whom or which a certain response is sought. A typical response may be adoption, purchase, usage or consumption. By using tracking software, individual EWOM reviews written about a book (the social object in this case), were tracked and analysed. The insights gained from analysing EWOM communication patterns can be used by e-commerce enterprises and book publishers to understand their users better, as well as to design more effective marketing strategies and website usability, to ultimately increase future online sale levels.

This chapter is organised into several sections. In the first section the background of the study is introduced, by examining enterprise management as well as WOM and EWOM communication. The limitations of existing solutions are then discussed, followed by the objectives of the study and the proposed research methodology. The chapter concludes with a brief overview of the study’s orientation.

1.2 BUSINESS AND MARKETING MANAGEMENT

Business management's field of study is enterprises that are established in the economy to satisfy the needs of consumers, and communities, and also the needs of other enterprises, while generating a profit for its owners (Badenhorst, Cant, De J Cronje, Du Toit, Erasmus, Grobler, Krüger, Machado, De K Marais, Marx, Strydom and Mpofu, 2003:12-14). Over the years, several business management functions have developed optimally to realise this goal, namely general and strategic, productions and operations, finances, purchasing, human resources, information, public relations, and marketing functions (Bruijs, Crous, De Villers, De Wit, Du Plessis, De V Maasdorp, Mahabir, Marx, Nieman, Schreuder and Van Vuuren, 1998:29-32).

According to Strydom, Cant, Jooste, Koekemoer, Bennet, Schreuder, Brink and Machado (2000:23), marketing can be regarded as a primary function in an enterprise because of its influence on the enterprise's profitability and growth. Marketing can be defined as "a combination of management tasks and decisions aimed at meeting opportunities and threats in a dynamic environment in such a way that its market offerings lead to the satisfaction of consumers' needs and wants so that the objectives of the enterprise, the consumer and society are achieved" (Cant, Van Heerden, Ngambi, Nieuwennhuizen, Roberts-Lombard, Brink, Dayan, Wiid, Koekemoer, Botha, Human, Bothma and Machado, 2010:16). Marketing can be divided into two main focus-areas, the management philosophy that emphasises customer satisfaction, and the marketing function that comprises several tasks, which together, as the marketing process, are used to execute this latter management philosophy (Strydom *et al.*, 2000:4-5). The marketing process usually includes a thorough understanding of the enterprise's mission and the role of marketing in executing that mission, formulating marketing goals after market analysis, targeting a specific market segment, developing and implementing a marketing mix through an integrated marketing strategy, setting up performance matrices, conducting measurement and evaluations of marketing efforts, and finally, making any alterations if needed (Cant *et al.*, 2010:2-3). The four marketing instruments, otherwise known as the "marketing mix", consist of four "Ps" (product, place, price and promotion or marketing communication) which enable the enterprise to deliver its unique market offering to ultimately satisfy target consumers' needs (Strydom *et al.*, 2000:21). The present study focuses on the promotion concept by examining the informal communication of personal selling by means of EWOM.

The next three sections will explore the fields of WOM communication, EWOM communication, and the role that EWOM plays in the applied research field of consumer behaviour.

1.3 WORD-OF-MOUTH COMMUNICATION

WOM as a communication tool has been shown to influence several outcomes, such as awareness, perceptions and behaviour patterns (Buttle, 1998:242). One of the first definitions of WOM described it as “oral person-to-person communication between a receiver and a communicator whom the receiver perceives as non-commercial, regarding a brand, a product or a service” (Arndt, 1967, cited in Breazeale, 2009:297-298). The term “WOM” first appeared in an article called “The Web of Word of Mouth”, by the pioneer Whyte (1954 cited in Sormunen, 2009:14). The first academic research that was conducted on WOM appeared in the literature in the following year, when Katz and Lazarsfeld (1955) referred to the role of personal influence and the implications it has on various industries. They concluded that “the media only have limited effects on the process of mass persuasion”. WOM can be divided into public or private conversations (Sormunen, 2009:14-15). According to Hoffman and Novak (1996:50), interpersonal interactions usually occur in face-to-face or group meetings and online forums. A more in-depth discussion on electronic media channels and communication follows next

1.4 ELECTRONIC WORD-OF-MOUTH COMMUNICATION

1.4.1 The Internet and the implication it has on business management

Peterson, Balasubramanian and Bronnenberg (1997:331) define the Internet as a type of “global information infrastructure consisting of computer hardware and software that is characterized as both general and open”. In the generic sense, the Internet embodies a proficient platform for allowing the categorisation and exchange of information. According to Farhoomand and Lovelock (2001), the Internet can be characterised as both a market and a medium; the computer-mediated market allows buyers and sellers to connect, and as a medium facilitates business functions such as sales and distribution. Marketing on the Internet or in short, Internet marketing is defined as the process of establishing relationships through online operations, while facilitating the exchange of products and services that

ultimately satisfies the objectives of both the buyers and the sellers within the economic transaction (Imber and Toffler, 2000).

In the sphere of communication, the Internet has various unique characteristics that distinguish it from traditional communication media. The Internet embodies several characteristics such as multi-media techniques, global network, and effortless 24-hour access to the newest information (Pavlik, 1998, cited in Cho and Khang, 2006:143). Since the first e-commerce transaction at AOL and the first marketing contract between *Hotwired* and AT&T in 1994, the Internet's commercialisation as a business tool has grown exponentially around the world, as it connects and facilitates both small and major transactions (Zeff and Aronson, 1999, cited in Cho and Khang, 2006:143).

1.4.2 Digital channels and information technology

Information technology and digital channels are constantly changing and reinventing the way enterprises produce, operate, and communicate with their customers (Merisavo, 2008:6). Because of these dynamic changes, it has become a high priority for many enterprises not only to build, but also to maintain, an ongoing relationship with their end-users.

Digital TV, Internet, mobile phones and other developing media channels offer opportunities for regular, proficient and interactive communication between enterprises and their customers (Merisavo, 2008:6). The Internet has not only increased consumer targeting, but has also allowed for more effective tracking of consumer behaviour, customised marketing messages, and two-way interaction between marketers and end-users (Rafi, Fisher, Jaworski and Cahill, 2001).

The present study considers different marketing communication concepts and their relevance to this specific marketing field. These concepts include interactive marketing, online advertising, digital marketing, one-to-many marketing, one-to-one marketing, and e-marketing (Merisavo, 2008:19).

1.4.3 Electronic word-of-mouth communication

Until relatively recently, consumers had comparatively few tools that could be used to communicate their experiences to a broad audience, and as a result they had little power to

influence other consumers with their brand opinions (Pajuniemi, 2009:71-78). According to Pitt, Berthon, Watson and Zinkhan (2002:7), “the net is shifting the balance and giving consumers hitherto undreamed-of opportunities to be more forceful with enterprises. Websites allow better-informed consumers to interact, band together, become more aware of corporate shortcomings, and gain easier access to the legal system.”

Since the emergence of different web communication tools (e.g. online social networks, blogging platforms and chat rooms), consumers have found a wider range of communication technologies that allow them to communicate their opinions among online groups by engaging in EWOM (Riegner, 2007:437). Electronic word-of-mouth or EWOM can be defined as “any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet” (Hennig-Thurau, Gwinner, Walsh and Gremler, 2004:39). The easy accessibility of these online social tools has led to basic changes in people’s behaviour patterns, to share their information and opinions with a now much larger audience.

1.5 CONSUMER WORD-OF-MOUTH COMMUNICATION

WOM conversations about brands are seldom neutral. Therefore the “buzz” about a brand can either increase the acceptance of a brand or harm its market share (Rosen, 2000:5-6). Positive WOM, in particular, yields several advantages for enterprises. Sormunen (2009:15) suggests that enterprises can build loyal relationships with their customers by simply understanding the principles of WOM. On the opposite side of the spectrum is negative WOM. East, Hammond and Wright (2007:175) indicate that marketers consider negative WOM to be more common than positive WOM. The marketing literature classifies WOM as one of the key drivers that influence consumer decision-making (Chevalier and Mayzlin, 2004:2). If marketing enterprises accept that WOM takes place on e-commerce platforms, the next difficult challenge is to assess which factors of WOM play the biggest role in the purchase process.

1.6 STATEMENT OF THE PROBLEM

Kumar, Aaker and Day (2002:48-50) suggest that the first step in the marketing research process is to recognise and define the stated research problem. According to several previous studies (Chen *et al.*, 2004; Chevalier and Mayzlin, 2004; Chen, Dhanasobhon and

Smith, 2008; East, Hammond and Lomax, 2008), the main problem identified in the EWOM research field was the inability to measure the impact that EWOM communication had on an online review process. Two key focus areas that were identified as problematic, were measuring the impact of several EWOM factors on review credibility, and secondly, assessing the relationship of EWOM reviews with the sales levels of online retailers.

1.7 THE DEPENDENT AND INDEPENDENT VARIABLES

This study is divided into two sections, where two models are assessed: model 1 tested the relationship between review credibility and EWOM review factors on e-commerce platforms, while model 2 tested the relationship solely between *Amazon.com* sales rank and EWOM star rating factors.

In model 1 the dependent variable was REVIEW CREDIBILITY and the independent variables were STARS, EX-STARS, SENTIMENT, EX-SENTIMENT, TEXT LENGTH, TIME and PLATFORM.

- The REVIEW CREDIBILITY was measured by the total percentage of readers who found or voted the review helpful on the day of data collection. The degree to which an individual review is considered helpful was quantified as “found helpful degree”.
- The value for the STARS variable ranged from 1 to 5, representing the number of stars the reviewer gave the book with regards to his/her liking.
- EX-STARS was based on the contention that a review might make more of an impression (and therefore be more likely to be found helpful) if the rating is more extreme. In other words, a review of 1 star or 5 stars is more likely to be read (and found helpful) than one with the average of 3 stars. The distance (in stars) from a star rating of 3 is therefore calculated as: $EX-STARS = abs(3 - STARS)$, where *abs* is the absolute value function.
- The SENTIMENT variable was used to measure the positive sentiment of review text and assign a quantitative score between 0.51 – 1.0.
- Following the same reasoning as the above-mentioned EX-STARS, it is of interest whether a strong opinion is likely to be more helpful, and therefore EX-SENTIMENT variable is defined as: $EX-SENTIMENT = abs(0.5 - SENTIMENT)$.
- A review’s length (in number of words) was recorded in TEXT LENGTH.
- The TIME variable was calculated as the difference between publish date (1 October 2010 until 31 January 2011) and collection date (1 February 2011 until 28 March

2011); yielding the number of days the review was available to be judged helpful or not helpful by the users of the two e-commerce platforms.

- Lastly, the Boolean variable PLATFORM, assigned a value of 0 for reviews on *Barnesandnoble.com* and 1 for reviews on *Amazon.com*.

In model 2 the dependent variable was SALES RANK and the independent variables were AVERAGE BOOK STAR RATING, MOST HELPFUL REVIEW'S STARS and MOST RECENT REVIEW'S STARS. The SALES RANK was measured by the log of sales rank, while each of the independent variables was quantified by a rating between 1 and 5 stars. A more in-depth explanation of each of the models is presented in Chapter four.

1.8 CONTRIBUTION OF THE PRESENT STUDY

Identifying relevant, credible information often increases search costs, and prolongs the purchase cycle for customers, sometimes preventing them from completing actual sales (Chen *et al.*, 2004:711). In its most elementary form, "credibility" can mean the propensity to trust someone (Shimp, 1997:289). To enhance credibility and decrease search cost, some enterprises have started investing in product-review platforms to build credible review communities about their products. However, not much is known regarding the effectiveness of these efforts.

In the present study, the impact of consumer feedback (EWOM reviews) is empirically investigated, and the effect it has on online review credibility (as measured by the degree to which an individual review is considered to be helpful) and a book's sales levels (as measured by the sales rank value). A more in-depth discussion of both these concepts is offered in section 4.2.6.

Ultimately, the knowledge emanating from this study can assist e-commerce enterprises to gain a deeper understanding of their users for more effective target marketing and better future business strategies.

1.9 RESEARCH OBJECTIVES

The primary objective of this study is to examine perceived credibility of electronic word-of-mouth communication on USA-based e-commerce platforms.

The secondary objectives are:

- analysing book reviews in terms of the following criteria: star ratings, text length and text sentiment, and
- assessing the influence of time in the review process.

Hypotheses were necessary in order to address these objectives, and are formulated next.

1.10 HYPOTHESES

Chen *et al.* (2008:10) has found that there are usually two major challenges in a review system, namely, a moral hazard and an adverse selection problem. A moral hazard problem occurs when reviewers create unreliable reviews, while an adverse selection problem occurs when the product review is questioned by consumers after they have purchased the product being discussed. Positive WOM, quantified as higher product ratings, indicates that the general community likes a book (Chen *et al.*, 2008:10). When the focus is on group liking or “social validation”, Cialdini (2000) suggests that consumers would rather acquire a book with a higher social validation than a book that is liked by fewer people in the group. The present study’s first four hypotheses have been formulated to assess the influence of a review’s star rating, a review’s ex-star rating (the distance in stars from a 3-point star rating on a rating scale between 1 to 5 stars), a review’s sentiment rating, and its ex-sentiment rating (the distance from an average sentiment rating of 0.5 on a rating scale between 0.0 and 1.0 sentiment) on review credibility.

H0¹: A review’s credibility on an e-commerce platform is not influenced by the review’s star rating

H0²: A review’s credibility on an e-commerce platform is not influenced by the review’s ex-star rating

H0³: A review’s credibility on an e-commerce platform is not influenced by the review’s positive sentiment rating

H0⁴: A review’s credibility on an e-commerce platform is not influenced by the review’s ex-sentiment rating

The number of reviews and the length of the reviews (measured by word count) can be seen as a good indication of that book's market popularity, because the interaction with that book reflects how people felt about the book before that measurement was taken (Chevalier and Mayzlin, 2004:6). The fifth and sixth hypotheses are as follows:

H0⁵: A review's credibility on an e-commerce platform is not influenced by the review's text length

H0⁶: A review's credibility on an e-commerce platform is not influenced by the time the review appeared

To address the seventh hypothesis, the study assesses the influence of the e-commerce's platform on its review credibility. The degree to which a review is considered or "found helpful" serves as an indicator of the review's credibility, as well as the overall quality of the reviews posted at a given e-commerce platform (Chen *et al.*, 2008:2; Mudambi and Schuff, 2010). The e-commerce platforms which were investigated in this study were USA-based *Amazon.com* and *Barnesandnoble.com*.

H0⁷: A review's credibility is not influenced by the e-commerce platform upon which it takes place

According to Chen *et al.* (2008:10), positive WOM, quantified as higher product ratings, indicates that the given book is of high quality. Referring to PWOM, research shows that if all external factors are considered equal, and product quality uncertainty is reduced, there is a higher probability that consumers will rather purchase books with higher ratings than those with lower ratings (Chen *et al.*, 2004:713). The study's final hypothesis was formulated to assess the impact of main landing web page reviews or spotlight EWOM reviews, on book sales rank levels, seen as a proxy for online sales levels.

H0⁸: Sales rank levels of *Amazon.com* books are not influenced by spotlight EWOM reviews

1.11 METHODOLOGY

The purpose of this section is to introduce the research study's methodology in the light of the marketing research process. According to Blumberg, Cooper and Schindler (2008:4-56), research methods will make available the knowledge and skill set that is required to address the study's research questions, and is usually described as a process, which involves several steps.

1.11.1 Secondary research

“Secondary research” refers to information or data that has been previously collected for some purpose other than to address the stated research problem (Kotler and Keller, 2006:104). According to Zikmund and Babin (2010:163), secondary data offers various advantages; the most noteworthy of them is the money and time that can be saved in the research process by using readily available assembled data. The main disadvantage of secondary data is that it may not fit the research problem, and that further data filtering needs to be conducted in order to address the stated research objectives (Zikmund and Babin, 2010:164). Secondary data sources can be divided into data that originates in the enterprise (internal data) or data that are generated outside the private enterprise (external data) (Zikmund and Babin, 2010:163-165). In the present study, an extensive literature study was conducted to collect secondary data.

1.11.2 Primary research

Primary research can also be classified into two basic types, namely qualitative and quantitative research. According to Blumberg *et al.* (2008:191-192), qualitative studies base their accounts on words, sentiment and storylines, while quantitative studies rely on numbers, figures and statistics to answer the research objectives. For the purpose of the present study, both qualitative and quantitative primary data were collected in the form of EWOM reviews on e-commerce platforms by means of the observational research approach.

1.11.3 Sampling procedure

The next step in planning the research design was to determine which and how many events to observe, by identifying the targeted population, as well as selecting an appropriate sample to represent the target population (Blumberg *et al.*, 2008:69).

A “population” is any complete collection of units or individuals that contains a common trait (Zikmund and Babin, 2010:412). The target population that was investigated for this study was narrowed down to the EWOM review behaviour about the new best-seller book list on e-commerce platforms *Amazon.com* and *Barnesandnoble.com*. The book list consists of 100 new best-seller books at any given point in time.

According to Blumberg *et al.* (2008:235), sample participants can be selected either on a probability or a non-probability base. A probability sample is drawn by random selection, where each population participant has a known non-zero probability of selection, while a non-probability sample is drawn on a non-random basis and is usually subjective (Blumberg *et al.*, 2008:235).

The present study used a purposive non-probability sampling procedure, more specifically, judgement sampling. Judgement sampling occurs when a sample is selected by a researcher to conform to some criterion (Blumberg *et al.*, 2008:253). The three criteria according to which the experiment’s samples were chosen, were: e-commerce platform, books, and time period. A more in-depth discussion of this process can be found in section 4.2.5.2.

1.11.4 Data collection and analysis

100 new best-seller books from *Amazon.com* were chosen as the benchmark population, and the same 100 titles were also selected from *Barnesandnoble.com* according to their unique book IDs. Collected EWOM review data and book sales rank data were categorised into two types, data that were collected from backlog product pages, and data that were collected from product landing pages as well as backlog review pages. Review and sales rank data on landing web pages were automatically collected daily, while all the backlog customer review data were manually imported into Excel sheets only at the end of the data collection period. After data collection was completed, the data were processed, coded and analysed by means of three statistical software programmes, namely MATLAB, R and

STATISTICA. Both descriptive statistics and inferential statistics were applied to the data. Descriptive statistics were used to present and describe the different characteristics of the data, while inferential statistics were used to estimate significance in the variable relationships.

1.12 ORIENTATION OF THE STUDY

Chapter one has been compiled to present the specific theme and purpose of this study, which is to track, measure and assess the impact of different EWOM review factors on review credibility, and also to examine the associative impact that spotlight EWOM reviews exert on online book sales. Chapter one provides a background sketch to the study, formulates the research problem and objectives, and introduces the research design that has been adopted. In order to get a better indication of the rest of the layout of the study, an orientation is compiled to briefly describe the remaining chapters.

Chapters two and three provide an in-depth discussion about the theoretical background with regard to business and marketing management, WOM, EWOM and consumer WOM. In Chapter four the marketing research process will be discussed in detail, as well as the research methods used in the data collection, processing and analysis. The findings of the empirical research will be presented in Chapter five. These findings refer to the impact of EWOM reviews on online review credibility, as well as the associative role that EWOM reviews have played on book sales. Based on the research results in Chapter five, the findings are interpreted, conclusions drawn, and recommendations made, in Chapter six.

CHAPTER 2

BUSINESS MANAGEMENT

2.1 INTRODUCTION

The primary purpose of this study was to examine perceived credibility of electronic word-of-mouth communication on USA-based e-commerce platforms, *Amazon.com* and *Barnesandnoble.com*. The primary goal of this chapter is to introduce the discipline of business management, and to contextualise the study in terms of both business management and the sub-discipline of marketing.

This chapter commences with a review of the role of business in society, as well as society's underlying needs that drive economic activity and progress. An outline of the nature of business management and the various business management functions is provided. The marketing function will then be the departure point, as well as the link between this chapter and the next chapter on WOM and EWOM communication.

2.2 THE ROLE OF BUSINESS IN SOCIETY

Business, according to Badenhorst *et al.* (2003:3-4), is "the means by which society endeavors to satisfy its needs and improve its standard of living". Business in a market economy is a multifaceted system of individuals and enterprises, which transforms raw materials into products and services to satisfy the needs of people. Du Toit, Du Plessis and Nortje (1981:23-24) define business as "any individual productive unit in the national economy involved in 'producing' and 'supplying' goods and services for the purpose of consumption".

An enterprise can usually be characterised by several elements: production, human activities, economic exchanges and finally, a profit (Badenhorst *et al.*, 2003:3-4). Firstly, an "enterprise activity" can be described as a production process. The enterprise transforms resources into products (e.g. flour is converted into bread) and services (e.g. labour, beds and medicine are converted into a health service). Secondly, even with the advent of the information age, advanced computer systems and robotic technology, enterprises are still generally being managed by people. The important thing to realise is that without proper

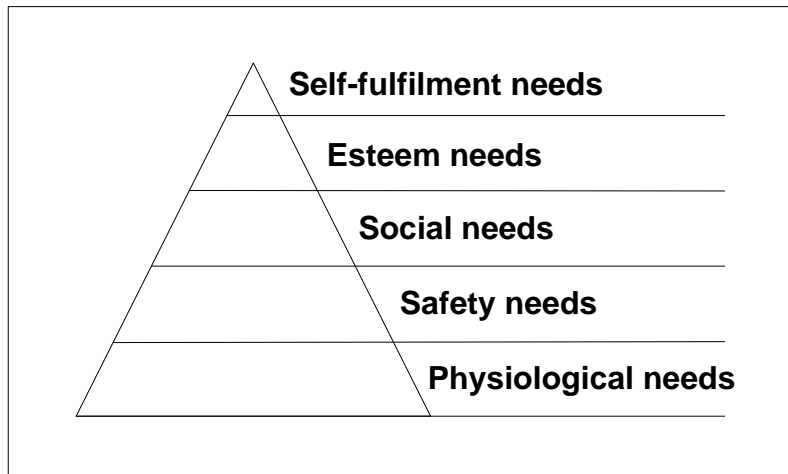
managerial decision-making, the effectiveness of the production process will remain at a low level, without the ability to serve its customers' needs optimally (Bruijs *et al.*, 1998:3-4). Thirdly, a transaction between two or more parties with regards to products or services, indicate an economic exchange, to ultimately make a profit. Finally, a "profitable enterprise" can be defined as an enterprise that obtains monetary rewards for satisfying the needs of its customers through its product or service offerings. It can also be that the enterprise exchanges its current products/services for other products/services. Profit in return allows the enterprise to pay for the daily operating expenses, and ultimately ensures its viability over the long term.

2.3 NEEDS AND NEED SATISFACTION

2.3.1 Hierarchy of human needs

In order to understand the need for the existence of enterprises in society, an examination of human needs will be used as a theoretical basis for this section. According to Bruijs *et al.* (1998:5-7), humankind has a boundless range of needs. In similar vein Badenhorst *et al.* (2003:9-10) point out that the continuous survival of humankind is driven by the perpetual satisfaction of its infinite needs. These needs can vary from simple to very complex. Needs may be few in the case of primitive and underdeveloped communities, or more complex in highly modern societies (Badenhorst *et al.*, 2003:9-10). The range of needs in a specific society will ultimately also reflect the size and the complexity of the enterprises that operate in that same society.

An American clinical psychologist, A.H. Maslow, first proposed that basic human needs can be structured in a sequence of importance (Badenhorst *et al.*, 2003:9-10). Before Maslow's hierarchy of human needs theory was developed, various other theories held that people's main drive to work was solely to fulfil their material needs (Glos, Steade and Lowry, 1980:101). Counter to other research done in psychology at that point in time, Maslow's theory identified five types of physiological and psychological needs. These various needs are explained as levels in a pyramid, organised according to the order in which people usually try to fulfil them (Bruijs *et al.*, 1998:5-7).

FIGURE 2.1: MASLOW'S HIERARCHY OF NEEDS

Source: Adapted from Glos *et al.* (1980:102)

Figure 2.1 shows the hierarchy of a needs pyramid. The need sequence is structured in such a way that the order of importance moves up from the most vital rudimentary physiological needs (e.g. hunger, thirst, shelter) at the bottom, to psychological needs (e.g. affection, belonging, self-esteem, self-actualisation) placed at the top of the pyramid (Badenhorst *et al.*, 2003:9-10). Whether in a community or as an individual, the most pressing needs will always be satisfied first, and only then will the focus shift to a higher level. As previously mentioned, people have unlimited needs, and as soon as an individual fulfils one need, a new need arises to form part of the ever-increasing complexity of modern day society (Glos *et al.*, 1980:100-102).

2.3.2 Needs satisfaction cycle

According to Badenhorst *et al.* (2003:11-12), the survival of humankind is driven by the cycle of continuous needs satisfaction, and the most important motivation for economic progress, is the fulfilment of these needs with limited available resources. These limited input factors, better known as production factors, are usually classified into four main groups: natural resources (e.g. industrial sites, residential stands, agricultural land, water), human resources (e.g. physical labour and mental skills of people), capital (e.g. buildings, machinery, computers), and entrepreneurship (e.g. individuals who initiate products and service offerings) (Badenhorst *et al.*, 2003:11-12). Economic value is created by entrepreneurs when scarce factors of production are used in the most economical way possible to fulfil the consumer's needs. Bruijs *et al.* (1998:6-7) refer to this latter statement as "the economic

principle". This principle is the reason for the existence of the discipline of business management.

According to Badenhorst *et al.* (2003:12-14), it is also important to note that when enterprises in a free-market system no longer satisfy the needs of the society in which they function, consumers will alter the economic system or else select an alternative need-fulfilling system. The starting-up and closing-down of enterprises illustrates this needs satisfaction cycle that underlies a free-market economy (Badenhorst *et al.*, 2003:12-14).

2.4 THE NATURE OF BUSINESS MANAGEMENT

The historical development of the discipline of business management can be best understood in terms of the development of the research field of management. In most places around the world, as well as in South Africa, industrial development was initiated at the start of World War I. During the 1920's the South African government's protectionist policy augmented this development, but the real progression only took place after the completion of World War II (De Klerk, Kroon, Jansen van Rensburg, Lotz, Steyn and Verreyne, 1997:57). The reason for the rapid development of management thinking was that people realised that a more skilled approach was needed to address prevailing enterprise-associated problems.

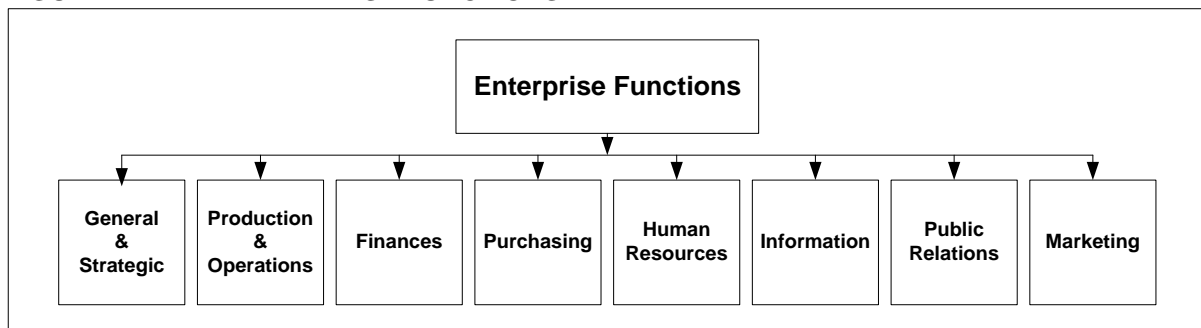
Business management, previously named "business economics", did not acquire its key features until the end of the 19th century. According to De Klerk *et al.* (1997:57), the evolution of business economics cannot be separated from the evolution of economics, but it inherited the same dynamics that shaped the science of economics.

The purpose of business management, as embodied in the economic principle, is to determine how an enterprise can realise the highest possible output units and maximise profits with the least possible input costs (Badenhorst *et al.*, 2003:25). Business management also includes the analysis of external or environmental factors, which can adversely affect an enterprise's success and long-term profitability (Badenhorst *et al.*, 2003:25). In order for any enterprise to fulfil its target users' needs and ultimately survive over the long term, management must ensure that all the various business management functions operate in an efficient and coordinated manner.

2.5 ENTERPRISE MANAGEMENT FUNCTIONS

For an enterprise to realise its goals, it needs to perform various complex functions and tasks in a synchronised manner (Bruijs *et al.*, 1998:26-27). The enterprise's function can be seen as a set of parallel activities or tasks that are performed to reach an end objective (Bruijs *et al.*, 1998:26-27). While functions are seen as broad and primary to an enterprise, enterprise tasks are more clearly formulated and derived from the main enterprise function itself. De Klerk *et al.* (1997:11-12) identify eight functions in an enterprise: general and strategic management, production, financial, purchasing, human resource, information, public relations, and marketing (see Figure 2.2).

FIGURE 2.2: THE ENTERPRISE FUNCTIONS



Source: Adapted from De Klerk *et al.* (1997:11-12)

Rather than viewing each enterprise function as independent, Bruijs *et al.* (1998:32) suggest that all the functions must complement each other, as well as actively serve to complement the whole enterprise. A similar viewpoint is held by Du Toit *et al.* (1981:4-5), who agree that enterprises that implement a holistic management approach over all functional departments, will generate greater long-term profits than functional departments of enterprises choosing to operate in an uncoordinated manner. General management principles can help to realise this purpose.

2.5.1 General and strategic management function

Eric Smit (*Google*), Mark Zuckerberg (*Facebook*), Soichiro Honda (*Honda*) and Anton Rupert (*Rembrandt*) are some of the successful enterprise leaders of the 19th and 20th centuries (Bruijs *et al.*, 1998:347). Many academic journals and articles have been published on leadership and management in the past 30 years, and many of them refer to the complexity of this topic (Bruijs *et al.*, 1998:347-348). One reason for the complexities, according to

Badenhorst *et al.* (2003:173), is the fact that an enterprise does not consist only of machines, but of both machines and people. Therefore a multifaceted approach is needed when leading the enterprise's taskforce to address its vision and mission.

Entrepreneurs typically spot market opportunities and start enterprises to produce the market's required products and services. Badenhorst *et al.* (2003:119) suggest that even though entrepreneurs are the driving force behind the operating entity, this is not the only thing needed to ensure a successful venture. They conclude that new projects or ventures need to be properly managed. The task of enterprise management is to strategise, organise, guide, coordinate and control (Badenhorst *et al.*, 2003:119). Du Toit *et al.* (1981:215-216) add that for the management function to be efficient, the enterprise requires people with leadership skills and a comprehensive understanding of how the whole enterprise operates.

2.5.2 Production function

Globally, countries are often divided into either already developed economies or those in the process of developing their economy according to certain criteria (Badenhorst *et al.*, 2003:479-480). One such criterion that is used to measure a country's growth is called the gross domestic product (GDP). The GDP represents the total value of all the final goods and services produced in a country in a given year (Nickels, McHugh and McHugh, 2005:53). The research field of production and operation management has two opposing descriptions. The production function is defined as the management of those technical processes to manufacture physical goods (e.g. motorcars, petroleum, electrical appliances) that are demanded by the market, while "operational management" encompasses a much broader view than the production process (Bruijs *et al.*, 1998:30). Operational management refers to the delivering of services by the enterprise to satisfy the needs of end-users (Machado, Strydom and Cant, 1999:144). Both concepts highlight the value that the enterprise adds to the market through its unique product/service, and is ultimately driven by the underlying demand of users' needs in the economy.

2.5.3 Financial function

The purpose of the financial function is to account for the financial activities and needs of an enterprise. It is important to consider the role of capital (short-term, medium-term and long-term) in the daily operations of the enterprise (Bruijs *et al.*, 1998:582).

In order for any enterprise to operate efficiently, raw materials and various assets (e.g. land, machinery and equipment) are needed. Through the use of a disposable budget, enterprises can purchase short-term operating assets and long-term fixed assets (Bruijs *et al.*, 1998:582-583). The people who take the risk of making these capital funds available for the business's operations (usually the owners or stakeholders), retain either full equity or shares of equity in the enterprise. The bank or loan institutions that supply funds, on the other hand, expect a return on the money that has been borrowed from them (Badenhorst *et al.*, 2003:393-395).

The financial management function is characterised by three main tasks (Du Toit *et al.*, 1981:173). Firstly, the financial manager needs to compile a budget for the enterprise's capital demands. Secondly, the financial manager needs to obtain capital by means of loans or investments at the best possible price. Lastly, these funds that have been obtained must be used in the best possible way to realise the enterprise's financial objectives, which are mainly to safeguard the liquidity of the enterprise.

2.5.4 Purchasing function

Bruijs *et al.* (1998:398) describe purchasing management and inventory building as "the gathering of adequate information and the making of decisions to ensure that the enterprise purchases the right quality raw materials or other requirements from the right suppliers, in the right quantities, at the right prices and at the right time".

For a more narrow view of what is required by the purchase management of an effective production process, Badenhorst *et al.* (2003:522-523) outline the following tasks:

- The quantity and quality of materials needed should be clearly defined;
- Select, purchase and transport materials from suppliers to the business;
- Apply warehousing control and determine the extent of inventory stock;
- Stock extra components for future service repair and maintenance needs.

The purchase function is, however, not restricted to for-profit enterprises, but effective purchase management is also a necessity for government institutions, churches, NGOs and universities (Bruijs *et al.*, 1998:398).

2.5.5 Human resources function

Human resources are a crucial production factor and human capital asset for any enterprise (Bruijs *et al.*, 1998:472). The practice of recruiting, developing and retaining people to form a talented workforce is described as one of the most difficult management functions (Badenhorst *et al.*, 2003:192-193). Bruijs *et al.* (1998:472) suggest this is probably because it is the production factor that is the most difficult to predict because of each person's individuality. Human resources include the following four aspects: formulating policy, consulting, service-rendering, and managing personnel affairs (Bruijs *et al.*, 1998:473). Ultimately, the human resource manager's task is to effectively use the human resources that are available to realise the enterprise's set objectives.

2.5.6 Information function

Even before the advent of the computer age, managers had to filter through large amounts of information to find reliable correct figures on how best to make a calculated enterprise decision. With the growing use of computer technology, mountains of information were generated. Over time, easier ways emerged to facilitate the processing of raw data into usable information. Cant *et al.* (2010:116) define the information function (with specific regard to marketing managers) as "an organised way of continually gathering and analysing data to provide marketing managers with information they need in order to make decisions". This system can also be described as an information-reporting dashboard that allows managers from different functional departments to access and use real-time statistics for individual department needs, as well as to realising enterprise-wide objectives (Bruijs *et al.*, 1998:31).

2.5.7 Public relations function

Every enterprise needs to adapt to its local environment in order to efficiently serve the community's needs (Badenhorst *et al.*, 2003:352-353). Bruijs *et al.* (1998:553-554) believe that this need for liaising between the above-mentioned community and the enterprise has arisen from the increased complexity and connectedness of the modern-day society. Public relations are defined by Strydom *et al.* (2000:351) as "the management, through communications, of perceptions and strategic relationships between an organization and its internal and external shareholders".

With the advent of the information age, and especially the 21st century media communication tools (e.g. *Facebook*, *Twitter*, *YouTube*), it has become so much easier to instantly broadcast one's opinion in the public domain (Bruijs *et al.*, 1998:553-554). These voices of stakeholders and consumers must be closely monitored in order to assess whether the enterprise is succeeding in serving current product/service demands. To examine these consumers' needs, the marketing management function, with a focus on personal selling, WOM, consumer opinion and sentiment analysis, follows next.

2.5.8 Marketing function

Marketing is the function which collects and analyses much of the needed enterprise information (Badenhorst *et al.*, 2003:279-280). This information is necessary to gain a better understanding of consumers' behaviour and to target future customers more effectively (Badenhorst *et al.*, 2003:279-280). A more in-depth explanation of the marketing function follows next to form a strong theoretical foundation on which the present study's research objectives are build.

2.5.8.1 Nature of the marketing function

It can be argued that, owing to its interaction with customers, the marketing department is the key department in the enterprise. The argument is based on the contention that management can only procure inventory, set up production capacity, and employ needed workforce after the market has been thoroughly evaluated (Badenhorst *et al.*, 2003:280). The marketing function serves as a bridge that connects the enterprise and its surrounding market, helping to manage its resources effectively in line with the market's demand (Badenhorst *et al.*, 2003:279-280).

The American Marketing Association defines marketing as "the activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large" (Rownd and Heath, 2008). Du Toit *et al.* (1981:142) refer to marketing as "those activities carried out by an individual firm, with the aim of establishing, on the selling market, exchange relationships which are of mutual benefit to both the customer and the firm".

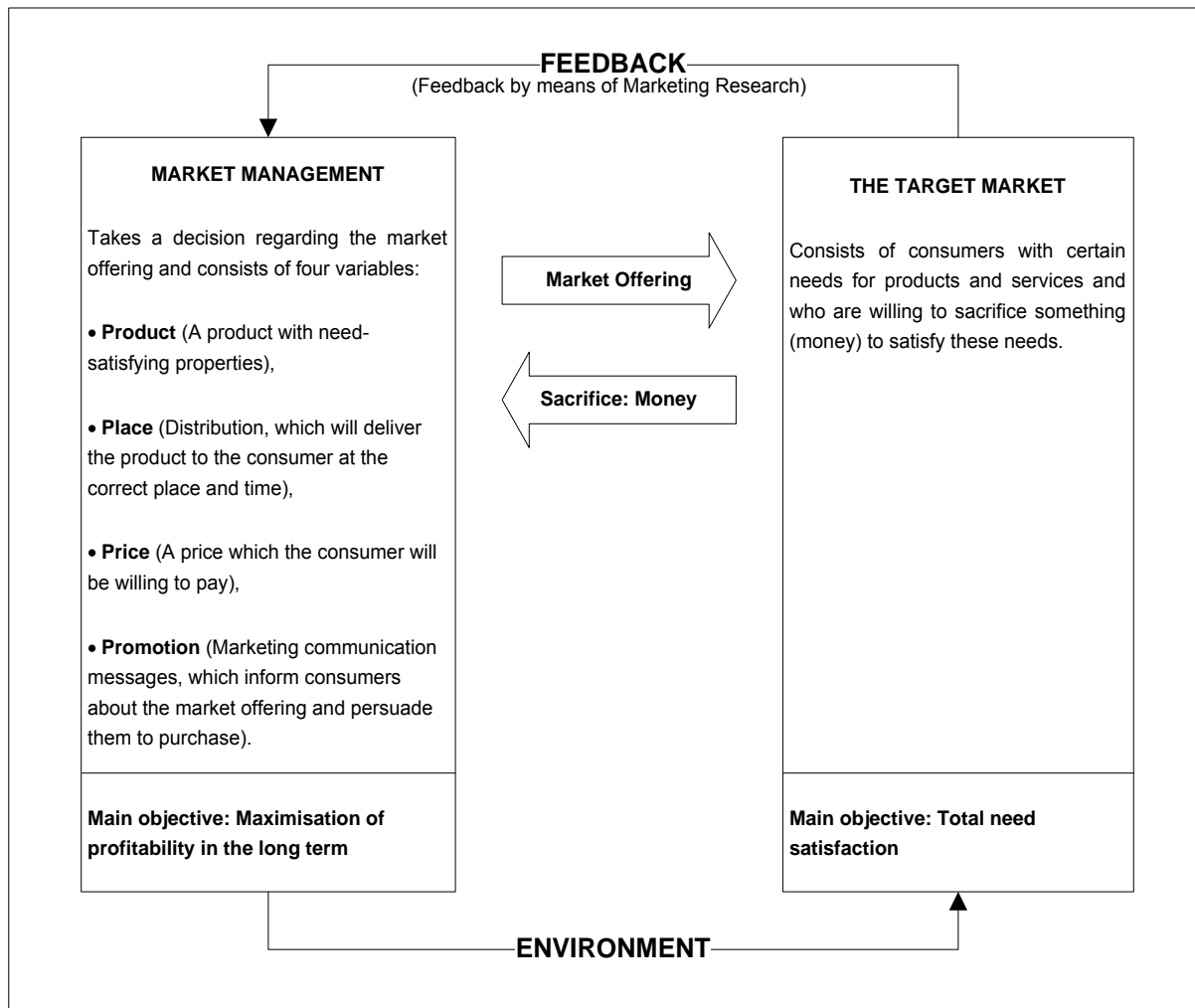
As previously specified, the industry in which the enterprise operates is also the place in which it resides as an economic value-adding entity (Bruijs *et al.*, 1998:522). Bruijs *et al.* (1998:522) have found that internal and external factors affect the marketing function in order to effectively exchange economic value with the enterprise's target customers. The environment surrounding the enterprise can be divided into three main parts (Bruijs *et al.*, 1998:522): the micro-environment, which consists of the internal enterprise forces (e.g. marketing budget) that influence the enterprise's marketing management; the market environment, that consists of all the market factors (e.g. individual and enterprises' needs) that affect the economic exchange with an enterprise; and finally, the macro-environment, that consists of all the factors (e.g. economic, technological, demographic, political and physical) outside the enterprise itself and its markets.

2.5.8.2 Marketing process

The marketing process (see Figure 2.3) can be described as the relocation of a product and service offering from one person to another (Badenhorst *et al.*, 2003:283-286). Even in primeval societies through time, one could see marketing activities being performed, such as bartering (Badenhorst *et al.*, 2003:283-286). People would barter meat for grain, or spice for gold. People would basically exchange the value of products or services that they owned, for relatively the same value of products or services that others could spare, which is a form of marketing.

Needs and the fulfilment of these needs through economic value-added exchanges are viewed as the fundamentals of the marketing function and the marketing process in particular (Badenhorst *et al.*, 2003:283-286). Bruijs *et al.* (1998:517) suggest that it is necessary for the marketing concept to reflect a consumer-centric philosophy, where the long-term success of an enterprise is determined by the forecast of a target market's needs and desires, and the ability to satisfy these needs more effectively than its competitors. The marketing concept therefore calls for synchronisation between all the enterprise's functions, to ensure consumer satisfaction and hopefully profitability (De Klerk *et al.*, 1997:165).

FIGURE 2.3: THE MARKETING PROCESS



Source: Adapted from Badenhorst *et al.* (2003:284)

Another aspect of the marketing process is market research. By scanning the internal and external environment, researchers can obtain accurate and relevant information for sound decision-making in the development of a tangible product or an intangible service offering (Badenhorst *et al.*, 2003:279-280). Products and service offerings are usually presented at a convenient place, at the right price, and after the target customer has received sufficient information about them (Badenhorst *et al.*, 2003:279).

Profit is the motive for the survival and the expansion of any profit-seeking enterprise (Badenhorst *et al.*, 2003:283-286). The marketing management of such an enterprise must therefore strive towards maximising the enterprise's profits by combining the key tasks of segmentation, targeting and marketing mix of the product/service offering.

2.5.8.2.1 Segmentation and target marketing

Cant *et al.* (2010:71) describe market segmentation as the process whereby marketing managers divide markets into smaller segments in order to target consumer needs more effectively. Three main segmenting approaches are generally used, namely market aggregation, a single-segment approach, and a multi-segment approach (Badenhorst *et al.*, 2003:299).

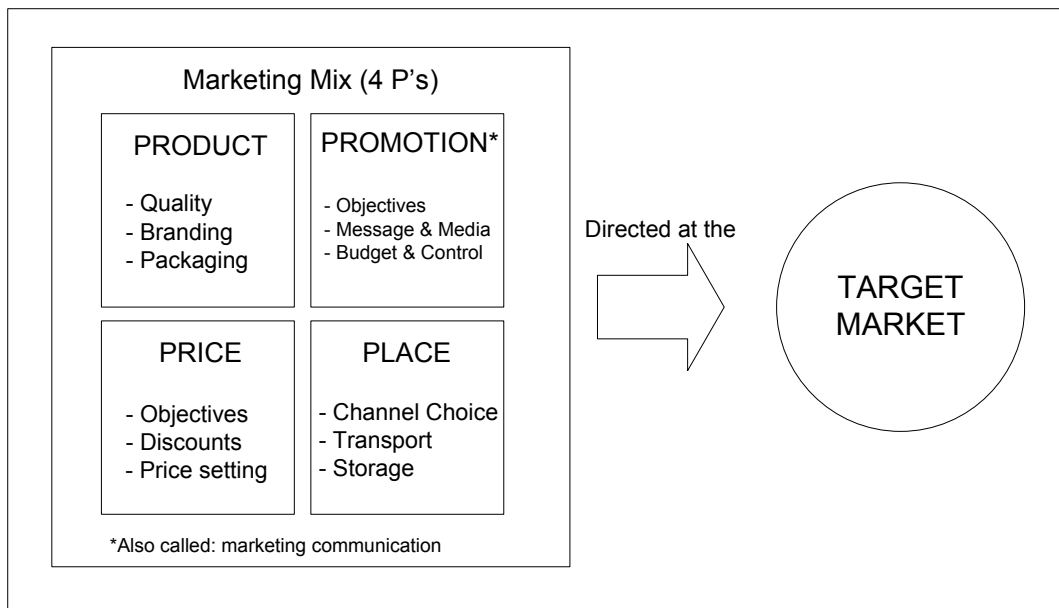
The benefits of market segmentation lie in a more calculated approach. According to Bruijs *et al.* (1998:535), it is better to first segment the market and target appropriate segments, rather than spending the budget on a consumer base that might not even be interested in the enterprise's product/service.

After the heterogeneous market has been partitioned into smaller homogeneous segments, the marketing management must identify a viable target market (Badenhorst *et al.*, 2003:304-305). This latter process is called target marketing, and it is initiated in order to reach attractive future customers by developing a competitive advantage (Bruijs *et al.*, 1998:535). A competitive advantage is something an enterprise does more efficiently and effectively than its competitors, through better execution of its marketing mix (De Klerk *et al.*, 1997:169).

2.5.8.2.2 The marketing mix

The marketing mix, also referred to as "marketing instruments", is a group of variables (4P's) that is used by a marketing manager when delivering a product/service to target consumers (Badenhorst *et al.*, 2003:307-308). The mix usually includes a need-fulfilling product/service, which is offered through a promotional message at a reasonable price to the target users (De Klerk *et al.*, 1997:169). The product must be distributed through various supply channels so that consumers can find it at the right place, at the correct time, in the right quantity.

FIGURE 2.4: TARGET MARKETING STRATEGY



Source: Adapted from De Klerk *et al.* (1997:170)

Badenhorst *et al.* (2003:307-308) describe marketing strategy (see Figure 2.4) as a key that has the power to unlock the needs of the market being targeted. The key is comprised of four marketing instruments, which must function in unison to satisfy the target market's needs.

The enterprise's "product offering" may consist of a tangible product/item or product ranges (Badenhorst *et al.*, 2003:308-309). Apart from the tangible offering, the product may also include various service elements or be offered primarily as a stand-alone service (Du Toit *et al.*, 1981:144-145). Marketing decisions relating to the product itself include product branding, features, quality and packaging decisions (De Klerk *et al.*, 1997:170).

"Price" may be explained as the benefit or economic value that a product or a service holds when it is exchanged (Badenhorst *et al.*, 2003:320-321). Before the price of products or services is estimated, both short-term considerations (e.g. operational costs) and long-term considerations aspects (e.g. infrastructure investment) need to be considered (Du Toit *et al.*, 1981:145-146).

There is a logistical side to marketing, in the sense that distribution channels are required to allow the enterprise to fulfil its tasks (Badenhorst *et al.*, 2003:324). Ultimately, it is the objective of the distribution policy to ensure that the product that is being offered is made

available to the target customer at the correct “place” and at the right time (Du Toit *et al.*, 1981:145).

“Promotion” (also called marketing communication) can be defined as “the personal and impersonal transmission of a message regarding any need-satisfying product or service, personality, institution, place or idea” (Bruijs *et al.*, 1998:543). Marketing communication is critical with regard to the process of educating, persuading and prompting the targeted consumers (Badenhorst *et al.*, 2003:329). Promotional communication can be divided into four broad elements, namely advertising, public relations, sales promotion, and personal selling (Badenhorst *et al.*, 2003:329). This thesis will focus on the informal communication of personal selling by means of EWOM (discussed further in Chapter three).

2.6 CONCLUSION

The purpose of this chapter was to provide a departure point for understanding the topic of marketing function in the broad research field of business management. First, the role of business in society, as well as the hierarchy of human needs and the needs satisfaction cycle was described to set the contexts for business management. Then, the nature of business management and the various business management functions were outlined. The focus finally shifted to the marketing function, which is also the primary focus of this research study.

The marketing management function can be divided into several proactive tasks to create long-term value and growth for the enterprise. It is necessary for the marketing manager not only to develop marketing strategies, shape market offerings and deliver value through strong brands, but also to capture market insights and connect with customers (Kotler and Keller, 2006:29-30). It is the primary goal of this thesis to understand e-commerce customers better, through mining their online brand opinions. Through tracking EWOM review behaviour about publishers’ books being sold on e-commerce platforms, the study will assess the impact that EWOM reviews exert on review credibility and sales levels. The following chapter focuses exclusively on WOM and EWOM communication in the applied research field of consumer behaviour.

CHAPTER 3

CONSUMER WORD-OF-MOUTH COMMUNICATION

3.1 INTRODUCTION

In the field of marketing, specifically consumer behaviour, various studies have been conducted that focus on offline WOM communication (East *et al.*, 2007:175; Sormunen, 2009:14-15). More recently, the focus has shifted to online WOM communication, where researchers analyse the communication patterns of participants on blogs, discussion forums, social networks, and online retailer websites (Hennig-Thurau *et al.*, 2004:39; Chen *et al.*, 2004:711; Sormunen, 2009:24).

Chapter three is divided into three main sections, namely offline, online and consumer WOM communication. The chapter begins with an overview of the nature of WOM and motives for participating in WOM communication. In addition, a more in-depth review of the communication process, as well as the input and output dynamics of WOM communication is provided. In the second section new advances in the fields of business, marketing and consumer behaviour, together with the differences between offline and online WOM communication and the technologies enabling online WOM communication, are discussed. In the last section the EWOM behaviour patterns on e-commerce product review pages are analysed, as well as the role that online WOM plays in the information search process before, during and after consumers undertake their purchase decision-making. This chapter contributes to the existing literature on online rating and review behaviour in e-commerce communities, and the impact that digitised WOM recommendation has on review credibility and sales levels.

3.2 WORD-OF-MOUTH COMMUNICATION

3.2.1 Defining word-of-mouth

Research in the area of consumer behaviour and information dissemination started from the process known as word-of-mouth or in short, "WOM" (Katz and Lazarsfeld, 1955; Al Leen, Ignatius, and Ramayah, 2009:28; Sormunen, 2009:14-15). Arndt (1967 cited in Breazeale, 2009:297-298) describes WOM as the unsolicited communication about a product or service between two or more people. WOM has been characterised as a method of transferring

information through verbal exchange in an informal face-to-face manner, rather than through mass media channels (Al Leen *et al.*, 2009:27). In the commercial arena, WOM involves consumers' advice and opinions about the products and services of enterprises (East *et al.*, 2007:175-176; Jansen, Zhang, Sobel, and Chowdury, 2009:2169-2172).

According to Carl (2006:605), WOM messages can be divided into two main types - "everyday" and "institutional" WOM. Everyday WOM is described as the natural exchange of information between two or more participants about a product or a service that they have experienced as very satisfactory or very unsatisfactory. Product or service information is usually transferred from the communicator to the receiver with the intention to recommend the best possible product or service offering. Institutional WOM, which is also known as "buzz" or "viral marketing" and usually occurs when consumers are driven by a financial incentive to promote a product or service to their peers (Carl, 2006:605; De Bruyn and Lilien, 2008:151-152).

3.2.2 Motives for participating in word-of-mouth communication

To gain a richer understanding of the complex nature of WOM and its impact on people's decision-making processes, the study focuses on the factors that motivate people to participate in WOM communication (Dichter 1966 cited in Sormunen, 2009:19-20). WOM participants do not recommend products or services randomly. They are likely to be driven by various reasons, mainly to exchange knowledge and their opinion about experiences of products and services. Three main motives for engaging in WOM are highlighted here: self-enhancement, altruism, and message involvement.

Self-enhancement as a motivation for WOM participation stems from consumers' personal characteristics and their inner desires to obtain positive recognition (Sormunen, 2009:20). Mazarol, Sweeney and Soutar (2007:1488) have found that some people are less willing to offer their opinions for fear that it will lead to a decrease in social validation by their peers. If the self-enhancement concept is applied to an online setting, recognition can involve peer evaluation in the form of contribution ratings and rankings (Hennig-Thurau *et al.*, 2004:43).

Secondly, altruism or the wish to assist others, also plays a role in WOM. According to Hennig-Thurau *et al.* (2004:40), altruism can originate from the desire to contribute to an enterprise, or the wish to help fellow consumers make decisions about product and service purchases. Altruism can also reach a point where it no longer reflects a willingness to help

out, but expresses a sense of responsibility towards actively serving a community fulltime (Mazzarol *et al.*, 2007). Tapscott and Williams (2008:xiv) suggest that the mass collaboration against global warming is one such initiative that reflects people's responsibility towards changing the future climate of the planet.

Lastly, message-involvement or message participation refers to the intriguing quality of a marketing message which gets people talking about brands, products or services (Dichter 1966 cited in Sormunen, 2009:20). With a decrease in the effectiveness of traditional advertising (e.g. television or newspaper ads), marketers are continually experimenting with communication strategies such as viral marketing (Leskovec, Adamic and Huberman, 2007:1-4). Viral marketing, as created by exponential brand exposure through customer conversations, leverages the fact that people like to become part of the social process of information exchange about products and services (Hennig-Thurau *et al.*, 2004:40; Leskovec *et al.*, 2007:1-4; Sormunen, 2009:48).

3.2.3 The communication process

Research over the past few decades confirms that WOM communication has had a major impact on the way people make decisions, as well as interacting and influencing other members of society (Buttle, 1998:242). Knowledge of interpersonal influence stretches back to ancient times when Aristotle discovered that the artistic aptitude of a speaker could be reflected by his or her ethical, emotional and logical appeals (Aristotle, trans. Roberts, 1924). At present, a plethora of research on interpersonal communication exists. Subsequently many enterprises have started to harness the power of WOM for the marketing communication process (Buttle, 1998:242).

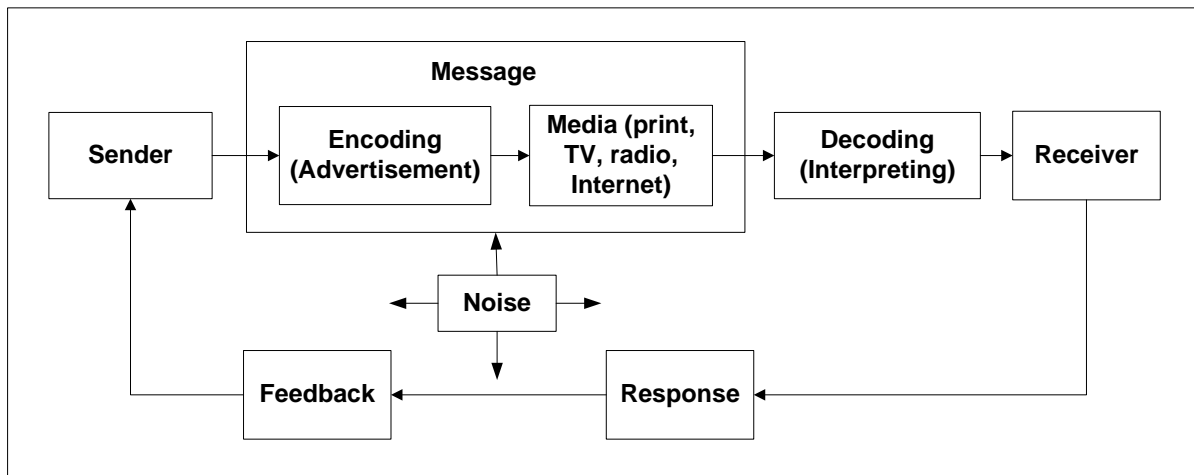
An important dynamic of human communication emerges when analysing information exchange between people. Sormunen (2009:21-23) has found that information is most frequently exchanged between those individuals who share similar interests, beliefs and status. Sormunen (2009:21-23) has also found that human beings want to connect with other human beings. By staying connected on a regular basis, individuals can form strong personal ties in, for instance, social networks.

In the mid-1950s a ground-breaking theory on public opinion formation was published, which radically departed from earlier studies on the understanding of media effects (Katz and Lazarsfeld, 1955). Unlike earlier studies that assumed a one-step information flow from the

media to the broad public, this theory suggested that a two-step flow of information exists (Hirokawa and Lowe, 2002:1-3). First, information and influence are transferred from the media to opinion leaders, and then second, information and influence are transferred from the opinion leaders to their individual social networks (Hirokawa and Lowe, 2002:1-3).

To obtain a better understanding of how messages about products or services are communicated by people, the marketing communication process can be considered. The marketing communication process (see Figure 3.1) can be viewed as a cycle where ideas, knowledge and attitudes are transferred to one person (the receiver or decoder) by another person known as the communicator or encoder (Badenhorst *et al.*, 2003:371).

FIGURE 3.1: THE MARKETING COMMUNICATION PROCESS



Source: Adapted from Brujjs *et al.* (1998:544)

The primary objective of communication is to transmit or encode a message in such a way that the receiver's reaction will be fairly equal to the perceived reaction by the sender (Badenhorst *et al.*, 2003:371). Brujjs *et al.* (1998:544) describe message-encoding as the process whereby text, images and sound combine in a relevant manner and are transmitted over appropriate media channels (e.g. newspapers, magazines, radio, television and the Internet).

The next step in the communication process, before the message is received by the target audience, is known as decoding. Brujjs *et al.* (1998:544) define decoding as the "interpretation of the message and the meaning the receiver attaches to the words, images and sounds". When composing a message, a communicator takes into account not only the decoding of the message, but also interference with reception of the message. Message-

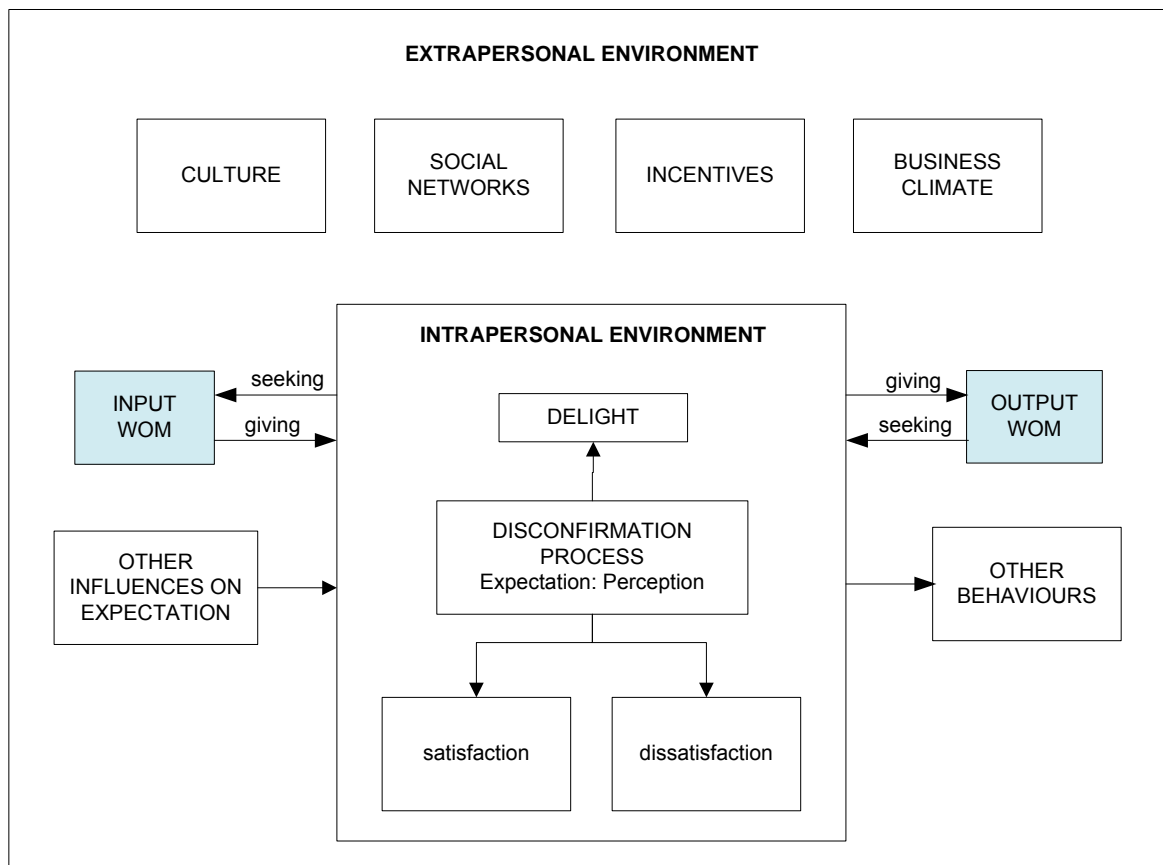
interference is usually referred to as signal or message “noise” (Badenhorst *et al.*, 2003:373).

After the message has been received, a reaction or a response usually occurs. This response determines whether the marketing communication process has successfully reached its goal or not (Brujjs *et al.*, 1998:544). The feedback section can be viewed as the last stage of the dynamic communication cycle, and reflects the receiver’s response to the marketing message (Badenhorst *et al.*, 2003:375).

3.2.4 The word-of-mouth communication model

The communication model can be extended to the two underlying WOM communication processes referred to as “input” and “output WOM”. The WOM model (see Figure 3.2) presented by Buttle (1998:246), will be used to describe the factors that influence consumers seeking input WOM and sharing output WOM.

FIGURE 3.2: THE WOM COMMUNICATION MODEL



Source: Buttle (1998:246)

3.2.4.1 Input word-of-mouth (consumers seeking information)

The process during which consumers actively seek product or service information before making a purchase decision can be referred to as input WOM (Buttle, 1998:245-246). Input WOM can be divided into opinion leaders and trust-level factors; trust-level factors can further be divided into credible sources, tie strength, source expertise, and perceived risk.

3.2.4.1.1 Opinion leaders and influencers

According to Buttle (1998:248), information first flows to key individuals called opinion leaders. Then, by means of WOM communication, the message is passed on to peers, where it impacts the receivers' attitudes and behaviours. Katz and Lazarsfeld (1955) originally defined an opinion leader as "a group member playing a key communications role". Opinion leaders cannot necessarily be classified as leaders of enterprises, but instead they can be seen as highly knowledgeable, connected individuals whose actions have a direct influence on the people who surround them (Watts and Dodds, 2007:441-443).

3.2.4.1.2 Trust-level factors

To better understand the conditions under which consumers actively seek information to make a purchase decision, the following areas in the trust-paradigm will be discussed: source credibility, source tie strength, sender's expertise, and lastly, the risk perceived by consumers (De Bruyn and Lilien, 2008:152-153).

WOM recommendations from credible sources such as colleagues, are considered to be more credible than solicited marketers' messages (Sormunen, 2009:18). It is believed that these WOM referrals have no wish to sell something, while WOM communication that emanates from marketers often has the sole intention to sell products or services to consumers (Sormunen, 2009:18).

Closely related to credibility, is tie strength. The role of the strength of interpersonal ties is well documented in the field of interpersonal influence, social networks and WOM networks (Sormunen, 2009:21-23). Interpersonal tie strength is the intensity of the social relationship between consumers, and can be classified into three broad areas: absent relationships, fleeting or weak tie strength, and firmly-established or strong ties (Bansal and Voyer, 2000:168). When the impact of personal tie strength on information transmission is

examined, weak ties seem to play a much bigger role than strong ties (Granovetter, 1983:203-233). Sormunen (2009:21-23) has found that when a rumour is spread among people, the message is more widely distributed among those with weak ties than those with strong ties, which may restrict the message distribution to the same people in the sender's social network.

Another factor that influences the trustworthiness and credibility of the communicator in the WOM process can be described as source expertise (Sweeney, Soutar and Mazzarol, 2008:347). Bansal and Voyer (2000:169-170) have found that information seekers do not necessarily obtain their information only from key influencers or opinion leaders, but also from people they perceive to be knowledgeable in the required product or service's category they are enquiring about.

Many researchers have found that WOM plays a central role in decreasing or totally removing consumers' uncertainty and perceived risk before making a purchase decision (Sormunen, 2009:34-35). Sweeney *et al.* (2008:349) believe that perceived risk is the main reason that drives consumers to usually opt for pre-purchase trials of products and services. It is also worth mentioning that perceived risk increases with a higher level of product/service importance, ultimately increasing the level of importance of WOM recommendations (Bansal and Voyer, 2000:168-169).

3.2.4.2 Output word-of-mouth (consumers giving information)

Human beings have an intrinsic need to exchange knowledge and opinions that others will comprehend and find relevant (Badenhorst *et al.*, 2003:371-375). When post-purchase product experiences are shared, people engage in what Buttle (1998:246) calls output WOM. Keller (2007:448-452) has found that 40% to 50% of receivers of WOM messages are likely to share this advice, signifying that consumers also spread other people's product or service experiences and not only their own.

Human beings express positive emotions through joy, or reduce tension by communicating negative feelings when an imbalanced situation occurs (Hennig-Thurau *et al.*, 2004:44). In the last decade, polarity or valence has been the focus of many researchers (Buttle, 1998:245-250; Mazzarol *et al.*, 2007:1477; East *et al.*, 2008:215-218). They broadly classify valence into two major types, positive WOM (or PWOM) and negative WOM (or NWOM).

The perception of products or services before a purchase determines the level of sentiment consumers express after the purchase. A higher level of positive perception drives a higher likelihood of communicating positive WOM (Buttle, 1998:247). The same dynamics apply when consumers experience product and service quality to be lower than initially perceived (Sormunen, 2009:36-37). Research conducted by Anderson (1998:5) indicates that the WOM satisfaction-dissatisfaction frequencies form a u-shaped curve, which suggests that consumers actively participate in WOM communication when they are either highly satisfied or highly dissatisfied (Sormunen, 2009:18).

Since WOM communication influences consumer purchase decision-making, it is crucial that marketers understand the nature and communication patterns of WOM (East *et al.*, 2007:175-177).

3.3 ELECTRONIC AND CONSUMER WORD-OF-MOUTH COMMUNICATION

The evolution of interpersonal communication in the Internet landscape is better known as “electronic word-of-mouth” or “EWOM” (Al Leen *et al.*, 2009:28). Besides EWOM, the occurrence has also been referred to as “Internet word-of-mouth” and “online WOM” (Hennig-Thurau and Walsh, 2004:67; Huang, Lin and Lin, 2009:161; Jansen *et al.*, 2009:2169). EWOM communication can be defined as “any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet” (Hennig-Thurau *et al.*, 2004:39). The study of online communication falls in the two major research fields of consumer behaviour and Internet marketing.

Schiffman and Kanuk (2007:3) define consumer behaviour as the “behaviour that consumers display in searching for, purchasing, using, evaluating, and disposing of products and services that they expect will satisfy their needs”. The examination of consumers and how they behave, stems from the marketing literature of the 1950s (Schiffman and Kanuk, 2007:4). Marketers found that a higher volume of goods could be sold more frequently when enterprises produced only those goods that were in higher demand by consumers, and not those goods that were produced without any customer input (Schiffman and Kanuk, 2007:4-5). Various dynamics of how consumers communicate through WOM, as well as the effect that these behaviours have on purchase decision-making, are integrated into the all-inclusive model of consumer behaviour (Schiffman and Kanuk, 2007:15-16).

It is important in the planning of a long-term e-commerce strategy that marketers cultivate an appreciation for the dynamics of consumer behaviour, as well as integrating offline media channels with the Internet and mobile channels (Chaffey *et al.*, 2009:74). Internet marketing is described as the “application of the Internet and related digital technologies in conjunction with traditional communication to achieve marketing objectives” (Chaffey *et al.*, 2009:9). Internet marketing is very similar to concepts like e-marketing, online advertising, Internet advertising, and digital marketing (Sormunen, 2009:40-41). Unlike Internet marketing, alternative concepts like electronic marketing cover a broader scope, by focusing on other digital media like wireless media and electronic customer relationship management systems (Chaffey *et al.*, 2009:9).

3.3.1 The power shift in business, marketing, and online consumer behaviour

When the evolution of the Internet is examined, several key turning points should be noted. Aldridge, Forcht and Pierson (1997:161) point out that the Internet was originally established in the 1960's to protect the United States's mainframe against attacks in the Cold War. From there it grew into a network that shared academic and research data in the 1970's, until it finally phased into mainstream adoption as the information highway in the 1990's.

The first class of web applications (e.g. websites, email, instant messaging, shared workspaces and web conferencing) was developed from 1990 onwards. Then a second growth and innovation spurt took place in 2004, with the initiation of web 2.0 applications including blogs, wikis, social networking, social bookmarking, tagging, RSS, mashups and widgets (Paxhia, 2008:10-12). Web 3.0 applications followed next. Chaffey *et al.* (2009:14) describe web 3.0 applications as the “next generation web” that will incorporate “high-speed connectivity, complex cross-community interactions and an intelligent or semantic web where automated applications can access data from different online services to assist searchers perform complex tasks of supplier selection”. In the present study, the focus will be only on the web 2.0 environment.

The Internet can be described as having a dual function. It not only serves as the platform that allows consumers to communicate through web applications, but it also allows enterprises and marketers to take advantage and narrow down their target market to promote a strategic promotional message to each individual consumer. With the advent of the Internet, a paradigm shift took place, especially in the way consumers conduct their search and purchase decision-making. This new consumer power in return influences the

way marketers and enterprises allocate their resources (e.g. time, money, assets) and strategise their offline and online operations.

3.3.1.1 The Internet's impact on consumer behaviour

Current media channels are blurring the traditional lines between interpersonal and mass communication (Jansen *et al.*, 2009:2170; Sormunen, 2009:25). Today consumers, friends and family can connect from anywhere in the world and communicate their opinions to a larger group of people instantaneously, or follow like-minded people through virtual communities (Schiffman and Kanuk, 2007:11-13; Sormunen, 2009:25).

Changes in consumers' information search and purchase decision-making has brought about greater customer empowerment, price sensitivity and bargaining power (Kotler and Keller, 2006:13-15). Through digital technology on the Internet, consumers can, in the comfort of their homes, obtain information on a much broader range of products and services, as well as instantaneously comparing the different price levels of these offerings (Kotler and Keller, 2006:13-15; Schiffman and Kanuk, 2007:11-13; Chaffey *et al.*, 2009:3). Owing to an increase in enterprise productivity and faster lifestyles of human beings, consumers are continually demanding a higher degree of customised product and service solutions, as well as the way in which these products and services are marketed to them (Aldridge *et al.*, 1997:163; Kotler and Keller, 2006:14). When the focus shifts from the buy-side to the sell-side of e-commerce, it can be seen that sellers can bypass intermediaries and advertise directly to other consumers via online auction houses such as *Ebay.com* (Schiffman and Kanuk, 2007:12).

3.3.1.2 Business and marketing management

Because the Internet shifted the dynamics between consumers and enterprises, it directly and indirectly changed the way customers perceived and experienced product and service offerings (Rowley, 2004:132). The most noteworthy were price changes, broader product ranges, better customer targeting, and more customised marketing messages.

As access to a broader range of product and service information improved, many enterprises became more aware of consumer price awareness. This price sensitivity led to business and marketing management spending more time on devising strategies to ultimately deliver more

competitively-priced products and services (Schiffman and Kanuk, 2007:11-13). According to Brynjolfsson, Hu and Smith (2006), the greatest value from online shopping is not necessarily the lower prices, but rather the greater product and service ranges that are now available to individual consumers. Leskovec *et al.* (2007) suggests that not only is the ability to host a wider variety of products online easier and cheaper, but they have found that a greater number of lesser-known products are bought by online customers.

If online marketing messages are compared to offline print and TV commercials, consumers are no longer bombarded with generic messages, but instead the web offers opportunities for a more targeted approach to individual consumer marketing (Aldridge *et al.*, 1997:164; Merisavo, 2008). Online retailers such as *Amazon.com* and *Ebay.com* are investing in new technologies (e.g. cloud computing, rating and review platforms, and recommender systems) to collect complex data about individual consumers, as well as putting systems in place to analyse information in real time in order to effectively recommend targeted marketing messages (Kotler and Keller, 2006:28; Schiffman and Kanuk, 2007:11-13). In an attempt to better understand online consumer behaviour, a comparison between offline WOM and online WOM communication is drawn.

3.3.2 A comparison between word-of-mouth and electronic word-of-mouth communication

When WOM and EWOM communications are compared to one another, previous researchers have found that the motivation for participating in WOM offline is approximately the same as for online WOM (Hennig-Thurau *et al.*, 2004:40; Huang *et al.*, 2009:160-161). Sormunen (2009:29) suggests that in the midst of these similarities, four main differences can be identified: verbal WOM compared to written WOM, conversation taking place face-to-face compared to indirect interaction, identification versus anonymity, and narrow reach versus broad dispersion of WOM.

3.3.2.1 Verbal word-of-mouth compared to written word-of-mouth

The key difference between offline WOM and online WOM is that traditional offline communication occurs through speech, while EWOM conversations take place in the form of written words (Sormunen, 2009:30). Speech consists of utterances, for example when people talk to one another with the intention to educate, warn or assert (Jansen *et al.*,

2009:2170). Conversely, online conversations are personal opinions that have been written on web or mobile pages (Sormunen, 2009:30-31). Unlike WOM that is spoken and vanishes in a spontaneous manner, online written WOM has permanence until the user decides to delete it (Stern, 1994:7).

3.3.2.2 Face-to-face conversation compared to indirect interaction through electronic word-of-mouth

One of the major strengths of the Internet and the digital environment is that it permits communication to evolve from a one-way exchange to a bi-directional, interactive, interpersonal exchange (Dellarocas, 2003:1407; Merisavo, 2008). Chaffey *et al.* (2009:12-14) suggest that interaction through participation in the creation of user-generated content has always been at the heart of web 2.0. It might be that EWOM is less personal than offline WOM, because communication does not occur through face-to-face conversation, or it might also be that a new level of personal interaction is experienced online (Jansen *et al.*, 2009:2170).

3.3.2.3 Identification versus anonymity

In traditional WOM conversations people are aware of the communicator's identity, while in EWOM conversations people can choose to communicate anonymously (Sormunen, 2009:31-32). Hennig-Thurau *et al.* (2004:43) suggest that the ability to communicate one's opinions in an anonymous manner is an important characteristic of EWOM. Ultimately, allowing participants to communicate their opinions to others, normally would not occur in an offline setting. Even though people now have the ability to access a broader range of opinions and potentially a richer source of expertise, new challenges arise from online interaction and exchange between anonymous identities. According to Huang *et al.* (2009:160-161) and Dellarocas (2003:1410), challenges such as manipulation and misrepresentation owing to a lack of contextual cues might occur in the online environment, which would not necessarily have happened in the traditional offline setting.

3.3.2.4 Narrow reach versus broad dispersion of word-of-mouth

Not only does the Internet allow voices to unite immediately over a significant reach, but it also provides people with a place to publish their product and service experiences where

they will be accessible and useful to a broad online community which is seeking other people's opinions and recommendations (Hennig-Thurau *et al.*, 2004:42; Sormunen, 2009:32; Huang, Hsiao and Chuang, 2009:1-3). The reason for this unparalleled scale of communication can be ascribed to the Internet's low cost and bi-directional information exchange capabilities (Dellarocas, 2003:1407; Huang *et al.*, 2009:160–161).

3.4 TECHNOLOGIES ENABLING ELECTRONIC WORD-OF-MOUTH COMMUNICATION

The platform on which EWOM communication occurs is often cited as the new media environment as opposed to traditional media platforms such as books, radio and television (Neuman, 1991). New media, otherwise known as “the web 2.0 environment” is described by Chaffey *et al.* (2009:12-14) as a “collection of web services that facilitate interaction of web users with sites to create user-generated content and encourage behaviours such as community or social network participation, mashups, content rating, use of widgets and tagging”. Li and Bernoff (2008) divide web 2.0 technologies into four broad areas of focus: creative technologies; collaborative technologies; RSS feed, widgets and tags; and lastly, common connective technologies.

In the current Internet environment, technologies made for creating and editing text, audio and video, are user-friendly, relatively inexpensive, and can be accessed by anyone with a mobile phone or computer (Sormunen, 2009:26-28). Creative web technologies include blogging, micro-blogging, video sharing, and podcasts. Weblogs (e.g. *Blogger.com*) can be described as a many-to-many publishing platform to share people's experiences on various interest categories, as well as using it as a WOM advertising tool when advocating focused interests, such as advertising in political elections (Thorson and Rodgers, 2006:40-43; Huang *et al.*, 2009:160-161). Another form of many-to-many publishing is described as “micro blogging” (Jansen *et al.*, 2009). This newer form of EWOM publishing allows users to diffuse ideas to a multitude of people with a limit of 140 text-characters (e.g. *Twitter.com*). Audio and video editing of special interest fields are usually performed on and distributed through platforms like *Apple's iTunes.com* and *Google's Youtube.com* (Chaffey *et al.*, 2009:11).

In the past, knowledge management systems of enterprises focused on accumulating documents into separate knowledge bases, rather than facilitating collaboration among individuals (Richter and Koch, 2008:87-90). Two key collaboration platforms that have played a critical role in knowledge sharing are so-called “wikis” (e.g. *Wikipedia.com*), and

open-source software (e.g. *Linux*). Wikis are a technology that enables a network of individuals or employees to share knowledge and strives to build collective intelligence by means of integrated collaboration (Sormunen, 2009:28). With the advent of the Internet, programming-code contributions between individuals eventually led to open-source software projects. The Linux project was initially established to allow software developers to contribute to a niche open-source operating system, but eventually grew into a widely followed phenomenon that changed forever the dynamics of the software industry (Tapscott and Williams, 2008:xi).

The third set of web technologies that enables online users to categorise, edit, follow and share knowledge about various data-types, is classified into Tags, RSS and Widgets. Sormunen (2009:29) describes tagging (e.g. *delicious.com*) as a way of classifying different web content into various categories through the use of words. “RSS” stands for Really Simple Syndication, and is a way for users who are subscribed to blogs, websites, or news articles, to receive updates through an XML-feed to reader services such as *Google Reader* (Chaffey *et al.*, 2009:11). Lastly, widgets can be described as mini-application services that allow users to obtain very specific information such as weather updates via a weather-widget (Sormunen, 2009:29).

Web technologies that create a common platform for their users to connect to one another are classified into broad communication tools, virtual communities, social networks, and finally, opinion platforms (Sormunen, 2009:28). Tools used for one-to-one online communication between employees, family and friends can be further categorised into services like instant messaging (e.g. *AOL's instant messenger* and *Google's g-talk*), as well as email services such as *Outlook.com* (Sormunen, 2009:28; Huang *et al.*, 2009:160). Virtual communities (e.g. *Flickr.com*) and social networks (e.g. *Facebook.com* and *LinkedIn.com*) on the other hand, are platforms that enable information exchange between individuals and various user-groups (Chaffey *et al.*, 2009:11; Huang *et al.*, 2009:160–161). Web users also have the ability to connect, exchange interests and share product and service experiences on discussion boards, forums, and online retailers' websites, such as *Amazon.com* (Sormunen, 2009:28).

The popularity of virtual communities and the WOM reviews that occur in them reflects the importance for further research consideration (Brown, Broderick and Lee, 2007:3).

3.5 THE SEARCH COST AND CREDIBILITY OF INFORMATION ON REVIEW WEB PAGES OF E-COMMERCE PLATFORMS

Human beings daily make several hundred if not thousands of decisions concerning all aspects of their intricate lifestyles. These decisions are generally made without having to stop and think about the actual process itself (Schiffman and Kanuk, 2007:526). According to Schmidt and Spreng (1996:246), the purchase decision-making process that is undertaken by consumers consist out of several steps that progress from recognising a problem, to information search, to evaluating the alternative choices, to the actual procurement decision, and finally the behaviour of consumers after the procurement.

Schiffman and Kanuk (2007:533) suggest that information search is initiated by consumers after a perceived need might be satisfied by the procurement and consumption of a product or service. The information search process can be divided into internal or external search: internal search take place when consumers can recall past experiences that provides sufficient information to make a present choice, while external search take place when consumers seek information in their current environment as no previous experience were acquired or past experiences are unable to be recalled from their long-term memory (Schmidt and Spreng, 1996:246; Schiffman and Kanuk, 2007:533).

Other than search cost the degree to which consumers perceive risk also influences the information search stage of the purchase decision-making process (Schiffman and Kanuk, 2007:533). According to Schiffman and Kanuk (2007:533), situations where consumers perceive risk to be low, they usually utilise simple search and assessment techniques, while in high-risk scenarios, they tend to utilise more complex search and assessment techniques.

In an effort to enhance the efficiency of purchase decision-making in online shopping environments, as well as to ensure higher credible product and service information for consumers, many e-commerce retailers (e.g. *Amazon.com*, *Ebay.com* and *Barnesandnoble.com*) have invested in technologies to cultivate online review and rating communities that produce more trusted product and service information (Chen *et al.*, 2008:1-7). E-commerce reviews not only improve the perception of usefulness for online consumers, but also lead to increased website traffic, and user time spent on product pages (Mudambi and Schuff, 2010:185-186). Customers who read these digitised reviews experience higher levels of satisfaction when the perceived product quality which was gleaned from reading online recommendations, is closer to their post-purchase experiences (Chen *et al.*, 2008:1-7; Sormunen, 2009:34-35; Mudambi and Schuff, 2010:185-186). Not only do review systems

allow consumers to read product and service reviews, but it also enables online users to rate the credibility of a review message through a mechanism named “found helpful” (Mudambi and Schuff, 2010:186). The helpfulness of customer reviews can be defined as “peer-generated product evaluation that facilitates the consumer’s purchase decision process” (Mudambi and Schuff, 2010:186). These “found helpful” ratings are usually placed alongside each review on product pages, and can also be used through a filtering algorithm to display to consumers the top three most helpful reviews.

One of the first and most critical factors that plays a role in the evaluation process of products by consumers, is that product pages need to include several reviews (Sormunen, 2009:36). With the motivation to deliver customer reviews which online users perceive to be valuable and credible, online retailers such as *Amazon.com* post detailed guiding principles for supplementing a review’s star rating with quality review text, as well as encouraging comments at the bottom of each review (Mudambi and Schuff, 2010:186). Sormunen (2009:36) and Jansen *et al.* (2009:2180) have also found that when users browse online product pages, more attention is paid to the relevance of the reviews’ content and comprehensiveness than just the polarity of the rating.

It is therefore essential to appreciate the importance that consumers attach to online reviews when initiating their purchase decision-making, as well as to measure which of those particular factors play the biggest role in influencing consumer’s behaviour (Hennig-Thurau and Walsh, 2004:52).

3.6 CONCLUSION

The objective of this chapter was to introduce the processes of WOM and EWOM communication in the broad research field of consumer behaviour and Internet marketing. First, the nature and the motivation drivers of WOM, as well as the basic communication model and the WOM communication model were described to set the contexts for offline WOM. Then the focus shifted to an online environment where EWOM communication was introduced. The underlying themes that were discussed started at the power shift that was brought about with the advent of the Internet, and moved to where the difference between WOM and EWOM communication was examined, and finally the attention was focused on the technologies that enabled the creation, editing and distribution of EWOM. One particular technology application, e-commerce rating and review platforms, was chosen as the central focus of this study.

In past research it has been suggested that EWOM communication has an important impact on online consumer behaviour and purchase decision-making (Chevalier and Mayzlin, 2004:2-12). The present study draws on previous literature to develop a theoretical understanding of the dynamics of WOM recommendations. Two hypothetical models were created to examine the relationship of EWOM review factors on review credibility, as well as to assess the impact of EWOM patterns on e-commerce sales levels over time. The model was empirically tested by using freely available public data from online retailers *Amazon.com* and *Barnesandnoble.com*.

This research study aims to bridge the gap between information technology, online consumer behaviour, user experience design, and Internet marketing. Overall, the analysis serves to help online retailers, user experience designers and e-commerce book-publishers to better understand the online behaviour of their users, and to apply the research study's findings to future e-commerce and marketing strategies.

CHAPTER 4

RESEARCH METHODOLOGY

4.1 INTRODUCTION

One unique characteristic that reflects an effective 21st century enterprise is the way in which the enterprise conducts its marketing research process (Badenhorst *et al.*, 2003:286). The same process allows the enterprise to investigate its electronic marketing channels in order to better serve its future customers through unique product and service offerings.

The primary purpose for conducting research in the field of Internet marketing and online consumer behaviour, with a specific focus on EWOM communication, is to develop better decision-making practices by the marketing management department of an e-commerce enterprise (Glos *et al.*, 1980:130).

The first step in the execution of the present marketing research process was to formulate the research problem, objectives and hypotheses. Secondly, the design of the research study together with the primary research method was chosen, to direct the sampling procedure that was used in drawing samples from the target population. Extracted sample data were first processed before any data analyses were conducted. The final stage in the process was to prepare the research's results and address the formulated research hypotheses.

4.2 THE MARKETING RESEARCH PROCESS

Marketing research, whether it is at scholar or at enterprise level, is defined by Kotler and Keller (2006:102) as the "systematic design, collection, analysis, and reporting of data and findings relevant to a specific marketing situation". Marketing research is a fundamental part of an organisation's intelligence system, and provides managers with relevant and correct information to improve decision-making (Kumar *et al.*, 2002:3). The research process consists of inter-related steps which need to be taken to collect and analyse the necessary marketing information (Bruijs *et al.*, 1998:523). Each consecutive step of the process is illustrated in Figure 4.1.

FIGURE 4.1: THE MARKETING RESEARCH PROCESS

Source: Adapted from Blumberg *et al.* (2008:57)

4.2.1 Step 1: Formulate the research problem

Recognising and defining the research problem is an important first step in the marketing research process (Kumar *et al.*, 2002:50). Enterprises can view this either as a problem that needs to be solved, or as a calculated opportunity that should be seized. According to Kotler and Keller (2006:104), marketing researchers often fall into the trap of defining the research problems too narrowly or too broadly. Dillon, Madden and Firtle (1993:25) suggest that the probability of designing research studies that will provide more accurate information is enhanced when the research problem and the objectives are well defined. Only when they are accurately defined can the research project be designed in such a way as to produce relevant information (Zikmund and Babin, 2010:106-119).

Theories, concepts, principles and objective-based evidence should always be the foundation of research (Malhotra, 2002:53). Theories enable researchers to, firstly, be directed in the empirical research, and secondly, discern which specific variables should be examined (Morris, 2010:60). The most important theories that will be examined for this study in Internet marketing and online consumer behaviour will be consumer search cost, EWOM review credibility, and EWOM review influence on sales levels.

According to previous researchers (Chevalier and Mayzlin, 2004; Chen *et al.*, 2004; East *et al.*, 2007; Chen *et al.*, 2008; East *et al.*, 2008), the most important limitation of earlier EWOM research was the inability to measure the impact of EWOM review communication in an online setting. The primary research problem of the present study was to investigate the influence that EWOM review communication has on online review credibility, as well as to test the relationship between spotlight EWOM reviews and e-commerce platform sales rank values, which serve as a proxy for online sales levels. Thus in this study, the dependent variable for model 1 was review credibility, and the independent variables were stars, ex-stars, sentiment, ex-sentiment, text length, time, and platform, while the dependent variable assigned to model 2 was sales rank and the independent variables were average book star rating, most helpful review stars, and most recent review stars.

4.2.2 Step 2: Formulate the research objectives and hypotheses

The research objective is a statement that guides the researcher to obtain the exact information that is required to satisfy the research purpose, which is to examine perceived credibility of electronic word-of-mouth communication on e-commerce platforms. The research objectives can be divided into two main components: research questions and the development of hypotheses (Kumar *et al.*, 2002:54).

4.2.2.1 Research questions

Research questions flow directly from the research objectives, and specify the information that decision-makers need to obtain to ultimately realise the purpose of the research study (Dillon *et al.*, 1993:25). The purpose of this study was to examine EWOM review communication on e-commerce product review pages. The research questions were formulated to assess¹:

- a review's *star rating* influence on its *credibility*,
- a review's *ex-star rating* influence on its *credibility*,
- a review's *sentiment rating* influence on its *credibility*,
- a review's *ex-sentiment rating* influence on its *credibility*,
- a review's *text length* influence on its *credibility*,
- a review's *publish date* influence on its *credibility*,

1) A more in-depth explanation of each of the review questions follow in section 4.2.6.2.

- a review's *e-commerce platform* influence on its *credibility*,
- the influence that *EWOM review communication* exerts on the *sales rank* values of new best-seller books.

Kumar *et al.* (2002:56) suggest that when research questions are being formulated, it is sometimes challenging to articulate those questions. The development of hypotheses allows the researcher to further refine the research questions into measurable statements that can be empirically addressed.

4.2.2.2 Hypothesis development

Hypothesis development is the generation of a list of formal statements that are constructed to analyse the relationship between two or more variables, to ultimately address the research objectives and obtain answers to the stated research questions (Kumar *et al.*, 2002:56; Coldwell and Herbst, 2004:86). Research questions flow directly from the research objectives, and are again linked to the related hypotheses.

As stated in Chapter one, the purpose of this study was to examine perceived credibility of electronic word-of-mouth communication on USA-based e-commerce platforms. In this study, EWOM book reviews were categorised into two main models: 1) the effect of individual online EWOM factors on a review's credibility, and 2) the association of EWOM review communication with the sales levels of new best-seller books.

Model 1's hypotheses, which focused on review credibility, were partially replicated from comparative studies conducted by Mudambi and Schuff (2010:189) as well as Tsang and Prendergast (2009:1271). A further hypothesis was formulated to test the association between EWOM review communication and sales levels. This latter hypothesis is based on a study conducted by Chen *et al.* (2004:713-714) that examined the impact of online recommendations and consumer feedback on the sales levels of e-commerce platforms.

The research questions and hypotheses were initially formulated to include the following e-commerce platforms: *Amazon.com* (American-based), *Barnesandnoble.com* (American-based) and *Kalahari.com* (South African-based). Owing to insufficient available review data, it was later decided to exclude *Kalahari.com* from the study.

4.2.2.2.1 The effect of individual electronic word-of-mouth factors on review credibility

To address the question whether review credibility is influenced by various individual EWOM factors, the following research questions and hypotheses were formulated.

- *Review question one: Does a review's star rating influence its review credibility?*

H0¹: A review's credibility on an e-commerce platform is not influenced by the review's star rating

- *Review question two: Does a review's ex-star rating (the distance in stars from a 3-point star rating on a rating scale between 1 and 5 stars) influence its review credibility?*

H0²: A review's credibility on an e-commerce platform is not influenced by the review's ex-star rating

- *Review question three: Does a review's sentiment rating influence its review credibility?*

H0³: A review's credibility on an e-commerce platform is not influenced by the review's positive sentiment rating

- *Review question four: Does a review's ex-sentiment rating (the distance from an average rating of 0.5 on a rating scale between 0.0 and 1.0 sentiment) influence its review credibility?*

H0⁴: A review's credibility on an e-commerce platform is not influenced by the review's ex-sentiment rating

- *Review question five: Does a review's text length influence its review credibility?*

H0⁵: A review's credibility on an e-commerce platform is not influenced by the review's text length

- Review question six: Does a review's publish date influence its review credibility?

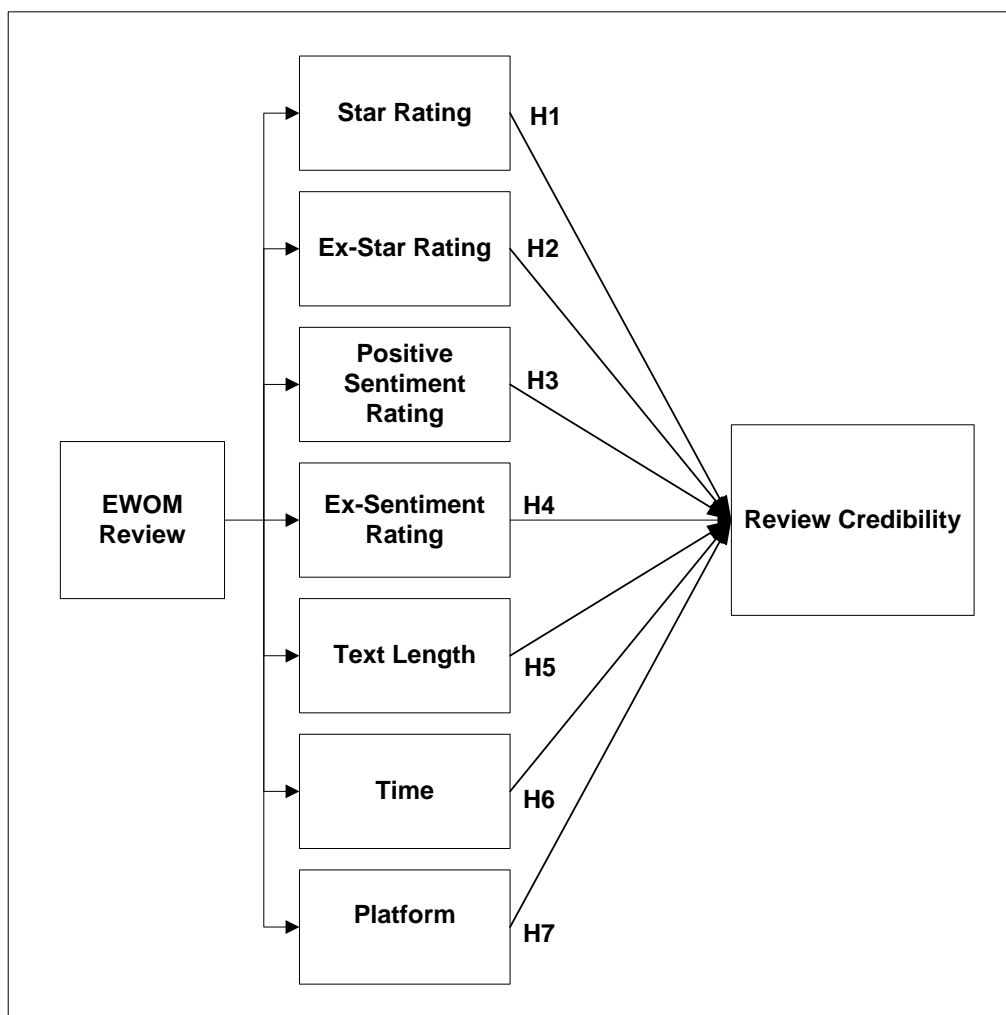
H0⁶: A review's credibility on an e-commerce platform is not influenced by the time the review appeared

- Review question seven: Does a review's e-commerce platform influence its review credibility?

H0⁷: A review's credibility is not influenced by the e-commerce platform upon which it takes place

These hypotheses can be graphically depicted as shown in Figure 4.2.

FIGURE 4.2: EWOM REVIEW FACTORS AND REVIEW CREDIBILITY (MODEL 1)



4.2.2.2 The association of electronic word-of-mouth review communication with the sales levels of *Amazon.com*'s new best-seller books

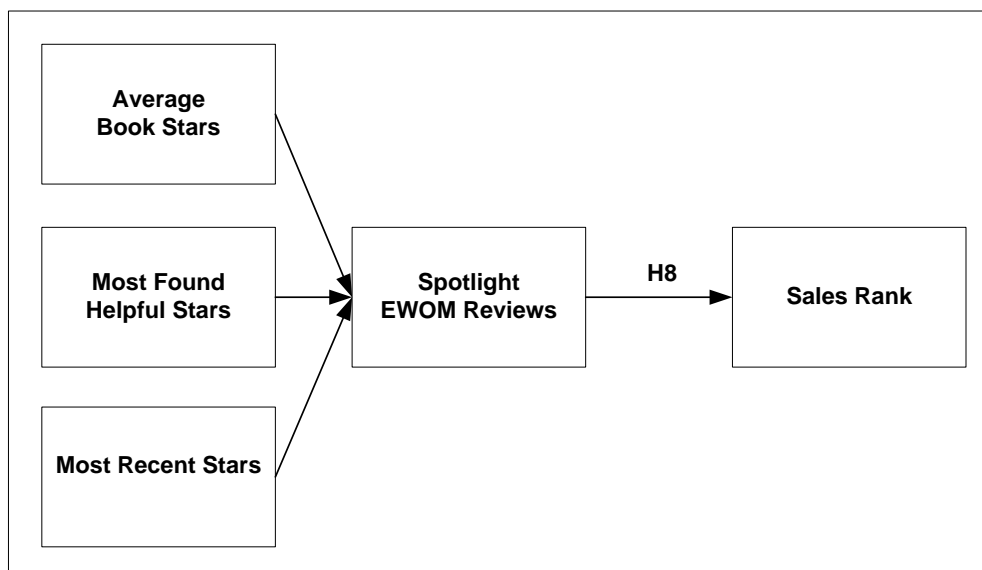
To address the question whether EWOM influences the sales levels of new best-seller books, the following research question was formulated.

- Review question eight: Does more EWOM about new best-seller books on *Amazon.com* have an impact on its sales rank levels?

H0⁸: Sales rank levels of *Amazon.com* books are not influenced by spotlight EWOM reviews

This hypothesis is graphically illustrated in Figure 4.3.

FIGURE 4.3: SPOTLIGHT EWOM REVIEWS AND SALES RANK LEVELS (MODEL 2)



4.2.3 Step 3: Conduct the research design

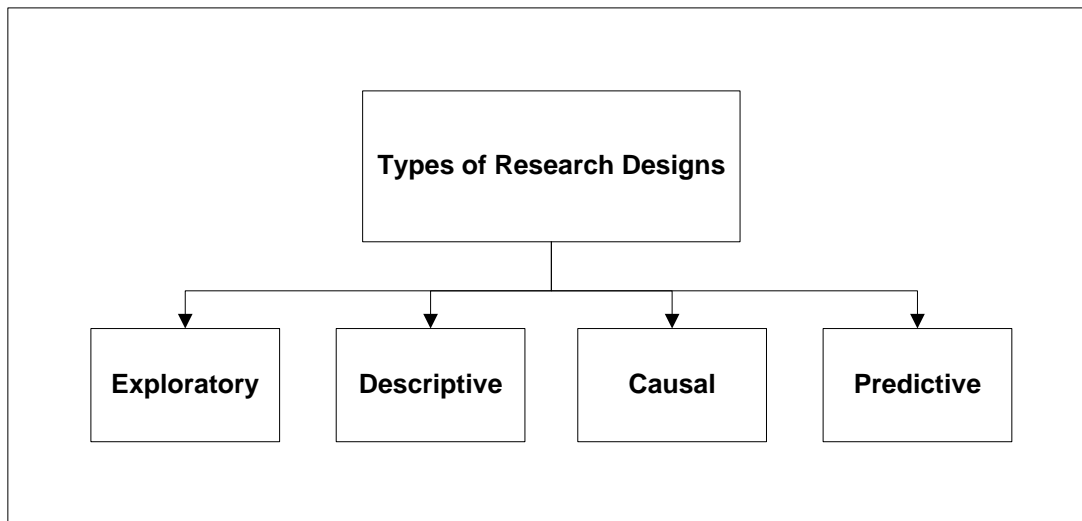
Blumberg *et al.* (2008:195) describe research design as the strategy that needs to be formulated to obtain the answers to the research questions under investigation. According to Kumar *et al.* (2002:73), a wide variety of data collection methods exist from which researchers can choose, either individually or collectively. The first question that needs to be

asked is whether the current research problem can be answered with information from primary or secondary data sources (Dillon *et al.*, 1993:33).

Even though secondary data are collected for a different purpose from addressing a current question being investigated, it may still contain valid information that can be used to answer a research problem (Zikmund and Babin, 2010:163). Secondary data can usually be collected from the enterprise's internal information system, databanks external to the enterprise, and other data repositories such as consumer purchase panels (Kumar *et al.*, 2002:74). Zikmund and Babin (2010:163) suggest that a lot of time and money can be saved if, before the research project is launched, researchers first conduct a thorough investigation of currently available secondary data sources. If data are available only in a format that is unsuitable, or if there are no available secondary data sources that can provide information to address the current research purpose, the researcher is left to collect primary data to specifically answer the stated research problem (Zikmund and Babin, 2010:164-165). For the purpose of this study, secondary data in the form of academic literature and primary data from *Amazon.com* and *Barnesandnoble.com* were collected.

Research design can be categorised into four general types of research (see Figure 4.4), namely exploratory, descriptive, causal and predictive research (Bruijs *et al.*, 1998:524).

FIGURE 4.4: TYPES OF RESEARCH DESIGNS



Source: Adapted from Bruijs *et al.* (1998:524)

According to Bruijs *et al.* (1998:524), exploratory research allows researchers to derive better insights and ideas about broadly defined research problems. As its name suggests,

this research design is used to explore research problems that are unclear or problems that researchers have little prior knowledge about (Aaker, Kumar and Day, 2001:72).

Descriptive research designs permit researchers to obtain answers to topic-related questions (e.g. who, what, when, where and how) (Brujij *et al.*, 1998:524). Churchill and Iacobucci (2002:91) suggest that descriptive studies often measure frequencies of occurrences to gain an accurate description of the situation.

The concept of cause is the principal theme of causal research (Brujij *et al.*, 1998:524). Causal research hypotheses are usually well defined to specifically determine which changes in one variable (e.g. decrease in oil prices) causes change in another variable (e.g. increase in consumer lending) (Morris, 2010:63).

According to Brujij *et al.* (1998:524), predictive research refers to the forecasting of future values. Examples of predictive research design include the forecasting of future enterprise sales in a given industry (Brujij *et al.*, 1998:524).

In the present study, a two-stage research design was adopted, in which problem exploration becomes a first stage, and assessing online review behaviour by means of an observation study (see section 4.2.4) the second stage (Blumberg *et al.*, 2008:344).

4.2.4 Step 4: Select a primary research method

Malhotra (2002:43) defines primary research data as that data that originates from the researcher for the specific purpose of addressing the stated research problem. Field work involves the actual collection of data according to specific research approaches (see Table 4.1), such as observational research, survey research, experimental research, behavioural data research, and focus group research (Brujij *et al.*, 1998:525).

TABLE 4.1: PRIMARY RESEARCH APPROACHES

Primary research approaches	Brief description
Observational research	Gathers data by observing the relevant actors and settings.
Survey research	Gathers data about people's preferences, beliefs and satisfaction, and measures their degree in the overall population.
Experimental research	Captures the cause-and-effect relationships of the observed data.
Behavioural data research	Tracks consumer behavioural patterns by means of in-store scanning data, customer databases and catalogue purchases.
Focus group research	Gathers data from a group of usually 6 to 10 people on various topics of interest.

Source: Adapted from Kotler and Keller (2006:105-106)

For the purpose of the present study, quantitative primary data were collected by means of an electronic observational research approach. Behaviour as it occurs in the natural setting can be observed directly, indirectly, disguised, undisguised, structured, unstructured, manually, or mechanically (Glos *et al.*, 1980:128; Burns and Bush, 2003:208). More specifically, EWOM review communication was collected by an automatic computer programme that was written to capture data from the two selected e-commerce platforms in a given time period. A more in-depth discussion of data collection, process and analysis can be found in section 4.2.6.

4.2.5 Step 5: Conduct the sampling procedure

Zikmund and Babin (2010:412) define a population as any complete collection of units or individuals that possess a common trait. These units of analysis are usually derived from the research question itself (Blumberg *et al.*, 2008:224). The target population that was investigated for the present study were narrowed down to the review communication about 100 new best-seller books on e-commerce platforms *Amazon.com* and *Barnesandnoble.com*.

4.2.5.1 Sample size

A sample size is a list of elements drawn from a population to equally represent the target population (Kotler and Keller, 2006:110-111). According to Zikmund and Babin (2010:412), an element list could consist of people, geographic areas, institutions, sales, or other units. In the present study, the sampling procedure did not take place in the traditional way, but instead the sample size was guided by the period of analysis. During the four-month period of data collection, the final sample size that was analysed for model 1 reached 14819 observations, while 9529 observations were analysed for model 2.

4.2.5.2 Sample selection process

According to Blumberg *et al.* (2008:235), the units of a sample are selected on either a probability- or non-probability basis. Probability samples are selected on the basis that each unit of the population has a known, non-zero chance of being chosen, while non-probability samples are chosen on a subjective basis and have an unknown non-zero chance of being selected (Blumberg *et al.*, 2008:235). Non-probability samples can further be divided into four main types: convenience, judgement, quota and snowball samples (Blumberg *et al.*, 2008:235; Zikmund and Babin, 2010:432).

For the purpose of collecting EWOM review samples on e-commerce platforms, a non-probability judgement sampling procedure was used. Judgement samples are hand-picked samples to purposefully address the stated research questions (Zikmund and Babin, 2010:432). According to Blumberg *et al.* (2008:253), sample units are chosen to conform to certain criteria. The three criteria on which the observation's samples were chosen in this study, were e-commerce platforms, books, and time period.

Amazon.com and *Barnesandnoble.com* were not chosen only because they are the biggest book e-commerce platforms in the world, but also because they invested substantially in reviewing systems that enabled them to cultivate what are now two of the most active book reviewing communities online (Chevalier and Mayzlin, 2004; Chen *et al.*, 2008:4). *Amazon.com*'s 100 new best-sellers' book-list on 2010/09/21 at 1:28pm was used as the benchmark book-list, from where the same new best-sellers were chosen from the *Barnesandnoble.com* platform. Secondly, the only product category that was focused on for this study was books. Book review pages allow reviewers to rate books and share their opinion and sentiment using the review's text. More specifically, only new best-seller books

were analysed, because EWOM review communication and sales rank values could be tracked from the beginning of the book's life-cycle until the end of the data collection period. The only requirement for new best-seller books to be included in the study was that a book needed to elicit at least one written review in the specified data collection period. Lastly, the time period when data collection took place was chosen to fall between 1 October 2010 and 31 January 2011. This time period was strategically chosen, because it represented the four busiest online browsing and shopping months of the year, during which many e-commerce users purchase and return Christmas gifts.

When undertaking a research study, it is essential to indicate not only the reliability of the variables and measurement units, but also to assess the validity of the research design (Struwig and Stread, 2001:130). To ensure that the findings from the study were trustworthy, the reliability of the methods were tested, and will be discussed when the data collection, processing and analysis sections are dealt with. According to Blumberg *et al.* (2008:403), "validity" is defined as the accuracy with which the research's measuring tool truthfully measures what it sets out to measure. More specifically, the external validity, which measures if a causal relationship can be generalised across the dataset, was fairly high owing to the real-world public data that were captured from *Amazon.com* and *Barnesandnoble.com* (Blumberg *et al.*, 2008:403).

4.2.6 Step 6: Collect, process and analyse the data

4.2.6.1 Data collection

100 new best-seller books from *Amazon.com* (see Annexure A: Table 4.2) were chosen as the benchmark population; the same 100 titles were selected from *Barnesandnoble.com* according to their ISBN numbers (unique ID). For example, the number 1 best-seller on the first day of data collection, *Freedom - The Novel*, has a unique ID of ISBN-13 978-0312576462 in both *Amazon.com* and *Barnesandnoble.com*.

Even though e-commerce platforms *Amazon.com* and *Barnesandnoble.com* offer essentially the same review experience for their individual users, each platform's review community reflects its own unique review patterns as a product's life-cycle evolves over time (Chevalier and Mayzlin, 2004:11). According to Chen *et al.* (2008:12-13), *Amazon.com* highlights its most helpful reviews as spotlight reviews, and prioritises them before others on main product landing pages. Together with spotlight reviews, most recent reviews are also allocated on

both *Amazon.com* and *Barnesandnoble.com* (see Annexure B: Figure 4.5 - 4.7) and are updated as new reviews are casted by online users (Chen *et al.*, 2008:12-13). As new reviews are created, older reviews are backlogged, and can only be accessed by navigating to each individual review. At the time of data collection, customer reviews at *Barnesandnoble.com* were allocated on the main landing product page, while with *Amazon.com*, the user would first have to click through from the main landing product page and navigate to a new page that shows all the customer reviews (see Annexure B: Figure 4.8 - 4.9).

Collected review and sales rank data were categorised into two different types: 1) data that were collected for model 1 from backlog product review pages, and 2) data that were collected for model 2 from main product landing pages as well as backlog review pages. Model 1's data more specifically consisted of ISBN, book title, platform, review posted timestamp, review collection timestamp, found helpful degree, star rating and review text. Model 2's data consisted of ISBN, book title, platform, review posted timestamp, review collection timestamp, number of customer reviews, star ratings, most helpful star ratings, most recent star ratings, found helpful degree and sales rank. All spotlight review and sales rank data were automatically collected on a daily basis by means of an automated web crawler, while all the backlog customer review data were manually imported into Excel sheets only at the end of the data collection period.

According to Liu (2011:311), the Internet's information consists of billions of dynamic web pages that are being served by literally millions of web servers² globally. It is these web pages that an automated web crawler visits, and from which it collects information to be analysed. Applied to this research study, a GNU wget crawler, formerly known as *Geturl*, was used to automatically retrieve file content from *Amazon.com* and *Barnesandnoble.com*'s web servers (Free Software Foundation Inc., 2011). Another software programme called *cron*, which allows its users to automatically execute a given script at any given time/date, was created to schedule the wget crawler to collect data four times a day for all 100 books between 1 October 2010 and 31 January 2011 (The Open Group, 2008).

2) A web server is the remote machine on which data are stored (Liu, 2008:1).

The reliability of each programme was first tested before it was implemented to start collecting data on 1 October 2010. Each *Amazon.com* and *Barnesandnoble.com* book's unique ID was included into the crawler so as to identify the unique path³ for each individual book's data extraction (see Annexure C: Table 4.3 and 4.4). Each downloaded web page was saved under a unique file name according to the following format: ProductID.datetimestamp.html (e.g. 1934781703.20110315-1206.html).

The total number of web pages scraped from *Amazon.com* was 67073, while 66513 web pages were scraped from *Barnesandnoble.com*. All web pages were downloaded to a research web server and later processed on a local personal computer (see Annexure B: Figure 4.10).

It was the study's objective not only to include data from main product landing pages, but also to aggregate all published EWOM reviews on backlog product review pages (see Annexure B: Figure 4.11 - 4.12). Initially Microsoft Excel's web page import function was used for data collection, but owing to slow and poor performance of this approach, it was decided to switch to manually importing all backlog product review pages directly into the designated Excel files.

Even though the backlog EWOM reviews were posted between 1 October 2010 and 31 January 2011, they were only imported after 1 February 2011. Data from each import were validated by its timestamps on the e-commerce platform. If the dates on the Excel sheet corresponded with the review dates that were posted on the e-commerce platform, the next import was undertaken.

It is worth mentioning that on certain *Barnesandnoble.com* books' review pages some individual reviews were incorrectly chronologically ordered according to their individual timestamps. It was assumed that the timestamps were correct, but needed to be ordered chronologically. This step ensured that all the collected data remained valid and reliable for further data processing.

3) The url or the universal resource locator indicates the unique address of a page on the web (Liu, 2008:3).

4.2.6.2 Data processing

After data collection had been completed, data processing can commence. According to Bruijs *et al.* (1998:525), it is necessary that collected data be “edited, verified and coded to facilitate the data capture on computer”. In order to clean and code data correctly, it is essential to identify the specific dependent and independent variables and measurement units. Data processing was divided into two sections to identify firstly, the variables and measures for model 1 and model 2, and secondly, to indicate the manner in which the data of each model were edited, verified and encoded. Altogether 15 book titles were discarded from the original target of 100 new best-seller book titles on *Amazon.com* and *Barnesandnoble.com*. The discarded list included 14 book titles (see Annexure A: Table 4.2, far right column) that had no reviews on *Barnesandnoble.com* in the allocated four-month period, and one book called *Freedom* (not to be confused with number 1 on the list, *Freedom - The Oprah book club*). These latter two books are essentially the same book; the only difference is that the one has an Oprah book club validation, while the other one does not.

4.2.6.2.1 Model 1: Processing of backlog product review data

Model 1 was constructed to assess the relationship between review credibility (as measured by found helpful degree) and EWOM factors on *Amazon.com* and *Barnesandnoble.com*, and ultimately address the first seven research questions. Altogether a total of 14819 *Amazon.com* and *Barnesandnoble.com* EWOM reviews (see Annexure B: Figure 4.11 - 4.12) were published by reviewers in the four-month period. The only significant difference at the time of data collection, was that *Barnesandnoble.com* found helpful statistics were shown at the bottom of every review, instead of at the top of every *Amazon.com* review. Model 1's dependent variable, independent variables and their specific measuring units are listed in Table 4.5.

TABLE 4.5: THE RELATIONSHIP BETWEEN REVIEW CREDIBILITY AND EWOM REVIEW FACTORS (MODEL 1)

VARIABLE	MEASURE
Dependent variable:	
REVIEW CREDIBILITY	“Found helpful degree”, or the total amount of people who found the review helpful divided by the total amount of people who voted
Independent variables:	
STARS	1 to 5 stars
EX-STARS	Distance in stars from a 3-point star rating
SENTIMENT	Positive sentiment rating
EX-SENTIMENT	Distance from 0.5 average sentiment rating
TEXT LENGTH	Word count per review
TIME	Elapsed time (collection date - publish date)
PLATFORM	<i>Amazon.com</i> or <i>Barnesandnoble.com</i>

The dependent variable named REVIEW CREDIBILITY was measured by the total percentage of readers who found the review helpful on the day of data collection. The value of the STARS variable ranged from 1 to 5, and represented the number of stars the reviewer gave the book. The EX-STARS variable was based on the contention that a review might make more of an impression (and therefore be more likely to be found helpful) if the rating is more extreme. In other words, a review of 1 star or 5 stars is more likely to be read (and found helpful) than one with an average of 3 stars. The distance (in stars) from a star rating of 3 is therefore calculated as: $EX-STARS = abs(3 - STARS)$, where *abs* is the absolute value function. The SENTIMENT variable was used to measure the review text’s positive sentiment, and assigned a quantitative score between 0.51 – 1.0. Following the same reasoning as the above-mentioned EX-STARS, it is of interest whether a strong opinion is likely to make more of an impression and be found more helpful, therefore EX-SENTIMENT variable is defined as: $EX-SENT = abs(0.5 - SENTIMENT)$. The TEXT LENGTH variable provides an indication of how many words each review consisted of. The TIME variable was calculated as the difference between publish date (1 October 2010 until 31 January 2011) and collection date (1 February 2011 until 28 March 2011); yielding the number of days the review was available to be judged helpful or not helpful by the users of the two e-commerce

platforms. Lastly, the Boolean variable PLATFORM assigned a value of 0 for reviews on *Barnesandnoble.com*, and a value of 1 for reviews on *Amazon.com*.

The first step in model 1's data processing dealt with categorising review data into specified columns on individual Excel sheets using Excel's filtering algorithms (see Annexure B: Figure 4.13 - 4.18). Based on a keyword search that identifies certain text characters (e.g. "found helpful"-text string), either a 0 or a 1 was allocated next to the latter cells (see Annexure B: Figure 4.19). The next step in the process was to use Excel's sort and filter-function to temporarily filter out all the cells that were equal to 1, and categorise the needed review data below one another. *Amazon.com* book's review data were sorted into three preliminary columns: found helpful, star rating and publish date, and lastly the review text (see Table 4.6). The 85 *Barnesandnoble.com* books' review data were further sorted into three preliminary columns, but were still in a different format. The three columns were: found helpful, star rating, and lastly, reviews' publish date and review text (see Annexure B: Table 4.7).

TABLE 4.6: AMAZON.COM FOUND HELPFUL, STAR RATING, PUBLISH DATE AND REVIEW TEXT (MODEL 1)

	A	B	C
1	Found Helpful	Star rating and publish date	Review text
2	1 of 1 people found the following review helpfu	4.0 out of 5 stars Great but Short, January 30, 2011	A great short high-level discussion of western physical sc
3	2 of 2 people found the following review helpfu	3.0 out of 5 stars The Title Is Somehow Big for Its T	I am a Hawking fan ever since I have so eagerly read and l
4	3 of 6 people found the following review helpfu	2.0 out of 5 stars Disconnect, January 25, 2011	It is hard to know who does the writing, to avoid the feel
5	14 of 23 people found the following review help	1.0 out of 5 stars Surprisingly facile and ill-informe	Sure, these are two very knowledgeable scientists and th
6	2 of 2 people found the following review helpfu	4.0 out of 5 stars briefer that a brief history of tim	Stephen Hawking is the Einstein of his generation and thi

In the second step, book review data were further processed by refining the data categories into new Excel sheet columns. Review text lengths were also calculated with Excel's count function, and represent the number of words per review (see Annexure B: Figure 4.20). *Amazon.com and Barnesandnoble.com's* new column names included published date, found helpful, star rating, review text, and review text length (see Table 4.8 and also Annexure B: Table 4.9).

TABLE 4.8: AMAZON.COM PUBLISH DATE, FOUND HELPFUL, STAR RATING, REVIEW TEXT AND TEXT LENGTH (MODEL 1)

	A	B	C	D	E
1	publish date	FH	stars	review text	text length
2	January 30, 2011	1 of 1 people found	4.0 out of 5 stars	A great short high-level discussio	108
3	January 26, 2011	2 of 2 people found	3.0 out of 5 stars	I am a Hawking fan ever since I ha	208
4	January 25, 2011	3 of 6 people found	2.0 out of 5 stars	It is hard to know who does the v	78
5	January 23, 2011	14 of 23 people foun	1.0 out of 5 stars	Sure, these are two very knowled	306
6	January 22, 2011	2 of 2 people found	4.0 out of 5 stars	Stephen Hawking is the Einstein	232

In the third step, book review texts were processed to obtain a text sentiment score for each of the *Amazon.com* and *Barnesandnoble.com* EWOM reviews. *Uclassify*, an open-source sentiment classifier⁴, was used to classify review text by giving each review a sentiment score between extremely negative 0.0 and extremely positive 1.0 (Kågström, Karlsson and Kågström, 2011). In order to classify each review’s text, each of the 85 *Amazon.com* and *Barnesandnoble.com* Excel files first had to be converted to comma separated values (csv-format). The next step in the process was to develop an algorithm in *Python*⁵ (see Annexure C: Table 4.10) that would automatically run through every book’s csv file and connect to *Uclassify*’s servers through their *API*⁶ to classify all the review text. After each Excel file was analysed, each review-cell’s positive and negative sentiment score was aligned with its original *Amazon.com* and *Barnesandnoble.com* review (see Table 4.11 and also Annexure B: Table 4.12).

TABLE 4.11: AMAZON.COM PUBLISH DATE, FOUND HELPFUL, STAR RATING, REVIEW TEXT, TEXT LENGTH, POSITIVE AND NEGATIVE SENTIMENT (MODEL 1)

	A	B	C	D	E	F	G
1	publish date	FH	stars	review text	text length	pos sent	neg sent
2	January 30, 2011	1 of 1 people found the following review helpful:	4.0 out of 5 stars	A great short high-level disc	108	0.153616	0.846384
3	January 26, 2011	2 of 2 people found the following review helpful:	3.0 out of 5 stars	I am a Hawking fan ever sin	208	0.522935	0.477065
4	January 25, 2011	3 of 6 people found the following review helpful:	2.0 out of 5 stars	It is hard to know who does	78	0.644089	0.355911
5	January 23, 2011	14 of 23 people found the following review helpful:	1.0 out of 5 stars	Sure, these are two very kn	306	0.574037	0.425963
6	January 22, 2011	2 of 2 people found the following review helpful:	4.0 out of 5 stars	Stephen Hawking is the Ein	232	0.243299	0.756701

4) It is also referred to as “classification, which aims to learn a classification function from data that are labeled with pre-defined classes or categories” (Liu, 2011:8).

5) Python is a dynamic programming language for virtually any computational task (Parkin, 2012).

6) By means of an API or application programming interface third party developers are able to build web services on top of current software architecture(s) (Sangiovanni-Vincentelli, Carloni, De Bernardinis and Sgroi, 2004:410).

During the fourth step the review publish and collection dates were categorised and the found helpful degree for each review calculated by dividing the number of readers who found the review helpful, by the total number of readers who voted. All the reviews were screened one more time to ensure that only the reviews that were published in the allocated time period would be included in the dataset. The final review dataset for model 1 consisted of 11520 *Amazon.com* reviews and 3299 *Barnesandnoble.com* reviews with the following column names: platform name, book title, review publish date, found helpful collection date, found helpful degree, stars rating, review text length, negative sentiment, and positive sentiment (see Table 4.13).

TABLE 4.13: AMAZON.COM AND BARNESANDNOBLE.COM BOOK TITLE, PUBLISH DATE, COLLECTION DATE, FOUND HELPFUL, STAR RATING, TEXT LENGTH AND SENTIMENT (MODEL 1)

	A	B	C	D	E	F	G	H	I
11514			Date Review	Date FH	Found Helpful		Text	Neg	Pos
11515	Platform	Book title	was Posted	was Collected	Value	Stars Rating	Length	Sentiment	Sentiment
11516	Amazon	Knuffle Bunny Free: An Unexpected Di	2010/12/08	2011/03/30	0.67	3	117	0.72	0.28
11517	Amazon	Knuffle Bunny Free: An Unexpected Di	2010/12/18	2011/03/30	0.00	5	104	0.86	0.14
11518	Amazon	Knuffle Bunny Free: An Unexpected Di	2010/12/30	2011/03/30	0.00	5	43	0.95	0.05
11519	Amazon	Knuffle Bunny Free: An Unexpected Di	2010/12/31	2011/03/30	0.00	5	76	0.91	0.09
11520	Amazon	Knuffle Bunny Free: An Unexpected Di	2011/01/01	2011/03/30	0.00	5	38	0.06	0.94

During the fifth and final step, a book index was incorporated into the model to facilitate data analysis. Book names were validated and checked for any discrepancies such as alternative names, spaces and spelling (see Annexure B: Table 4.14). The elapsed time (measured in days) was calculated by subtracting the publish date from the collection date (see Annexure B: Table 4.15). Dates were number-formatted in Excel (e.g. 2010/10/01 becomes 40452) in order to be analysed in the statistical software package (MATLAB). Two final columns were added to the left of the already existing columns which represented the EX-SENTIMENT and EX-STARS-variables (see Annexure B: Table 4.16).

The accuracy of the data processing was tested by going through each review to make sure that no double copies of reviews existed. Any reviews that were incorrectly categorised by the Excel filtering algorithms could be identified when the reviews were aligned below one another and their timestamps did not match. If an error did occur, the appropriate steps were taken to go back to the original dataset to correct the problem. Lastly, the accuracy of the sentiment scores was tested by comparing the review's star rating to the sentiment score, as well as by aligning the calculated sentiment score next to the original review text column. If

this approach did not work, the sentiment programme was run again to determine where the error occurred, so that it could be corrected.

4.2.6.2.2 Model 2: Processing of product landing page data

Model 2 was constructed to address the eighth research question and ultimately assess whether EWOM influences the sale rank levels of new best-seller books. Even though data collection took place on both e-commerce platforms, it was decided that model 2 (see Table 4.17) would focus exclusively on *Amazon.com*, because there were not sufficient data from *Barnesandnoble.com* to do a meaningful data analysis (see Annexure B: Figure 4.21 - 4.23).

TABLE 4.17: THE RELATIONSHIP BETWEEN AMAZON.COM SALES RANK AND EWOM STAR RATING FACTORS (MODEL 2)

VARIABLE	MEASURE
Dependent variable:	
SALES RANK	Log sales rank (automatic scrapes)
Independent variables:	
AV: Average book star rating	Stars rating between 1 and 5 (automatic scrapes and manual data combined)
MH: Most helpful review's stars	Stars rating between 1 and 5 (automatic scrapes)
MR: Most recent review's stars	Stars rating between 1 and 5 (automatic scrapes and manual data combined)
EWOM star ratings = Average (AV + MH + MR)	

After model 2 was constructed, the collected data were categorised into four general types to be processed, namely books' sales rank, stars from most helpful reviews, average book stars, and lastly stars from most recent reviews.

Instead of showcasing the actual sales number, *Amazon.com* provides sales rank data on main product landing pages (Hu, Liu and Zhang, 2008:207-208). A sales rank is calculated by *Amazon.com* in descending order (see Table 4.18) and is therefore negatively related to product sales (Hu *et al.*, 2008:207-208). For example, the best-selling book will reflect a sales rank of 1 and a book that has had fewer sales will be reflected by a higher sales rank

number. Owing to the large differences in sales ranks, the log-transformation of the sales rank is generally used as proxy for product sales (Chen *et al.*, 2004:716).

TABLE 4.18: ESTIMATED SALES RANK UPDATE FREQUENCY

Platform	Sales Rank	Update-frequency
Amazon.com	1 – 10 000	Once per hour
	10 001 - 100 000	Once per day
	> 100 001	Once per month
Barnesandnoble.com	-----	All rankings are updated on a daily basis

Source: Adapted from Chevalier and Goolsbee (2003:5-6)

The 85 books from *Amazon.com*'s sales ranks were extracted from originally scraped web pages by means of another Python software programme (see Annexure C: Table 4.19) and imported into a separate Excel sheet (Richardson, 2011). Data on the Excel sheet were further categorised into five columns, namely platform, date collected (date the sales rank data were scraped from the web), time collected (specific time of the day the sales rank data were scraped from the web), book title, and sales rank value (see Table 4.20).

TABLE 4.20: AMAZON.COM SALES RANK DATA (MODEL 2)

	A	B	C	D	E
1	Platform	Date Collected	Time Collected	Book title	Sales Rank
2	Amazon	2010/10/01	00:08	I Shall Wear Midnight	80.00
3	Amazon	2010/10/01	06:07	I Shall Wear Midnight	72.00
4	Amazon	2010/10/01	12:08	I Shall Wear Midnight	68.00
5	Amazon	2010/10/01	18:08	I Shall Wear Midnight	70.00
6	Amazon	2010/10/02	00:09	I Shall Wear Midnight	67.00

Further data processing was conducted on the *Amazon.com* Excel sheet. The platform column was deleted, since only *Amazon.com* was used. To enable further analysis in MATLAB, dates had to be changed to Excel's number format. After the range of the timestamps was checked for accuracy, the time column was deleted, because it was no longer needed for further analysis. The book title index was again incorporated into model 2, ranging from 1 to 85 to facilitate analysis.

The sales rank column was also checked for any discrepancies, and it was found that all values were positive integers ranging from 1 to 253529. MATLAB programmes (see

Annexure C: Table 4.21) were written to calculate the following: natural logarithm of the sales ranks (LSR), the average daily LSR, and the seven-day moving average of the average daily LSR (MA7).

The reason for using the average over a week is that the day-to-day change of sales rank values revealed excessive fluctuation. In the analysis it was required to gauge the effect of some independent variables on the sales rank. In order to determine the change in the sales rank on day X, the MA7 was calculated up to day (X-1) as well as starting on day (X + 1). A portion of the resulting data matrix is displayed in Table 4.22.

TABLE 4.22: BOOK, DATE, PREVIOUS AND FOLLOWING WEEK SALES RANK DATA (MODEL 2)

	A	B	C	D
1	Book	Date	Previous Week	Following week
2	1	40456	6.8638	7.6409
3	1	40457	6.9369	7.7664
4	1	40458	7.0110	7.7772
5	1	40459	7.0872	7.7919
6	1	40460	7.1740	7.7866

The first line of the Table 4.22 can be interpreted as follows: for book 1 (first alphabetically of the 85 books), on short date 40456 (corresponding to 5 October 2010) the MA7 of the previous week was 6.8638 (the average of the average daily LSR over short dates 40449 to 40455). The following week's MA7 was 7.6409 (the average of the average daily LSR over short dates 40457 to 40463)⁷. Sales rank data were available from short date 40449 until 40455, but no previous week value could be calculated for it. Fortunately this only occurred at dates in the beginning of the series since, collecting the sales rank data continued even after the observation of the other variables (e.g. average star rating) ceased.

The stars' data from spotlight most found helpful reviews were also processed in the same way as the above-mentioned sales rank values. All the data that were extracted from the web pages by means of a second Python programme (see Annexure C: Table 4.23), were imported into separate Excel sheets. After the platform column and the time column were deleted, the Excel sheet was categorised into five columns, namely book title, date collected (date the found helpful data were scraped) and spotlight most helpful star ratings (FH1, FH2 and FH3) (see Table 4.24). The columns FH1, FH2 and FH3 refer to the number of stars of

⁷) Note that, since a moving average is used, there is some loss in information.

the three most helpful reviews that were given to the book. The dataset contained four observations per day.

TABLE 4.24: AMAZON.COM SPOTLIGHT MOST HELPFUL STAR RATINGS (MODEL 2)

	A	B	C	D	E
1	Book	Date	FH1	FH2	FH3
2	1	2010/10/01	5	5	5
3	1	2010/10/01	5	5	5
4	1	2010/10/01	5	5	5
5	1	2010/10/01	5	5	5
6	1	2010/10/02	5	5	5
7	1	2010/10/02	5	5	5
8	1	2010/10/02	5	5	5
9	1	2010/10/02	5	5	5

Only the first row of each date was retained since the three most helpful reviews changed relatively seldom. Next, the date was number-formatted, and the average of the final three columns was calculated to obtain the daily average number of stars for the three most helpful reviews, which served as inputs for further analysis of model 2 (see Table 4.25).

TABLE 4.25: AMAZON.COM DAILY AVERAGE NUMBER OF STARS FOR THE THREE MOST HELPFUL REVIEWS (MODEL 2)

	A	B	C	D	E	F
1	Book	Date	FH1	FH2	FH3	Average FH
2	1	40452	5	5	5	5
3	1	40453	5	5	5	5
4	1	40454	5	5	5	5
5	1	40455	5	5	5	5
6	1	40456	5	5	5	5

The original goal was to program an automatic extraction method to collect not only the sales rank and most found helpful review star data from main landing web pages (as above-mentioned), but also to extract *average star ratings* and *most recent star ratings* from the same scraped web pages (see Annexure B: Figure 4.22 - 4.23). As the data scraping period commenced, the structure of the web page⁸ was frequently changed by *Amazon.com*, which continually rendered the scraping programmes insufficient for the specific data extraction.

8) More specifically, the HTML-tags, in which the data occurs, were changed too frequently.

Instead of discarding these two variables from the model, proxy values were imputed from the already obtained automatic and manual data collection process. The proxy values for the very first day (1 October 2010) had to be calculated first, and only then could the proxy values for the rest of the data collection period be imputed by means of several MATLAB programmes (see Annexure C: Table 4.26).

Proxy values on the first day

Specific data were extracted from each of the 85 *Amazon.com* books' main landing web pages, which were scraped on 1 October 2010⁹. After the extraction process all the values were imported into a single Excel sheet. The Excel sheet (see Annexure B: Table 4.27) was categorised into 23 columns, consisting of: book number, date of first scrape, time of first scrape, number of reviews, average stars, number of 5 stars, number of 4 stars, number of 3 stars, number of 2 stars, number of 1 stars, number of stars at each of the three most helpful reviews (MH stars) and number of stars at each of the ten most recent reviews (MR stars). For ease of explanation, Table 4.27 has been divided into two sections to explain how firstly, the *average star ratings* and secondly, the *most recent star ratings* were imputed.

Imputed average star ratings for the entire data collection period

From the original average star rating data (see Table 4.28), the dates were number-formatted and the valid range checked. The number of reviews column was checked to correspond to the sum of all 1-to-5 star ratings. The average stars were also checked, to correspond with the following formula:

$$[5*(\text{number of 5 stars}) + 4*(\text{number of 4 stars}) + 3*(\text{number of 3 stars}) + 2*(\text{number of 2 stars}) + 1*(\text{number of 1 stars})]/\text{Number of Reviews}.$$

As both these latter checks yielded no inconsistencies, columns 1-to-5 stars were deleted, and the number of reviews and the average stars columns were ready to be used for the next step in the data analysis process.

9) If no reviews were posted on 1 October, the web page of the very first day when a review was posted, was used.

TABLE 4.28: AMAZON.COM AVERAGE STAR RATING DATA (MODEL 2)

	A	B	C	D	E	F	G	H	I	J
1	Book	Date of 1st scrape	Time	Number of Reviews	Average Stars	# 5 Stars	# 4 Stars	# 3 Stars	# 2 Stars	# 1 Stars
2	1	2010/10/01	00:10	10	5	10	0	0	0	0
3	2	2010/10/01	00:02	17	4	11	1	2	0	3
4	3	2010/10/01	00:10	15	4.1	9	2	2	1	1
5	5	2010/10/01	00:08	61	3.9	26	16	10	7	2
6	6	2010/10/01	00:05	6	4.7	5	0	1	0	0

To calculate the average stars for the entire four-month period, daily star ratings were needed, and because the study already determined those values for model 1 (see section 4.2.6.2.1: Table 4.13), they were again used here. Unnecessary columns were deleted, dates were number-formatted and book names indexed (see Table 4.29).

TABLE 4.29: AMAZON.COM STAR RATING DATA (ADAPTED FROM MODEL 1)

	A	B	C
1	Book	Date posted	Stars
2	1	40454	5
3	1	40469	2
4	1	40469	5
5	1	40469	1
6	1	40472	5

Data from Table 4.29 were then used to update the average stars rating each day, from the first scrape (see Table 4.28) till the end of the study (see Table 4.30). For instance, book 1’s first scrape was on 40452 (1 October 2010), by which time it had ten reviews with an average of five stars and a total of 50 stars (e.g. 10 reviews * average stars of 5).

TABLE 4.30: AMAZON.COM AVERAGE STAR RATING CALCULATION (MODEL 2)

Book	Date posted	Total stars	# Reviews	Av Stars
1	40452	50	10	5
1	40453	50	10	5
1	40454	55	11	5
1	40455	55	11	5
:	:	:	:	:
1	40468	55	11	5
1	40469	63	14	4.5000
1	40470	63	14	4.5000
1	40471	63	14	4.5000
1	40472	68	15	4.5333

Imputed most recent star ratings for the entire data collection period

Additional columns from Table 4.27 (Annexure B) were used to calculate *most recent reviews' star ratings* over the four-month period (see Table 4.31).

TABLE 4.31: AMAZON.COM MOST RECENT REVIEWS' STAR RATINGS (MODEL 2)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Book	Date	NR	MH1	MH2	MH3	MR1	MR2	MR3	MR4	MR5	MR6	MR7	MR8	MR9	MR10
2	1	40452	10	5	5	5	5	5	5	5	5	5	5	#N/A	#N/A	#N/A
3	2	40452	17	5	5	3	3	4	5	5	1	1	5	5	5	1
4	3	40452	15	5	4	5	5	3	5	5	2	5	5	5	3	4
5	5	40452	61	4	3	5	4	2	5	3	5	4	3	4	5	2
6	6	40452	6	5	5	5	5	5	3	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
7	8	40452	33	5	5	5	1	1	4	5	5	1	4	3	5	1
8	10	40452	43	5	5	5	5	5	5	5	4	5	5	5	5	5
9	11	40452	38	5	5	5	5	5	5	5	5	5	4	5	4	5
10	12	40452	101	5	4	2	5	5	5	3	4	4	5	4	5	5
11	13	40452	177	5	5	5	5	5	5	4	5	1	5	5	5	5
12	15	40452	21	4	3	4	4	4	5	1	5	4	2	3	5	5
13	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:

For books that had ten or more recent reviews for the first scrape on 1 October 2010 (e.g. books 2, 3, 5, 8, 10, 11, 12, 13 and 15 in Table 4.31: column A), the averages of the MR1 to MR10 columns were calculated, to provide the average star ratings for the ten most recent reviews on date 40452 (1 October 2010). From there the same data that were used to update the average star rating in Table 4.29, were used to update the average most recent reviews' star ratings. For instance, suppose two reviews were added on a given day, these two star ratings would be classified as MR1 and MR2. The values previously in columns MR1 to MR8 would then move to MR3 to MR10, and the average of the ten most recent reviews would be calculated.

For books that had fewer than ten reviews on the 1 October 2010 (e.g. books 1 and 6 in Table 4.31: column A), a modification to the procedure had to be made. For instance for book 1, which had only seven MR reviews on the 1 October, zeros were assigned to the not applicable cells (MR8 to MR10). The average was then calculated as the average of all nonzero MR-values.

In some instances where no reviews had yet been posted by 1 October (as indicated by the first scrape on 40452), the average MR could not be calculated in the usual fashion, as there were no non-zero values. In such a case an average of 3 was assigned, since 3 reflects a

neutral star rating (Assigning a value of zero, or using the most helpful ratings when the MR values are all zero, does not make sense. A zero value will cause the MR average star rating to have a negative influence on the total EWOM. Assigning the most helpful rating will double the effect of that review for the initial period).

4.2.6.3 Data analysis

Each EWOM review and sales rank observation (model 1 and model 2) was examined for any errors and eligibility and was assigned a unique code which was in turn used for recording and interpreting the data (Zikmund and Babin, 2010:491). Stellenbosch University's Centre for Statistical Consultation provided assistance with encoding and analysis of the data by using software programmes MATLAB, STATISTICA and R. The data analysis process can further be categorised into two statistical techniques, namely descriptive and inferential analysis.

Descriptive statistics is defined by Elifson, Runyon and Haber (1990:10) as "rules and procedures for presenting information about a variable or the relationship between variables in a more useable and meaningful form". Basic characteristics of model 1 and model 2's data were graphically described using pie charts, histograms, frequency distribution tables, scatter plots and time series analysis (Zikmund and Babin, 2010:516-532).

Inferential statistics are used to test the nature of relationships between variables and infer from the sample of the population whether any statistical significant differences might exist and how big these differences are (Zikmund and Babin, 2010:538-539). According to Cooper and Schindler (2011:454), a statistically significant difference can be described as results that are unlikely to have occurred by chance, but are significantly indicated by the statistical tests conducted on the population's samples. Diamantopoulos and Schlegelmilch (1997:139) suggest that an appropriate significance level (α) is a good indication of the degree of risk that a researcher is willing to accept when rejecting the true null hypothesis. For the purpose of this research study, a significance level of 0.05 was used. The following statistical methods were used to conduct inferential statistics on the data:

- Chi-square test. According to Elifson *et al.* (1990:410-416), the chi-square test (χ^2) "assesses statistical significance particularly for hypotheses about frequencies which are arranged in a frequency or contingency table". In other words, it is a statistical method that can be used to standardise data in a contingency table by equating the

relationships between the actual cell frequencies and the expected cell frequencies (Hair, Anderson, Tatham and Black, 1998:520).

- Correlation. Two correlation measures were used in this study, namely a correlation matrix and a coefficient of determination (R^2). A correlation matrix can be defined as a summary of relationships between all possible variable pairs, and is visually represented by means of a table (Elifson *et al.*, 1990:206-207), while the coefficient of determination (R^2) “gives us the proportion of variance in the dependent variable explained by the entire set of independent variables” (Treiman, 2009:111).
- Regression analysis. Berenson and Levine (1992:605) suggest that the objective of a regression analysis is to develop a statistical model that can be used to forecast the value of a dependent variable based on at least one given independent variable. Simple linear regression is a regression analysis when only one independent variable (X) is used to forecast the dependent variable (Y), while a multiple regression model uses multiple explanatory variables (X_1, X_2, \dots, X_p) to forecast a quantitative dependent variable (Y) (Berenson and Levine, 1992:605-607).
- Residual analysis. Residual analysis is used to assess the validity of the proposed model and examine possible errors that may have been committed against the assumptions of normality, homoscedasticity and independence of error (Berenson and Levine, 1992: 635-636).

Chi-square tests, a correlation matrix, coefficient of determination (R^2), multiple stepwise regression analyses and residual analyses were used to test the first seven hypotheses (H_0^1 - H_0^7), while the eight hypothesis (H_0^8) was tested by applying correlation matrixes and chi-square tests. The results of the tests are reported in Chapter five.

4.3 CONCLUSION

In order to effectively reach a research study's outcome, a systematic process with interrelated business management tasks has been developed by marketing professionals. The present study's focus was particularly on the business and marketing management tasks that were needed to investigate the EWOM review communication in an online retail platform. Two main questions were formulated to measure, firstly, the independence of individual EWOM factors on a review's credibility, and secondly, to assess the association between EWOM review behaviour and sales rank of new best-seller books. In the beginning of the research marketing process, the research problem, research design, research methods, sampling procedure and the collection, processing and analysis of data were

addressed. The final stages of the research process are directed at report preparation, which consists of representing the research findings (descriptive and inferential), addressing the stated research hypotheses, and discussing the study's conclusions and recommendations (Chapters five and six).

CHAPTER 5

RESEARCH RESULTS

5.1 INTRODUCTION

The previous chapter described the research methodology of this study and how it was designed to address the stated objectives. These were to assess the influence of EWOM factors on the credibility of reviews, as well as to test the association between EWOM communication and online sales rank. Two models were proposed to test both the outcomes. Model 1's data were manually collected, while model 2's data consisted of both manual and automatically captured data. MATLAB, STATISTICA, and R software programmes were used to process and prepare the data of both models on Excel sheets, after which descriptive and inferential statistical analysis was done.

The purpose of Chapter five is to explain the data that has been analysed, rather than to interpret and reach conclusions about it (Blumberg *et al.*, 2008:566). The descriptive statistics are explained first, and then the results of the hypothesis tests are presented in the inferential statistics section.

5.2 RESULTS

These research results are divided into two sections, detailing the descriptive and inferential statistics which were used to address the research objectives and hypotheses.

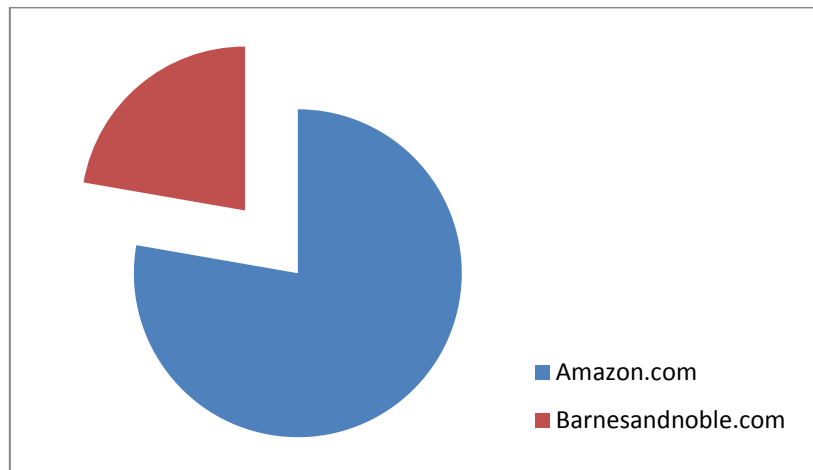
5.2.1 Descriptive statistics

According to Blumberg *et al.* (2008:566), quantitative data can be presented meaningfully with charts, graphics and tables. Pie charts, histograms, frequency distribution tables and scatter plots have been used as descriptive statistics for model 1, and time series analysis has been used for model 2.

5.2.1.1 Model 1's descriptive statistics

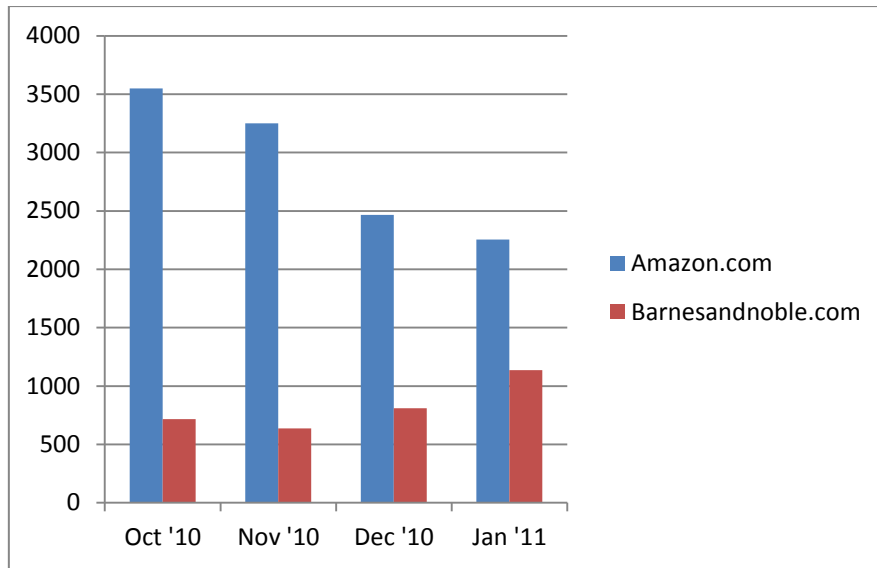
In model 1, a total number of 14819 EWOM review observations were manually recorded from e-commerce platforms' backlog review pages, and processed for statistical analysis (see section 4.2.6.2.1). Over the four-month period, 11520 reviews (78%) were published on *Amazon.com* and 3299 reviews (22%) were published on *Barnesandnoble.com* (see Figure 5.1).

FIGURE 5.1: EWOM REVIEW DATA ACCORDING TO E-COMMERCE PLATFORM



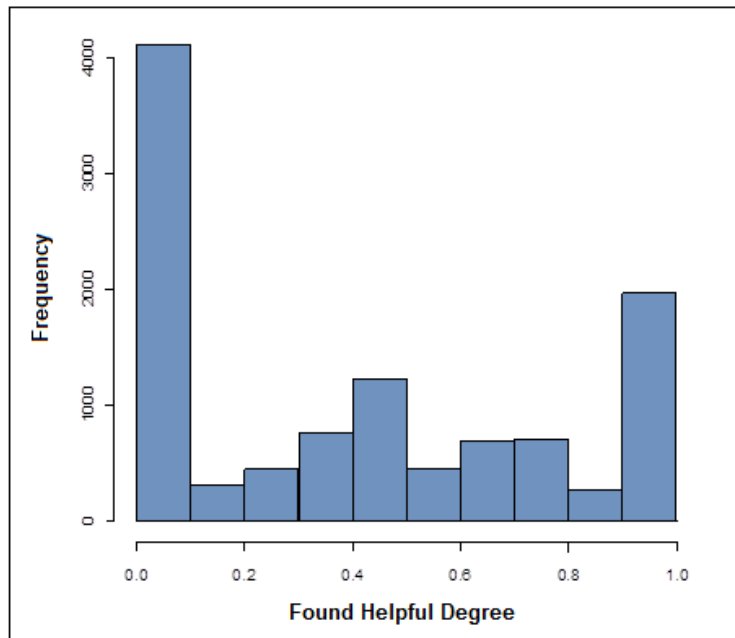
The time dimension of each review was also captured to analyse the frequency of generated review content over the specified four-month period (see Figure 5.2). From the distribution of EWOM reviews per e-commerce platform, it can be seen that the highest number of *Amazon.com* reviews (3548 reviews) were created in the first month (1 – 31 October 2010) and then gradually decreased each month until the end of January 2011 to the number of 2255 reviews. A contrasting pattern can be seen from *Barnesandnoble.com* where 717 reviews were posted in the first month, followed by a slight decrease to 637 reviews, and thereafter increased through December 2010, and ultimately reached the highest level of 1136 reviews by the end of January 2011.

FIGURE 5.2: THE DISTRIBUTION OF EWOM REVIEW DATA OVER TIME



When the found helpful frequency of *Amazon.com* reviews was analysed, the results indicated that most of the EWOM reviews (6197 reviews or 54%) had a found helpful degree of either 0.00 – 0.10 or 0.90 - 1.00 (see Figure 5.3). The interval with the third highest found helpful degree was located between 0.40 - 0.50, which accounted for 1283 reviews (11%).

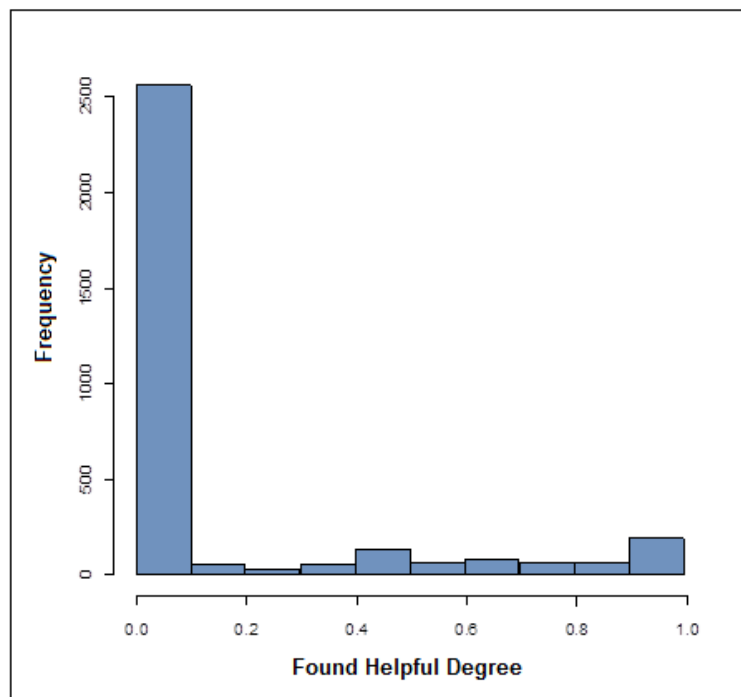
FIGURE 5.3: BI-MODAL DISTRIBUTION OF REVIEW FOUND HELPFUL DEGREE AT AMAZON.COM



Compared to *Amazon.com*, the found helpful frequency of *Barnesandnoble.com* reviews (see Figure 5.4) yielded a completely different graph. Most of the EWOM reviews (2562 reviews or 78%) have a 0.00 – 0.10 found helpful degree. Other than 0.00 – 0.10 interval, the most significant found helpful degree occurred between the following intervals: 0.90 – 1.00 (193 reviews or 6%) and 0.40 - 0.50 (136 reviews or 4%).

For both *Amazon.com* and *Barnesandnoble.com* it should, however, be noted that the 0.00 found helpful degree indicated one of two things: either until the review was collected it was not yet rated, or no-one who had read the review perceived it as helpful.

FIGURE 5.4: BI-MODAL DISTRIBUTION OF REVIEW FOUND HELPFUL DEGREE AT BARNESANDNOBLE.COM

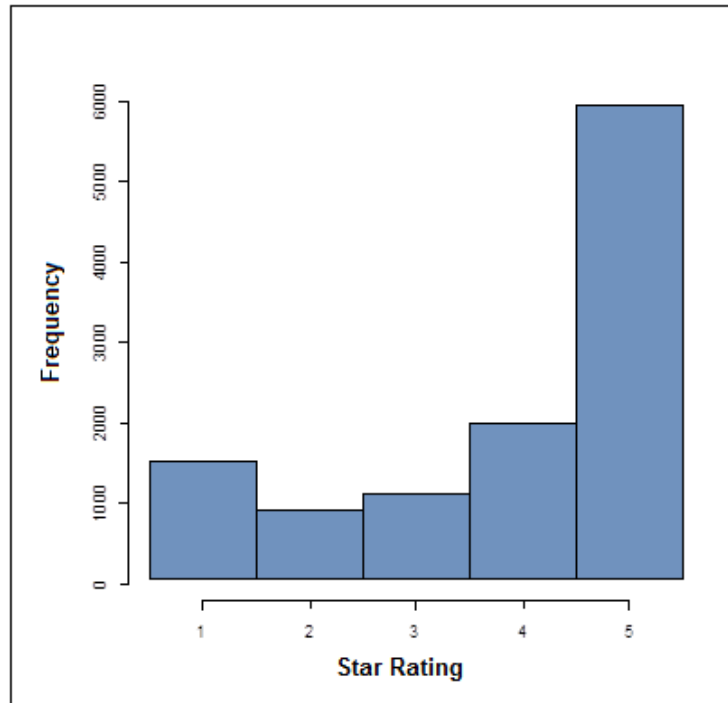


To indicate the sentiment of an online review, the study examined both review star ratings and review text sentiment.

Figure 5.5 and 5.6 indicates that the readers at both e-commerce platforms gave new best-seller books very high star ratings. This extremely positive opinion is indicated by 52% or 5954 *Amazon.com* reviews and 57% or 1881 *Barnesandnoble.com* reviews having a five-star rating. Of the remaining reviews posted at *Amazon.com*, 17% (2004 reviews) were four-

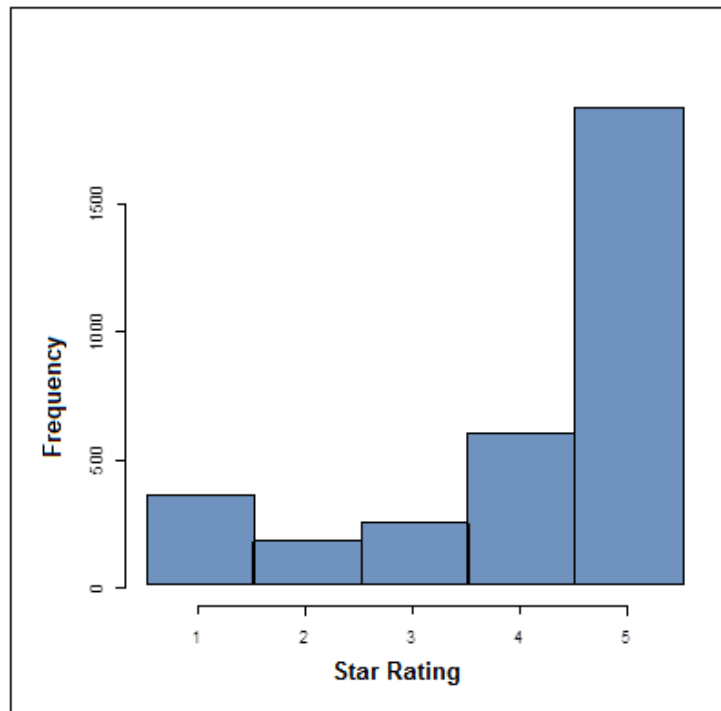
star ratings, 10% (1118 reviews) were three-star ratings, 8% (908 reviews) were two-star ratings, and 13% (1536 reviews) were one-star ratings.

FIGURE 5.5: BI-MODAL DISTRIBUTION OF REVIEW STAR RATINGS AT AMAZON.COM



Barnesandnoble.com reviews indicated more or less the same trend as *Amazon.com*, with the exception of a smaller number of two- and three-star ratings (see Figure 5.6). Of the remaining reviews posted at *Barnesandnoble.com*, 18% (607 reviews) were four-star ratings, 8% (257 reviews) were three-star ratings, 6% (187 reviews) were two-star ratings and 11% (367 reviews) were one-star ratings.

FIGURE 5.6: BI-MODAL DISTRIBUTION OF REVIEW STAR RATINGS AT BARNESANDNOBLE.COM



When the positive sentiment¹⁰ about review text was analysed, the histogram distributions (see Figure 5.7 and 5.8) indicated that review sentiment at *Amazon.com* were more bell-shaped than the sentiment scores at *Barnesandnoble.com*. The standard average or “mean” of the distribution at *Barnesandnoble.com* were higher at a sentiment score of 0.55 than *Amazon.com* (0.53), as well as for the median distribution, which was calculated at a sentiment score of 0.57 for *Barnesandnoble.com* and 0.55 for *Amazon.com*.

10) It must be noted that because sentiment analysis is a relatively new research area of Natural Language Processing, sentiment classification software has not yet been extensively developed and tested. The data from new software programmes such as Uclassify should be retested as sentiment classifiers become more advanced.

FIGURE 5.7: BI-MODAL DISTRIBUTION OF REVIEW TEXT SENTIMENT AT AMAZON.COM

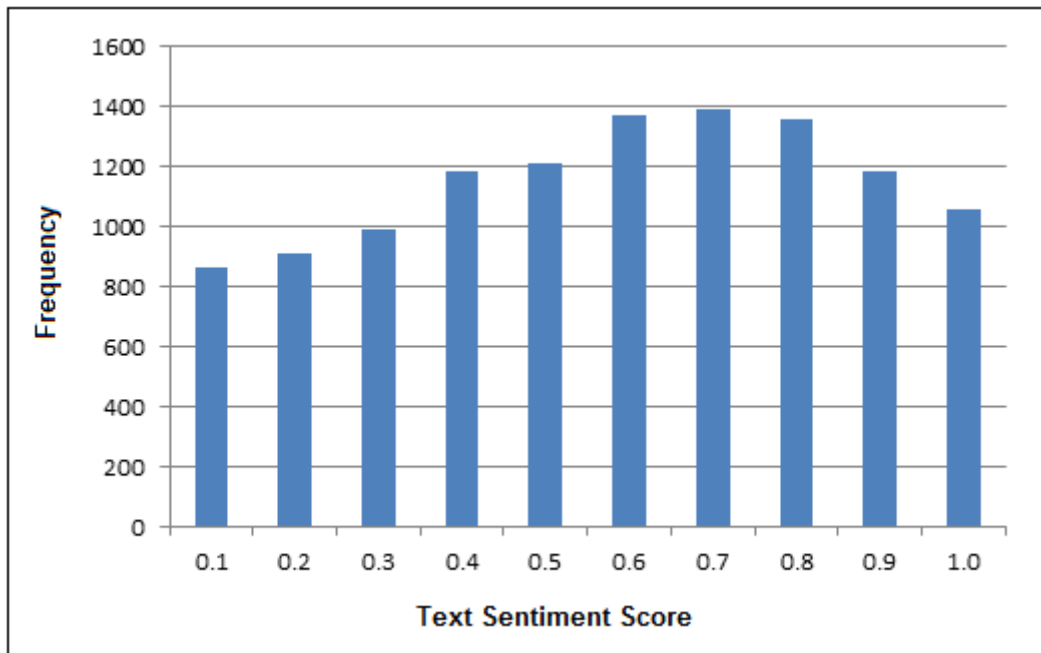
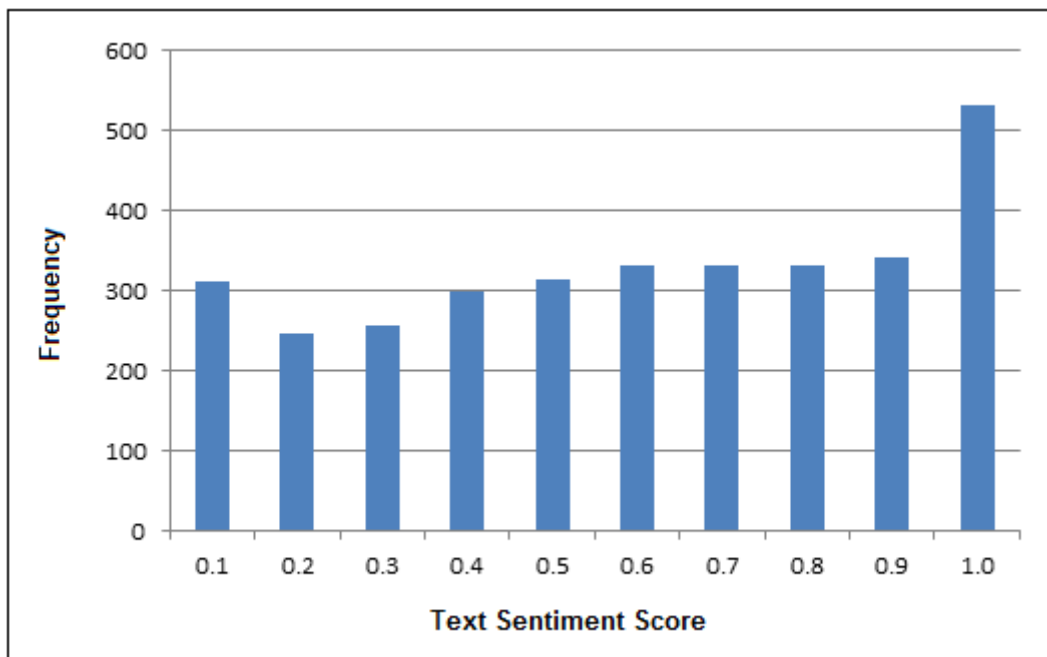


FIGURE 5.8: BI-MODAL DISTRIBUTION OF REVIEW TEXT SENTIMENT AT BARNESANDNOBLE.COM



The last EWOM review factor in model 1 that will be discussed is review text length. Both Figures 5.9 and 5.10 indicate that a high number of reviews had shorter review text lengths, and a smaller number of reviews had longer review text lengths. A higher number of *Amazon.com* reviewers used fewer than 500 words per review, while a higher number of *Barnesandnoble.com* reviewers used fewer than 200 words per review.

FIGURE 5.9: BI-MODAL DISTRIBUTION OF REVIEW TEXT LENGTH AT AMAZON.COM

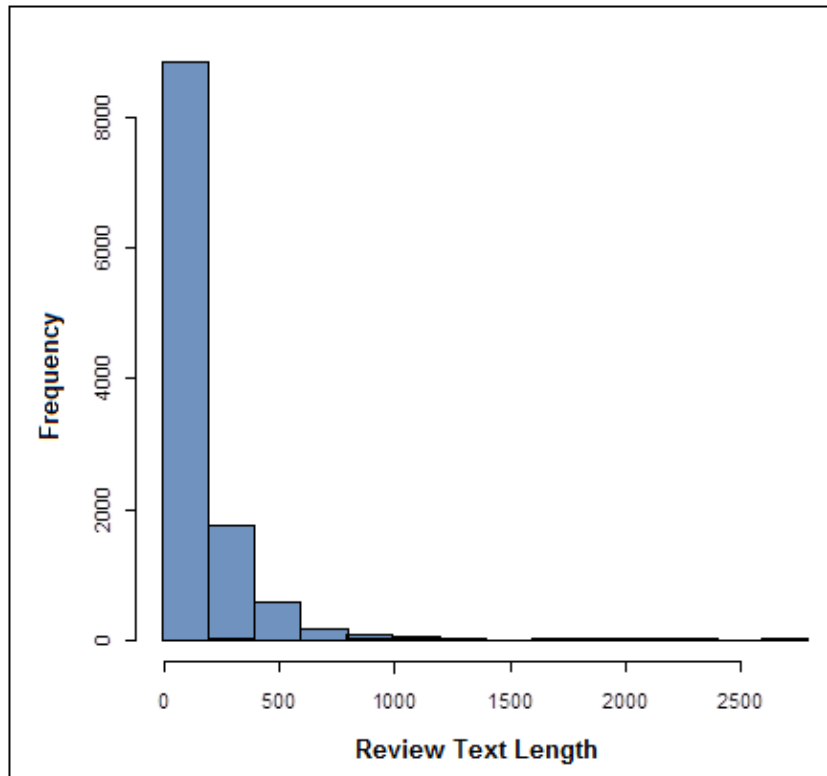
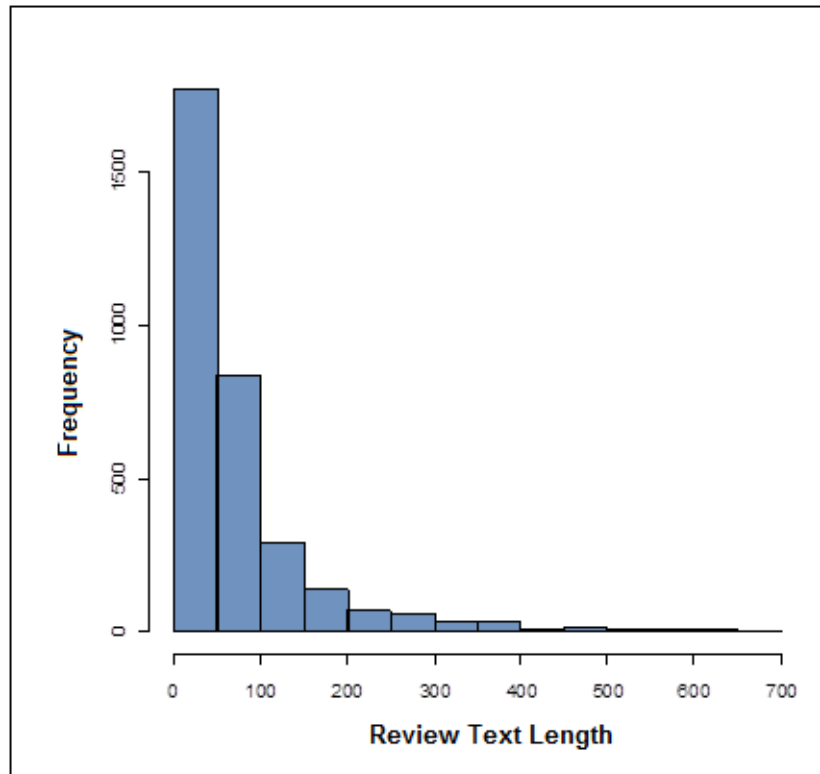


FIGURE 5.10: BI-MODAL DISTRIBUTION OF REVIEW TEXT LENGTH AT BARNESANDNOBLE.COM



Even though the review text lengths of both the e-commerce platforms had similar distributions, upon closer inspection of both platforms' boxplot diagrams (see Annexure B: Figure 5.11 and 5.12) the following differences emerged:

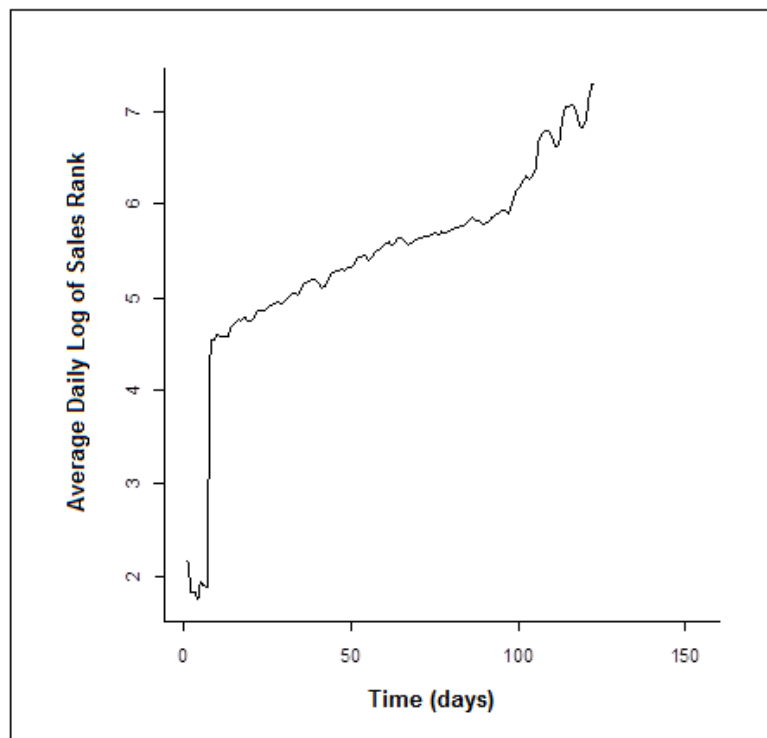
- The smallest observation (q_1) for *Amazon.com* was 12 words, while it was 0 words for *Barnesandnoble.com* (the review consisted of only a star rating without any review text),
- The median for *Amazon.com* was 88 words, while it was 46 words for *Barnesandnoble.com*,
- The largest observation (q_5) for *Amazon.com* was 2795 words, while it was 660 words for *Barnesandnoble.com*.

A frequency table was constructed to assess the relationship between the average reviews' text length and their star ratings at *Amazon.com* and *Barnesandnoble.com* (see Annexure B: Table 5.1). The results indicate that *Amazon.com* hosted on average the longest reviews per star rating. Another interesting observation is that for both *Amazon.com* and *Barnesandnoble.com* the text length values were normally distributed around the 3-star rating level.

5.2.1.2 Model 2's descriptive statistics

Only one descriptive statistical test, a time series plot, was constructed to present data from model 2. Figure 5.13 reveals an important pattern that explains the targeted 85 *Amazon.com* books' average daily log of sales rank values over the 122-day collection period. In the first couple of days of October 2010 the average daily values decreased, after which *Amazon.com* experienced double exponential growth until the end of the first two weeks in October 2010. For the period until the end of December 2010, the sales rank values increased at a constant rate and then again experienced exponential growth until 31 January 2011. It should also be kept in mind (see section 4.2.6.2.2) that the natural logarithm of sales rank can be viewed as a proxy for online sales, and is inversely related to the actual sales of a book. This means that when the *sales rank* of a given book is increasing, its sales levels are in fact decreasing, which explains why the *Amazon.com* books' sales levels were decreasing over the long term.

FIGURE 5.13: AVERAGE DAILY LOG OF SALES RANK VALUES OVER TIME



5.2.2 Inferential statistics

Inferential statistics allows for sound decision-making practices based on the estimation of population characteristics (Berenson and Levine, 1992:3). In this section the chi-square, a correlation matrix, coefficient of determination (R^2), multiple stepwise regression analyses and residual analyses were used to test the hypotheses implied by model 1 and model 2.

5.2.2.1 Model 1: The effect of individual electronic word-of-mouth review factors on review credibility

According to Mudambi and Schuff (2010:185-186), consumers access thousands of product reviews daily when searching for and examining product offerings online. While past researchers conducted various experiments on the benefits of EWOM reviews' presence on online retailers, the specific factors that influence a review's found helpful degree in online decision-making has received comparatively little attention (Mudambi and Schuff, 2010:185-186). Each of the following review factors has the potential to influence the review's credibility (as measured by review found helpful degree): star rating, ex-star rating, sentiment rating, ex-sentiment rating, text length, time, and e-commerce platform upon which the review took place. The hypotheses formulated to test the impact of various EWOM review factors (the independent variables) on review credibility (the dependent variable), were as follows:

H0¹: A review's credibility on an e-commerce platform is not influenced by the review's star rating

H1^a: A review's credibility on an e-commerce platform is influenced by the review's star rating

H0²: A review's credibility on an e-commerce platform is not influenced by the review's ex-star rating

H2^a: A review's credibility on an e-commerce platform is influenced by the review's ex-star rating

- H0³: A review's credibility on an e-commerce platform is not influenced by the review's positive sentiment rating
- H3^a: A review's credibility on an e-commerce platform is influenced by the review's positive sentiment rating
- H0⁴: A review's credibility on an e-commerce platform is not influenced by the review's ex-sentiment rating
- H4^a: A review's credibility on an e-commerce platform is influenced by the review's ex-sentiment rating
- H0⁵: A review's credibility on an e-commerce platform is not influenced by the review's text length
- H5^a: A review's credibility on an e-commerce platform is influenced by the review's text length
- H0⁶: A review's credibility on an e-commerce platform is not influenced by the time the review appeared
- H6^a: A review's credibility on an e-commerce platform is influenced by the time the review appeared
- H0⁷: A review's credibility is not influenced by the e-commerce platform upon which it takes place
- H7^a: A review's credibility is influenced by the e-commerce platform upon which it takes place

To address the hypotheses associated with model 1, multiple stepwise regression analysis was fitted to the EWOM review dataset. A significance level of 0.05 was used to assess statistical significance.

The starting model's equation was:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \varepsilon.$$

The dependent variable Y is REVIEW CREDIBILITY, X_1 is PLATFORM (*Amazon.com* or *Barnesandnoble.com*), X_2 is TEXT LENGTH (number of words per review), X_3 is TIME (elapsed time between collection date and publish date), X_4 is STARS (number of stars per review), X_5 is EX-STARS (distance in stars from a 3-point star rating), X_6 is SENTIMENT (review sentiment score) and X_7 is EX-SENT (distance from 0.5 average sentiment rating). The error term (ε) for the purpose of computation is assumed to have a mean of 0 and a standardised regression coefficient (β) indicates the relative significance of accompanying X values in the model (Blumberg *et al.*, 2008:789). According to Berenson and Levine (1992: 668 - 672), β_1 is the slope of Y with variable X_1 holding variables $X_2, X_3, X_4, X_5, X_6, X_7$ constant; β_2 is the slope of Y with variable X_2 holding variables $X_1, X_3, X_4, X_5, X_6, X_7$ constant;.... β_7 is the slope of Y with variable X_7 holding variables $X_1, X_2, X_3, X_4, X_5, X_6$ constant.

The stepwise regression model's equation can be executed in three different ways, namely forward selection, backward elimination, and both forward selection and backward elimination (Blumberg *et al.*, 2008:833). Forward selection starts with the constant (β_0) and adds variables that allows for the biggest increase in R^2 . Backward elimination includes all the independent variables in a model and eliminates those variables that allow for the least change in R^2 . Lastly, stepwise selection consists of a composite of both forward selection and backward elimination. For the purpose of conducting stepwise regression analysis on model 1, both forward selection and backward elimination were used in this study.

The independent variable that contributes the most to explaining the dependent variable was added first. Subsequent variables were included based on their incremental contribution over the first variable and whether they met the criterion (a significance level of 0.05) for entering the equation. Variables may be removed at each step if they meet the removal criterion, which is a larger significance level than for entry (a significance level of 0.1). The first step in the multiple stepwise regression analysis was to determine which variables significantly affected Y (Blumberg *et al.*, 2008:789). Only these variables were retained in the model. The

rest of the variables were removed from the model. In the interest of simplicity in describing the multiple stepwise regression procedure, the first four steps of the model will not be described here.

The fifth step in the multiple stepwise regression model included the five independent variables (in order of importance), namely PLATFORM, TEXT LENGTH, TIME, STARS and EX-STARS. The regression equation was: $Y = 0.0765 + 0.2107X_1 + 0.0004X_2 + 0.0015X_3 - 0.0218X_4 - 0.0158X_5$ (see Table 5.2). The constant's value (β_0) of 0.0765 indicates that the regression plane intersected the Y-axis at 0.0765. The multiple regression coefficient associated with X_1 (PLATFORM) indicates that REVIEW CREDIBILITY increased 0.2107 Y-units with the known possibility of one unit increase of X_1 (PLATFORM) while holding X_2 , X_3 , X_4 and X_5 constant. The multiple regression coefficient associated with X_2 (TEXT LENGTH) indicates that REVIEW CREDIBILITY increased 0.0004 Y-units with the known possibility of one unit increase of X_2 (TEXT LENGTH) while holding X_1 , X_3 , X_4 and X_5 constant. The multiple regression coefficient associated with X_3 (TIME) indicates that REVIEW CREDIBILITY increased 0.0015 Y-units with the known possibility of one unit increase of X_3 (TIME) while holding X_1 , X_2 , X_4 and X_5 constant. The multiple regression coefficient associated with X_4 (STARS) indicates that REVIEW CREDIBILITY decreased by 0.0218 Y-units with the known possibility of one unit increase of X_4 (STARS) while holding X_1 , X_2 , X_3 , X_5 constant. Finally, the multiple regression coefficient associated with X_5 (EX-STARS) indicates that REVIEW CREDIBILITY decreased by 0.0158 Y-units with the known possibility of one unit increase of X_5 (EX-STARS) while holding X_1 , X_2 , X_3 , X_4 constant.

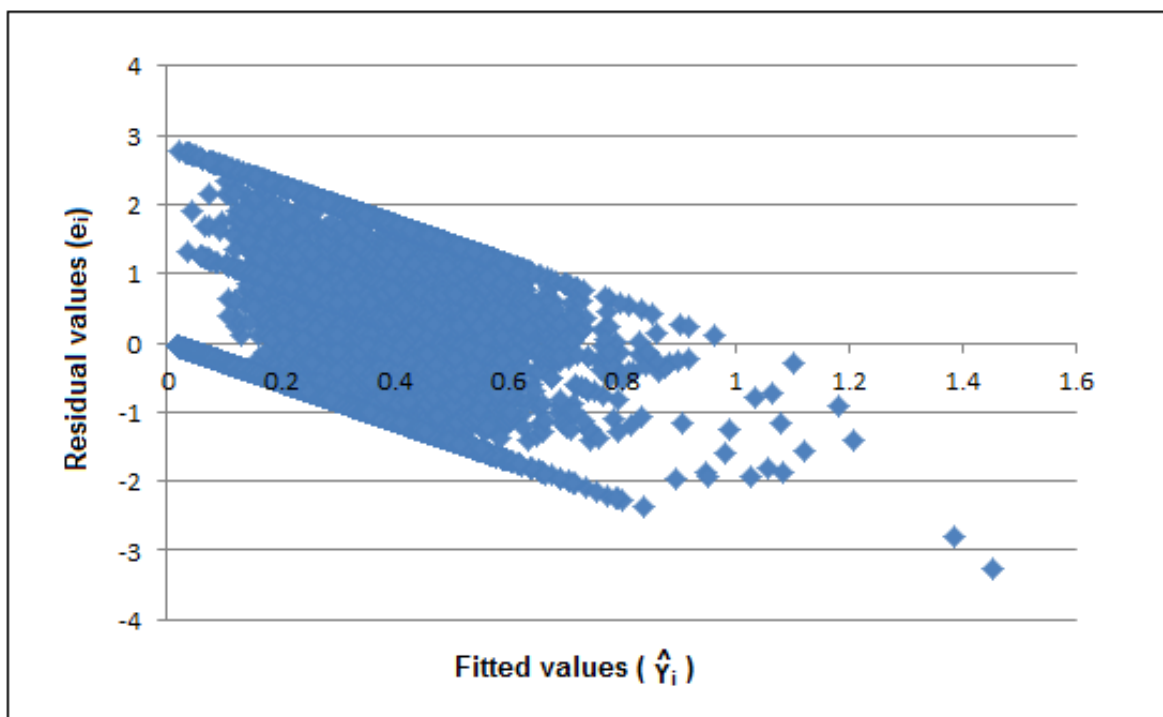
Squaring the multiple correlation coefficient (R) of 0.3771 resulted in a coefficient of multiple determination (R^2) of 0.1422, or an 14.22 % variation in the review credibility can be explained by the variation in the independent variables. The R^2 was adjusted to a number equal to 0.1419.

TABLE 5.2: THE MULTIPLE STEPWISE REGRESSION RESULTS (MODEL 1)

Summary measures				
		Change	% Change	
Multiple R	0.3771	0.0009	%0.2	
R-Square (R^2)	0.1422	0.0007	%0.5	
Adj R-Square	0.1419	0.0006	%0.4	
StErr of Est	0.3525	-0.0001	%0.0	
Regression coefficients				
	Coefficient	Std Err	t-value	p-value
Constant	0.0765	0.0146	5.2484	0.0000
PLATFORM	0.2107	0.0072	29.2462	0.0000
TEXT LENGTH	0.0004	0.0000	20.1142	0.0000
TIME	0.0015	0.0001	18.0709	0.0000
STARS	-0.0218	0.0021	-10.2602	0.0000
EX-STARS	-0.0158	0.0047	-3.3822	0.0007

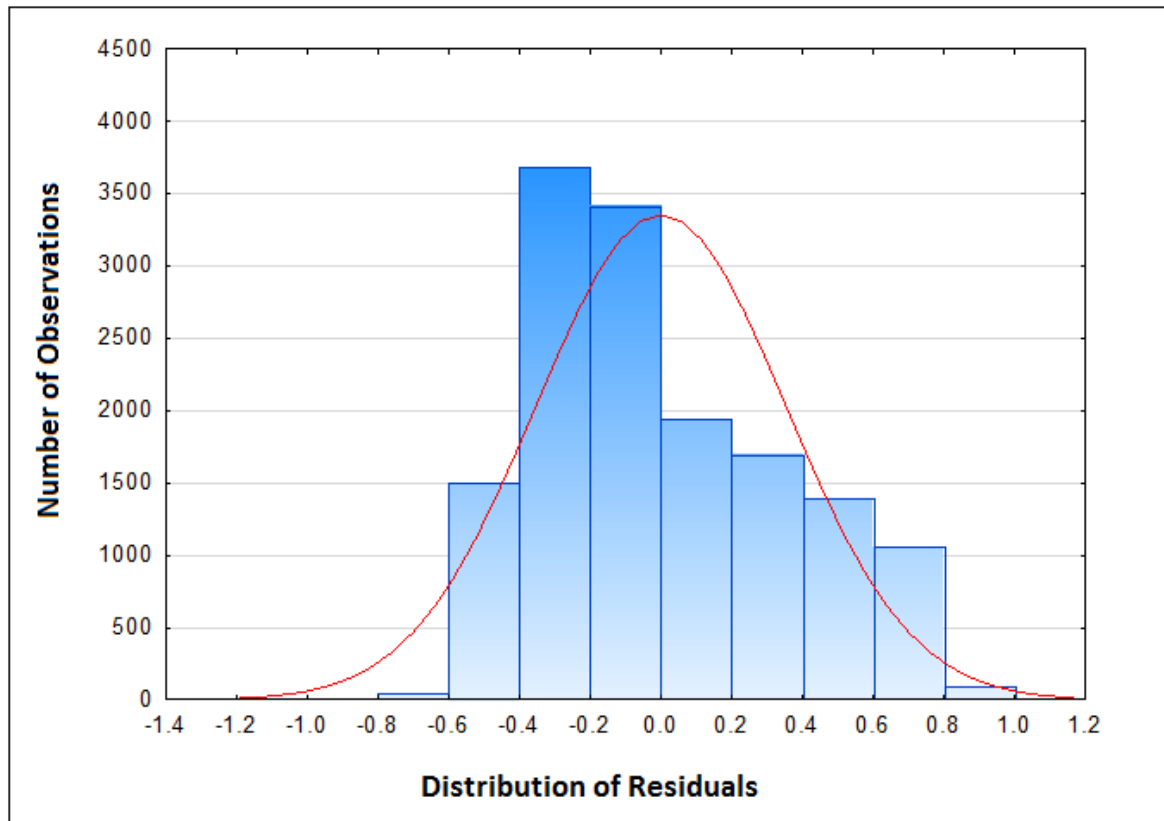
After the stepwise regression model was constructed, a residual analysis test was used to verify the suitability of the model that had been fitted. From Figure 5.14 it can be seen that a clear pattern is present, which indicates that the residuals (e_i) were not independent of the fitted values (\hat{Y}_i) and that the assumption of independence of the residuals was violated.

FIGURE 5.14: FITTED AND RESIDUAL VALUES (MODEL 1)



For the qualitative test of normality (see Figure 5.15), STATISTICA was used to fit the normal distribution (indicated by the red line) to the distribution of the residuals (indicated by the blue bars). The distribution of the residuals, which is represented by the histogram, is clearly skewed to the right and therefore the assumption of normality of the residuals is not met either.

FIGURE 5.15: THE NORMAL AND RESIDUAL DISTRIBUTION (MODEL 1)



The chi-squared test of normality is a quantitative test of goodness-of-fit for the normal distribution. Table 5.3 indicates that the null hypothesis of normality yielded a p-value of zero to four decimal places. Thus, quantitatively the assumption of normality is also rejected.

TABLE 5.3: CHI-SQUARED TEST OF NORMALITY (MODEL 1)

	A	B	C	D	E	F	G
1		<i>Column 1</i>		<u>Intervals</u>	<u>Probability</u>	<u>Expected</u>	<u>Observed</u>
2	Mean	-1.22286E-15		(z <= -2)	0.02275	337.13225	11
3	Standard deviation	0.3525		(-2 < z <= -1)	0.135905	2013.976195	2615
4	Observations	14819		(-1 < z <= 0)	0.341345	5058.391555	6021
5	chi-squared Stat	1456.1876		(0 < z <= 1)	0.341345	5058.391555	3280
6	df	3		(1 < z <= 2)	0.135905	2013.976195	2568
7	p-value	0.0000		(z > 2)	0.02275	337.13225	324
8	chi-squared Critical	7.8147					

When the assumptions regarding the residuals are not met, transformations of the dependent variable (Y) can be made in an attempt to solve the problem (Berenson and Levine, 1992:705). Applying the square root and square transformations on the variable REVIEW CREDIBILITY did not solve the problem. Removing all reviews for which REVIEW CREDIBILITY was zero on the day of data collection, also did not lead to the assumption being met. For this reduced dataset (REVIEW CREDIBILITY all positive), the square, square root, log, and reciprocal transformation were all attempted, none of which solved the problem of non-normality of the residuals. Non-normality was accepted and dealt with in the most appropriate way.

Howell (1999:208-211) suggests that even though deviation from normality may occur, the p-values of the t-test can still be used to give an indication of which variables are most significant (see the p-values in Table 5.2). A p-value of zero to four decimal places will indicate significance even in the presence of substantial deviation from normality.

For practical reasons the final independent variable, EX-STARS, was left out of the model. The reasons were fourfold: the EX-STARS-variable was the least significant of the five, and the only variable for which the p-value of the t-test was not zero to four decimal places; the STARS-variable from which EX-STARS was derived, already appears in the model; the EX-STARS-variable may be considered artificial and the interpretation of the EX-STARS-variable's coefficient a bit too complex.

After the independent variable EX-STARS was removed, an ordinary regression (not stepwise) model was constructed using only the independent variables PLATFORM, TEXT LENGTH, TIME, and STARS (see Table 5.4).

The regression equation of the model is indicated below and is interpreted at the end of this section:

$$\hat{y} = 0.0607 + 0.2108X_1 + 0.0004X_2 + 0.0015X_3 - 0.0239X_4$$

Note that Table 5.4 results in exactly the same outcome which will be obtained when terminating the stepwise regression after the first four steps. Ordinary regression was performed simply because variable selection had already taken place. Stepwise regression had taken care of variable selection, and one of the major advantages of the technique is that two variables which are dependent (or highly correlated) will not both appear in the final model (Treiman, 2009:108). STATISTICA was used to verify that the same variables led to the same coefficients and respective p-values.

TABLE 5.4: ORDINARY REGRESSION RESULTS (MODEL 1)

Regression Statistics ¹¹				
Multiple R	0.37626			
R Square (R ²)	0.14157			
Adjusted R Square	0.14134			
Standard Error	0.35265			
Observations	14819			
	Coefficients	Standard Error	t Stat	p-value
Intercept	0.0607	0.0138	4.3965	0.0000
PLATFORM	0.2108	0.0072	29.2440	0.0000
TEXT LENGTH	0.0004	0.0000	20.8492	0.0000
TIME	0.0015	0.0001	17.9109	0.0000
STARS	-0.0239	0.0020	-11.8209	0.0000

A correlation matrix (see Table 5.5) was also constructed to examine all of the possible relationships between the four independent variables PLATFORM, TEXT LENGTH, TIME and STARS. According to Elifson *et al.* (1990:206-207), the computation of the coefficients can only be determined when one of the four variables is available for all of the 14819 EWOM review observations. The variables included in the model were poorly correlated. The highest correlation, measured according to absolute values, had a correlation of 18.90% between PLATFORM and TIME, as well as a correlation of 18.77% between PLATFORM

11) The confidence intervals' outputs are not shown in Table 5.4 since the confidence intervals are only valid if the normality assumption is met.

and TEXT LENGTH. Even though these values were relatively higher than the rest of the values in the matrix, when compared to a benchmark of 25% (1st quarter correlation) the overall relationship among the variables was weak. The data were also analysed to identify whether any multicollinearity existed. Multicollinearity can be described as the occurrence of two or more independent variables that are highly correlated, to the extent that it causes an adverse result on multiple regression (Blumberg *et al.*, 2008:836). Since the highest correlation (in absolute value) was below 19%, multicollinearity did not influence the results of the analysis.

TABLE 5.5: CORRELATION MATRIX OF THE INDEPENDENT VARIABLES (MODEL 1)

	A	B	C	D	E
1		PLATFORM	TEXTLENGTH	TIME	STARS
2	PLATFORM	1			
3	TEXTLENGTH	0.18771	1		
4	TIME	0.18902	0.08843	1	
5	STARS	-0.05300	-0.01681	-0.07501	1

In the final analysis, EX-STAR ($H0^2$), SENTIMENT ($H0^3$) and EX-SENTIMENT ($H0^4$) were removed, and the relationship between the dependent variable (REVIEW CREDIBILITY) and remaining independent variables (PLATFORM, TEXT LENGTH, TIME and STARS) tested. The null hypotheses were then reformulated in the context of the modified regression model.

In the model:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$; where: Y = REVIEW CREDIBILITY, X_1 = PLATFORM, X_2 = TEXT LENGTH, X_3 = TIME, X_4 = STARS, the:

$H0^1: \beta_4 = 0$ and $H1^a: \beta_4 \neq 0$.

Given the low p-value in Table 5.4 $H0^1$ is rejected (and $H1^a$ accepted). It can be concluded that a review's credibility on both *Amazon.com* and *Barnesandnoble.com* was indeed influenced by the review's star rating.

$H0^5: \beta_2 = 0$ and $H5^a: \beta_2 \neq 0$.

Given the low p-value in Table 5.4 $H0^5$ is rejected (and $H5^a$ accepted). It can be concluded that a review's credibility on both *Amazon.com* and *Barnesandnoble.com* was indeed influenced by the review's text length.

$H0^6: \beta_3 = 0$ and $H6^a: \beta_3 \neq 0$.

Given the low p-value in Table 5.4 $H0^6$ is rejected (and $H6^a$ accepted). It can be concluded that a review's credibility on both *Amazon.com* and *Barnesandnoble.com* was indeed influenced by the date the review appeared.

$H0^7: \beta_1 = 0$ and $H7^a: \beta_1 \neq 0$.

Given the low p-value in Table 5.4 $H0^7$ is rejected (and $H7^a$ accepted). It can be concluded that a review's credibility was indeed influenced by the e-commerce platform upon which it took place.

5.2.2.2 Model 2: The relationship between electronic word-of-mouth review behaviour and the sales rank of new best-seller Amazon.com books

The marketing literature often refers to the importance of interpersonal communication, in the form of WOM, as a possible driver of consumer decision-making (Chevalier and Mayzlin, 2004:2). According to Chen *et al.* (2004:713-714), customers potentially reduce the quality uncertainty and potential risk of buying products when they read or listen to other users' feedback, allowing them to gain more confidence in their decision-making when considering buying a product.

Related to EWOM in the buying of books online, conflicting results have also been reported by various authors. One study conducted by Chen *et al.* (2004:713-714) suggests that if a specific book is not browsed, then the ratings for the particular book cannot be read by readers, thus having no influence on the consumer's decision-making process. Other results suggest that if ratings are not perceived to be trustworthy or meaningful, consumers will not take these ratings into consideration while conducting online shopping (Chen *et al.*, 2004:713-714). Limited star rating variance across books and a relatively higher number of positive book ratings on e-commerce platforms can potentially be explained by authors and publishers who have a higher incentive to submit more biased reviews for their own book titles, and may potentially cause a decrease in the credibility of the review experience (Godes and Mayzlin, 2003; Chevalier and Mayzlin 2004).

Similar to the comparative study that was conducted by Chen *et al.* (2004:713-714), the present study does not postulate any explicit hypothesis on consumer reviews and sales, but

instead follows the conventions of empirical research. The following hypothesis was formulated:

H0⁸: Sales rank levels of *Amazon.com* books are not influenced by spotlight EWOM reviews

H8^a: Sales rank levels of *Amazon.com* books are influenced by spotlight EWOM reviews

In order to test this hypothesis, the previously prepared daily observed variables were combined into one sheet (see Table 5.6). The column EWOM was added, which was a single value representing the spotlight EWOM of the book at a given time. Spotlight EWOM was calculated as the average of average stars, average recent review stars, and average most helpful star ratings, thus allocating a weight of 0.3333 to each of the factors (e.g. $EWOM = 0.3333 AV + 0.3333 FH + 0.3333 MR$).

TABLE 5.6: AMAZON.COM SALES RANK AND EWOM DATA (MODEL 2)

M	N	O	P	Q	R	S	T
Book	Date	Sales rank Previous Week	Sales rank Following week	Average Stars	Average Recent Stars	Average Found Helpful	EWOM
1	40484	6.862	7.431	4.588	4.3	5	4.629
1	40485	6.997	7.397	4.588	4.3	5	4.629
1	40486	7.108	7.404	4.588	4.3	5	4.629
1	40487	7.192	7.405	4.611	4.3	5	4.637
1	40488	7.300	7.444	4.632	4.3	5	4.644
1	40489	7.410	7.469	4.182	3.6	5	4.261
1	40490	7.506	7.628	4.182	3.6	5	4.261
1	40491	7.463	7.799	4.182	3.6	3.667	3.816
1	40492	7.431	7.904	4.182	3.6	3.667	3.816
1	40493	7.397	7.989	4.182	3.6	3.667	3.816
1	40494	7.404	8.084	4.182	3.6	3.667	3.816
:	:	:	:	:	:	:	:

On each date, it needs to be determined whether or not sales rank¹² relative to EWOM remained unchanged (same), decreased (worse) or increased (better) from the previous date. Therefore the following table rows (see Table 5.7) were generated, by using all the data across all dates for the entire population.

12) Measured by the seven day average log of sales rank (see section 4.2.6.2).

TABLE 5.7: CHANGES IN EWOM COMMUNICATION (SAME, WORSE AND BETTER)

		Sales Rank		
		Worse	Same	Better
EWOM	Worse	Worse*Worse	Worse*Same	Worse*Better
	Same	Same*Worse	Same*Same	Same*Better
	Better	Better*Worse	Better*Same	Better*Better

One problem which was encountered was how to define a change in sales rank. The ever-changing nature of the sales rank, combined with the fact that an average was calculated over a long time (seven days), meant that the previous week's sales rank and the following week's sales rank were seldom the same.

When change in the sales rank value was insignificant, it was classified as "same". Clearly a change in sales rank from 11000th to 11010th was insignificant, and it could not be interpreted to reflect a decrease in sales (for which we would like to determine a cause or influencing factor). In the total number of 9529 observations that were made, only 29 observations had the same sale rank for the previous and following weeks. This was when book 14, *Cake Pops: Tips, Tricks, and Recipes for More Than 40 Irresistible Mini Treats*, was the top best-seller for quite a while, yielding 0 seven-day average log sales rank for both the previous and the following week.

A factor of 1.1 was applied to indicate a significant decrease (worse) or increase (better). This conversion translated into a difference of 10% as a rule of thumb. If SR_A and SR_B are two consecutive sales ranks differing by the minimum significant difference of 10%, then:

$$\frac{SR_A}{SR_B} = 1.1 \rightarrow \log\left(\frac{SR_A}{SR_B}\right) = \log(SR_A) - \log(SR_B) = \log(1.1) = 0.09531.$$

Therefore sales rank will be regarded as unchanged (same), if abs (previous week sales rank – following week sales rank) < 0.09531. The results of these sales rank changes over the four-month period are displayed in Table 5.8.

TABLE 5.8: OBSERVED SALES RANK AND EWOM CHANGES

		Sales Rank		
		Worse	Same	Better
EWOM	Worse	1307	375	245
	Same	2881	1084	923
	Better	1742	575	397

Next, the expected frequencies were calculated by changing table values from counts to probabilities (see Table 5.9). This was done simply by dividing all values by the grand total of 9529 observations.

TABLE 5.9: PROBABILITIES OF SALES RANK AND EWOM CHANGES

		Sales Rank		
		Worse	Same	Better
EWOM	Worse	0.1372	0.0394	0.0257
	Same	0.3023	0.1138	0.0969
	Better	0.1828	0.0603	0.0417

Thereafter the rows and column totals were calculated to obtain marginal probabilities. In Table 5.10 the marginal probability values are added to the very bottom and very right of the original Table 5.9.

TABLE 5.10: MARGINAL PROBABILITIES OF SALES RANK AND EWOM CHANGES

		Sales Rank			
		Worse	Same	Better	
EWOM	Worse	0.1372	0.0394	0.0257	0.2022
	Same	0.3023	0.1138	0.0969	0.5130
	Better	0.1828	0.0603	0.0417	0.2848
		0.6223	0.2135	0.1642	1.0000

If sales rank and EWOM are statistically independent (as stated under H_0^8), then the joint probability of a cell (see Table 5.11) can be obtained by multiplying the respective marginal probabilities. For instance the expected probability for EWOM worse and sales rank better is calculated as: $0.2022 \times 0.1642 = 0.0332$.

TABLE 5.11: JOINT PROBABILITIES OF SALES RANK AND EWOM CHANGES

		Sales Rank			
		Worse	Same	Better	
EWOM	Worse	0.1258	0.0432	0.0332	0.2022
	Same	0.3192	0.1095	0.0842	0.5130
	Better	0.1772	0.0608	0.0468	0.2848
		0.6223	0.2135	0.1642	1.0000

Multiplying all the probabilities (Table 5.11) by the total of 9529 observations returns the expected sales rank counts (see Table 5.12). For instance, the expected count for EWOM worse and sales rank worse is calculated as: $0.2022 \times 0.6223 \times 9529 = 1199.19$. The further these are from the observed counts, the more evidence there is against independence.

TABLE 5.12: EXPECTED SALES RANK AND EWOM CHANGES

		Sales Rank		
		Worse	Same	Better
EWOM	Worse	1199.19	411.33	316.48
	Same	3041.86	1043.36	802.78
	Better	1688.95	579.31	445.74

The null hypothesis (H_0^8) assumes that sales rank is not influenced by spotlight EWOM reviews. A significance level (α) at a minimum of 0.05 was used, implying that a p-value of $p < 0.05$ indicates statistical significance (Howell, 1999:129). As a measure of deviation the Chi-Squared (χ^2) test statistic was used (Howell, 1999:375). Here n denotes the number of cells in the table, O_i denotes the observed frequency in cell i and E_i the expected frequency of cell i (if the null hypothesis is true):

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Next, the value of $\frac{(O_i - E_i)^2}{E_i}$ for each cell was then calculated and added up (see Table 5.13).

TABLE 5.13: CHI-SQUARE TEST RESULTS

		Sales Rank			
		Worse	Same	Better	
EWOM	Worse	9.6918	3.2080	16.1452	29.0449
	Same	8.5061	1.5829	18.0025	28.0915
	Better	1.6662	0.0321	5.3285	7.0268
		19.8642	4.8229	39.4762	64.1633

The distribution of the chi-squared statistic under the null hypothesis is stated as the chi-squared distribution with m degrees of freedom (χ_m^2), where the null hypothesis is rejected for a large value of the statistic and the m -value equals four (e.g. $m = [\text{number of rows} - 1][\text{number of columns} - 1] = 4$). The p -value is therefore calculated as the probability of a χ_4^2 random variable exceeding the observed value. The outcome was as follows:

$p(\chi_4^2 > 64.1633) = 3.8611\text{E-}13 \ll 0.05$. H_0^8 is therefore rejected (and H_8^a accepted). It can ultimately be concluded that *Amazon.com* sales rank was influenced by spotlight EWOM.

The above-mentioned chi-squared test was also performed for several combinations of weights and change-factors that could potentially alter the outcome of the p -value (see Table 5.14).

The weights of average book stars, average most found helpful stars and average most recent stars were not only tested at an equal weight of 33.33% for each variable (e.g. 0.3333 AV, 0.3333 FH and 0.3333 MR), but were also changed to 0.55 AV, 0.1 FH and 0.35 MR to take into consideration the importance that online users gave to spotlight EWOM review factors, with regards to the location on the web page. According to Chen *et al.* (2008:12-13), spotlight reviews on product landing pages are much easier to access for online users, and possibly carry a higher weight in consumers' decision-making.

As indicated in Table 5.14, all other combinations of sales rank-factors and weights yielded even smaller p -values, except a sales rank factor at 1.1 with EWOM weights equal to 0.55 AV, 0.1 FH and 0.35 MR, which yielded a p -value equal to 4.27E-13. Since this was only a subjective estimate, the original factor of 1.1 (or 10% rule of thumb) and more neutral weights of 33.33% were considered for the final model.

TABLE 5.14: WEIGHTS, SALES RANK CHANGE FACTORS AND P-VALUES

EWOM Weights			Sales Rank Factor	p-value
Average book stars	Average most found helpful stars	Average most recent stars		
0.33333	0.33333	0.33333	1	4.59E-19
			1.1	3.86E-13
			1.2	7.17E-16
			1.3	1.45E-18
Average book stars	Average most found helpful stars	Average most recent stars		
0.55	0.1	0.35	1	5.3E-19
			1.1	4.27E-13
			1.2	4.3E-16
			1.3	7.26E-19

5.3 CONCLUSION

The results chapter analysed specific data which were collected to address the stated research hypotheses as well as reflect the findings of both the descriptive and inferential statistical analysis. The research objectives of this study were to assess the influence of EWOM factors on the credibility of reviews (model 1) as well as to test the association which spotlight EWOM reviews had with online sales rank (model 2). It was concluded that EWOM review factors (platform, text length, time, stars) did influence reviews' credibility (measured by found helpful degree) at *Amazon.com* and *Barnesandnoble.com*, and secondly, that spotlight EWOM reviews did in fact influence the sales rank of the entire population of new best-seller books at *Amazon.com*. In the next and final chapter the research study concludes with implications for the industry, as well as limitation of this study and recommendations for future research.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

6.1 INTRODUCTION

A conclusion is a summary of the inferences drawn from a study's findings (Blumberg *et al.*, 2008:567-568). This study was designed to assess the influence that EWOM reviews exert on review credibility and online retailers' sales levels, by formulating two hypothetical models and then empirically testing the models on actual customer review and sales rank data from *Amazon.com* and *Barnesandnoble.com*.

This concluding chapter is organised into several sections. Firstly, the research findings for model 1 and model 2 will be interpreted, and secondly, industry recommendations will be suggested based on the empirical results. Thereafter, the limitations of this study as well as future research directions will be discussed. Lastly, the chapter concludes with a reconciliation of the study's research objectives, together with final remarks.

6.2 CONCLUSIONS FROM THE STUDY

6.2.1 The target population and realised samples

The target population that was investigated was narrowed down to the EWOM review communication about 100 new best-seller books on e-commerce platforms from 1 October 2010 until 31 January 2011. Two hypothetical models were proposed to test the impact of EWOM review attributes on review credibility, as well as to investigate the association between spotlight EWOM reviews and sales rank values of new best-seller books. Non-probability judgement sampling was used (Zikmund and Babin, 2010:432). The three criteria, on which the observation's samples were chosen, were e-commerce platforms, books, and time period.

USA-based e-commerce platforms *Amazon.com* and *Barnesandnoble.com* started out in the mid 1990's and today host two of the world's most active book reviewing communities (Chevalier and Mayzlin 2004; Chen *et al.*, 2008:4). When the results of both e-commerce enterprises were compared, it was found that *Amazon.com*'s net sales had increased 38% to

\$9.86 billion in the first quarter of 2011 (Amazon.com press release, 2011), while *Barnesandnoble.com* sales amounted to \$217 million for the quarter and \$858 million for the entire financial year (Barnes and Noble press release, 2011). It should also be noted that *Barnes and Noble* had originally started as a brick-and-mortar enterprise, and eventually extended their product offering online as *Barnesandnoble.com*. Even though *Barnesandnoble.com*'s online sales were comparatively lower than *Amazon.com*'s in 2011, its offline sales together with online sales at *BarnesandNoble.com* reached a record of \$7 billion at the financial year end (Barnes and Noble press release, 2011).

The books category was the only product-category that was focused on for this study, partly because it is the most frequently studied in applied e-commerce research, and also because it is one of the leading online sales categories for e-commerce enterprises (Chevalier and Goolsbee, 2003:2). Individual product review pages can be browsed through to obtain needed information about a book, as well as allow readers to rate books and share their experiences and sentiment through the reviews' text. More specifically, only new best-seller books were analysed. Chen and Wu (2007:1) found that best-seller books compared to less-reviewed and less-popular books usually had more information available and were more accessible to online buyers. The only requirement for new best-seller books to be included in the study was that each book needed to accumulate at least one written review in the specified four-month data collection period.

The time period when data collection took place was chosen because it represented the four busiest online shopping months of the year (1 October 2010 and 31 January 2011). When US retail e-commerce sales estimates for the last quarter in 2010 and the first quarter in 2011 are examined (see Annexure D: Table 6.1), it is interesting to note that online retail spending reached a record of \$43.4 billion for the fourth quarter in 2010 (11% increase from a year earlier) and \$38.0 billion for the first quarter in 2011 (12% increase from a year earlier).

6.2.2 Individual electronic word-of-mouth review factors and review credibility (model 1)

While conducting an online consumer behaviour study, Mudambi and Schuff (2010:185) found that even before consumers purchased items from e-commerce enterprises, they first turned to the vast number of product reviews available on e-commerce platforms, as well as elsewhere on the web. One particular area that was recommended for future research

studies is the examination of factors in the purchase cycle that makes online reviews (more) credible for consumers. The factors that were of particular interest to this study were star rating, ex-star rating, sentiment rating, ex-sentiment rating, text length, the date the review appeared, and lastly, the e-commerce platform on which the review took place.

6.2.2.1 Review frequency distribution

To indicate the differences in the characteristics of EWOM reviews between *Amazon.com* and *Barnesandnoble.com*, the study focused on the frequency distribution of the collected reviews. The first notable difference across the two e-commerce platforms was that *Amazon.com* had more reviews published about its books than *Barnesandnoble.com*.

100 new best-seller *Amazon.com* books and the same 100 books on *Barnesandnoble.com* were identified as the target population for this study on 21 September 2010. Upon the completion of the data collection period, it emerged that 15 books from *Barnesandnoble.com* should be discarded, because there were no reviews published about them between 1 October 2010 and 31 January 2011. The same 15 books were also discarded at *Amazon.com*, which left the target population at 85 new best-seller books for both e-commerce platforms.

From those 85 books that were analysed for model 1, *Amazon.com* had a mean number of 408 reviews per book, a minimum of 10 reviews per book, and a maximum of 805 reviews per book. *Barnesandnoble.com* on the other hand had a mean number of 176 reviews per book, a minimum of 0 reviews per book, and a maximum of 352 reviews per book.

When this study's results are compared to a study done by Chevalier and Mayzlin (2004:32) with a sample consisting of 2394 observations over a two-day period in May of 2003, their results indicate that *Amazon.com* had a mean of 61.27 reviews per book, a minimum of 0 reviews per book, and a maximum of 4457 reviews per book. *Barnesandnoble.com* had a mean of 12.87 reviews per book, a minimum of 0 reviews per book, and a maximum of 956 reviews per book. A different set of results was found by Chen *et al.* (2004:716), who collected and analysed data from 610 *Amazon.com* book observations in December 2003. They found that each *Amazon.com* book had a mean of 96 reviews, a relatively big standard deviation of 279 reviews per book, and a median of 17 reviews per book.

6.2.2.2 Found helpful frequency distribution

The found helpful degree for each review on *Amazon.com* and *Barnesandnoble.com* was calculated by dividing the numerator (readers who voted the review helpful) with the denominator (total number of readers who voted). Most of the reviews on *Amazon.com* (see Figure 5.3) indicate either an extremely low found helpful degree between 0.00 – 0.10 (36%), or an extremely high found helpful degree between 0.90 – 1.00 (18%). Compared to *Amazon.com*, the found helpful frequency distribution of *Barnesandnoble.com* reviews (see Figure 5.4) indicates that most of the reviews have only an extreme low found helpful degree between the intervals of 0.00 – 0.10 (78%), as well as at higher levels of 0.90 – 1.00 (6%) and 0.40 – 0.50 (4%). The extremely positive and negative found helpful degree indicates that people mostly found all reviews of both platforms either very helpful or not helpful at all.

6.2.2.3 The frequency distribution of star sentiment

Each user of the rating and review systems on both *Amazon.com* and *Barnesandnoble.com* has the ability to not only rate other reviewers' reviews, but also to reflect their own positive or negative sentiment through a star rating and review text.

The mean as well as the median for review text sentiment scores were higher at *Barnesandnoble.com* than at *Amazon.com* (see Figures 5.7 and 5.8), while the reviewers' star ratings at *Amazon.com* and *Barnesandnoble.com* both reflected positive scores about the new best-seller books (see Figures 5.5 and 5.6). One reason for the relatively high ratings in the book category at *Amazon.com* and *Barnesandnoble.com* could be that the target population of new best-seller books enjoys popularity from around the world, causing people to be less negative in their opinions. Reviewers who previously would have given a book that was not part of the new best-sellers list a lower rating, feel less inclined now to hold a contrasting opinion. The same results were found by Chevalier and Mayzlin (2004:33), Chen *et al.* (2004:716), and Mudambi and Schuff (2010:193), who indicated that the average consumer star ratings of *Amazon.com* and *Barnesandnoble.com* books were quite positive.

The research done on WOM satisfaction-dissatisfaction frequencies by Anderson (1998:5), indicate a u-shaped curve of how people communicate their WOM at extremely positive or extremely negative levels. Even though Figures 5.5 and 5.6 do not form u-shaped curves on the rating scale between 1 and 5 stars, they do form a partial u-shaped curve on a rating

scale between 1 and 4 stars for both *Amazon.com* and *Barnesandnoble.com*, also indicating extreme positive or negative rating levels.

6.2.2.4 Review text length frequency distribution

Beyond the sentiment of review text, additional information can also be extracted from the review text (Chevalier and Mayzlin, 2004:12-41). One such measure which was included in model 1 to indicate a reviewer's effort to write a review, was the size of the review's text as measured in word count.

Out of the total number of 14819 unique EWOM reviews that were analysed for *Amazon.com* and *Barnesandnoble.com*, the average text length of an *Amazon.com* review was a mean of 88 words relative to a mean of 46 words per *Barnesandnoble.com* review.

When the average text length per review's star rating (1 to 5 stars) was analysed, the results indicated that reviewers at *Amazon.com* posted longer reviews at all of the star ratings than did reviewers at *Barnesandnoble.com* (see Table 5.1). According to Chevalier and Mayzlin (2004:34), who also found a higher on average word count at all five star levels from *Amazon.com*, the higher word count may be ascribed to *Amazon.com*'s efforts to stimulate the creation of more meaningful reviews from its reviewer community.

6.2.2.5 The relationship between electronic word-of-mouth review factors and review credibility

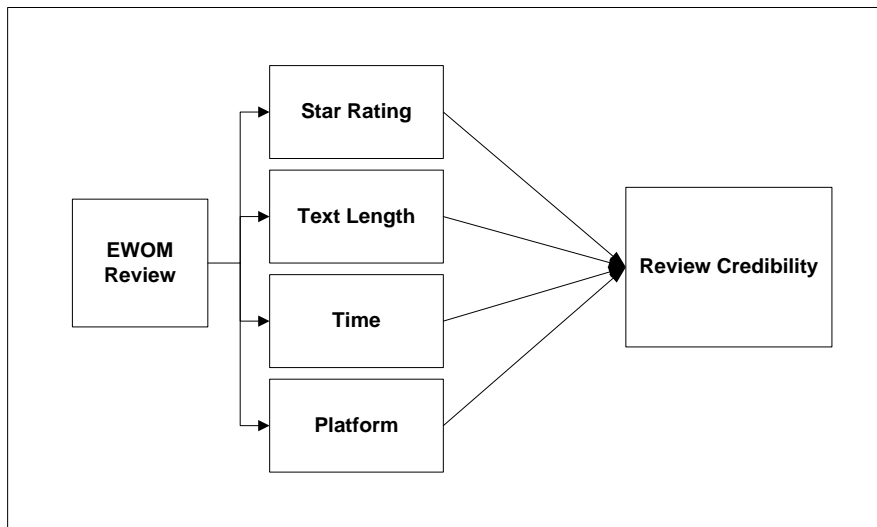
To better understand the conditions under which consumers actively seek credible EWOM review information to make an informed purchase decision, the relationship between review credibility and different EWOM review factors (star rating, ex-star rating, sentiment rating, ex-sentiment rating, review text length, review publish date and reviews' e-commerce platforms) was examined.

To examine model 1, multiple stepwise regression analyses were fitted to the EWOM review dataset to test whether independent variables STAR, EX-STAR, SENTIMENT, EX-SENTIMENT, TEXT LENGTH, TIME AND PLATFORM can concurrently serve as predictors of the dependent variable REVIEW CREDIBILITY. A significance level of 0.05 was used to indicate statistical significance in hypothesis testing.

After stepwise selection was applied to the selected variables, the independent variables EX-STAR (X_5), SENTIMENT (X_6) and EX-SENTIMENT (X_7) were removed because they led to the least change in R^2 . H_0^2 , H_0^3 , H_0^4 could thus not be empirically tested.

In the final model, the association between the dependent variable (REVIEW CREDIBILITY) and the remaining independent variables (PLATFORM, TEXT LENGTH, TIME, and STARS) were tested (see Figure 6.1).

FIGURE 6.1: THE RELATIONSHIP BETWEEN EWOM REVIEW FACTORS AND REVIEW CREDIBILITY



Star rating

Consumers typically express negative emotions through NWOM when an unfortunate situation occurs, or communicate PWOM when positive emotions are felt (Hennig-Thurau *et al.*, 2004:44). The perception of products before a purchase determines the level of valence consumers usually express after the purchase. Reviews' star ratings were analysed to assess the impact of the valence of new best-seller books on *Amazon.com* and *Barnesandnoble.com*'s review credibility (measured by found helpful degree). H_0^1 was rejected, and it can be concluded that STAR is a significant variable and that for every additional STAR, the REVIEW CREDIBILITY decreases on average by 2.39%, given that all other factors remain unchanged. This means that, in general, a review which gives a poor rating for the book is found more helpful. At the risk of over-interpretation, it seems that when consumers are interested in buying a book, they generally feel a review is more helpful if it gives them a reason not to buy the book.

Text Length

With a low p-value, H_0^5 was also rejected, and it can be concluded that TEXT LENGTH is a significant variable and that for each additional word the review TEXT LENGTH increases, the REVIEW CREDIBILITY increases on average by 0.04%, given that all other factors remain unchanged. Similar results were also reported by Mudambi and Schuff (2010:195) in their analysis of 1587 *Amazon.com* reviews across two product types (search and experience goods). They found that review depth had a positive effect on the credibility of the review, and that review depth is a highly significant ($p < 0.000$) predictor of the degree to which a review was found helpful for both experience goods (e.g. mp3 players, music CDs and PC video games) and search goods (e.g. cell phone, digital camera and laser printer). Mudambi and Schuff (2010:192-193) suggest that one reason why people are more likely to vote “found helpful” for a longer review might be that additional EWOM review content has a higher potential to stimulate a reaction from a reader than a shorter review. Another reason might be that consumers believe longer reviews are written by more knowledgeable people from whom they are likely to gain knowledge before a purchase decision (Bansal and Voyer, 2000:166-167).

Time

The time period when data collection took place covered the four busiest online browsing and shopping months of the financial year. H_0^6 was rejected, and it can be concluded that TIME is a significant variable, and that for every day the review is available for scrutiny, the REVIEW CREDIBILITY increases on average by 0.15%, given that all other factors remain unchanged. This result may seem counter-intuitive, but this outcome might be because scanning through a number of older reviews, the potential buyer focuses only on worthwhile reviews, and is most likely to react only on them.

Platform

The final variable, PLATFORM, was tested under H_0^7 and rejected. It can be concluded that PLATFORM is a significant variable in predicting review credibility. The PLATFORM's results showed that REVIEW CREDIBILITY is on average 21.08% higher for a review on *Amazon.com* than one on *Barnesandnoble.com*, assuming that all other factors remain unchanged. This result is potentially very important for *Amazon.com*'s business-to-business and business-to-customers operation, especially the way they can market themselves to

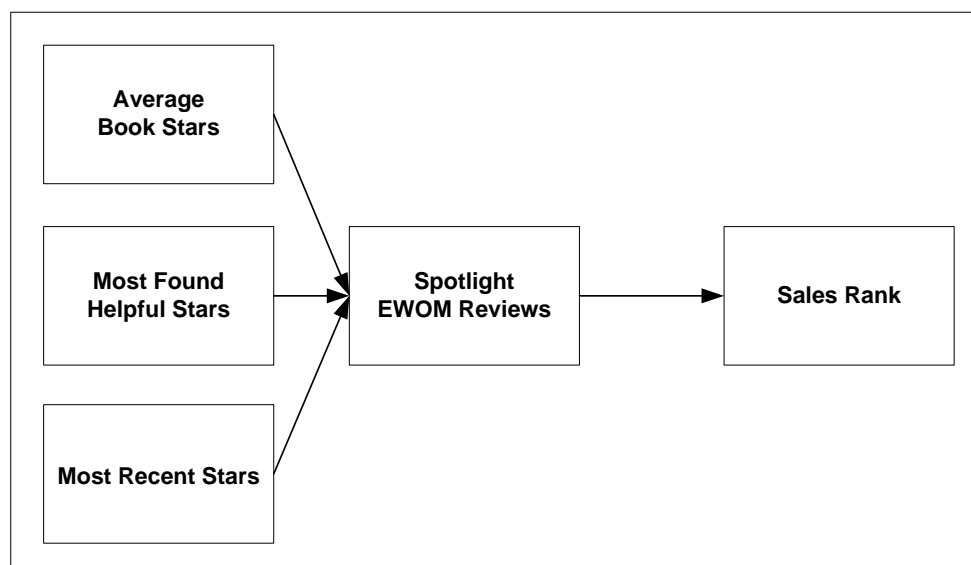
future online readers and book publishers as one of the most trusted e-commerce platforms on the Internet.

It can also be seen from Table 5.4 that the *adjusted R*² value is 0.1413. This means that only 14.13% of the variation in REVIEW CREDIBILITY is explained by variation in PLATFORM, TEXT LENGTH, TIME and STARS. The remaining 85.87% of variation is explained by variables which are not included in the model. Despite the fact that the model explains only limited variation, it is still helpful in the sense that it quantifies the effect of four variables on REVIEW CREDIBILITY, all of which are significant. One must also bear in mind that the purchase decision-making process, as well as the review process, is in essence a complex psychological exercise, which is predominantly influenced by non-quantifiable extraneous factors, such as emotional state (e.g. anger or confusion), legibility of the reader, and his/her financial means, which are in practical terms impossible to measure and model without adequate information.

6.2.3 Spotlight electronic word-of-mouth and sales rank levels (model 2)

Previous studies have shown that WOM communication is a possible driver of consumer decision-making, but the process of influence is generally dynamic and complex (Chevalier and Mayzlin, 2004:2; Sormunen, 2009:33-34). Against this background, a model was constructed (see Figure 6.2) to test whether the sales rank values of *Amazon.com* books are influenced by spotlight EWOM reviews.

FIGURE 6.2: THE RELATIONSHIP BETWEEN SPOTLIGHT EWOM AND SALES RANK



On each day of the data collection period it was determined whether or not EWOM and sales rank had changed from the previous date. The present study applied a factor of 1.1 to indicate a significant change. A Pearson's chi-square (χ^2) test was used to test whether variable independence did exist for model 2. The p-value was therefore calculated as the probability of a χ_4^2 random variable exceeding the observed value, and resulted in $p(\chi_4^2 > 64.1633) = 3.8611\text{E-}13 \ll 0.05$. Based on this result H_0^8 was rejected, and it can be concluded that *Amazon.com*'s sales rank values are influenced by spotlight EWOM.

If model 2's results are compared with the results of two previously reported studies, different outcomes are found. In the first study, Chen *et al.* (2004:713) examined the influence of ratings and reviews on *Amazon.com*'s sales levels, and found that even though the published number of reviews was positively related to sales levels, consumer ratings were not related to sales levels. Chen *et al.* (2004:722) also suggest that causality may not simply be implied, but rather that the results indicate that if sales data are not available, the number of published reviews may be used as a potential proxy for online sales. The results reported by Chevalier and Mayzlin (2004:32), concurred with the findings of model 2 of the present study, which revealed that consumer WOM indeed influences consumer decision-making at both *Amazon.com* and *Barnesandnoble.com*.

It should also be noted that these two previous studies used different samples and models to test whether EWOM communication had an impact on sales levels. The first study's sample consisted of EWOM reviews that were published about 610 *Amazon.com* books in December 2003 (Chen *et al.*, 2004:715-722), while the second study examined 2394 books from both *Amazon.com* and *Barnesandnoble.com* over a two-day period in May 2003 (Chevalier and Mayzlin, 2004:32). Both studies were conducted over a much shorter time period than the four-months, during which the present study's data was collected.

6.3 IMPLICATIONS AND RECOMMENDATIONS

The following excerpt was taken from *comScore, Inc.*, an online research enterprise. The paragraph highlights the importance of e-commerce as an added distribution channel for producers, as well as a value-adding service for consumers, who are continually driven by an almost unlimited set of needs.

"Faced with rapidly rising gas prices and stubbornly high unemployment, consumers continued to take advantage of the Internet's lower prices by shifting their spending from

offline retail stores. In fact, in the first quarter, the growth in e-commerce spending was roughly double that observed at offline retail. While we would expect online buying to dampen slightly if gas prices continue to eat into discretionary spending, it's clear that e-commerce has become a mainstay in consumer behavior, driven by the attraction of both lower prices and convenience" (Fulgoni, 2011).

Most established enterprises operate in a holistic manner across several management functions (e.g. general and strategic management, production and operation, financial, purchasing, human resources, information, public relations and marketing) to use its given resources and produce a product or deliver a service that will ultimately address a consumer need. The focus of this study was not only to provide a departure point for understanding the topic of marketing function in the broad research field of business management, but also to investigate management practices connected with the marketing mix, specifically the informal communication of personal selling by means of electronic word-of-mouth communication.

Marketing communication in an e-commerce environment was focused on to assess the effect of personal transmission of marketing messages on credibility and purchase behaviour (Brujjs *et al.*, 1998:543). The results of this study revealed that informal communication of personal selling through the means of EWOM does indeed influence online sales rank levels. It was also confirmed that several EWOM factors (platform, text length, time and stars) influenced the review credibility and that other EWOM factors (sentiment, ex-sentiment and ex-stars) did not play that significant role according to the results.

The present study offers several industry recommendations for e-commerce enterprises without any formal review system, as well as already established e-commerce enterprises who wish to enhance EWOM in their online review communities.

6.3.1 Usability and user experience design

The presence of EWOM reviews on e-commerce platforms not only decreases consumer uncertainty by their social presence, but also serves to decrease information overload by allowing only quality product information (Chen *et al.*, 2008:18-25; Sormunen, 2009:34-35). This ultimately implies that through proper engineering of an online review system, an e-commerce platform could leverage their most helpful reviews to add greater value for online shoppers (Brynjolfsson *et al.*, 2006; Mudambi and Schuff, 2010:185-186).

When considering showcasing EWOM reviews to readers, it is recommended that time, text length and stars have a higher significance for found helpful degree than the review sentiment of new best-seller books. Thus, a higher precedence can be given to time (review posted), stars and text length than for sentiment when designing the visualisation of data on product review pages.

When representing reviews on e-commerce platforms, the study's results also indicate that it is important not only to show the positive star ratings, but also to reflect the negative star ratings that a given book has received. Tsang and Prendergast (2009:1277) have found that when enterprises do not hide negative outliers, they communicate to readers that they have no hidden agendas, which ultimately enhances credibility.

6.3.2 Forecasting, review insights and benchmark statistics

EWOM cultivated in an online review system are not only important to build platform trust and decrease search cost for online consumers, but is also important for the producers who sell their products through e-commerce vendors. By analysing positive and negative sentiment from review text, marketers can potentially forecast a product's production, inventory, and sales cycles, and adapt management functions accordingly (Tsang and Prendergast, 2009:1270).

Depending on the product-type, further review insights may also be collected by mining review text (Tsang and Prendergast, 2009:1271). A book's author may go through the statistics of review factors of his/her book to gain a deeper understanding of how his/her readers responded to the book. The given feedback can in turn be used to better formulate ideas for future book titles.

E-commerce enterprises may also gain insight into whether the services they provide are indeed satisfying their users. Users not only share their negative shopping experiences on online review platforms like *Epinion.com*, but also provide negative feedback in the form of a review about poor service delivery on product pages itself. By mining the feedback from these reviews, e-commerce enterprises can better satisfy the needs of their customers.

The results of this study, in particular model 1's findings, can be used as a benchmark statistic when comparing the found helpful degree of e-commerce enterprises with another. The review credibility (measured by found helpful degree) was on average 21.08% higher for

a review on *Amazon.com* than a review that is posted on *Barnesandnoble.com* at the time of the study. This result is important for *Amazon.com*'s business-to-business operation, especially the way they can market themselves to future book publishers and writers. *Amazon.com* can also leverage this benchmark statistic in marketing campaigns to retain current customers and attract new customers.

6.4 LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

The purchase decision-making, as well as the online review process, is in essence a complex exercise which may demand more comprehensive research models. Any research study usually has a set of limitations, which can serve as departure points for future research projects. The most common limitations refer to the study's sample (section 6.4.1) and measurement units (section 6.4.2) that were used to address the stated hypotheses (Blumberg *et al.*, 2008:567).

6.4.1 Future samples

Although this study's samples were extracted from a targeted book population across many genres, the results are only generalisable to those specific new best-seller book titles. One possible future direction of research may include not only books, but also a wider range of experience and search-product categories, such as mp3, DVDs and CDs (Mudambi and Schuff, 2010:192). It would be interesting to examine whether spotlight EWOM reviews, found helpful degree and sales rank levels differs significantly across product types.

Another limitation in terms of sample selection is a lack of comparative e-commerce platforms. This study only assessed EWOM communication and its relationship to review found helpful degree and sales rank levels on two e-commerce platforms, while other studies such as Pavlou and Dimoka (2006:392) also studied other e-commerce enterprises, such as *Ebay.com*. Future research in EWOM communication and natural language processing may also include other e-commerce vendors, not necessarily only from one country, but also from other countries across the world. Global e-commerce enterprises might include: *Buscapé.com.br* (Brazil), *Allegro.pl* (Eastern Europe), *Ricardo.ch* (Western Europe), and *Sulit.com.ph* (Philippines).

6.4.2 Future measurement units

The limitations with regards to measurement units that must be noted, are that no reviewer demographics, price changes or shipping estimates were included in either model 1 or model 2. The regression model (model 1) could be made more comprehensive by including other possible review antecedents, such as reviewer identity, demographics, and review status (Mudambi and Schuff, 2010:196). Future studies might include not only spotlight EWOM reviews in the hypothetical model (model 2), but also both product prices and shipping estimates, and ultimately test their impact on e-commerce sales rank values (Chevalier and Goolsbee, 2003:16- 27).

One other limitation of model 1 is the fact that no comparative sentiment classification software packages were used to test review text sentiment. Even though an alternative text analytics software package (e.g. Leximancer) was investigated, the study only used review sentiment scores obtained through *Uclassify's API* (Exon-Taylor, Smith and McFadden, 2011). It is suggested that future studies could compare the sentiment scores of different software packages with one another, and use the end-results of the software package that performed the best.

6.5 CONCLUSION

The primary objective of this study was to examine the influence of EWOM communication on the online review process on USA-based e-commerce platforms. Several secondary objectives were also identified while conducting the study. The secondary objectives were formulated to assess:

- a review's *star rating* influence on its *credibility*,
- a review's *ex-star rating* influence on its *credibility*,
- a review's *sentiment rating* influence on its *credibility*,
- a review's *ex-sentiment rating* influence on its *credibility*,
- a review's *text length* influence on its *credibility*,
- a review's *publish date* influence on its *credibility*,
- a review's *e-commerce platform* influence on its *credibility*,
- the influence that *EWOM review communication* exerts on the *sales rank* values of new best-seller books.

By using a non-probability judgement sampling procedure, it was observed that EWOM reviews do indeed influence the sales levels (as measured by log of sales rank) of e-commerce platform *Amazon.com*, and that certain review factors (platform, text length, time and stars) significantly influenced credibility (as measured by found helpful degree) of *Amazon.com* and *Barnesandnoble.com* reviews. It can be concluded that the primary objective, as well as the secondary objectives of the study, were appropriately addressed.

Ultimately, the results of this study can assist e-commerce enterprises to gain a greater understanding of their online users for better future strategies, such as user experience design, sales and inventory forecasting and benchmark statistics, for marketing campaigns.

LIST OF REFERENCES

Aaker, D.A., Kumar, V. & Day, G.S. 2001. *Marketing Research*. 7th ed. New York, USA: John Wiley & Sons Inc.

Aldridge, A., Forcht, K. & Pierson, J. 1997. Get linked or get lost: marketing strategy for the Internet. *Internet Research: Electronic Networking Applications and Policy*, 7(3):161-169.

Al Leen, J.Y., Ignatius, J. & Ramayah, T. 2009. Ranking Preferred Web 2.0 Applications of eWOM: An AHP Proposed Research Agenda. *American Journal of Scientific Research*, 5:27-32.

Amazon.com press release. 2011. *Q1 2011 Financial Results*. [Online]. Available: <http://bit.ly/gMYlr> [22 May 2012].

Anderson, E.W. 1998. Customer satisfaction and word of mouth. *Journal of Service Research*, 1(1):5-17.

Aristotle, trans. Roberts, W.R. 1924. *Rhetoric*. New York, USA: Oxford University Press.

Badenhorst, J.A., Cant, M.C., De J Cronje, G.J., Du Toit, G.S., Erasmus, B.J., Grobler, P.A., Krüger, L.P., Machado, R., De K Marais, A., Marx, J., Strydom, J.W. & Mpofo, R.T. 2003. *Introduction to Business Management*. 6th ed. Cape Town, South Africa: Oxford University Press Southern Africa.

Bansal, H.S. & Voyer, P.A. 2000. Word-of-mouth processes within a services purchase decision context. *Journal of Service Research*, 3(2):166-177.

Barnes and Noble press release. 2011. *Financial Results on 21 June 2011*. [Online]. Available: <http://bit.ly/ketMvU> [22 May 2012].

Berenson, M.L. & Levine, D.M. 1992. *Basic Business Statistics: Concepts and Applications*. 5th ed. New Jersey, USA: Prentice-Hall, Inc.

Blumberg, B., Cooper, D.R. & Schindler, S. 2008. *Business Research Methods*. 2nd European ed. London, UK: McGraw-Hill.

Breazeale, M. 2009. Word of mouse: An assessment of electronic word-of-mouth research. *International Journal of Market Research*, 51(3):297-318.

Brown, J., Broderick, A.J. & Lee, N. 2007. Word of mouth communication within online communities: conceptualizing the online social network. *Journal of Interactive Marketing*, 21(3):2-20.

Brujls, A., Crous, M.J., De Villers, J.A., De Wit, P.W.C., Du Plessis, A.P., De V Maasdorp, E.F., Mahabir, M., Marx, A.E., Nieman, G.H., Schreuder, A.N. & Van Vuuren, J.J. 1998. *Business Management*. 2nd ed. Pretoria, South Africa: Van Schaik Publishers.

Brynjolfsson, E., Hu, Y. & Smith, M. 2006. From Niches to Riches: Anatomy of the Long Tail. *Sloan Management Review*, 47(4):67-71.

Burns, A.C. & Bush, R.F. 2003. *Marketing Research: Online Research Applications*. 4th ed. New Jersey, USA: Pearson Prentice Hall.

Buttle, F.A. 1998. Word of Mouth: Understanding and managing referral marketing. *Journal of Strategic Marketing*, 6:241–254.

Cant, M.C., Van Heerden, C.H., Ngambi, H.C., Nieuwenhuizen, C., Roberts-Lombard, M., Brink, A., Dayan, O., Wiid, J., Koekemoer, L., Botha, J., Human, G., Bothma, C.H. & Machado, R. 2010. *Marketing Management: A South African Perspective*. Cape Town, South Africa: Juta & Co. Ltd.

Carl, W.J. 2006. What's all the buzz about? *Management Communication Quarterly*, 19(4):601-634.

Chaffey, D., Ellis-Chadwick, F., Mayer, R. & Johnston, K. 2009. *Internet Marketing*. 4th ed. Edinburgh, UK: Pearson Education Limited.

Chen, P., Wu, S. & Yoon, J. 2004. *The Impact of Online Recommendations and Consumer Feedback on Sales*. [Online]. Available: <http://bit.ly/hu9q6J> [16 February 2011].

Chen, P. & Wu, S. 2007. *Does Collaborative Filtering Technology Impact Sales? Empirical Evidence from Amazon.com*. [Online]. Available: <http://bit.ly/Nkhfpl> [27 June 2012].

Chen, P., Dhanasobhon, S. & Smith, M.D. 2008. *All Reviews are Not Created Equal: The Disaggregate Impact of Reviews and Reviewers at Amazon.com*. [Online]. Available: <http://bit.ly/OqaXCj> [16 February 2011].

Chevalier, J. & Goolsbee, A. 2003. *Measuring prices and price competition online: Amazon and Barnes and Noble*. [Online]. Available: <http://bit.ly/g6FIRX> [16 February 2011].

Chevalier, J. & Mayzlin, D. 2004. *The Effect of Word-of-Mouth on Sales: Online Book Reviews*. [Online]. Available: <http://bit.ly/gpt8c8> [18 July 2010].

Cho, C. & Khang, H. 2006. The State of Internet-Related Research in Communications, Marketing and Advertising: 1994-2003. *Journal of Advertising*, 35(3):143-165.

Churchill, G.A. & Iacobucci, D. 2002. *Marketing Research: Methodological Foundations*. 8th ed. Ohio, USA: Thomson South-Western.

Cialdini, R.B. 2000. *Influence: Science and Practice*. 4th ed. New York, USA: HarperCollins.

Coldwell, D. & Herbst, F. 2004. *Business Research*. 2nd ed. Cape Town, South Africa: Juta & Co. Ltd.

Cooper, D.R. & Schindler, P.S. 2011. *Business Research Methods*. 11th ed. New York, USA: McGraw-Hill/Irwin.

De Bruyn, A. & Lilien, G.L. 2008. A multi-stage model of word-of-mouth influence through viral marketing. *International Journal of Research in Marketing*, 25:151-163.

De Klerk, G.J., Kroon, J., Jansen van Rensburg, L.R., Lotz, J.O., Steyn, T.F.J. & Verreyne, M. 1997. *Business Management, an Introduction*. Pretoria, South Africa: Van Schaik Publishers.

Dellarocas, C. 2003. The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science*, 49(10):1407-1424.

Diamantopoulos, A. & Schlegelmilch, B.B. 1997. *Taking the fear out of data analysis: A step-by-step approach*. London, UK: The Dryden Press.

Dillon, W.R., Madden, T.J. & Firtle, N.H. 1993. *Essentials of Marketing Research*. USA: Irwin/McGraw-Hill.

Du Toit, C.M., Du Plessis, A.P. & Nortje, J.D. 1981. *Fundamental Business Economics*. Durban, South Africa: Butterworth & Co (PTY) Ltd.

East, R., Hammond, K. & Wright, M. 2007. The relative incidence of positive and negative word of mouth: A multi-category study. *International Journal of Research in Marketing*, 24(2):175-184.

East, R., Hammond, K. & Lomax, W. 2008. Measuring the impact of positive and negative word of mouth on brand purchase probability. *International Journal of Research in Marketing*, 25(3):215-224.

Elifson, K.W., Runyon, R.P. & Haber, A. 1990. *Fundamentals of Social Statistics*. 2nd ed. USA: McGraw-Hill, Inc.

Exon-Taylor, M., Smith, A. & McFadden, T. 2011. *Leximancer*. [Online]. Available: <http://bit.ly/N0hDM4> [11 July 2012].

Farhoomand, A. & Lovelock, P. 2001. *Global E-commerce*. Singapore: Prentice-Hall, Pearson Education Asia Pte Ltd.

Free Software Foundation, Inc. 2011. *GNU Wget 1.13.4 Manual*. [Online]. Available: <http://bit.ly/uk52m> [24 May 2012].

Fulgoni, G. 2011. *comScore Reports \$38 Billion in Q1 2011 U.S. Retail E-Commerce Spending, Up 12 Percent vs. Year Ago*. [Online]. Available: <http://bit.ly/j79qfm> [22 May 2012].

Glos, R.E., Steade, R.D. & Lowry, J.R. 1980. *Business: Its Nature and Environment*. 9th ed. Ohio, USA: South-Western Publishing Co.

Godes, D. & Mayzlin, D. 2003. *Using Online Conversations to Study Word of Mouth Communication*. [Online]. Available: <http://bit.ly/NWKY8X> [29 June 2012].

Granovetter, M. 1983. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1:203-233.

Hair, J.F., Anderson, R.E., Tatham, R.L. & Black, W.C. 1998. *Multivariate Data Analysis*. 5th ed. New Jersey, USA: Prentice-Hall.

Hennig-Thurau, T., Gwinner, K.P., Walsh, G. & Gremler, D.D. 2004. Electronic Word-of-Mouth Via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet? *Journal of Interactive Marketing*, 18(1):38-52.

Hennig-Thurau, T. & Walsh, G. 2004. Electronic Word of Mouth: Motives for and Consequences of Reading Customer Articulations on the Internet. *International Journal of Electronic Commerce*, 8(2):51-74.

Hirokawa, R.Y. & Lowe, J.B. 2002. *Katz and Lazarsfeld Revisited: Using Intermedia Theory to Enhance Health Campaigns*. [Online]. Available: <http://bit.ly/S70aR9> [27 June 2011].

Hoffman, D.L. & Novak, T.P. 1996. Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations. *The Journal of Marketing*, 60(3):50-68.

Howell, D.C. 1999. *Fundamental statistics for the behavioral sciences*. 4th ed. California, USA: Brooks/Cole Publishing Company.

Huang, J., Hsiao, T. & Chuang, S. 2009. *The role of virtual community in disseminating electronic word of mouth*. [Online]. Available: <http://bit.ly/LJ1L0I> [29 June 2012].

Huang, C., Lin, T. & Lin, K. 2009. Factors affecting pass-along email intentions (PAEIs): Integrating the social capital and social cognition theories. *Electronic Commerce Research and Applications*, 8:160-169.

Hu, N., Liu, L. & Zhang, J.J. 2008. Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Journal Information Technology and Management*, 9(3):201–214.

Imber, J. & Toffler, B. 2000. *Dictionary of Marketing Terms*. 3rd ed. Hauppauge, NY: Barrons Business Dictionaries.

Jansen, B.J., Zhang, M., Sobel, K. & Chowdury, A. 2009. Twitter Power: Tweets as Electronic Word of Mouth. *Journal of the American Society for Information Science and Technology*, 60(11):2169-2188.

Kågström, J., Karlsson, R. & Kågström, E. 2011. *Uclassify*. [Online]. Available: <http://bit.ly/M3nmPC> [11 July 2012].

Katz, E. & Lazarsfeld, P.F. 1955. *Personal influence: The part played by people in the flow of mass communications*. Illinois, USA: Free Press.

Keller, E. 2007. Unleashing the Power of Word of Mouth: Creating Brand Advocacy to Drive Growth. *Journal of Advertising Research*, 47(4):448-452.

Kotler, P. 1972. A Generic Concept of Marketing. *Journal of Marketing*, 36(2):46-54.

Kotler, P. & Keller, K.L. 2006. *Marketing Management*. 12th ed. New Jersey, USA: Pearson Education Inc.

Kumar, V., Aaker, D.A. & Day, G.S. 2002. *Essentials of Marketing Research*. 2nd ed. New York, USA: John Wiley & Sons Inc.

Laudon, K.C. & Traver, C.G. 2010. *E-commerce 2010*. 6th ed. New Jersey, USA: Pearson Education Inc.

Leskovec, J., Adamic, L.A. & Huberman, B.A. 2007. The Dynamics of Viral Marketing. *ACM Transactions on the Web*, 1(1):1-46

Li, C. & Bernoff, J. 2008. *Groundswell: Winning in a world transformed by social technologies*. Massachusetts, USA: Harvard Business School Publishing.

Liu, B. 2011. *Web Data Mining: Exploring Hyperlinks, Contents and Usage Data*. 2nd ed. Berlin, Germany: Springer-Verlag.

Machado, R., Strydom, J.W. & Cant, M.C. 1999. *The Foundations of Business*. Cape Town, South Africa: Juta & Co. Ltd.

Malhotra, N.K. 2002. *Basic Marketing Research: Applications to Contemporary Issues*. New Jersey, USA: Pearson Education, Inc.

Mazzarol, T., Sweeney, J.C. & Soutar, G.N. 2007. Conceptualizing word-of-mouth activity, triggers and conditions: An exploratory study. *European Journal of Marketing*, 41(11):1475-1494.

Merisavo, M. 2008. *The Interaction between Digital Marketing Communication and Customer Loyalty*. [Online]. Available: <http://bit.ly/h4Z0KI> [14 July 2010].

Morris, A.L. 2010. *Assessing brand image transfer in sponsorship*. Stellenbosch, South Africa: University of Stellenbosch. (MComm, Business Management).

Mudambi, S.M. & Schuff, D. 2010. What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, 34(1):185-200.

Neuman, W.R. 1991. *The Future of the Mass Audience*. Cambridge, UK: Cambridge University Press.

Nickels, W.G., McHugh, J.M. & McHugh, S.M. 2005. *Understanding Business*. 7th ed. New York, USA: The McGraw-Hill Companies Inc.

Pajuniemi, J. 2009. The Revolution of Brand Marketing: The Era of Virtual Consumer Communities, in: M. Koskela, & M. Vinnari (eds.), *Proceedings of the Conference "Future of the Consumer Society"*, Tampere, Finland, May 28-29, 2009, p.71-80.

Parkin, T. 2012. *Python Software Foundation*. [Online]. Available: <http://bit.ly/1njFvk> [16 July 2012].

Pavlou, P. & Dimoka, A. 2006. The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation. *Information Systems Research*, 17(4):392-414.

Paxhia, S. 2008. The Business Side of Social Networks. *Seybold Report: Analyzing Publishing Technologies*, 8(17):10.

Peterson, R.A., Balasubramanian, S. & Bronnenberg, B.J. 1997. Exploring the Implications of Internet for Consumer Marketing. *Journal of the Academy of Marketing Science*, 25(4):329-346.

Pitt, L.F., Berthon, P.R., Watson, R.T. & Zinkhan, G.M. 2002. The Internet and the birth of real consumer power. *Business Horizons*, 45(4):7-14.

Rafi, M., Fisher, R.J., Jaworski, B.J. & Cahill, A.M. 2001. *Internet Marketing: Building Advantage in a Networked Economy*. Boston, USA: McGraw-Hill/Irwin.

Richardson, L. 2012. *Beautiful Soup Documentation*. [Online]. Available: <http://bit.ly/Kvnc> [26 May 2012].

Richter, A. & Koch, M. 2008. *Functions of Social Networking Services*. [Online]. Available: <http://bit.ly/6v1W7C> [12 July 2012].

Riegner, C. 2007. *Word of Mouth on the Web: The Impact of Web 2.0 on Consumer Purchase Decisions*. [Online]. Available: <http://bit.ly/ii7snG> [15 July 2010].

Rosen, E. 2000. *The Anatomy of Buzz: How to Create Word-of-Mouth Marketing*. London, UK: Harper Collins Business.

Rowley, J. 2004. Online Branding. *Online Information Review*, 28(2):131-138.

Rownd, M. & Heath, C. 2008. *The American Marketing Association Releases New Definition for Marketing*. [Online]. Available: <http://bit.ly/jwl4tq> [29 Augustus 2012].

Sangiovanni-Vincentelli, A., Carloni, L., De Bernardinis, F. & Sgroi, M. 2004. *Benefits and Challenges for Platform-Based Design*. [Online]. Available: <http://bit.ly/MgqJTb> [16 July 2012].

Schiffman, L.G. & Kanuk, L.L. 2007. *Consumer Behavior*. 9th ed. New Jersey, USA: Pearson Education Inc.

Schmidt, J.B. & Spreng, R.A. 1996. A Proposed Model of External Consumer Information Search. *Journal of the Academy of Marketing Science*, 24(3):246-256.

Shimp, T.A. 1997. *Advertising, Promotion, and Supplement Aspects of Integrated Marketing Communications*. 4th ed. Florida, USA: The Dryden Press.

Sormunen, V. 2009. *International Viral Marketing Campaign Planning and Evaluation*. [Online]. Available: <http://bit.ly/LfS60n> [15 July 2010].

Stern, B. 1994. A revised model for advertising: multiple dimensions of the source, the message, and the recipient. *Journal of Advertising*, 23(2):5-16.

Struwig, F.W. & Stread, G.B. 2001. *Planning, designing and reporting research*. Cape Town, South Africa: Pearson Education South Africa.

Strydom, J., Cant, M., Jooste, C., Koekemoer, L., Bennet, A., Schreuder, A., Brink, A. & Machado, R. 2000. *Marketing Management*. Cape Town, South Africa: Juta & Co. Ltd.

Sweeney, J.C., Soutar, G.N. & Mazzarol, T. 2008. Factors influencing word of mouth effectiveness: receiver perspectives. *European Journal of Marketing*, 42 (3):344-364.

Tapscott, D. & Williams, A.D. 2008. *Wikinomics: How mass collaboration changes everything*. 2nd ed. USA: Penguin Group, Inc.

The Open Group. 2008. *Crontab - schedule periodic background work*. [Online]. Available: <http://bit.ly/Ozv4y3> [12 July 2012].

Thorson, K.S. & Rodgers, S. 2006. Relationships between Blogs as eWOM and Interactivity, Perceived Interactivity, and Parasocial Interaction. *Journal of Interactive Advertising*, 6(2):39-50.

Treiman, D.J. 2009. *Quantitative Data Analysis: Doing Social Research to Test Ideas*. San Francisco, USA: John Wiley & Sons, Inc.

Tsang, A.S. & Prendergast, G. 2009. Is a "star" worth a thousand words? *European Journal of Marketing*, 43(11):1269-1280.

Watts, D.J. & Dodds, P.S. 2007. Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*, 34:441-458.

Zikmund, W.G. & Babin, B.J. 2010. *Exploring Marketing Research*. 10th ed. South-Western: Cengage Learning.

ANNEXURE A: BOOK LIST

TABLE 4.2: TARGET POPULATION

Number	Title	Author	Included into final population (Yes or No)
1	Freedom: A Novel (Oprah's Book Club)	Jonathan Franzen	
2	The Orange Revolution	Gostick, Elton	No
3	Freedom: A Novel	Jonathan Franzen	No
4	The Grand Design	Hawking, Mlodinow	
5	Jon Stewart Presents Earth (The Book)	Jon Stewart	
6	Fall of Giants (The Century Trilogy)	Ken Follett	
7	Diary of a Wimpy Kid: The Ugly Truth	Jeff Kinney	
8	Safe Haven	Nicholas Sparks	
9	Mockingjay	Suzanne Collins	
10	Room: A Novel	Emma Donoghue	
11	The Lost Dogs	Jim Gorant	
12	White House Diary	Jimmy Carter	
13	The Paleo Solution	Robb Wolf, Loren Cordain	
14	Crimes Against Liberty	David Limbaugh	
15	The Wave	Susan Casey	
16	The Heroes of Olympus, Book One	Rick Riordan	
17	Towers of Midnight (Wheel of Time)	Jordan, Sanderson	
18	Strategic Stock Trading	Michael Swanson	
19	The Confession: A Novel	John Grisham	
20	Pinheads and Patriots	Bill O'Reilly	
21	A Journey: My Political Life	Tony Blair	
22	The Exile: An Outlander Graphic Novel	Gabaldon, Nguyen	
23	Happy Ever After (Bride Quartet)	Nora Roberts	
24	Hunger Games Trilogy Boxset	Suzanne Collins	
25	Mini Shopaholic: A Novel	Sophie Kinsella	
26	Mad As Hell	Rasmussen, Schoen	
27	American Assassin	Vince Flynn	
28	Making Our Democracy Work	Stephen Breyer	No
29	Wicked Appetite	Janet Evanovich	
30	Bad Blood: a Virgil Flowers novel	by John Sandford	
31	Tracy Anderson's 30-Day Method	Anderson, Paltrow	
32	Gunn's Golden Rules	Tim Gunn, Ada Calhoun	
33	The Reversal (Harry Bosch)	Michael Connelly	
34	Proofiness	Charles Seife	
35	Getting to Happy	Terry McMillan	
36	Barefoot Contessa How Easy Is That?	Ina Garten	
37	Naked Heat (Nikki Heat)	Richard Castle	
38	Half Broke Horses: A True-Life Novel	Jeannette Walls	
39	Worth Dying For	Lee Child	
40	Halo: Reach Limited Edition Guide	BradyGames	No
41	Hell's Corner	David Baldacci	
42	True Prep: It's a Whole New Old World	Lisa Birnbach, Chip Kidd	No
43	The Warmth of Other Suns	Isabel Wilkerson	
44	Her Daughter's Dream (Marta's Legacy)	Francine Rivers	
45	Cake Pops	Bakerella, Angie Dudley	
46	The Roots of Obama's Rage	Dinesh D'Souza	
47	Squirrel Seeks Chipmunk	David Sedaris	
48	My Mommy Hung the Moon: A Love Story	Jamie Lee Curtis, Laura Cornell	No
49	The Book of the Dead	John Mitchinson, John Lloyd	No
50	Overhaul	Steven Rattner	No
51	Assholes Finish First	Tucker Max	
52	Third World America	Arianna Huffington	No
53	The Chronicles of Vladimir Tod	Heather Brewer	
54	Clockwork Angel (Book 1)	Cassandra Clare	
55	Power Thoughts	Joyce Meyer	
56	In the Company of Others	Jan Karon	

57	Ranger's Apprentice, Book 9	John Flanagan	
58	Torment (Fallen)	Lauren Kate	
59	Of Thee I Sing: A Letter to My Daughters	Barack Obama	
60	The Tiger	John Vaillant	
61	Civilization V Official Strategy Guide	BradyGames	No
62	Half a Life	Darin Strauss	No
63	Last Sacrifice (Book 6)	Richelle Mead	
64	The Fall: Book Two of the Strain Trilogy	Del Toro, Hogan	
65	Decision Points	George W. Bush	
66	Zero History	William Gibson	
67	I Shall Wear Midnight	Terry Pratchett	
68	Outlive Your Life	Max Lucado	
69	The 39 Clues, Book 10	Margaret Peterson Haddix	
70	Ape House: A Novel	Sara Gruen	
71	Forever Young	Nicholas Perricone	
72	Too Big to Fail	Andrew Ross Sorkin	
73	Awakened (House of Night)	P. C. Cast, Kristin Cast	
74	Bob Dylan In America	Sean Wilentz	No
75	Young Guns	Cantor, Ryan, McCarthy	
76	Choosing to SEE	Chapman,C.Chapman, Vaughn	
77	Wolf Hall: A Novel	Hilary Mantel	
78	The Twelfth Imam	Joel C. Rosenberg	
79	The Widower's Tale: A Novel	Julia Glass	
80	Aftershock	Robert B. Reich	
81	Fat Witch Brownies	Patricia Holding	
82	The Calculus Diaries	Jennifer Ouellette	No
83	Full Dark, No Stars	Stephen King	
84	Don't Blink	Patterson, Roughan	
85	Warlord: An Alex Hawke Novel	Ted Bell	
86	The Way of Kings	Brandon Sanderson	
87	100 Simple Things You Can Do	Jean Carper	
88	Guinness World Records 2011	Guinness World Records	
89	Fallout	Ellen Hopkins	
90	Never Let Me Go	Kazuo Ishiguro	
91	Waiting for "SUPERMAN"	Karl Weber	
92	Storyteller: Roald Dahl	Donald Sturrock	No
93	The Fort	Bernard Cornwell	
94	Side Jobs	Jim Butcher	
95	No Mercy (Dark-Hunter)	Sherrilyn Kenyon	
96	Dexter Is Delicious	Jeff Lindsay	
97	The Art of Non-Conformity	Chris Guillebeau	
98	Muhammad	Deepak Chopra	
99	Rules Compendium	Wyatt, Collins, Heinsoo	No
100	Knuffle Bunny Free	Mo Willems	

ANNEXURE B: WEB PAGE DATA

FIGURE 4.5: THE TOP-HALF OF THE SPOTLIGHT AMAZON.COM BOOK REVIEW PAGE (FREEDOM - AN OPRAH NOVEL)

Freedom: A Novel (Oprah's Book Club) [Hardcover]
 Jonathan Franzen (Author)
 ★★★★★ (182 customer reviews)

List Price: ~~\$28.00~~
 Price: **\$14.00** & eligible for **FREE Super Saver Shipping** on orders over \$25.
 You Save: **\$14.00 (50%)**

Product Details

Hardcover: 576 pages
Publisher: Farrar, Straus and Giroux (August 31, 2010)
Language: English
ISBN-10: 0312600844
ISBN-13: 978-0312600846
Product Dimensions: 9 x 6.1 x 1.8 inches
Shipping Weight: 2 pounds ([View shipping rates and policies](#))
Average Customer Review: ★★★★★ (182 customer reviews)
Amazon Bestsellers Rank: #5 in Books ([See Top 100 in Books](#))

Would you like to [update product info](#) or [give feedback on images](#)?

Source: Amazon.com (Captured at the beginning of October 2010)

FIGURE 4.6: THE BOTTOM-HALF OF THE SPOTLIGHT AMAZON.COM BOOK REVIEW PAGE (FREEDOM - AN OPRAH NOVEL)



Source: Amazon.com (Captured at the beginning of October 2010)

FIGURE 4.7: THE SPOTLIGHT BARNESANDNOBLE.COM BOOK REVIEW PAGE (FREEDOM - AN OPRAH NOVEL)

Freedom
by Jonathan Franzen
(Hardcover)

Reader Rating: ★★★★★ (404 ratings)
[See All Reviews](#) [Write a Review](#)

Pub. Date: September 2010
 ISBN: 9780312500546
 Sales Rank: 23

OTHER FORMATS

eBook	\$14.99
Hardcover - First Edition	\$16.99
Compan Disc - Unabridged, 9 CDs	\$43.99

NEW FROM BN.COM SPEND \$25, GET FREE SHIPPING
 \$16.99 Online Price (You Save 41%)
[Add to Cart](#)

NEW & USED FROM OUR TRUSTED SELLERS
 New starting at \$16.00 (You Save 2%)
 Used starting at \$24.00
[See All Prices](#)

PICK ME UP
 Reserve & pick up in 60 minutes at your local store.
 Enter your zip:
[Find In-Store](#)

Customers who bought this also bought

Overview Editorial Reviews Customer Reviews Meet the Writer Features

Customer Reviews

Reader Rating: ★★★★★
 Ratings: 404 Reviews: 117
[See All Reviews](#)

WRITE A REVIEW

Be the first to write a review!

Just too much...
by MattC1
 Reader Rating: ★★★★★
 Ratings: 1036 Reviews: 53
[See All Reviews](#)
 The entire time I wanted to love it, and just couldn't.

not recommended
by sereckviller
 Reader Rating: ★★☆☆☆
 See Detailed Ratings: 5
 How did it make it to Oprah's list. The sample looked good, so I bought it. Up until about page 137 it is fine. I was really relating to the characters, then it drags on and on with the author expressing his thoughts. The spell is broken - I do not think will finish the one.

Extremely disappointing!
by Mammatac

Synopsis
 From the National Book Award-winning author of *The Corrections*, a daily comedy novel about family.

The Barnes & Noble Review
 Franzen frames his story through the eyes of neighbors, observers physically close but emotionally at a remove. What's impressive to my mind about Franzen's choice to begin and end at this distance, so that the retrospective judgment of neighbors turns out to be pretty good and yet always askew. But for the rest of us all our privileged closer vantage points, Franzen has included a lot of observational data — more than 500 pages worth — so that we can see those little markings that make all the difference, while trying to hold moral binoculars up far so long can make the characters wobble between likability and loathsomeness, the development of the muscles of sympathy and judgment it encourages is all to the good — if we can avoid relaxing into the comforts of snugness which Franzen sometimes also puts on display. It's our choice, readers, after all, have certain freedoms, too.

Read the Full Review

Excerpt — Benjamin Alsop
 ...a great novel. While his contemporaries content themselves with small books about nothing much or big books about cinema, Franzen delivers the massive, old-school yarn. It's not that Franzen's prose makes other writers seem unlettered; it's that he makes them seem so lazy, so irrelevant, as lacking in the kind of Utopian we once expected here our best authors. Franzen doesn't name check War and Peace for nothing. It's making a claim for shelf space among the kind of books that the big dogs used to write. The kind they called important. The kind they called great.

[More Reviews and Recommendations](#)

Source: *Barnesandnoble.com* (Captured at the beginning of October 2010)

FIGURE 4.8: BACKLOG AMAZON.COM RATINGS AND REVIEWS (FREEDOM - AN OPRAH NOVEL)

The screenshot displays the Amazon.com product page for the book 'Freedom' by Jonathan Franzen. The page features a star rating system with 4.5 stars, a 'Customer Reviews' section with a search bar, and a list of reviews. The reviews are dated from July 2010 to July 2011. The first review is by G. Cole 'The Bookish' (Indanepoli, IN) dated July 29, 2010. The second review is by Lady Jane (Berford, VA United States) dated July 28, 2011. The third review is by Elizabeth Miller (Pittsburgh, Pennsylvania) dated July 28, 2011. The fourth review is by chadla 'The Pittsburgher' (Pittsburgh, Pennsylvania United States) dated July 14, 2011. The reviews discuss the book's themes, including family, class, and the American dream, and mention the author's previous work, 'The Corrections'.

Source: Amazon.com (Captured at the beginning of October 2010)

FIGURE 4.9: BACKLOG BARNESANDNOBLE.COM RATINGS AND REVIEWS (FREEDOM - AN OPRAH NOVEL)

Sort by: [Most Recent](#) | [Highest to Lowest](#) | **[Most Helpful](#)** Showing 1 - 5 of 176 | [Next](#)

★★★★☆
Bookbabe53
September 15, 2010

Sad, but still an interesting read

Not as great as *The Corrections*, but...there is something about Franzen's writing that makes it difficult for me to put it down. This story was so realistic in many ways to my experiences and those of my friends (predominately boomers), as well as our adolescent and young adult kids. He really gets family ties and the effects of depression on relationships. I felt like I knew Patty. You feel like you are right there... watching or experiencing everything, especially the effects of 9/11 on everything (I live near DC and he gets everything right about it). I would read him again and again. Recommend for book clubs.

41 of 48 people found this review helpful

Was this review helpful? [Yes](#) | [No](#) | [Report this review](#)

★☆☆☆☆
onlydlk
September 17, 2010

Priced too high...

I agree with crabsomatic on not purchasing until the price comes down to \$9.99. I can wait or get it at the library even though I do want to read it now. An e-book that has NOT been printed with ink(\$\$) on paper(\$\$) bound (\$\$)transported (\$\$) and handled numerous times en route (\$\$) is outrageously over priced at \$12.99!

35 of 71 people found this review helpful

Was this review helpful? [Yes](#) | [No](#) | [Report this review](#)

★☆☆☆☆
Tammy01
September 24, 2010

What'chu talkin 'bout Oprah

Not sure why Oprah picked this book, it was hard to read, I put it down and re-started it five times, I finally finished it, but kind of wished I would not have. Very boring, self serving story about a not so normal family. I would not recommend this to anyone, sorry Oprah. If you are looking for a good read, try *Unconditional*. It is well written story, with well developed characters. The plot twist and ending will certainly surprise you, and I guarantee that you wont put it down until you reach the end.

26 of 33 people found this review helpful

Was this review helpful? [Yes](#) | [No](#) | [Report this review](#)

I Also Recommend:

★★★★★
SHARON39
September 18, 2010

There is no Fairy Tale here.

FREEDOM is the story of our time, our concerns. Beautifully written, this amazing novel has a lot of entertaining satire, witty metaphors, deep insights, and a smattering of politics, (all sides). He works in what makes people tick, self-awareness and growth.human nature. The baby boomer generation, in search of the almighty self-gratification, apparently, have destroyed

Source: *Barnesandnoble.com* (Captured at the beginning of October 2010)

FIGURE 4.10: SCREENSHOT OF THE SCRAPED WEB PAGES ON THE PERSONAL COMPUTER

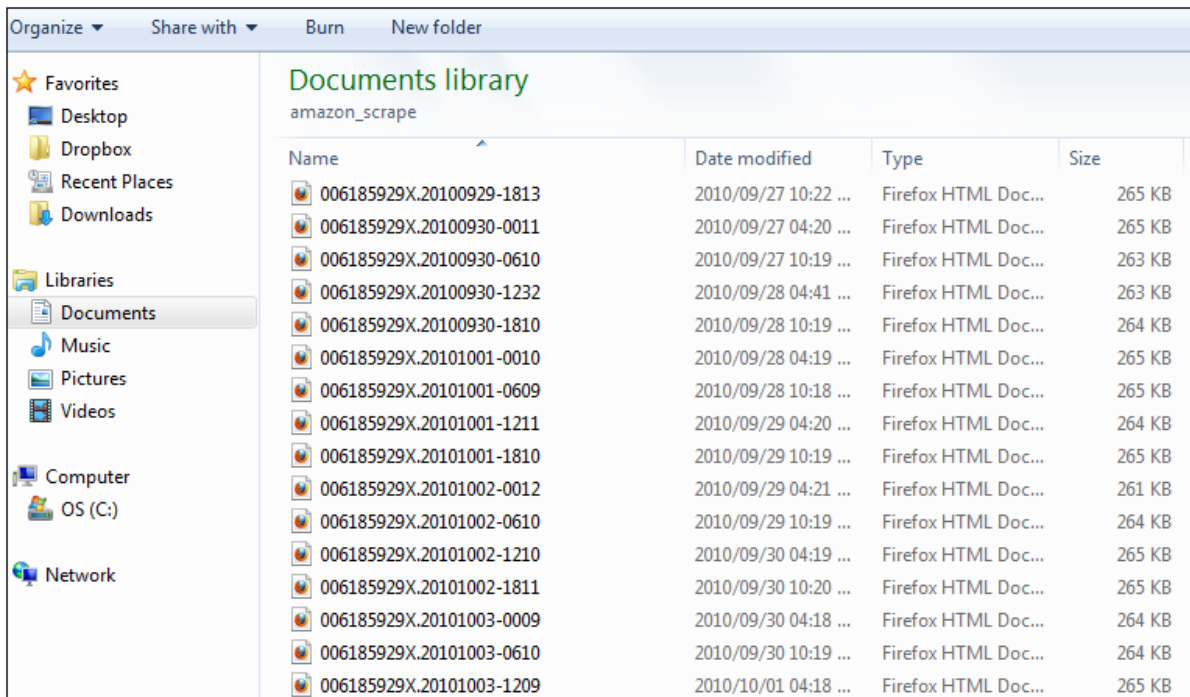


FIGURE 4.12: PROCESSING OF BACKLOG BARNESANDNOBLE.COM REVIEW DATA

The screenshot displays a list of reviews from BarnesandNoble.com, sorted by 'Most Helpful'. The interface includes a sorting menu at the top with options for 'Most Recent', 'Highest to Lowest', and 'Most Helpful', and a pagination indicator showing 'Showing 1 - 5 of 176' with a 'Next' button.

The first review is by 'Bookbabe53' (5 stars), dated September 15, 2010. The title is 'Sad, but still an interesting read'. The text discusses Franzen's writing style and its realism. It states '41 of 48 people found this review helpful' and includes a 'Was this review helpful?' section with 'Yes', 'No', and 'Report this review' options.

The second review is by 'onlydik' (4 stars), dated September 17, 2010. The title is 'Priced too high...'. The text complains about the high price of the e-book compared to the library and paper versions. It states '35 of 71 people found this review helpful' and includes the same helpfulness options.

The third review is by 'Tammy01' (4 stars), dated September 24, 2010. The title is 'What'chu talkin 'bout Oprah'. The text expresses frustration with Oprah's book choice and recommends 'Unconditional'. It states '26 of 33 people found this review helpful' and includes the helpfulness options. Below the review, there is a section for 'I Also Recommend:' with a small book cover image.

The fourth review is by 'SHARON39' (5 stars), dated September 18, 2010. The title is 'There is no Fairy Tale here.'. The text discusses the novel 'FREEDOM' and its satirical and insightful nature. It does not show a helpfulness count or options for this review.

Source: *Barnesandnoble.com* (Captured at the beginning of October 2010)

FIGURE 4.13: EXCEL ALGORITHM FOR IDENTIFYING FOUND HELPFUL (MODEL 1-STEP 1)

	FH	stars	begin review	end review	review text	review text for filter
1						
2	1	0	0	0	f 1 people found the following r	0
3	0	1	0	0	owing review helpful: 4.0 out of	0
4	0	0	0	0	wing review helpful: 4.0 out of 5	0
5	0	0	0	0	ut of 5 stars Great but Short, Jan	0
6	0	0	0	0	stars Great but Short, January 30,	0

=IF(OR(AND(ISNUMBER(SEARCH("By",A4)), ISNUMBER(SEARCH("out of",A3))),AND(ISNUMBER(SEARCH("found the following review helpful",A2)), ISNUMBER(SEARCH("out of",A3)))),1,0)

FIGURE 4.14: EXCEL ALGORITHM FOR IDENTIFYING STARS (MODEL1-STEP 1)

	FH	stars	begin review	end review	review text	review text for filter
1						
2	1	0	0	0	f 1 people found the following r	0
3	0	1	0	0	owing review helpful: 4.0 out of	0
4	0	0	0	0	wing review helpful: 4.0 out of 5	0
5	0	0	0	0	ut of 5 stars Great but Short, Jan	0
6	0	0	0	0	stars Great but Short, January 30,	0

=IF(B1=1,1,0)

FIGURE 4.15: EXCEL ALGORITHM FOR IDENTIFYING WHERE A REVIEW BEGINS (MODEL 1-STEP 1)

	FH	stars	begin review	end review	review text	review text for filter
1						
2	1	0	0	0	f 1 people found the following r	0
3	0	1	0	0	owing review helpful: 4.0 out of	0
4	0	0	0	0	wing review helpful: 4.0 out of 5	0
5	0	0	0	0	ut of 5 stars Great but Short, Jan	0
6	0	0	0	0	stars Great but Short, January 30,	0

=IF(ISNUMBER(SEARCH("This review is from:",A2)),1,0)

FIGURE 4.16: EXCEL ALGORITHM FOR IDENTIFYING WHERE A REVIEW ENDS (MODEL 1-STEP 1)

	A	B	C	D	E	F	G	H
1		FH	stars	begin review	end review	review text	review text for filter	
2	1 of 1 people found the following review helpful:	1	0	0	0	1 of 1 people found the following r	0	
3	4.0 out of 5 stars Great but Short, January 30, 2011	0	1	0	0	owing review helpful: 4.0 out of	0	
4	By	0	0	0	0	wing review helpful: 4.0 out of 5	0	
5	Phil Virgo (Gaithersburg, MD) - See all my reviews	0	0	0	0	ut of 5 stars Great but Short, Jan	0	
6	(REAL NAME)	0	0	0	0	stars Great but Short, January 30,	0	

=IF(AND(ISNUMBER(SEARCH("Help other customers find the most ",A2)),ISNUMBER(SEARCH("Was this review helpful to you?",A3))),1,0)

FIGURE 4.17: EXCEL ALGORITHM FOR IDENTIFYING A REVIEW'S ENTIRE TEXT STRING (MODEL 1-STEP 1)

	A	B	C	D	E	F	G	H
1		FH	stars	begin review	end review	review text	review text for filter	
2	1 of 1 people found the following review helpful:	1	0	0	0	1 of 1 people found the following r	0	
3	4.0 out of 5 stars Great but Short, January 30, 2011	0	1	0	0	owing review helpful: 4.0 out of	0	
4	By	0	0	0	0	wing review helpful: 4.0 out of 5	0	
5	Phil Virgo (Gaithersburg, MD) - See all my reviews	0	0	0	0	ut of 5 stars Great but Short, Jan	0	
6	(REAL NAME)	0	0	0	0	stars Great but Short, January 30,	0	

=IF(D1=1,A2,F1&"&A2)

FIGURE 4.18: EXCEL ALGORITHM FOR CATEGORIZING THE REVIEW'S TEXT STRING (MODEL 1-STEP 1)

	A	B	C	D	E	F	G	H
1		FH	stars	begin review	end review	review text	review text for filter	
2	1 of 1 people found the following review helpful:	1	0	0	0	1 of 1 people found the following r	0	
3	4.0 out of 5 stars Great but Short, January 30, 2011	0	1	0	0	owing review helpful: 4.0 out of	0	
4	By	0	0	0	0	wing review helpful: 4.0 out of 5	0	
5	Phil Virgo (Gaithersburg, MD) - See all my reviews	0	0	0	0	ut of 5 stars Great but Short, Jan	0	
6	(REAL NAME)	0	0	0	0	stars Great but Short, January 30,	0	

=IF(E3=1,1,0)

FIGURE 4.19: EXCEL'S SORT AND FILTER FUNCTION (MODEL 1 - STEP 1)

B2		=IF(OR(AND(ISNUMBER(SEARCH("By",A4)), ISNUMBER(SEARCH("out of",A3 SEARCH("out of",A3)))),1,0)				
	A	B	C	D	E	
1		found helpful	stars	begin review	end review	review text
2		1	0	0	0	review text
3	5.0 out of 5 stars Very educational if I w	0	1	0	0	review text 5.0 out of 5 stars
4	By	0	0	0	0	review text 5.0 out of 5 stars
5	D. Schaffer (Charlotte, NC USA) - See all	0	0	0	0	review text 5.0 out of 5 stars
6	(REAL NAME)	0	0	0	0	review text 5.0 out of 5 stars
7	Amazon Verified Purchase(What's this?	0	0	0	0	review text 5.0 out of 5 stars
8	This review is from: The Daily Show wit	0	0	1	0	review text 5.0 out of 5 stars
9	I keep this book at work and read a few	0	0	0	0	I keep this book at work and r
10	Help other customers find the most hel	0	0	0	1	I keep this book at work and r
11	Was this review helpful to you? Yes No	0	0	0	0	I keep this book at work and r
12	Report abuse Permalink	0	0	0	0	I keep this book at work and r
13	Comment Comment	0	0	0	0	I keep this book at work and r

TABLE 4.7: BARNESANDNOBLE.COM DATA (MODEL 1 - STEP 1)

	A	B	C
1	Found Helpful	Stars rating	Review text
2	1 out of 1 people found this review helpful.	5	can I possibly say than that it was just an amazing
3	1 out of 1 people found this review helpful.	5	kind of hard to follow in the beginning, but overall it
4	1 out of 1 people found this review helpful.	5	gave me great insight as to what its like to be a
5	4 out of 5 people found this review helpful.	2	Spoilers* For those who are not aware of the premise,
6	1 out of 1 people found this review helpful.	4	confusing begining to me and the ending..... but I
7	1 out of 1 people found this review helpful.	4	and like kids. i found jack to be advanced for his age,
8	1 out of 1 people found this review helpful.	5	was unlike anything I'd ever read before. I absolutely
9	1 out of 1 people found this review helpful.	4	BOOK....starts off kinda of weird but hang in there.
10	0 out of 1 people found this review helpful.	5	forward to read it any suggestions or comments
11	1 out of 1 people found this review helpful.	5	book with a first point view, from Jack's view. best
12	1 out of 1 people found this review helpful.	5	story of a Jack and his Ma who have been living in a
13	1 out of 1 people found this review helpful.	5	Book, Has Something About It That Makes You Keep
14	1 out of 1 people found this review helpful.	5	beautiful, wonderful book about a boy and his mom
15	1 out of 1 people found this review helpful.	3	interesting read.
16	2 out of 4 people found this review helpful.	2	told via a 5 year old was difficult to get into. The story

FIGURE 4.20: EXCEL ALGORITHM FOR WORD COUNT (MODEL 1 - STEP 2)

	A	B	C	D	E	F	G	H
1	1	22 March = dc	January 31, 2011, 7:24 PM	0.00	1	kept reading to see if it ever got better - never did! Not s	28	
2	2	22 March = dc	January 30, 2011, 4:52 PM	0.00	5	AFTER I'D FINISHED READING THIS NOVEL, I WAS LFEFT WA	59	
3	3	22 March = dc	January 28, 2011, 9:14 AM	0.00	2	If you are middle age this book is very depressing.	10	
4	4	22 March = dc	January 27, 2011, 10:01 PM	0.00	5	One of those books that you never want to end. Unforget	18	
5	5	22 March = dc	January 27, 2011, 12:51 PM	0.71	5	I am really surprised by some of the negative reviews of t	65	

=IF(LEN(TRIM(F1))=0,0,LEN(TRIM(F1))-LEN(SUBSTITUTE(F1," ",""))+1)

TABLE 4.9: BARNESANDNOBLE.COM DATA (MODEL 1 - STEP 2)

	A	B	C	D	E
1	published date	FH	Stars	review text	text length
2	January 31, 2011, 7:24 PM	0.00	1	kept reading to see if it ever got better - never did! Not s	28
3	January 30, 2011, 4:52 PM	0.00	5	AFTER I'D FINISHED READING THIS NOVEL, I WAS LFEFT WA	59
4	January 28, 2011, 9:14 AM	0.00	2	If you are middle age this book is very depressing.	10
5	January 27, 2011, 10:01 PM	0.00	5	One of those books that you never want to end. Unforget	18
6	January 27, 2011, 12:51 PM	0.71	5	I am really surprised by some of the negative reviews of t	65
7	January 25, 2011, 8:02 AM	0.00	5	Awesome book ...everybody must read this book it may b	16
8	January 24, 2011, 2:12 PM	0.00	1	Even if you agree with the liberal preaching, this book is e	51
9	January 22, 2011, 3:16 PM	0.00	1	I can not understand why anyone would recommend this	78
10	January 22, 2011, 1:51 PM	0.00	2	Very disappointing. I finished it but only because I kept l	24
11	January 20, 2011, 7:33 PM	0.00	3	Long, tedious, uninspired wreck of a novel bogged down	103
12	January 18, 2011, 10:48 AM	0.00	3	I came to this book expecting a chronicling of a young cou	180
13	January 16, 2011, 7:14 PM	0.00	5	Jonathan Franzen doesn't write books to make people ha	78
14	January 13, 2011, 4:06 PM	0.00	1	None of the characters had a single redeeming quality; as	24
15	January 13, 2011, 2:47 PM	0.00	5	This author is hard to put down. The only criticism I have	47
16	January 13, 2011, 11:39 AM	0.00	3	Although it took me a while to get into the story, I ended	98

TABLE 4.12: BARNESANDNOBLE.COM DATA (MODEL 1 - STEP 3)

1	publish date	FH	stars	review text	text length	pos sent	neg sent
2	January 31, 2011, 7:24 PM	0.00	1	kept reading to see if it ever got better - never did! Not s	27	0.637724	0.362276
3	January 30, 2011, 4:52 PM	0.00	5	AFTER I'D FINISHED READING THIS NOVEL, I WAS LFEFT WA	58	0.0174993	0.982501
4	January 28, 2011, 9:14 AM	0.00	2	If you are middle age this book is very depressing.	9	0.410734	0.589266
5	January 27, 2011, 10:01 P	0.00	5	One of those books that you never want to end. Unforget	17	0.18415	0.81585
6	January 27, 2011, 12:51 P	0.71	5	I am really surprised by some of the negative reviews of t	64	0.400384	0.599616
7	January 25, 2011, 8:02 AM	0.00	5	Awesome book ...everybody must read this book it may b	15	0.0361803	0.96382
8	January 24, 2011, 2:12 PM	0.00	1	Even if you agree with the liberal preaching, this book is e	50	0.928172	0.0718283
9	January 22, 2011, 3:16 PM	0.00	1	I can not understand why anyone would recommend this	77	0.804954	0.195046
10	January 22, 2011, 1:51 PM	0.00	2	Very disappointing. I finished it but only because I kept l	23	0.702842	0.297158
11	January 20, 2011, 7:33 PM	0.00	3	Long, tedious, uninspired wreck of a novel bogged down	102	0.843726	0.156274
12	January 18, 2011, 10:48 A	0.00	3	I came to this book expecting a chronicling of a young cou	179	0.498592	0.501408
13	January 16, 2011, 7:14 PM	0.00	5	Jonathan Franzen doesn't write books to make people ha	77	0.329797	0.670203
14	January 13, 2011, 4:06 PM	0.00	1	None of the characters had a single redeeming quality; as	23	0.995485	0.00451456
15	January 13, 2011, 2:47 PM	0.00	5	This author is hard to put down. The only criticism I have	46	0.111584	0.888416
16	January 13, 2011, 11:39 A	0.00	3	Although it took me a while to get into the story, I ended	97	0.118446	0.881554

TABLE 4.14: AMAZON.COM AND BARNESANDNOBLE.COM INDEX (MODEL 1 – STEP 5)

	A	B
1	Book titile	Index nmr
2	100 Simple Things You Can Do to Prevent Alzheimer's and Age-Related Memory	1
3	A Journey: My Political Life	2
4	Aftershock: The Next Economy and America's Future	3
5	American Assassin: A Thriller (Mitch Rapp)	4
6	Ape House: A Novel	5
7	Assholes Finish First	6
8	Awakened (House of Night)	7
9	Bad Blood: a Virgil Flowers novel	8
10	Barefoot Contessa How Easy Is That?: Fabulous Recipes & Easy Tips	9
11	Cake Pops: Tips, Tricks, and Recipes for More Than 40 Irresistible Mini Treats	10
12	Choosing to SEE: A Journey of Struggle and Hope	11
13	Clockwork Angel (The Infernal Devices, Book 1)	12
14	Crimes Against Liberty: An Indictment of President Barack Obama	13
15	Decision Points	14
16	Dexter Is Delicious	15
17	Diary of a Wimpy Kid: The Ugly Truth	16

TABLE 4.15: ELAPSED TIME CALCULATION (MODEL 1 – STEP 5)

	A	B	C	D	E	F	G	H	I	J
1	Platform	Book name	Date Review	Date FH	Found Helpful	Stars Rating	Text	Neg	Pos	Time
2			was Posted	was Collected	Value		Length	Sentiment	Sentiment	available
3										
7	2	72	40501	40626	0	1	0	0.5	0.5	125
8	2	43	40512	40625	0.052631579	1	0	0.5	0.5	113
9	2	37	40535	40625	0	2	0	0.5	0.5	90
10	2	14	40518	40625	0	4	0	0.5	0.5	107
11	2	38	40542	40624	0	4	0	0.5	0.5	82
12	2	43	40542	40625	0	5	0	0.5	0.5	83
13	2	49	40551	40624	0	5	0	0.5	0.5	73
14	2	57	40540	40624	0	5	0	0.5	0.5	84
15	2	57	40566	40624	0	5	0	0.5	0.5	58
16	2	63	40552	40624	0	5	0	0.5	0.5	72
17	2	38	40542	40624	0	1	0	1	1.80093E-18	82
18	2	56	40546	40625	0	5	0	0.0020697	0.99793	79
19	2	17	40530	40626	0	3	0	0.0501116	0.949888	96
20	2	63	40545	40624	0	4	0	0.0501116	0.949888	79

TABLE 4.16: EX-SENT AND EX-STARS CALCULATION (MODEL 1 – STEP 5)

	A	B	C	D	E	F	G	H
1	FH	PLATFORM	STARS	TEXTLENGTH	SENT	TIME	EXSENT	EXSTARS
2	0.4	1	5	13	0.783479	166	0.283479	2
3	1	1	5	13	0.113309	143	0.386691	2
4	0	1	5	13	0.00127959	139	0.49872041	2
5	1	1	5	13	0.932649	167	0.432649	2
6	0	1	5	14	0.999113	138	0.499113	2
7	0.395604396	1	5	14	0.0959424	140	0.4040576	2
8	1	1	5	14	0.0106394	132	0.4893606	2
9	0.6	1	5	14	0.999136	171	0.499136	2
10	0	1	5	14	0.996196	144	0.496196	2
11	1	1	5	14	0.981626	164	0.481626	2
12	0.5	1	5	14	0.910114	140	0.410114	2
13	0	1	5	14	0.900844	165	0.400844	2
14	0	1	5	14	0.891238	168	0.391238	2
15	0.25	1	5	14	0.880472	134	0.380472	2
16	0	1	5	14	0.852617	156	0.352617	2

FIGURE 4.21: PROCESSING OF AMAZON.COM MAIN LANDING WEB PAGE DATA - TOP HALF OF THE WEB PAGE (MODEL 2)

Freedom: A Novel (Oprah's Book Club) [Hardcover]
 Jonathan Franzen (Author)
 ★★★★★ (182 customer reviews)

List Price: ~~\$28.00~~
 Price: **\$14.00** & eligible for **FREE Super Saver Shipping** on orders over \$25.
 You Save: **\$14.00 (50%)**

Product Details
Hardcover: 576 pages
Publisher: Farrar, Straus and Giroux (August 31, 2010)
Language: English
ISBN-10: 0312600844
ISBN-13: 978-0312600846
Product Dimensions: 9 x 6.1 x 1.8 inches
Shipping Weight: 2 pounds ([View shipping rates and policies](#))
Average Customer Review: ★★★★★ (182 customer reviews)
Amazon Bestsellers Rank: #5 in Books ([See Top 100 in Books](#))

Would you like to [update product info](#) or [give feedback on images](#)?

Source: Amazon.com (Captured at the beginning of October 2010).

FIGURE 4.22: PROCESSING OF AMAZON.COM MAIN LANDING WEB PAGE DATA - LEFT HAND SIDE ON THE BOTTOM HALF OF THE WEPAGE (MODEL 2)



Source: Amazon.com (Captured at the beginning of October 2010).

FIGURE 4.23: PROCESSING OF AMAZON.COM MAIN LANDING WEB PAGE DATA - RIGHT HAND SIDE ON THE BOTTOM HALF OF THE WEB PAGE (MODEL 2)

Most Recent Customer Reviews

★★★★★ These are my peeps!
Freedom is one of the finest, most enjoyable works of fiction I've read in a long time. The voices of the characters are distinct and engaging. [Read more](#)
Published 2 hours ago by Greenbyoo

★★★★★ Believe the hype - mainly
This book is so great; I just wish it got a little more publicity. (Tip your waitresses; I'll be here all week). [Read more](#)
Published 4 hours ago by Scott Huizenga

★★★★★ I hate this book.
I absolutely hate this book. I had to force myself to complete it. It is not a modern day classic as Oprah proclaimed.
Published 13 hours ago by Lori Weatherford

★★★★★ Best book I've read in years
Whether you trust Oprah's recommendations or not, this is the book everyone over the age of 30 should probably read this year. [Read more](#)
Published 15 hours ago by Mom of Two

★★★★★ In A Word: Dreadful
This is what passes for important fiction these days? God help us all.

Ridiculous characters (a 19 y/o UVA student involved in the Iraqi arms trade !!!!)
[Read more](#)
Published 17 hours ago by Reviewer

★★★★★ Amazing, lives up to hype
A wonderful read that is just as good as the praise. Excellent writing and beautiful storytelling. An easy five stars.
Published 1 day ago by M. J. Hansen

★★★★★ A Magnificent Epic
I just finished reading Freedom and although, like some other reviewers, found that it dragged in places, but all in all, it is a towering achievement of fiction writing. [Read more](#)
Published 1 day ago by Eric Gross

★★★★★ Eh ...
Yeah, it's a recommended book, but upon actually reading the thing I'm left wondering why I bothered.
Published 1 day ago by John Addington

★★★★★ A bunch of non interesting people talking no end about their silly issues.
This book is tedious, the characters' behaviors are confusing, and the story sinks after the beginning of Patty's memoirs (page 25, more or less), as a reviewer has already... [Read more](#)
Published 1 day ago by So many books, so little time

★★★★★ Poetry in motion
Look around you and see the view??

Franzen is poetry in motion.

Conundrum ... ??
Published 2 days ago by Lisa Meyer

Search Customer Reviews

 Only search this product's reviews
[> See all 250 customer reviews...](#)

Source: *Amazon.com* (Captured at the beginning of October 2010)

TABLE 4.27: AMAZON.COM AVERAGE STAR RATING, NUMBER OF REVIEWS, MOST HELPFUL REVIEW'S STARS AND MOST RECENT REVIEW'S STARS (MODEL2)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Book	Date of 1st scrape	Time	Number of Reviews	Average Stars	# 5 Stars	# 4 Stars	# 3 Stars	# 2 Stars	# 1 Stars	MH1	MH2	MH3	MR1	MR2	MR3	MR4	MR5	MR6	MR7	MR8	MR9	MR10
2	1	2010/10/01	00:10	10	5	10	0	0	0	0	5	5	5	5	5	5	5	5	5	5	#N/A	#N/A	#N/A
3	2	2010/10/01	00:02	17	4	11	1	2	0	3	5	5	3	3	4	5	5	1	1	5	5	5	1
4	3	2010/10/01	00:10	15	4.1	9	2	2	1	1	5	4	5	5	3	5	5	2	5	5	5	3	4
5	5	2010/10/01	00:08	61	3.9	26	16	10	7	2	4	3	5	4	2	5	3	5	4	3	4	5	2
6	6	2010/10/01	00:05	6	4.7	5	0	1	0	0	5	5	5	5	5	3	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
7	8	2010/10/01	00:03	33	3.9	18	6	3	1	5	5	5	5	1	1	4	5	5	1	4	3	5	1
8	10	2010/10/01	00:05	43	4.9	40	3	0	0	0	5	5	5	5	5	5	5	4	5	5	5	5	5
9	11	2010/10/01	00:09	38	4.7	30	6	2	0	0	5	5	5	5	5	5	5	5	5	4	5	4	5
10	12	2010/10/01	00:06	101	4.4	60	22	16	3	0	5	4	2	5	5	5	3	4	4	5	4	5	5
11	13	2010/10/01	00:01	177	4.2	129	15	4	3	26	5	5	5	5	5	5	4	5	1	5	5	5	5
12	15	2010/10/01	00:11	21	3.9	6	9	4	1	1	4	3	4	4	4	5	1	5	4	2	3	5	5
13	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:

FIGURE 5.11: BOXPLOT OF AMAZON.COM REVIEW TEXT LENGTH (MODEL 1)

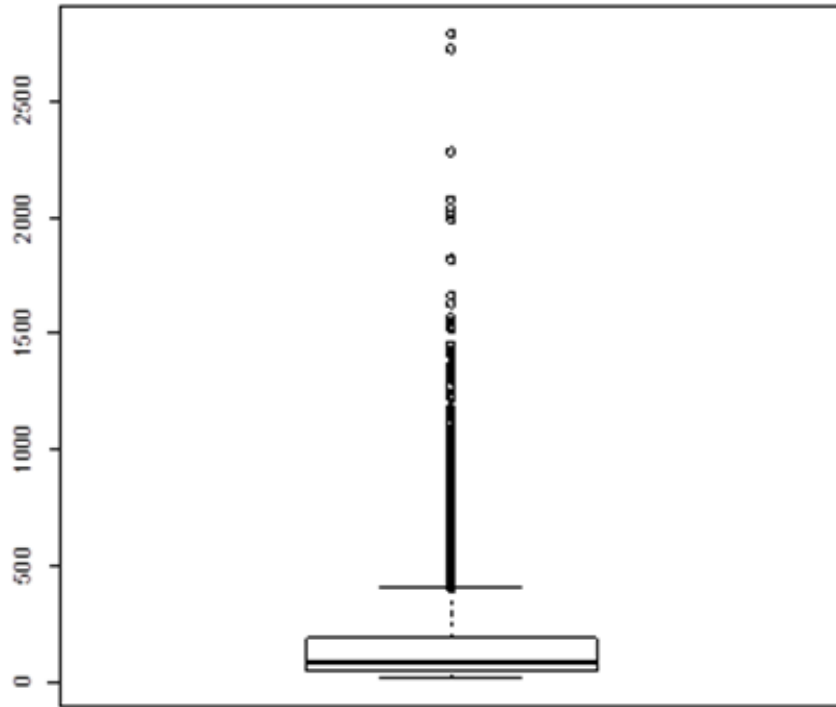


FIGURE 5.12: BOXPLOT OF BARNESANDNOBLE.COM REVIEW TEXT LENGTH (MODEL 1)

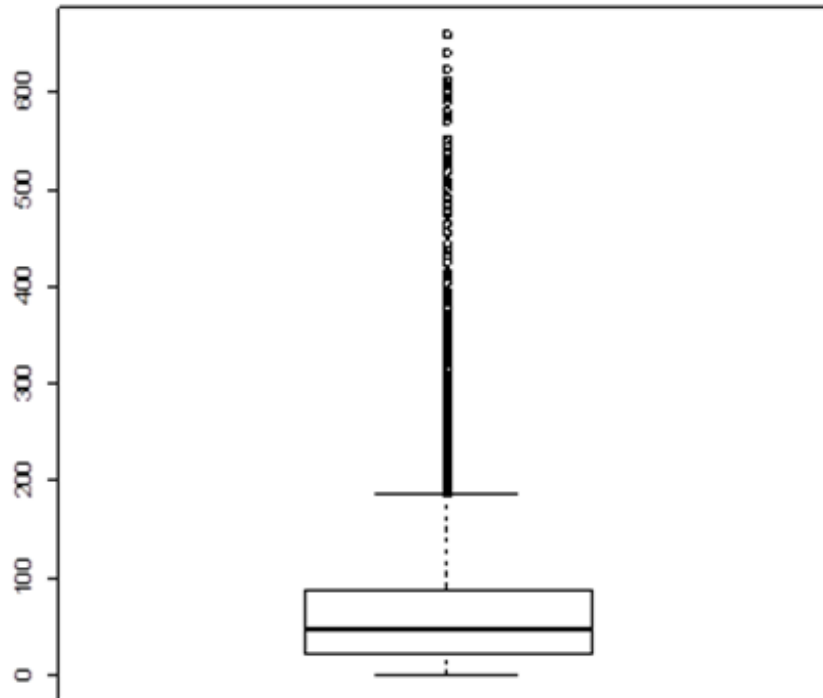


TABLE 5.1: AVERAGE REVIEW TEXT LENGTH PER STAR RATING (MODEL 1)

Stars	Barnes	Amazon
1	63	115
2	104	161
3	132	205
4	130	219
5	92	162

ANNEXURE C: SOFTWARE PROGRAMMES

TABLE 4.3: AMAZON.COM WEB SCRAPING PROGRAMMES

Step 1 - Extract each Amazon.com book's unique ID from its *url*:

http://www.amazon.com/Freedom-Novel-Oprahs-BookClub/dp/0312600844/ref=pd_nr_b_1?ie=UTF8&s=books

E.g. "Freedom: A Novel" has an ID of 0312600844.

Step 2 – Import all unique IDs from text file into the WGET crawler:

```
#!/bin/sh
for ID in $(cat IDs.txt)
do
    wget http://www.amazon.com/x/dp/$ID -O $ID.$(date +%Y%m%d-%H%M).html
done
```

Step 3 – Schedule the WGET crawler to extract *Amazon.com* web page content by means of a cron-programme that executes every six hours:

```
0 */6 * * * /home/djbosman/test.pl
```

TABLE 4.4: BARNESANDNOBLE.COM WEB SCRAPING PROGRAMMES

Step 1 - Extract each Barnesandnoble.com book's unique ID from its *url*:

<http://search.barnesandnoble.com/books/product.aspx?EAN=9780312600846>

E.g. "Freedom: A Novel" has an ID of 9780312600846.

Step 2 – Import all unique IDs from text file into the WGET crawler:

```
#!/bin/sh
for ID in $(cat IDs.txt)
do
    wget http://www.barnesandnoble.com/x/dp/$ID -O $ID.$(date +%Y%m%d-%H%M).html
done
```

Step 3 – Schedule the WGET crawler to extract Barnesandnoble.com web page content by means of a cron-programme that executes every six hours:

```
0 */6 * * * /home/djbosman/test.pl
```

TABLE 4.10: THE PYTHON PROGRAMME THAT CONNECTS TO UCLASSIFY'S API FOR SENTIMENT CLASSIFICATION (MODEL 1)

```

import glob
import httpLib
import urllib
import json
import csv
import os.path

INPUT_FOLDER = "text examples"
OUTPUT_FOLDER = "output"
API_KEY="pPxmAqRrcwhk58LEZStglvoMQvg"
SERVICE = "/browse/uClassify/Sentiment/ClassifyText"

CSVFiles = glob.glob(INPUT_FOLDER+"/*.csv")
print CSVFiles

conn = httpLib.HTTPConnection("uclassify.com")

for full_filename in CSVFiles:

    (path, filename) = os.path.split(full_filename)

    f_input = open(full_filename)
    f_output = open(OUTPUT_FOLDER+"/"+filename, "wb")
    writer = csv.writer(f_output)

    csvreader = csv.reader(f_input, quotechar='"')

    for line in csvreader:
        if len(line) == 0:
            continue
        requestString = urllib.urlencode(("readkey", API_KEY), ("text", line), ("output", "json"), ("version", "1.01"))
        conn.request("GET", SERVICE+"?" + requestString)
        r1 = conn.getresponse()
        response = r1.read()
        try:
            data = json.loads(response)
        except:
            print("WARNING: Could not parse entry in "+filename)
            print data
            writer.writerow( [line, data["cls1"]["negative"], data["cls1"]["positive"]] )
        f_input.close()
        f_output.close()

conn.close()

```


TABLE 4.19: THE PYTHON PROGRAMME THAT EXTRACTS SALES RANK DATA FROM AMAZON.COM WEB PAGES (MODEL 2)

```

import glob
import csv
import re
import datetime
from BeautifulSoup import BeautifulSoup

CSVFile = open("amazon.csv", "wb")
writer = csv.writer(CSVFile)

# Net Oktober: HTMLFiles = glob.glob("*.201010??-????.html")
HTMLFiles = glob.glob("*.html")

for HTMLFile in HTMLFiles:
    try:
        print HTMLFile
        r = re.search("(.*).\.(.*)-(.*)\.html", HTMLFile)
        bookid = r.group(1);
        date = r.group(2)[0:4]+'-'+r.group(2)[4:6]+'-'+r.group(2)[6:8]
        time = r.group(3)[0:2]+'-'+r.group(3)[2:4]

        f = open(HTMLFile)
        soup = BeautifulSoup(f)

        try:
            title = str(soup.html.body.findAll("div", "buying")[1].h1.span.contents[0])
        except:
            title = "ERROR"
        #
        try:
            author = str(soup.html.body.findAll("div", "buying")[1].a.contents[0])
        except:
            author = "ERROR"
        #
        try:
            sales_rank = soup.html.body.find("li", {"id": "SalesRank"}).contents[2]
        except:
            sales_rank = "ERROR"

        print("TITLE: "+title)
        print
        print("AUTHOR: "+author)
        print
        ## print
        index = sales_rank.index('in')
        print("# RANK: "+sales_rank[0:index].strip().lstrip('#'))
        print
        ## print
        #-----
        writer.writerow( [date, time, title, author, bookid, sales_rank[0:index].strip().lstrip('#')] )
    except Exception as err:
        print "Error in file: ",
        print HTMLFile
        print err
        print "*****"
CSVFile.close()

```

TABLE 4.21: MATLAB PROGRAMMES USED TO CALCULATE SALES RANK VALUES (MODEL 2)

```

%Book, date, sales rank (already sorted by column 1, then 2).
Data = [1 40449 781.00
1 40449 638.00
1 40450 1035.00
1 40450 923.00
1 40450 880.00
1 40450 753.00
1 40451 922.00
1 40451 878.00
: : :
85 40616 10295.00
85 40616 8725.00
85 40617 11163.00
85 40617 14448.00
85 40617 12904.00];

% Calculates previous and following week average log sales rank
% On day 8, previous week is average of days 1 - 7, next is average of
% days 9 - 15.
% Calculates average daily log SR
clear
SR; % Data: Book; Date; SR already sorted by 1,2
Data(:,3) = log(Data(:,3));
AvDay = zeros(size(Data));
NBooks = max(Data(:,1)); % Actually largest book number
RowNo = 0;
for CountBook = 1:NBooks
    Vec = Data(:,1);
    Pos = find(Vec == CountBook);
    if ~isempty(Pos)
        d2 = Data(Pos, :, :);
        d2 = d2(:, [2 3]);
        Done = 0;
        while ~Done
            Vec2 = d2(:,1);
            Pos2 = find(Vec2 == Vec2(1));
            d3 = d2(Pos2, :);
            RowNo = RowNo + 1;
            AvDay(RowNo, :) = [CountBook Vec2(1) mean(d3(:,2))];
            if length(Vec2) == length(Pos2)
                Done = 1;
            else
                d2 = d2((Pos2(end) + 1):end, :);
            end
        end
    end
end
AvDay = AvDay(1:RowNo, :);

% Calculates weekly moving average
Data = AvDay; % Data: Book; Date; Daily SR
MA7 = zeros(size(Data));
RowNo = 0;
for CountBook = 1:NBooks

```

%continues on next page

```

CountBook
Vec = Data(:,1);
Pos = find(Vec == CountBook);
if ~isempty(Pos)
    d2 = Data(Pos,[2 3]);
    Done = 0;
    while ~Done
        try
            ThisMA = mean(d2(1:7,2));
            RowNo = RowNo + 1;
            MA7(RowNo,:) = [CountBook d2(1,1) ThisMA];
            d2 = d2(2:end,:);
        catch
            Done = 1;
        end
    end
end
end
MA7 = MA7(1:RowNo,:);

Data = MA7; % Book; date; 7-day MA starting at that date
[NRows, NCols] = size(Data);
SRPrevNext = zeros(NRows,NCols + 1);
RowNo = 0;
for CountBook = 1:NBooks
    CountBook
    Vec = Data(:,1);
    Pos = find(Vec == CountBook);
    if length(Pos) >= 9
        d2 = Data(Pos,[2 3]);
        for d2Pos = 8:(length(Pos) - 1)
            RowNo = RowNo + 1;
            PrevSR = d2(d2Pos - 7,2);
            NextSR = d2(d2Pos + 1,2);
            SRPrevNext(RowNo,:) = [CountBook d2(d2Pos,1) PrevSR NextSR];
        end
    end
end
end
SRPrevNext = SRPrevNext(1:RowNo,:);

%PRINTMAT(DATA,FNAME,PRECISION)
PrintMat(SRPrevNext,'c:\matlab\work\SRPrevNext.txt',8);

```

TABLE 4.23: THE PYTHON PROGRAMME THAT EXTRACTS SPOTLIGHT FOUND HELPFUL STAR RATINGS FROM AMAZON.COM WEB PAGES (MODEL 2)

```

import glob
import csv
import re
import datetime
from BeautifulSoup import BeautifulSoup

CSVFile = open("amazon.csv", "wb")
writer = csv.writer(CSVFile)
# Net Oktober: HTMLFiles = glob.glob("*.201010??-?????.html")
HTMLFiles = glob.glob("*.html")
for HTMLFile in HTMLFiles:
    print HTMLFile
    r = re.search("(.*).*(.*)\.html", HTMLFile)
    bookid = r.group(1);
    date = r.group(2)[0:4]+'-'+r.group(2)[4:6]+'-'+r.group(2)[6:8]
    time = r.group(3)[0:2]+'-'+r.group(3)[2:4]
    f = open(HTMLFile)
    soup = BeautifulSoup(f)

    try:
        title = str(soup.html.body.findAll("div","buying")[1].h1.span.contents[0])
    except:
        title = "ERROR"
        #
    try:
        found_helpful1 = str(soup.html.body.findAll("div", style="margin-bottom:0.5em;")[0].contents[0])
    except:
        found_helpful1 = "ERROR"
        #-----
    try:
        stars1 = str(soup.html.body.findAll("span", style="margin-right:5px;")[0].span.span.contents[0])
    except:
        stars1 = "ERROR"
        #-----
    try:
        found_helpful2 = str(soup.html.body.findAll("div", style="margin-bottom:0.5em;")[4].contents[0])
    except:
        found_helpful2 = "ERROR"
        #-----
    try:
        stars2 = str(soup.html.body.findAll("span", style="margin-right:5px;")[1].span.span.contents[0])
    except:
        stars2 = "ERROR"
        #-----
    try:
        found_helpful3 = str(soup.html.body.findAll("div", style="margin-bottom:0.5em;")[8].contents[0])
    except:
        found_helpful3 = "ERROR"
        #-----
    try:
        stars3 = str(soup.html.body.findAll("span", style="margin-right:5px;")[2].span.span.contents[0])
    except:
        stars3 = "ERROR"
        #-----
    print("TITLE: "+title)
    print("MHCR1 FH: "+found_helpful1.strip())
    print("MHCR1 Stars: "+stars1.strip())
    print("MHCR2 FH: "+found_helpful2.strip())
    print("MHCR2 Stars: "+stars2.strip())
    print("MHCR3 FH: "+found_helpful3.strip())
    print("MHCR3 Stars: "+stars3.strip())
    writer.writerow( [date, time, bookid, title, found_helpful1.strip(), stars1.strip(), found_helpful2.strip(), stars2.strip(),
found_helpful3.strip(), stars3.strip()] )
    CSVFile.close()

```

TABLE 4.26: MATLAB PROGRAMMES USED TO CALCULATE AVERAGE STAR RATINGS AND MOST RECENT REVIEWS' STAR RATINGS (MODEL 2)

```

% Indicating first scrape of every book
% Book number, date of first scrape, # of reviews, ave star, MHx3, MRx10
AV3 = [1 40452 10 5 5 5 5 5 5 5 5 5 5 5 5 0 0 0
2 40452 17 4 5 5 3 3 4 5 5 1 1 5 5 5 1
3 40452 15 4.133333333 5 4 5 5 3 5 5 2 5 5 5 3 4
4 40463 2 4.5 5 4 0 0 0 0 0 0 0 0 0 0 0
5 40452 61 3.93442623 4 3 5 4 2 5 3 5 4 3 4 5 2
: : : : :
80 40452 10 2.3 5 5 4 1 1 1 2 1 2 1 0 0 0
81 40452 78 2.679487179 2 1 2 2 1 1 2 4 5 4 2 4 2
82 40452 315 3.466666667 5 3 5 5 3 5 2 5 1 4 4 4 3
83 40467 5 4 4 5 4 3 4 0 0 0 0 0 0 0 0
84 40452 16 4.0625 4 5 5 5 4 5 5 5 5 4 1 5
85 40452 27 3.888888889 3 5 3 4 1 5 4 1 5 2 2 5 5];

% Programme for data on every review
% Per review sheet in Project 3 Amazon
% Book, date of review, FH, star rating
RevData = [1 40454 0.6 5
1 40469 0.454545455 2
1 40469 0.833333333 5
1 40469 0.74 1
1 40472 0.6 5
1 40475 0.833333333 5
: : :
85 40562 0.333333333 1
85 40562 1 3
85 40563 1 1
85 40566 0.666666667 2
85 40568 1 1
85 40572 0 2];

% Average star rating per day
clear
AV3Data; % AV3: Book, date, #Reviews, Av.Stars (, others)
PerReview; % RevData: Book, date, (FH,) stars
AV3 = AV3(:,1:4);
RevData = RevData(:, [1 2 4]);
AV3 = SortBy(AV3,1);
RevData = SortBy(RevData, [1 2]);
ResMat = zeros(120*85,3); % Book, date, average star
RowNo = 0;
for CountBook = 1:85
    CountBook

%Initial
    ThisDate = AV3(CountBook,2);
    NRev = AV3(CountBook,3);
    TotalStars = round(AV3(CountBook,4)*NRev);
    RowNo = RowNo + 1;
    ResMat(RowNo,:) = [CountBook ThisDate TotalStars/NRev];
    disp('Initial')
    [CountBook ThisDate TotalStars/NRev]
    %pause

%continues on next page

```

```

% Review(s) on first date
Vec = RevData(:,1);
Pos = find(Vec == CountBook);
if ~isempty(Pos)
    d2 = RevData(Pos,2:3);
    Vec2 = d2(:,1);
    Pos2 = find(Vec2 == ThisDate);
    if ~isempty(Pos2)
        disp('On first date')
        d2 = d2(Pos2,2);
        TotalStars = TotalStars + sum(d2)
        NRev = NRev + length(d2)
        ResMat(RowNo,3) = TotalStars/NRev;
        TotalStars/NRev
        %pause
    end
end

disp('Rest')
% Rest of dates till 31 Jan 2011
while ThisDate < 40574 % 31 Jan 2011
    ThisDate = ThisDate + 1
    Vec = RevData(:,1);
    Pos = find(Vec == CountBook);
    if ~isempty(Pos)
        d2 = RevData(Pos,2:3);
        Vec2 = d2(:,1);
        Pos2 = find(Vec2 == ThisDate);
        if ~isempty(Pos2)
            TotalStars = TotalStars + sum(d2(Pos2,2))
            NRev = NRev + length(Pos2)
        end
    end
    RowNo = RowNo + 1;
    ResMat(RowNo,:) = [CountBook ThisDate TotalStars/NRev];
    [CountBook ThisDate TotalStars/NRev]
    %pause
end
end
ResMat = ResMat(1:RowNo,:);

%PRINTMAT(DATA,FNAME,PRECISION)
PrintMat(ResMat,'c:\matlab\work\AvStar.txt',8);
%continues on next page

```

```

% Average daily most recent star rating per day
clear
AV3Data;
PerReview;
AV3 = AV3(:, [1 2 8:17]); % AV3: Book, date, MR1, ..., MR10
RevData = RevData(:, [1 2 4]); % RevData: Book, date, stars
RevData = SortBy(RevData, [1 2]);
ResMat = zeros(120*85,3); % Book, date, average recent star
RowNo = 0;
for CountBook = 1:85
    CountBook
    %Initial
    ThisDate = AV3(CountBook,2);
    MRVec = AV3(CountBook,3:12);
    RowNo = RowNo + 1;
    M2 = MRVec(find(MRVec ~= 0));
    ResMat(RowNo,:) = [CountBook ThisDate mean(M2)];

    % Review(s) on first date
    Vec = RevData(:,1);
    Pos = find(Vec == CountBook);
    if ~isempty(Pos)
        d2 = RevData(Pos,2:3);
        Vec2 = d2(:,1);
        Pos2 = find(Vec2 == ThisDate);
        if ~isempty(Pos2)
            d2 = d2(Pos2,2)
            MRVec = [d2' MRVec];
            MRVec = MRVec(1:10);
            M2 = MRVec(find(MRVec ~= 0));
            ResMat(RowNo,3) = mean(M2);
        end
    end
end

% Rest of dates till 31 Jan 2011
while ThisDate < 40574 % 31 Jan 2011
    ThisDate = ThisDate + 1;
    Vec = RevData(:,1);
    Pos = find(Vec == CountBook);
    if ~isempty(Pos)
        d2 = RevData(Pos,2:3);
        Vec2 = d2(:,1);
        Pos2 = find(Vec2 == ThisDate);
        if ~isempty(Pos2)
            AddVec = d2(Pos2,2)
            MRVec = [AddVec' MRVec];
            MRVec = MRVec(1:10);
        end
    end
    RowNo = RowNo + 1;
    M2 = MRVec(find(MRVec ~= 0));
    ResMat(RowNo,:) = [CountBook ThisDate mean(M2)];
end
end
ResMat = ResMat(1:RowNo,:);
ResMat(find(isnan(ResMat))) = 3;

PrintMat(ResMat, 'c:\matlab\work\AvRecStar.txt',8);

```

ANNEXURE D: EXTERNAL SOURCES

TABLE 6.1: RETAIL E-COMMERCE (NON-TRAVEL) GROWTH RATES FOR 2011, 1ST QUARTER

Quarter	E-Commerce Spending (\$ Millions)	Y/Y Percent Change
Q1 2007	\$27,970	17%
Q2 2007	\$27,176	23%
Q3 2007	\$28,441	23%
Q4 2007	\$39,132	19%
Q1 2008	\$31,178	11%
Q2 2008	\$30,581	13%
Q3 2008	\$30,274	6%
Q4 2008	\$38,071	-3%
Q1 2009	\$31,031	0%
Q2 2009	\$30,169	-1%
Q3 2009	\$29,552	-2%
Q4 2009	\$39,045	3%
Q1 2010	\$33,984	10%
Q2 2010	\$32,942	9%
Q3 2010	\$32,133	9%
Q4 2010	\$43,432	11%
Q1 2011	\$38,002	12%

Source: Gian Fulgoni, *comScore Inc.*, Q1 2011 results