



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



GORDON INSTITUTE
OF BUSINESS SCIENCE
University of Pretoria

The impact of long tail distribution in keyword selection on the effectiveness of sponsored search advertising.

Justinus Adriaanse
97081028

A research project submitted to the Gordon Institute of Business Science,
University of Pretoria, in partial fulfilment of the requirements for the degree of
Master of Business Administration.

11 November 2009

Abstract

Search engines have revolutionised the access to information to the general public. Today search engines are the most important promotional method on the Internet. Sponsored search dominates the revenue model behind this growth. The rise in popularity and the auction pricing mechanism of sponsored advertising have increased the average cost-per-click. Marketing managers need tools to enable them to increase return on investment in this medium.

The application of Anderson's (2004) long tail distribution holds great promise to solve this dilemma. The current study used causal research in a two by two factorial design. Here data from an online property portal in a developing market was collected in order to examine the effect of a long tail (LT) distribution in keyword selection on return on investment (ROI) with sponsored search. Sponsored search allows for individualised targeting of the users behaviour. The application of the long tail (LT) enables further matching the advert text to the users search query.

The results provide strong support for the significant impact on cost-per-click and by implication the return on investment that keyword selection and targeted advert text have when used in conjunction with the principles of the long tail. The interaction of the independent variables of long tail and sponsored search is significant, contributing to a 430% increase in click-through (CTR) rates and 61% reduction in cost-per-click, translating into a 61% increase in return on investment.

Declaration

I declare that this research project is my own work. It is submitted in partial fulfilment of the requirements for the degree of Masters of Business Administration at the Gordon Institute of Business Science, University of Pretoria. It has not been submitted before for any degree or examination in any other University. I further declare that I have obtained the necessary authorisation and consent to carry out his research.

Justinus Adriaanse

11 November 2009

Acknowledgements

I would like to make the following acknowledgements to people who have been a key part of my MBA and research journey:

- My wife for supporting me through the two years, being patient with me when the stress levels were near boiling point, and for being such a wonderful mother to Christopher.
- Kerry Chipp, my supervisor, for taking the stress I placed on her in the last two weeks in her stride.
- The company that provided the data. Thank you for making this research possible.
- GIBS: for believing in me and giving me the opportunity to do the MBA.
- My fellow MBA student: I am really going to miss you all.



TABLE OF CONTENTS

1. INTRODUCTION TO THE RESEARCH PROBLEM	1
1.1. Research title	1
1.2. Research problem	1
1.3. Research objectives	4
1.4. Research aim	5
2. LITERATURE REVIEW	7
2.1. Introduction	7
2.2. Targeted advertising	7
2.3. One-to-one advertising	10
2.4. Online advertising	13
2.5. Sponsored search	17
2.6. Long tail distribution	23
2.7. Conclusion	28
3. RESEARCH HYPOTHESES	29
3.1. Introduction	29
3.2. Objective 1	29
3.3. Objective 2	29
3.4. Objective 3	30
3.5. Objective 4	30
4. RESEARCH METHODOLOGY	32
4.1. Introduction	32
4.2. Research design	32
4.3. Unit of analysis	35
4.4. Universe	35
4.5. Sampling method and size	36
4.6. Data gathering	38
4.7. Data analysis	38
5. RESULTS	40
5.1. Introduction	40
5.2. Cost-per-click: Hypothesis 1	40

5.3.	Click-through rate: Hypothesis 2	43
5.4.	Targeted and generic advert text: Hypotheses 3 and 4	46
5.5.	Pages views: Hypothesis 5	51
5.6.	Time spent: Hypothesis 6	53
5.7.	Bounce rate: Hypothesis 7	56
5.8.	The long tail	59
5.9.	Total clicks and cost per keyword	60
5.10.	Cost-per-click for targeted and generic ad text	62
5.11.	Sales and rental keywords	65
5.12.	Conclusion	71
6.	DISCUSSION OF RESULTS	73
6.1.	Introduction	73
6.2.	Cost-per-click: Objective 1	74
6.3.	Click-through rate: Objective 2	77
6.4.	Targeted advert text: Objective 3	81
6.5.	User comparison: Objective 4	85
6.6.	Sales and rental keywords	89
6.7.	Research question	91
7.	CONCLUSION	92
7.1.	Introduction	92
7.2.	Findings	92
7.3.	Recommendations	94
7.4.	Limitations	97
7.5.	Suggestions for future research	98
	REFERENCES	100

1. INTRODUCTION TO THE RESEARCH PROBLEM

1.1. Research title

The impact of long tail distribution in keyword selection on the effectiveness of sponsored search advertising

1.2. Research problem

“Not since Gutenberg invented the modern printing press more than 500 years ago, making books and scientific tomes affordable and widely available to the masses, has any new invention empowered individuals, and transformed access to information as profoundly as Google.” (Vise, 2005)

Sponsored search advertising on search engines has completely revolutionised online advertising (Jansen and Mullen 2008; Xu, Chen and Whinston 2009). Today, search engines are the most important promotional method on the Web, and by far the most common way for new e-commerce business to market themselves (Jansen and Molina, 2006).

Google dominates online sponsored search advertising with more than US\$21 billion in advertising revenue (Google, 2009). Future predictions for online advertising are as high as US\$106.6 billion in the United States alone by 2011 (IDC, 2008). Search revenue is currently making up 45% of online advertising and is the main driver of online advertising growth; it is predicted to be more than \$33 billion in 2010 (PricewaterhouseCoopers, 2009; Ghose and Yang, 2009).

In 2005, sponsored search overtook display (banner) advertising as the biggest form of online advertising (Rutz and Bucklin, 2007). Sponsored search is where the user's desire for relevant information meets the advertiser's need to target a user at the exact point they are looking for their product, thereby giving the advertiser the ability to target a customer's behaviour, not his/her demographic (Ghose *et al.*, 2009; Jansen *et al.*, 2008).

Sponsored search success is, in part, a function of the ability advertisers have to target customers on an individual level, as well as the extremely low cost, transparency and ease of use of the cost-per-click (CPC) pricing method (Edelman, Ostrovsky and Schwarz 2007; Pedersen 2008; Wilbur and Zhu 2009). Goto.com was the first to combine the CPC and auction pricing model in 1998, and this was later copied and improved by Google (Fain and Pederson 2006; Goldfarb and Tucker 2007; Jansen and Spink 2007; Jansen *et al.*, 2008; Wilbur *et al.*, 2009).

The 'generalised second-price' (GSP) auction-based pricing model employed by search engines is aimed at maximising their revenue and reducing strategic bidding by rival advertisers (Jansen *et al.*; 2008; Edelman *et al.*, 2007). Since prices are ultimately influenced by the competition for a keyword, and with more and more companies competing for popular keywords, the auctioned-based pricing model is making it increasingly difficult to maintain or improve the advertising return on investment (ROI) (Pedersen 2008).

If the Pareto Principle, or 80/20 rule, holds true for keyword selection, 80% of the traffic will be achieved from 20% of the keywords in a campaign. Herein lies the dilemma: since this 20% of keywords are always in high demand, it results in an escalation of advertising costs. With the ever increasing pressure to improve return on investment (ROI), this makes the marketer's job

very difficult. The problem is to continuously create the most effective online advertising campaign in order to maximize the return on investment (ROI) for every rand spent, in an environment where increased competition will consistently put upward pressure on the cost of each keyword and, therefore, each click. The online blogging environment are filled with desperate please for lower cost per clicks, on one United Kingdom based blog McCubbin (2009) suggest the cost-per-click CPC Grinch might steel advertisers Christmas profits.

Despite the growth in sponsored search, the academic literature on the topic has been rather limited (Animesh, Ramachandran and Viswanathan 2005; Edelman and Ostrovsky 2006; Goldfard *et al.*,2007; Feng, Bhargava and Pennock 2007; Ghose and Yang 2007; Jansen *et al.*, 2008; Edelman *et al.*, 2007; Kumar, 2008; Xu *et al.*, 2009). Laffey (2007), Fain *et al.* (2006) and Jansen *et al.* (2008) provided an introduction to sponsored search and some general guidelines for new users.

Initial concerns about the quality of sponsored search links resulted in some research about its relevance to search queries by Jansen and Resnick (2006), Jansen (2007a) and Jansen *et al.* (2007). There has also been research on electronic auctions, which determine the cost-per-click (CPC), and bidding strategies to optimise performance by Goldfard *et al.* (2007), Xu *et al.* (2009), Edelman *et al.* (2006), Edelman *et al.* (2007) and Animesh *et al.* (2005).

Limited research on a keyword level has been done by Kumar (2008), Rutz *et al.* (2007), Rutz and Bucklin (2008) and Ghose *et al.* (2007). Ghose *et al.* (2007) specifically examined some of the attributes of sponsored search (like ranking, retailer presence in ad copy, brand information and length of keyword) on such metrics as click-through rates (CTRs), conversion rates, bid prices and keyword ranking.

Rutz *et al.* (2007) developed a model to measure keyword conversion rates that took into account performance of keywords that did not generate any response, and showed it to outperform existing management strategies. Kumar (2008) added to Ghose *et al.*'s (2007) research by including organic results to compare keyword metrics, including click-through rates, conversion rates and bid prices. Rutz *et al.* (2008) examined the relationship between generic and branded keywords. There appears to be a void in the literature around an approach to selecting keywords and a strategy that can be used as a tool to decrease cost-per-click (CPC) and, therefore, increase return on investment (ROI). This research aims to fill this void.

Strategies for keyword selection could come from theories and principles other than traditional marketing ones. Chris Anderson (2004) created a sensation in the media business with the long tail concept (McDonald, 2008). Anderson (2004) showed how the Internet is changing the effect of the Pareto principle by allowing firms to profit from products in the long tail even if it sells only once per year. Anderson's (2004) long tail concept could possibly be used as a strategy in keyword selection to help managers improve the return on investment (ROI) for sponsored search advertising.

1.3. Research objectives

The fundamental question this research aims to answer is: *“Can a long tail distribution in keyword selection help improve the effectiveness of a sponsored search online advertising campaign?”*

The main objectives of the research will be:

- Objective 1: to determine if a long tail approach to keyword selection can improve the return on investment (ROI) of the campaign by bringing the overall average cost-per-click down significantly.
- Objective 2: to explore if the long tail keywords will yield significantly higher click through rates than head keywords.
- Objective 3: to see if targeted ad text, related to the keyword used, will significantly increase the click through rate for either head or tail keywords.
- Objective 4: to compare the behaviour of users that click on head and long tail keywords by comparing commonly used measures like: pages viewed, bounce rate and time spent.

1.4. Research aim

This research aims to add to the small but growing body of knowledge about sponsored search on a keyword level. It specifically seeks to address some of the lack of guidance available about keyword selection and techniques to reduce overall cost-per-click (CPC) and thereby increase return on investment (ROI).

In addition, the paper aims to confirm that the increased targeting that results from a more granular (long tail) approach, combined with the very targeted ad copy, will increase the key



metrics for sponsored search advertising: click through rate (CTR). This will, in turn, contribute toward lowering the cost-per-click (CPC) and improving the return on investment (ROI).

Therefore, the aim is to find a theoretical approach to the selecting of keywords as part of the design of a sponsored search advertising campaign, that will enable sustainable results to be achieved, thereby giving the marketer tools to improve the return on investment (ROI) achieved with online sponsored search marketing.

2. LITERATURE REVIEW

2.1. Introduction

The theory reviewed in this section is broken into three distinct sections: targeted advertising; online advertising; and the long tail (LT). Firstly, targeted advertising is discussed, exploring the benefits it has for firms, and the techniques used to achieve it. Targeted advertising is taken further by looking at one-to-one advertising and how the Internet is making it a reality. Secondly, the evolution of online advertising is explored, recounting some of the debates about the value of online advertising, the decline in click through rates, and the search for the ideal pricing model. Sponsored search is then discussed. As the biggest part of online advertising today, it powers the online search engines and ultimately the Web, and enables targeting to an unprecedented level by targeting behaviour, not demographics. Lastly, the long tail (LT) is explained, reasons for its occurrence are stated, and studies are mentioned that both seem to confirm and contradict the concept.

Combining the long tail (LT) concept with the power of search engines through sponsored search advertising provides a means to achieve one-to-one behavioural targeting, and could reduce the cost per click (CPC) by finding less popular keywords, thereby increasing the return on investment (ROI) achieved with online advertising.

2.2. Targeted advertising

Targeted advertising leads to higher equilibrium profits and gives firms a more sustainable advantage, unlike that of most marketing initiatives which can be easily copied by

competitors, thereby leading to a higher cost of doing business for both firms (Iyer, Soberman and Villas-Boas, 2005). Targeted advertising also allows firms to eliminate ‘wasted’ advertising. Iyer *et al.* (2005) showed that the ability to target advertising has a bigger impact on the profitability of a firm than the ability to set target pricing. Iyer *et al.* (2005) argued that the increasing ability to target advertising comes from two advances in the marketing environment:

- Better information available about consumers, what their preferences are, and their media consumption habits; and
- The fragmentation of existing media, and the addition of a vast array of new advertising media, including the Internet.

Iyer *et al.* (2005) concluded that in an environment where advertising is expensive, firms have low levels of advertising, leaving potential demand unrealised; but once targeting has been used, the firm will get better results, allowing it to increase its marketing budget.

Early research on targeting by Rossi, McCulloch and Allenby (1996) showed a substantial improvement in the return on direct coupon mailings when using historical data for targeting. As little as one observation about the purchase history of a household increased the revenue generated from a coupon mailing campaign by 50%, and a full model led to 2.5 times the revenue (Rossi *et al.*, 1996). Rossi *et al.* (1996) predicted that the exponential decline in the cost of gathering and storing customer data will lead to an increased ability to target advertising and promotions, resulting in an increase in its effectiveness; an ideal environment for the Internet. Anderson (2009, p13) suggested that processing, bandwidth and storage on

the Internet is increasing in capacity at about 100%, and decreasing in cost at nearly 50% per year. These properties of the Internet make it the ideal medium for targeted advertising.

Targeting is of great use to the direct marketer who, when armed with the correct information, can prevent lucrative customers from taking their business to competitors (Allenby, Leone and Jen, 1999). Allenby *et al.* (1999) illustrated that modelling and then predicting individual customer behaviour using the specific customer's historical data could assist firms in targeting customers most likely to cease trading with the firm. This enabled the firm to apply its limited resources to achieve maximum results (Allenby *et al.*, 1999). A major challenge in targeting advertising is thus finding the customers most likely to be interested in the product or service (Kim, Street, Russell and Menczer, 2005).

More targeted advertising increases advertisers' willingness to pay for such advertising, and allows media owners to maximise revenue (Chandra, 2009). Chandra (2009) showed that in more competitive markets, newspapers needed to drop their circulation prices but were able to charge a substantial premium on advertising as a result of their increased ability to better segment readers according to location and other demographical data. Chandra (2009) went further by showing that any advertising medium – print, radio, television and the Internet – is dependent on reaching a core, targeted audience to maximise the return in placing advertising in the particular medium. These media increase their rates with their ability to target.

Kim *et al.* (2005) used artificial neural networks and algorithms, ever increasing in complexity, to mine huge consumer data to identify households most likely to purchase particular products, with great results. However, this is still based on the household's demographics and

possible only when offline companies conduct database management and/or loyalty programmes to collect such data. Traditional media gives everyone the same message and this no longer meets business's requirements, due to the growing sophistication of the consumer (Adams, 2004; Kazienko and Adamski, 2007). To increase results from advertising the right message should reach the right person at the right time in the right context (Adams 2004; Kazienko *et al.*, 2007). How can this be achieved? One-to-one advertising aims to take targeted advertising to that level.

2.3. One-to-one advertising

Personalisation of advertising (one-to-one) takes advertising to this individual level and is even more focussed than targeted advertising, which simply seeks to divide customers into segments based on demographics, including gender, age and geographical location (Adams 2004; Kazienko *et al.*, 2007). Personalisation is based more on the individual's behaviour than demographics (Adams 2004; Kazienko *et al.*, 2007). The increased ability of users to interact directly with firms online has enabled this shift from mass advertising to more personalised advertising (Ghose *et al.*, 2009).

Personalised or dialogue marketing (one-to-one) could mean the end of mass marketing, and it allows firms to communicate with all customers, not just the traditional 20% of top revenue earning customers (Ferguson and Hlavinka, 2006). However, in order to measure the effectiveness, the firm's strategic intention must be considered (Dong, Puneet and Chintagunta, 2009). Dong *et al.* (2009) stressed that ignoring the firm's strategic behaviour can severely underestimate the positive effect of individual level targeting, and argued that

although their results are limited to a specific data set, it is likely to be generalised beyond this single application.

Customised communications (one-to-one) attracts more customer attention and helps foster more customer loyalty; it also aids customer decisions and reduces information overload, all of which can be translated into increased revenue and profitability (Ansari and Mela, 2003). Dong *et al.* (2009) showed a remarkable improvement in a firm's profit (14%-23%) when employing strategic targeting at the individual customer level (one-to-one), compared with targeting at a segment level. This was achieved in a business-to-business setting in the pharmaceutical industry. The individualising involved allocation of the limited time of sales representatives to visits physicians in order to show new products, which is a very expensive and time consuming marketing channel.

The Internet, however, can produce true one-to-one advertising, which is exceedingly costly in other contexts; and because the medium is dynamic and highly addressable, true customisation is possible, which means the Internet has the ability to deliver the right content to the right person at the right time (Ansari *et al.*, 2003). Targeting can be based on the users' Web browsing history. Sherman and Deighton (2001) used users' browsing history from one website to identify other websites that they also visited, arguing that other browsers on these sites would be disproportionately likely to also be interested in the original website. Sherman *et al.* (2001) tested the prediction by placing banner advertising and achieved an average cost per response of nine times lower for high affinity grouped sites compared to low groups.

Targeting can also be based on the content of the Web page being viewed; Ansari *et al.* (2003) suggested that content targeted advertising could potentially increase click-through

rates (CTRs) by as much as 62%. Furthermore, Chickering and Heckerman (2003) showed that by targeting banner advertising to the specific context of the Web page being visited, their delivery system was able to increase click-through rates (CTRs), optimise revenue under multiple revenue models, including cost-per-thousand impressions (CPMs), cost-per-click (CPC), as well as hybrid models. Danaher (2007) used usage data from across multiple sites to target users in more than 3,000 advertising campaigns, and showed significant promise for a huge improvement in results.

Targeting can also be based consumers' search preferences. Bhatnagar and Papatla (2001) proposed a model that was based on consumer search behaviour, plotting categorised information on a search continuum in a multidimensional space, with the search point in the centre and a threshold envelope at its outer boundary. The search point consists of the focal group of products that a specific consumer might be interested in, and different consumers will have different focal groups of products (Bhatnagar *et al.*, 2001). The further you move away from the centre towards the threshold envelope, the likelihood of consumers being interested decreases, with very little or no interest on or outside the threshold envelope (Bhatnagar *et al.*, 2001).

Sponsored search advertising allows individual level targeting (one-to-one) to consumers as they enter the market for a product - the ultimate means of targeting behaviour, not demographics (Wilbur *et al.*, 2009). Sponsored search is the relevance-targeted text advertisements displayed above or next to organic search results generated by search engines (Fain *et al.*, 2006; Jansen *et al.*, 2007). With sponsored search, advertisers can track the consumers' actions online, which allows for accurate measurements of advertising

profitability (Wilber *et al.*, 2009). Sponsored search will be discussed at more length once online advertising is reviewed.

2.4. Online advertising

Ilfeld and Winer (2002) concluded that for Internet based firms, off-line advertising will increase site visitations through the significant influence on consumer awareness, while online advertising directly increases website traffic. The discussion on Internet advertising has been dominated by banner (display) advertising. Poiesz and Robben (1994, p27) stated, “advertising has for too long been visible in its appearance and highly invisible in its effects.” Rust and Varki (1996) predicted more than ten years ago that the Internet and other interactive media will bring a new era in accountability for advertising, and will allow marketers to finally solve the riddle of the wasteful ‘half’ of advertising, by offering more measurability.

The Internet allows for measurement of response, since, unlike most traditional media where the whole message is delivered in one stage, on the Internet the audience is in control of whether they want more information about the advertisement (Bhatnagar *et al.*, 2001). Bhatnagar *et al.* (2001) argued that the rapidly expanding market of the early 2000s in online advertising campaigns worked against the need to achieve click-throughs and, therefore, reduced its effectiveness.

Declining click-through rates (CTRs), combined with the inability to easily compare online advertising to traditional advertising metrics, fuelled scepticism about the value of advertising in the digital medium (Briggs and Hollis, 1997; Dreze and Zufryden, 1998; Hoffman and Novak, 2000a; Dreze and Husherr, 2003; Cho and Cheon, 2004; Hollis, 2005). It has also

been suggested that click-through rates (CTRs) would continue to decline, falling from around 7% in the 1990s to levels as low as 0.7% in 2002, 0.2% in 2006, and 0.1% in 2008 (Hoffmann and Novak, 2000b; Hoffmann and Novak, 2000c; Chatterjee, Hoffman and Novak, 2003; Rutz *et al.*, 2007; Anderson, 2009; Fulgoni and Morn, 2009). The declining click-through rates (CTRs) inhibited the opportunity for increased Web marketing effectiveness.

The lack of effectiveness was partly fuelled by little accountability in the pricing system of online advertising. Flat fee was the earliest Web advertising model; this model had no guarantees with regard to traffic levels, and was calculated at a constant cost per time period, for example, per month (Hoffmann *et al.*, 2000c). Flat fee later evolved into the more traditional impressions (CPMs), where impressions or exposures are guaranteed by the publisher in a given time period (Hoffmann *et al.*, 2000c).

This cost-per-thousand impressions (CPMs) model was principally driven by the broadcast theory, the belief being that exposure-based pricing will find equilibrium where all advertisers' response functions are taken into account, thereby providing a rational way to price Web advertising (Hoffmann *et al.*, 2000c). Contracts were negotiated on a one-on-one basis, therefore the minimum was large and entry was slow (Edelman *et al.*, 2007). This transfer of offline paradigm limited the application of the unique properties of the online arena.

The distinctive attributes of the Internet stem from its fundamental difference from traditional broadcasting. The Web is built on a many-to-many communication model, whereas traditional media is a one-to-many model (Hoffmann *et al.*, 2000c; Mangàni, 2004). The cost-per-thousand impressions (CPMs) pricing places all the emphasis on the banner or display

advertising and no importance on the target communication the advertiser wants the visitor to read (Hoffmann *et al.*, 2000c).

In an attempt for a more accountable metrics, pricing based on action in the form of click-through rates (CTRs) was introduced; cost-per-click (CPC) provides the first opportunity in a commercial medium to measure the consumer's response, not just assume it (Hoffmann *et al.*, 2000c). Cost-per-acquisition (CPA) or cost per inquiry takes the performance-based model one step further, where advertisers only pay for actual sales generated. Cost-per-acquisition (CPA) has been the driving force behind huge affiliate networks built by companies like amazon.com (Hoffmann *et al.*, 2000b, 2000c).

Publishers have pushed against performance-based pricing, however, arguing that click-through rates (CPRs) and resulting sales are at least partially influenced by the quality of the creativity in the advertising, which is completely under the advertisers' control, over which publishers have little or no influence (Hoffmann *et al.*, 2000c). In challenging economic times, such as those experienced during 2008 and 2009, pricing on performance became more paramount. Both advertisers and agencies appeared to be moving their online advertising campaigns from ones where pricing is based on exposure, to 'pay-for-performance' (CPC or CPA), where some action is required by the consumer (Fulgoni *et al.*, 2009).

Calls for performance-based pricing is countered by Briggs *et al.* (1997) and Yoo (2008), who showed that a single exposure to a banner is beneficial to advertisers even though it did not necessarily result in a click-through. Banners have been shown to influence brand recognition, aid advertising recall and increase repeat purchase probability (Dreze *et al.*, 2003; Manchanda, Dube, Goh and Chintagunta, 2006).

Fulgoni *et al.* (2009) also showed that despite a lack of clicks, display advertising can increase visits to the advertiser's website by up to 46%. It can also increase the odds of a consumer conducting a search using the advertiser's branded terms by 38% (Fulgoni *et al.*, 2009). Display advertising also increases the consumer's chance of purchasing the advertised brand online (27%) and offline (17%) (Fulgoni *et al.*, 2009). Hollis (2005) argued that click-through is, in fact, primarily a result of the brand building effort and the users' willingness to learn more about the brand.

However, Ilfeld *et al.* (2002) suggested that online advertising's focus should be placed on traffic-building, not brand building, as the latter will result from the former. If that suggestion holds true, online based firms should aim to maximise traffic to their websites with online advertising (Ilfeld *et al.*, 2002). Therefore, despite the vaunted benefits of exposure without click-through, some Internet functions encourage the cost-per-click (CPC) pricing model.

To determine the quality of click, Hoffman and Novak (2000b, 2000c) suggested that the measurability of online advertising should be extended to measure consumer behaviour after click-through, proposing measures such as frequency of visits, time spent on site (TS), bounce rate (BR) (percentage of user's session that views only one page), and number of pages viewed (PV) should be used as indicators of interactivity. If online firms then aim to maximise the traffic to their websites, return on investment (ROI), for the purposes of this research, will be defined as the number of clicks generated per rand spent. Therefore, to maximise return on investment (ROI), one should aim to get the maximum number of quality leads for a given budget by reducing the cost-per-click (CPC).

Sponsored search has spurred the proliferation of the cost-per-click (CPC) pricing model; first implemented by Goto.com in 1998 (Goto.com later became Overture, now part of Yahoo!) (Goldfarb *et al.*, 2007; Edelman *et al.*, 2007). In such searches, advertisers could specify keywords that were relevant to their products and enter the maximum they were willing to pay per click (Goldfarb *et al.*, 2007; Edelman *et al.*, 2007). The actual price paid per click was first determined by a Generalised First-Price Auction (GFP), which later evolved to a Generalised Second-Price Auction (GSP) (Edelman *et al.*, 2006; Edelman *et al.*, 2007). The ease of use, extremely low cost and transparency quickly made this pricing method a hit and led to the success of Goto.com, and to it becoming the advertising provider to major search engines, including Yahoo! and MSN, as well as the later rise to prominence of Google (Edelman *et al.*, 2007; Pedersen, 2008). Sponsored search has become the biggest part of online advertising (PricewaterhouseCoopers 2009; Ghose *et al.* 2009). The report now turns to this concept in detail.

2.5. Sponsored search

The Internet has already become indispensable in some parts of the world (Hoffman, Novak and Venkatesh, 2004). Search engines have facilitated this process, and are indispensable tools for browsers when interacting on the Internet (Laffey, 2007; Jansen *et al.*, 2008). Search engines have left people wondering how they ever survived without them (Economist, 2004). Today, search engines are the most important promotional method used by e-commerce sites, and also the most common way for new e-commerce businesses to market themselves (Jansen *et al.*, 2006). Fain *et al.* (2006), Laffey (2007), Jansen *et al.* (2007) and Jansen *et al.* (2008) stated that sponsored search provides the revenue streams for search engines to

deliver the service for free. Sponsored search, and as a result search engines, have enjoyed phenomenal commercial success (Chen and He, 2006).

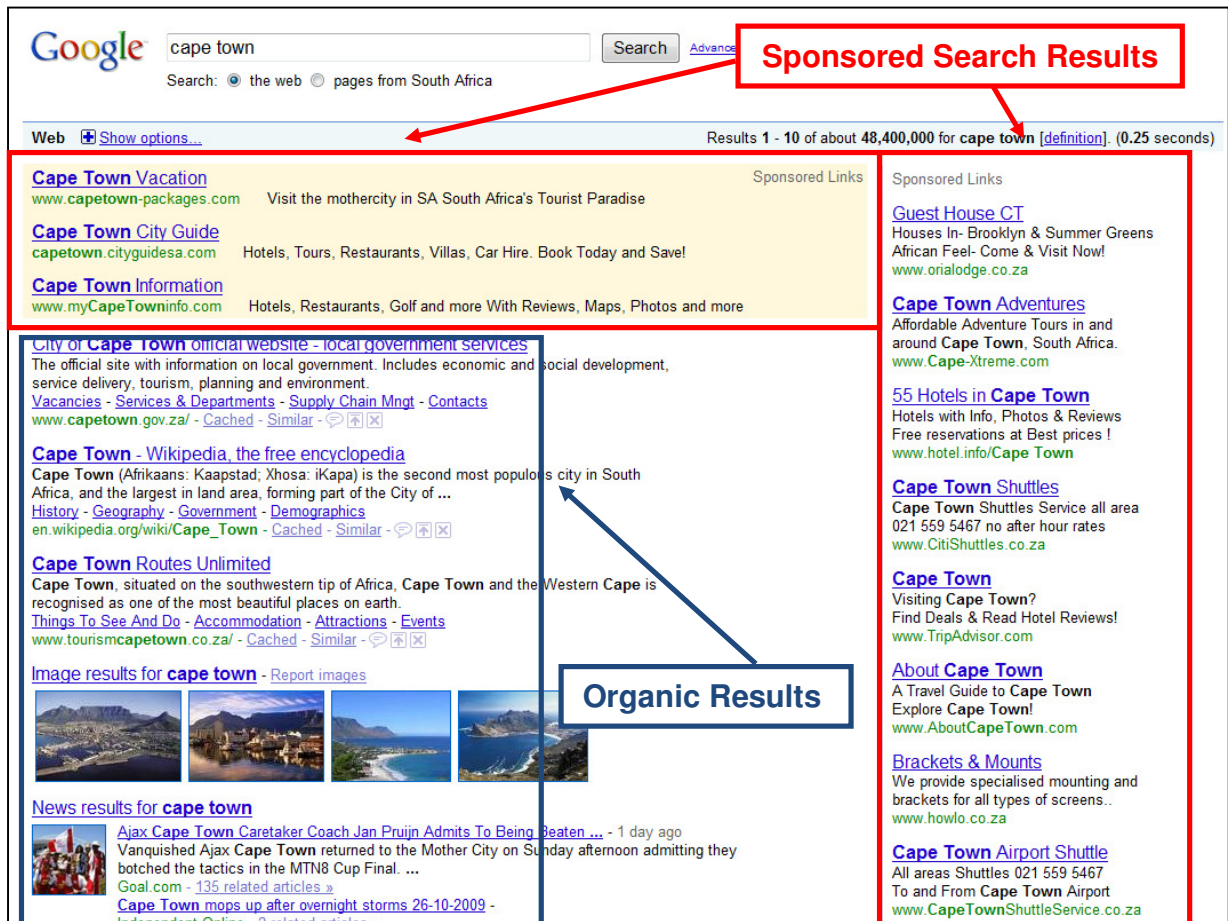


Figure 1: Example of organic and sponsored searched results for keyword 'Cape Town'

Fain *et al.* (2006) and Jansen *et al.* (2007) explained that sponsored search is the relevance-targeted text advertisements displayed above or next to organic search results generated by search engines (See Figure 1). Sponsored search has proven to be a successful business model for search engines, online vendors and advertisers, as well as a valuable way to deliver relevant content to searchers (Jansen *et al.*, 2007).

With the growing use of search engines, sponsored search gives advertisers the valuable opportunity to tap into search engines users' base, lured by the promise of increased traffic from good placement on search pages (Feng *et al.*, 2007). Display (banner) advertising dominated online advertising until 2005, when sponsored search overtook it for the first time (Rutz *et al.*, 2007). Sponsored search is now the predominant form of advertising on the Internet (Chen *et al.*, 2006). Sponsored search matches Internet users' desire for relevant information with advertisers' need for targeted advertisements, resulting in traffic to their websites (Ghose *et al.*, 2009). Sponsored search gives advertisers the ability to target Internet users at the exact point they search for the specific keyword, which is thus related to the users' behaviour, not his/her demographics (Jansen *et al.*, 2008). Sponsored search text ads match the context of the existing goal of the browser and are, therefore, not perceived as a goal impediment (Rutz *et al.*, 2007). This increased ability to target users' intent exactly enables advertisers to reach their markets with much smaller budgets through sponsored search (Ghose *et al.*, 2009).

Bill Gross from Idealab was the first to introduce sponsored search with Goto.com, which incorporated two innovative concepts to search advertising that had not been applied before in advertising markets: 1) using an electronic auction to determine the cost-per-click (CPC), and: 2) the ability of advertisers to target specific search terms or 'keywords', and to set prices at this level (Fain *et al.* 2006; Goldfarb *et al.*, 2007; Jansen *et al.*, 2007; Jansen *et al.*, 2008; Wilbur *et al.*, 2009). Rutz *et al.* (2007) stated that advertisers in sponsored search need to make four basic decisions. The first decision is which keywords to be included in a campaign (Rutz *et al.*, 2007). Keywords are the individual search phrases entered by browsers when using a search engine, and can be one word or a combination of words. Next, the advertiser must decide what the maximum bid should be for each selected keyword, since the pricing

model work on a cost-per-click (CPC) basis (Rutz *et al.*, 2007). Then the text advert must be designed, this usually consisting of a heading or title, three lines of text and a display Uniform Resource Locators (URL) (Rutz *et al.* (2007). The final decision is the design of the landing page, which is the page a browser will be taken to when clicking on the advertisement (Rutz *et al.*, 2007).

Chen *et al.* (2006) argued that an advertiser will bid higher for keywords where his product is more relevant, resulting in the auction pricing model showing his results higher, thus contributing to the relevance of the sponsored search results. Jansen (2007a) showed that sponsored search links are indeed just as relevant to the search query as the organic results produced by search engine algorithms.

Despite the relevance, browsers do seem to have a negative bias towards sponsored search and perceive them as less relevant (Jansen *et al.*, 2006; Jansen *et al.*, 2007). However, because sponsored search advertisements are based on the users' own query, they are considered far less intrusive than online banner or pop-up ads (Ghose *et al.*, 2009).

Google has argued that the main motivation behind separate sets of ads, as well as prices for each separate search term, is to enhance the experience of the browser by increasing the relevance of the advertising to the search term entered (Goldfarb *et al.*, 2007). As stated earlier, Ansari *et al.* (2003) found that improved targeting, consistent with Google's claim, makes consumers and firms better off. Goldfarb *et al.* (2007) claimed, however, that while this is true, it also allows Google to be more effective in price discrimination, thereby countering claims by the Economist (2007) that Google's monopolistic clout is limited because the price of advertising is set by auction.

The auction pricing model relieves the search engine of explicitly assigning a price to each of the millions of keywords, each of which is independently valued by each would-be advertiser (Edelman *et al.*, 2006). The auction price method also helps the search engine to rank the advertisements, since the top position is more desirable because of the higher click-through rate (CTR) (Chen *et al.*, 2006; Edelman *et al.*, 2007; Rutz *et al.*, 2007; Jansen *et al.*, 2007).

Edelman *et al.* (2007) examined the Generalised Second-Price (GSP) auction used by search engines since 2002, and showed that it aims to maximise revenue for search engines. The Generalised Second-Price (GSP) auction type is unique in that it charges the advertiser the highest bid of the advertisement next in line; this is done to prevent strategic bidding by competitors (Chen *et al.* 2006; Edelman *et al.* 2007).

Rutz *et al.* (2008) showed that there is a relationship between generic and branded keywords: generic keywords positively affect branded searches, but there is no relationship the other way round. Conceptually, keyword selection can be argued to be a dynamic form of metatagging (metadata is data about data), which focuses on associating possible search terms with specific websites and pages within these websites (Jansen *et al.*, 2007). By further providing bid prices, degree of matching of search terms, time restrictions, geographical limits and budget amounts, advertisers need to be an active participant in the search process (Jansen *et al.*, 2007).

A prominent position or placement (for example, the slots listed at the top or highlighted in special colour) is commonly believed to be desirable, because of higher click-through rate (CTR) (Ansari *et al.*, 2003; Ghose *et al.*, 2007). Xu *et al.* (2009), however, showed that there

are situations when firms should not pay the premium required to get the top spot, and can get a better return on investment (ROI) by bidding to be placed second or third.

Advertisers would prefer a cost-per-acquisition (CPA) model, where the search engine bears all the risk; and search engines would ideally like a cost-per-thousand impressions (CPMs) model, where the advertiser bears all the risk. Cost-per-click (CPC) advertising has emerged as the middle ground between what advertisers want and what search engines would like as the pricing model, and risk is shared by both parties (Edelman *et al.*, 2007; Jansen *et al.*, 2008).

Search engines have also introduced quality factors, sometimes referred to as the quality score (QS), in assigning ranking of sponsored search. These factors include historical click-through rates (CTRs), keyword relevance, landing page and site quality (Edelman *et al.*, 2007; Jansen *et al.*, 2007; Jansen *et al.*, 2008; Li and Jhang-Li, 2009; Pedersen, 2008). The ranking is then determined by multiplying the keyword quality score (QS) and highest bid. This means that a higher quality score (QS) will result in lower cost-per-click (CPC) and, therefore, give a better return on investment (ROI) (Edelman *et al.*, 2007; Jansen *et al.*, 2008).

Sponsored search also has some problems that still need to be solved. Click fraud remains a substantial threat to the long term sustainability of the business model (Kitts, LeBlanc, Meech and Laxminarayan, 2006; Laffey, 2007; Jansen, 2007b; Jansen *et al.* 2007; Li *et al.*, 2009; Wilber *et al.*, 2009). Click fraud is when an advertisement is clicked with sole intent of generating a charge to the advertiser or exhausting their budget; this can be done by humans or by computer scripts (Jansen *et al.*, 2008; Wilber *et al.*, 2009).

Wilber *et al.* (2009) showed that the search engine industry will benefit from using an independent third party to audit the fraud detection algorithms. While the threat is very real and requires constant attention, Pedersen (2008) argued that it is not as great a threat as many suspect. Despite the potential downfalls, sponsored search is the engine driving much of the innovation from search engines, and provides a measurable approach to advertising on the Web (Jansen *et al.*, 2008). Sponsored search still provides a very high return on investment (ROI) compared to other marketing methods (Szymanski and Lee, 2006).

If, as stated earlier, return on investment (ROI) is defined as the number of quality clicks generated for every rand spent, then employing Anderson's (2004) long tail (LT) concepts to selecting keywords could provide a theoretical approach to help managers improve the return on investment (ROI) in sponsored search advertising by reducing the cost-per-click (CPC). The long tail (LT) also provides a means of making the advertising even more targeted.

These well understood and sometimes contested metrics have been understood to involve greater advertiser participation and improved 'quality score' (QS) determination. Models and tools from economics could potentially aid marketers in this quest.

2.6. Long tail (LT) distributions

The Pareto Principle, or 80/20 rule, has been applied to wealth distribution, city population, product sales and sales force management; however, the Internet has the potential to shift this balance (Brynjolfsson, Hu and Simester, 2007). Anderson (2004) coined and popularised the term 'the long tail' (LT) in describing the effect the Internet had on the sale of niche products and their increasing contribution to total sales.

The long tail (LT) phenomenon has occupied statisticians for many years, and refers to a frequency distribution described by the power law function, where the ‘head’ (HE) is very high but short, and the ‘tail’ low but long, as illustrated in Figure 2 (McDonald, 2008). This is also sometimes called the Pareto Distribution, after the famous Italian economist, Vilfredo Pareto, who determined that 20% of the population accrued 80% of the wealth (Brynjolfsson *et al.*, 2007; McDonald, 2008).

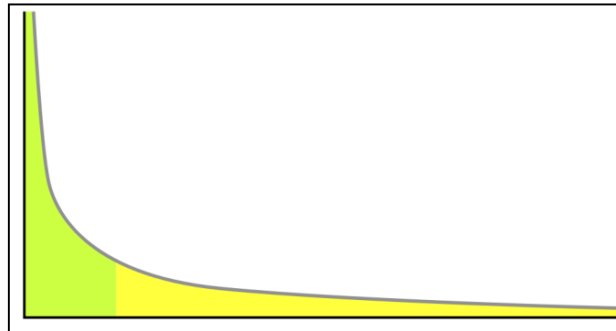


Figure 2: Power law distribution (the long tail (LT) in yellow, head (HE) in green) (Le Ray, 2008)

Anderson (2004) defined the head (HE) as the number of titles available in a traditional retail store, where shelf space is limited, and the tail as the hugely increased titles available in equivalent online retailers’ inventory. Traditional media businesses are built on the ‘hits’ located in the head (HE), a small number of very popular movies, songs, books or television series that are the driving force behind the industry (Anderson 2004; Brynjolfsson, Hu and Smith, 2006; Brynjolfsson *et al.*, 2007; McDonald, 2008). Head (HE) thus refers to the few choices popular with many, while the tail is the many choices popular to small groups of consumers (McDonald, 2008).

When applying the long tail (LT) to offline retailing, Sorensen (2008) argued that managing the relationship between around 400 head (HE) products most households buy once a year, and the more than 40 000 a normal supermarket stocks, will define retail success in the future. The long tail (LT) will also affect the future of media measurement; McDonald (2008) suggested that the 'natural monopolies' of media measurement firms will be threatened as consumers move towards media consumption on multiple platforms, thereby blurring the distinction between sectors. McDonald (2008) goes further and questions if the measurement industries' faith in random sampling and probabilistic inferential statistics may come under pressure in a long tail (LT) future.

In other offline studies, the application of the long tail (LT) principle to customer loyalty and retention showed companies can do a much better job retaining all their customers, not just the ones perceived to be in the top 20% of revenue earners (Ferguson *et al.* (2006). Ferguson *et al.* (2006) calls it dialogue marketing, which extends beyond the reach of direct marketing, and adds to Anderson's (2004) call that this might mean the end of mass marketing.

When the attributes of the Internet are introduced into this scenario, Anderson (2004) argues that improved search technology and recommendation systems will expose consumers to previously inaccessible products, thereby changing the shape of the power law distribution by reducing the size of the 'head' and making the 'tail' longer and fatter. Brynjolfsson *et al.* (2006) stated that both passive and active search tools are essential to move consumers down towards the tail. Anderson (2004) offered examples from online music (Rhapsody), books (Amazon.com) and movies (Netflix) as evidence of this phenomenon.

The impact of the long tail (LT) on the online environment has been furthered been by Brynjolfsson *et al.* (2006). These authors state that the long tail (LT) is driven by both supply and demand side changes, and that a positive feedback loop occurs on both sides, which will amplify its effect over time. Brynjolfsson *et al.* (2007) found evidence of the long tail (LT) when comparing a company's product sales from its website with the products ordered by telephone from its printed catalogue. The online offering was positively skewed towards the long tail (LT), with more customers choosing more obscure products on the Internet compared to catalogue sales (Brynjolfsson *et al.*, 2007).

One further online application of the long tail (LT) concept is to social networking, where users can stay in touch with a far greater number of contacts, even if a contact is only acted upon once a year (Enders, Hungenberg, Denker and Mauch, 2008). Enders *et al.* (2008) argues this combination of online technology and the long tail (LT) of friendship will help social networks build a sustainable business model in the long run.

Nevertheless, some economists have argued the opposite, asserting that the technological changes experienced over that last few decades will fuel a culture where popular products will make an even bigger share of sales, often referred to as 'double jeopardy' (Frank and Cook in Elberse 2008; Dawes 2009). 'The Economics of Superstars' predicted that a smaller and smaller numbers of 'hits', increasing in popularity, will perform even better relative to also-rans, and earn an ever increasing share of revenue and profits (Rosen, 1981). Technologies like the broad, fast Internet connections, as well as the ease with which media can be replicated, are making customers more likely to converge in their tastes and buying habits (Frank *et al.*, in Elberse, 2008). Social forces are also argued to contribute to this

phenomenon, because people find value in listening to the same music and watching the same movies as others (Elberse, 2008).

In support of technological hampering of the long tail (LT), Elberse (2008) studied an online retailer's sales data and concluded that retailers will find it hard to profit from the long tail (LT), and suggested that it remains to be seen if new media can make these previously unprofitable niches profitable. Elberse (2008) suggested that producers aiming to serve the long tail (LT) in a market should ensure that they keep costs low, since chances of success are still limited. According to Elberse (2008), the biggest opportunity in the long tail (LT) will be in selling more copies of old hits, and even goes as far as to warn retailers not to direct customers towards long tail (LT) products, suggesting that retailers risk customers becoming dissatisfied.

In practice, the long tail (LT) plays an enormous part in organic search engine marketing. It is made up of those thousands of keywords that are not searched for very often but, as an aggregate, can drive a significant amount of traffic to a website (Jones, 2008). More and more people are using long tail (LT) type search terms, looking for information related to their search instead of looking for a specific homepage of a website (Jones, 2008). Jansen *et al.* (2007) showed evidence of a classic long tail (LT) distribution when examining more than 1.8 million searches on a search engine. The top 100 terms (of 360,174 unique terms), or 0.003%, accounted for 18.5%, or more than 1 million searches (Jansen *et al.*, 2007).

The current research explored the effects of reaching these long tail (LT) terms on the return on investment (ROI) in sponsored search advertising by including the relevant keywords and targeting the advertising text to the keywords.

2.7. Conclusion

In order to maximise the return on advertising, it is key to reach a targeted audience (Chandra, 2009). By giving access to consumers at the exact point they are searching for a product or service, sponsored search takes targeted advertising to a new level (Jansen *et al.*, 2008). This could accelerate the shift away from mass marketing, and adding the concept of a long tail (LT) distribution allows for targeting on a truly one-to-one level, even when a keyword is only used once a year (Anderson, 2004; Ferguson *et al.*, 2006).

The auction pricing model employed by search engines, however, aims to maximise the revenue for search engines, thereby reducing the return for advertisers by increasing the cost-per-click (CPC) (Edelman *et al.*, 2007). If Anderson's (2004) long tail (LT) principle can be used to include keywords that are far less frequent in their occurrence but vast in numbers compared to the popular head (HE) keywords, this could give account managers a strategy to improved return on investment (ROI) in search campaigns.

The high level of targeting of users' search terms achieved by using long tail (LT) keywords should achieve higher click-through rates (CTRs) which, in turn, will increase the quality score (QS), giving the advertiser a sustainable advantage (Edelman *et al.*, 2007; Jansen *et al.*, 2007; Jansen *et al.*, 2008; Li *et al.*, 2009; Pedersen, 2008). It also allows advertisers to use more targeted advertising text instead of generic text when using broad matches on head (HE) keywords. This, in turn, should increase the click-through rates (CTRs) and, therefore, the quality score (QS) even further (Edelman *et al.*, 2007; Jansen *et al.*, 2007; Jansen *et al.*, 2008; Li *et al.*, 2009; Pedersen, 2008). This research aims to provide empirical evidence to support this argument.



3. RESEARCH HYPOTHESES

3.1. Introduction

In order to explore if the long tail approach to selecting keywords for a sponsored search campaign can offer marketers a tool to improve ROI, the research objectives are combined with the literature reviewed and, therefore, the following research hypotheses are proposed.

3.2. Objective 1

The null hypothesis under Objective 1 states that head (HE) keywords will achieve a similar cost-per-click (CPC) and, therefore, a similar return on investment than long tail (LT) keywords. The alternative hypothesis states that long tail (LT) keywords will achieve a lower cost-per-click (CPC) and, therefore, a higher return on investment (ROI) than the head (HE) keywords.

$$H1_0: CPC_{HE} - CPC_{LT} = 0$$

$$H1_A: CPC_{HE} - CPC_{LT} > 0$$

3.3. Objective 2

The null hypothesis under Objective 2 states that head (HE) keywords will achieve similar click through rates (CTR) than a long tail (LT) keywords. The alternative hypothesis states that long tail (LT) keywords will yield higher click through rates (CTR) than head (HE) keywords.

$$H2_0: CTR_{HE} - CTR_{LT} = 0$$

$$H2_A: CTR_{HE} - CTR_{LT} < 0$$

3.4. Objective 3

The first null hypothesis under Objective 3 states that head keywords without targeted text adverts (WOTA) will yield similar click through rates (CTRH) than head keywords with targeted text adverts (WTA). The alternative hypothesis states that head keywords with targeted adverts (WTA) will yield higher click through rates (CTRH) than head keywords without targeted adverts (WOTA).

$$H_{3_0}: CTRH_{WOTA} - CTRH_{WTA} = 0$$

$$H_{3_A}: CTRH_{WOTA} - CTRH_{WTA} < 0$$

The second null hypothesis under Objective 3 states that long tail keywords without targeted text adverts (WOTA) will yield similar click through rates (CTRL) than long tail keywords with targeted text adverts (WTA). The alternative hypothesis states that long tail keywords with targeted adverts (WTA) will yield higher click through rates (CTRL) than long tail keywords without targeted adverts (WOTA)

$$H_{4_0}: CTRL_{WOTA} - CTRL_{WTA} = 0$$

$$H_{4_A}: CTRL_{WOTA} - CTRL_{WTA} < 0$$

3.5. Objective 4

The first null hypothesis under Objective 4 states that the pages viewed (PV) by users from head (HE) keywords is equal to those from long tail (LT) keywords. The alternative hypothesis states that pages viewed (PV) by users from head (HE) keywords is not equal to those from long tail (LT) keywords.

$$H5_0: PV_{HE} - PV_{LT} = 0$$

$$H5_A: PV_{HE} - PV_{LT} \neq 0$$

The second null hypothesis under Objective 4 states that the time spent (TS) by users from head (HE) keywords is equal to those from long tail (LT) keywords. The alternative hypothesis states that times spent (TS) by users from head (HE) keywords is not equal to those from long tail (LT) keywords.

$$H6_0: TS_{HE} - TS_{LT} = 0$$

$$H6_A: TS_{HE} - TS_{LT} \neq 0$$

The third null hypothesis under Objective 4 states that the bounce rate (BR) of users from head (HE) keywords will be equal to those from long tail (LT) keywords. The alternative hypothesis states that the bounce rate (BR) for head (HE) keywords will not be equal that of long tail (LT) keywords.

$$H7_0: BR_{HE} - BR_{LT} = 0$$

$$H7_A: BR_{HE} - BR_{LT} \neq 0$$

4. RESEARCH METHODOLOGY

4.1. Introduction

Marketing managers make decisions every day based on assumed causal relationships. Since these assumptions cannot be justified, formal casual research is needed to examine the causal relationship (Malhotra, 2007, p89). Causal research is used to identify cause-and-effect relationships between variables (Zikmund, 2003, p56). In doing causal research, the researcher manipulates one or more independent variables to test the effect on the dependant variable (Malhotra, 2007, p81). Bagozzi in Hulland, Chow, and Lam (1996) suggested that there are four key advantages to causal models:

1. Causal models make the assumptions, constructs and hypothesised relationships in a researcher's theory clear.
2. Causal models add an amount of precision to a researcher's theory, since they require apparent definitions of constructs, operationalisations and the functional relationship between constructs.
3. Causal models permit a more comprehensive representation of complex theories.
4. Causal models provide a formal framework for constructing and testing both theories and measures.

4.2. Research design

Malhotra (2007, p91) stated that not every research needs to start with exploratory research. It depends on the precision with which the problem has been defined, and the researcher's

degree of conviction about the approach to the problem. This research aims to show such a cause-and-effect relationship between keyword distribution and targeted advertising text (both causal/independent variables) and return on investment (ROI) by examining the cost-per-click achieved (effect/dependent variable). Since the problem in this research has been well defined, this research seeks to establish evidence of a causal relationship. Malhotra (2007, p221) states that the conditions for causality are:

- Concomitant Variation, which is the extent to which a cause (keyword distribution, targeted text ad) and effect (cost per click (CPC)) occur together and vary together in the way predicted by the hypothesis under consideration.
- Time order of occurrence of variables, which states that the causing event (keyword distribution, targeted text ad) must occur either before or at the same time as the effect (lower CPC); it cannot occur afterwards.
- Elimination of other causal factors. These can never be completely excluded, but the hypothesis testing aims to provide statistical evidence.

This research will aim to meet two of these conditions, as it is impossible to meet the third condition. The research area is well understood and readily quantifiable, therefore, causal research, not exploratory research, is appropriate.

Within causal research types, a factorial design has been chosen. Malhotra (2007, p237) explained, "Factorial design is a statistical experimental design that is used to measure the effect of two or more independent variables at various levels and allows for interaction between variables." Interaction is when the combined effect of two variables is different from the sum of their individual effects (Malhotra 2007, p237; Zikmund 2003, p269, p283).

The design is a classic two by two design. There will be 2 x 2 levels to this research, the independent or causal variables are keyword distribution and targeted ad text, each of which have the two levels of head (HE) and long tail (LT). The dependent variables are cost-per-click (CPC) which will determine the return on investment (ROI). Factorial design, therefore, is appropriate (see Figure 3).

	Keywords Distribution	
Ad Text	Head Targeted	Long Tail Targeted
	Head Generic	Long Tail Generic

Figure 3: Causal variables – keyword distribution and advert text

A wealth of data will be available for individual keyword performance using both Google Adwords and Analytics reports, including impressions, clicks, click through rates, average cost per click, average time spent, and pages viewed. A single company’s online advertising campaign will be used for this research. Keywords selection, advert creation and setting bid prices were done using software programmes.

External validity refers to whether the cause-and-effect relationship found in the research can be generalized (Malhotra 2007, p225). Because the data was all gathered in actual market conditions or field environment, it enhances external validity (Malhotra, 2007, p225). The results should therefore be applicable in sponsored search in general.

Internal validity is a measure of the accuracy of the experiment, it measures if the independent variable did indeed cause the effect on dependent variables (Malhotra, 2007, p225). Extraneous variables are factors that influence internal validity (Malhotra, 2007, p226). Two extraneous variables that could have affected the results were history and mortality. History refers to external event that are external to the experiment event but occurred simultaneously to the experiment (Malhotra, 2007, p226). The economic conditions during the experiment could have influenced the results. Mortality is the loss of a test unit while experiment is in progress (Malhotra, 2007, p227). In this context mortality could be users disconnected while still browsing the website because of connectivity issues. This should however affect all subsets of the data.

4.3. Unit of analysis

In order to answer the hypotheses stated earlier, the unit of analysis for the research will be the keyword.

4.4. Universe

Keywords are the most basic building blocks of a sponsored search account; together with ad creative (advert text) they form ad groups. Ad Groups can have multiple keywords and multiple ad creatives. Cost-per-click (CPC), impressions, click-through rate and cost data are available in ad groups for individual keywords and ad creatives. Campaigns are made up of ad groups, and campaigns, in turn, make up the account. See Figure 4.

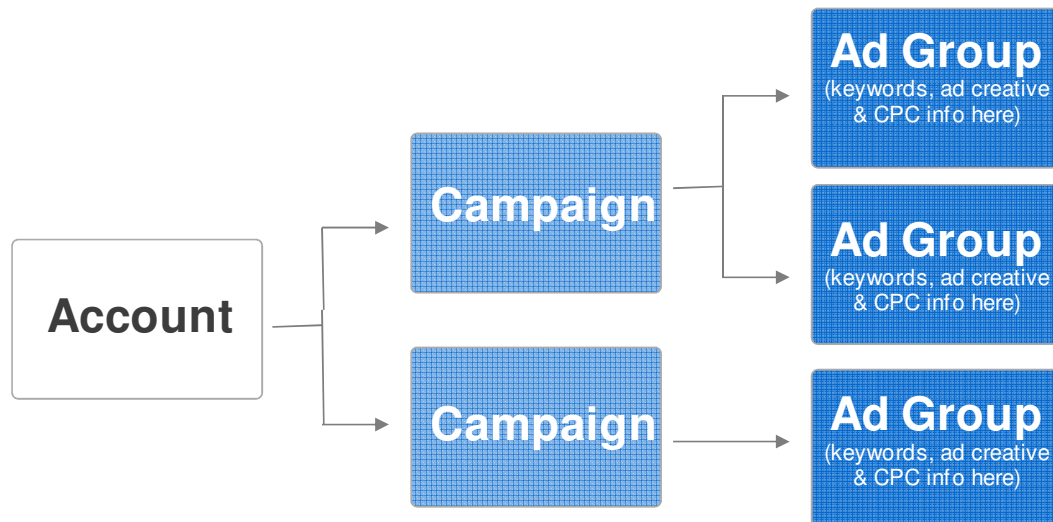


Figure 4: Google Adwords hierarchical setup (Google Adwords 2009)

The universe will be all the keywords in all four campaigns that received at least one click during the specific time used in sampling.

4.5. Sampling method and size

The research will focus on a single company's sponsored search advertising. The company operates an Internet portal for property in South Africa, and has been running sponsored search advertising for a few years. The account has been set up in four campaigns:

1. Sales related head keywords, including brand and generic terms.
2. Rental related head keywords.
3. Sales related long tail keywords.
4. Rental related long tail keywords.

Both rental and sales long tail campaign keywords were generated by combining suburb names with property related terms, for example, 'property for sale in Woodlands.' Property

related terms included property types, for example, ‘duplex’, ‘simplex’, ‘townhouse.’ Each ad group was based on a single suburb to allow for targeted ad text around the suburb (see Figure 5).

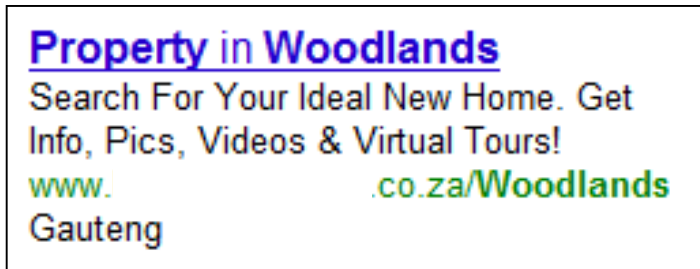


Figure 5: Example of ad text targeted around the suburb name

A census involves total enumeration of all the elements of the population, and therefore allows the parameters of the population to be calculated in a straightforward way once the census has been enumerated (Malhotra, 2007, p335). The number of keywords in the four campaigns exceeded 50,000, of which only around 7,000 got at least one click in July 2009, the month whose data is used for this research. All the ad text was targeted so, in order to gather data for research Objective 3, a random sample of keywords was selected to run an experiment during October 2009, without targeted text ad. A total of just over 1,000 keywords was selected for this experiment, with just under 200 head keywords and around 800 long tail keywords. The experiment was run in the same ad campaign to ensure that historical data that influence the quality score was maintained.

The total data set was a combination of a census of all keywords with a least one click during July 2009, combined with the random sample used in the experiment with generic

advert text during October 2009. During July 2009 the average user on the website used for this research, viewed 9.09 pages, spent 520s on the website, and 33.98% viewed only a single page. This data will be important when analysing the results to objective four.

4.6. Data gathering

Malhotra (2007, p203) states that there are five observation methods: personal, mechanical, audit, content analysis and trace analysis. Trace analysis data collection is based on physical traces and evidence of earlier behaviour, with Internet cookies being a prime example (Malhotra, 2007 p207). Cookies are a sophisticated means by which websites collect user information by storing text data in the user's browser. Google Adwords and Google Analytics use cookies to collect the data related to sponsored search advertising and the resulting traffic to the advertisers' websites. The current study, therefore, will collect data by means of a trace analysis.

Google Adwords gave the following data: impressions, clicks, click-through rate, average cost per click and total cost per keyword, for every keyword. Google Analytics provided the following data: pages per visit, bounce rate and time spent on site for every keyword. Both Adwords and Analytics are Internet based interfaces with the ability to export data to .csv files. These files were then merged to form one central database for analysis.

4.7. Data analysis

Data analysis for each hypothesis will be as follows:

- Hypothesis 1 needed to compare the means of two independent data sets; the most appropriate statistical technique to do this was a t-Test (Zikmund, 2003 p524; Albright, Winston and Zappe, 2006 p496).
- Hypothesis 2 needed to compare the means of two independent data sets; the most appropriate statistical technique to do this was a t-Test (Zikmund, 2003, p524; Albright *et al.*, 2006, p496).
- Hypotheses 3 and 4 were combined to form four data sets, each with a mean. The method needed to compare the means of more than two data sets was an analysis of variance (ANOVA) (Zikmund, 2003, p529, Albright *et al.*, 2006, p537).
- Hypothesis 5 needed to compare the means of two independent data sets; the most appropriate statistical technique to do this was a t-Test (Zikmund, 2003, p524; Albright *et al.*, 2006, p496).
- Hypothesis 6 needed to compare the means of two independent data sets; the most appropriate statistical technique to do this was a t-Test (Zikmund, 2003, p524; Albright *et al.*, 2006, p496).
- Hypothesis 7 needed to compare the means of two independent data sets; the most appropriate statistical technique to do this was a t-Test (Zikmund, 2003, p524; Albright *et al.*, 2006, p496).

5. RESULTS

5.1. Introduction

The results from the research will be presented in this chapter. The objective of this research was to explore if there is a causal relationship between both keyword selection and targeted ad text (independent variables) and return on investment (ROI) by examining the effect on cost-per-click (CPC) (dependent variable).

A total of 6,639 of the more than 50,000 keywords in the firm's account received at least one click during July 2009. Because all of these keywords had targeted text ad, 1,110 were randomly chosen to be used in an experiment during October of 2009 with generic ad text. Of these, 521 received at least one click during this week. Total dataset for analysis was thus 7,160 keywords.

The results will be presented per hypothesis, and then some interesting observations about the data that is not covered by the hypothesis are noted. The chapter will end with a summary of the results which will lead into the discussion about the results in Chapter 6.

5.2. Cost-per-click: Hypothesis 1

The null hypothesis under Objective 1 states that head (HE) keywords will achieve a similar cost-per-click (CPC) and, therefore, a similar return on investment than long tail (LT) keywords. The alternative hypothesis states that long tail (LT) keywords will achieve a lower cost-per-click (CPC) and, therefore, a higher return on investment (ROI) than the head (HE) keywords.

$$H_{10}: CPC_{HE} - CPC_{LT} = 0$$

$$H_{1A}: CPC_{HE} - CPC_{LT} > 0$$

The dataset to test this hypothesis was 7,160 keywords, of which 1,095 were head keywords and 6,065 were long tail keywords. The data was distributed as show in Figure 6.

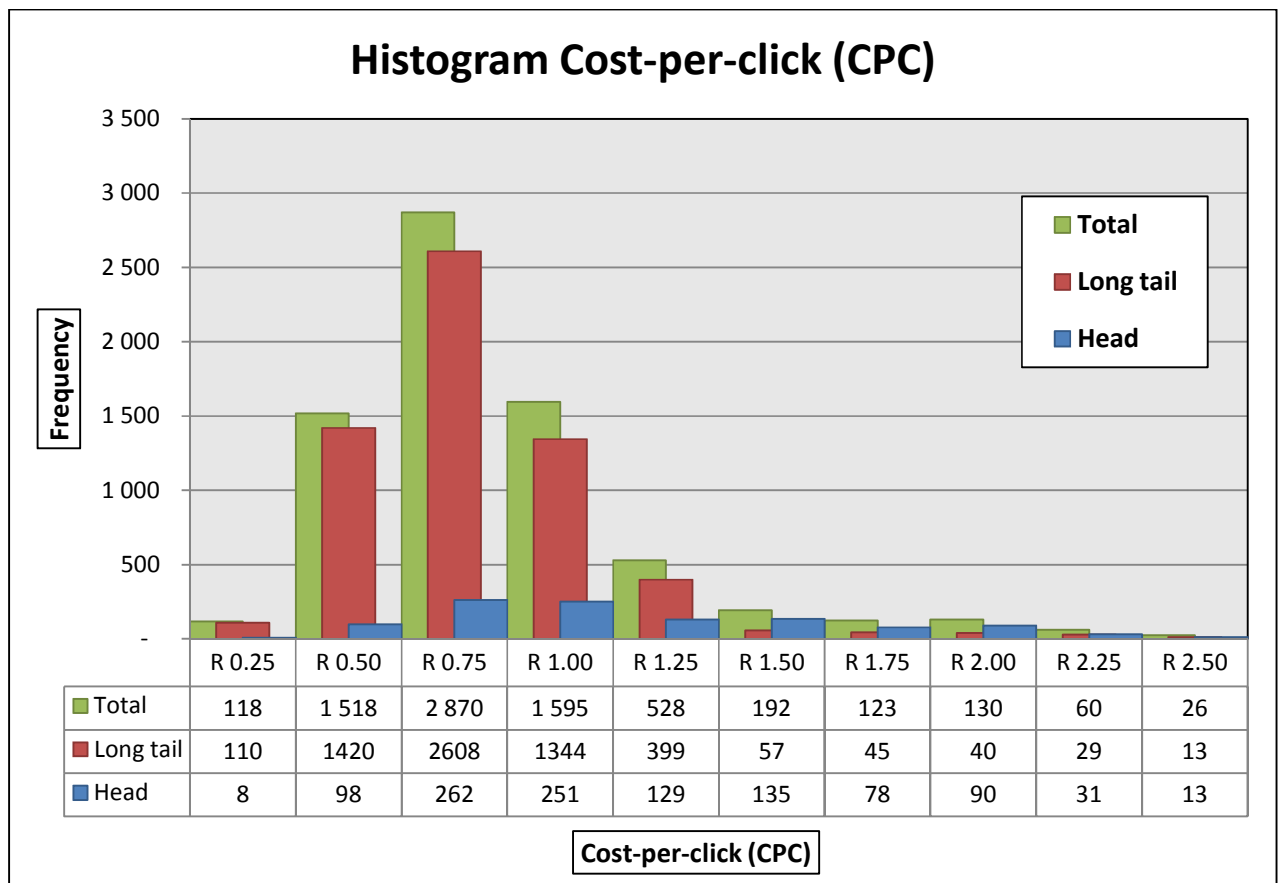


Figure 6: Histogram for cost-per-click (CPC) distribution

The mean of the head keywords' cost-per-click (CPC) was R1.04, with a standard deviation of R0.49. For long tail keywords the mean of the cost-per-click (CPC) was R0.68, with a standard deviation of R0.28. The complete descriptive statistics below in Table 1.

<i>Cost-per-click (CPC)</i>				
	Head		Long tail	
Mean	R	1.044	R	0.683
Standard Error	R	0.015	R	0.004
Median	R	0.912	R	0.647
Mode	R	0.350	R	0.540
Standard Deviation	R	0.490	R	0.288
Sample Variance		0.240		0.083
Kurtosis		-0.387		6.694
Skewness		0.667		1.852
Range	R	2.316	R	2.430
Minimum	R	0.154	R	0.070
Maximum	R	2.470	R	2.500
Sum		1143		4142
Count		1095		6065

Table 1: Descriptive statistics for cost-per-click (CPC)

Hypothesis 1 needed to compare the means of two independent data sets; the most appropriate statistical technique to do this was a t-Test (Zikmund, 2003 p524; Albright *et al.*, 2006, p496). The result of the t-Test is show below in Table 2 :

CPC	Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal Variances Assumed	829.11	.000	33.57	7158	.000	R.3607	R.0107	R.339	R.381
Equal Variances Not Assumed			23.62	1233	.000	R.3607	R.0153	R.330	R.390

Table 2: t-Test for equality of means in cost-per-click (CPC)

Levene's test shows a significance of 0.000, since this is less than 0.1 equal variance cannot be assumed. With no assumption of equal variance the t-value = 23.62, with degrees of

freedom = 1,233. The mean difference = R0.36, indicating that, on average, a head (HE) keyword will have a cost-per-click (CPC) R0.36 higher than that of a long tail (LT) keyword.

The p-value = 0.000 indicates there is a significant difference in the means of the samples, and the null hypothesis (H_{10}) should be rejected in favour of the alternative hypothesis (H_{1A}). Therefore, the data supports the following: H_{1A} : $CPC_{HE} - CPC_{LT} > 0$. Which means the cost-per-click (CPC) of long tail (LT) keywords is significantly lower than head (HE) keywords.

5.3. Click-through rate: Hypothesis 2

The null hypothesis under Objective 2 states that head (HE) keywords will achieve similar click-through rate (CTR) than a long tail (LT) keywords. The alternative hypothesis states that long tail (LT) keywords will yield higher click-through rates (CTR) than head (HE) keywords.

$$H_{20}: CTR_{HE} - CTR_{LT} = 0$$

$$H_{2A}: CTR_{HE} - CTR_{LT} < 0$$

The dataset to test this hypothesis was 7,160 keywords, of which 1,095 were head (HE) keywords and 6,065 were long tail (LT) keywords. The data was distributed as show in Figure 7. The mean of the head keywords click-through rate (CTR) was 10.42%, with a standard deviation of 14.8%. For long tail (LT) keywords, the mean of the click through rate (CTR) was 20.4%, with a standard deviation of 21.7%. The complete descriptive statistics are in Table 3.

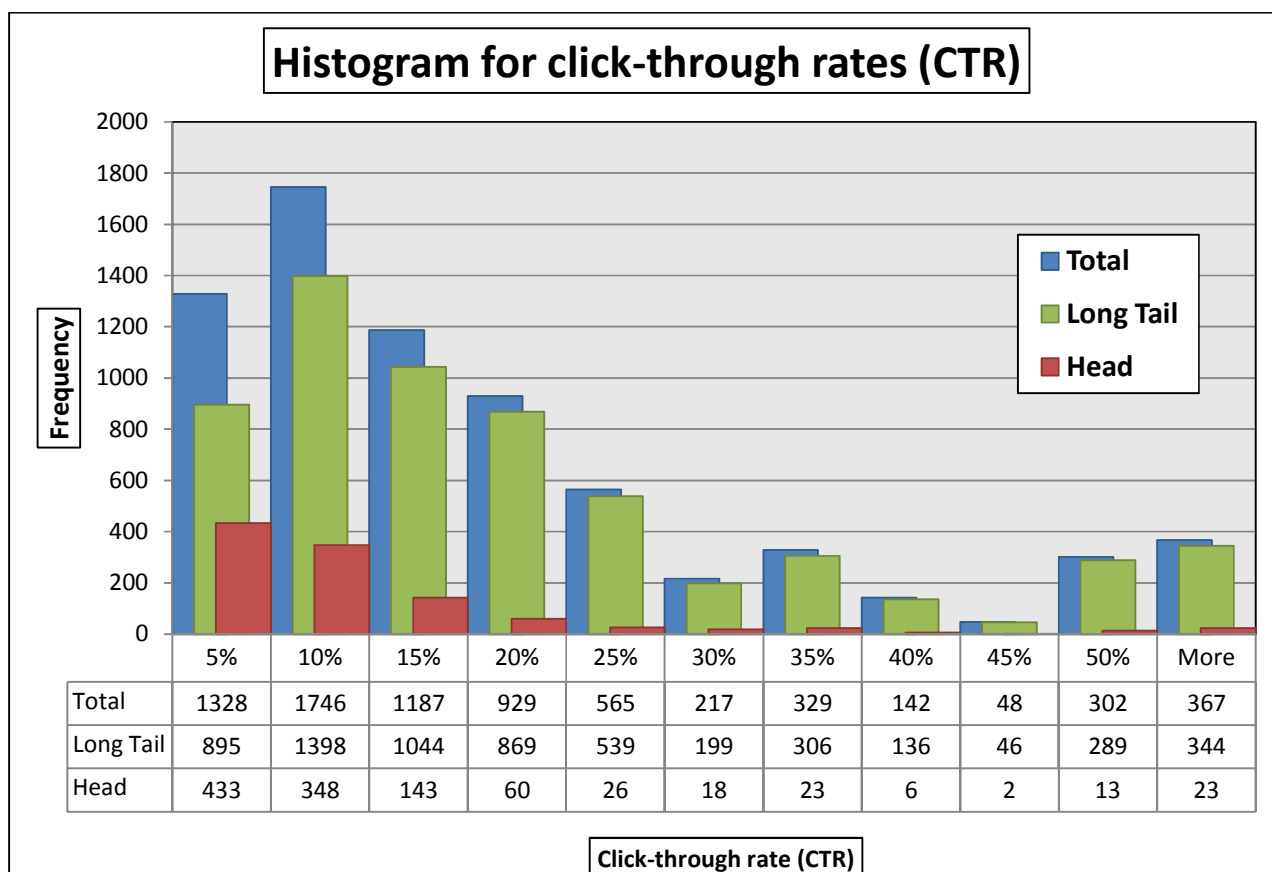


Figure 7: Histogram for click-through rate (CTR) distribution

<i>Click-through rate (CTR)</i>		
	Head	Long tail
Mean	10.42%	20.40%
Standard Error	0.00448	0.00279
Median	6.56%	13.33%
Mode	33.33%	50.00%
Standard Deviation	14.81%	21.72%
Sample Variance	0.0219	0.0472
Kurtosis	20.051	7.567
Skewness	4.097	2.522
Range	100%	200%
Minimum	0%	0%
Maximum	100%	200%
Sum	114.112	1237.48
Count	1095	6065

Table 3: Descriptive statistic for click-through rate (CTR)



Hypothesis 2 needed to compare the means of two independent data sets; the most appropriate statistical technique to do this was a t-Test (Zikmund, 2003, p524; Albright *et al.*, 2006, p496). The result of the t-Test is show below in Table 4.

CTR	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal Variances Assumed	149.0	.000	-14.60	7158	.000	-.09982	.00683	-.11322	-.08643
Equal Variances Not Assumed			-18.92	2052	.000	-.09982	.00527	-.11017	-.08948

Table 4: t-Test for equality of means in click through rate (CTR) for head and long tail keywords

Levene's test shows a significance of 0.000, since this is less than 0.1 equal variance cannot be assumed. With no assumption of equal variance the t-value = -18.92, with degrees of freedom = 2,052. The mean difference = - 9.9%, indicating that on average a head (HE) keyword will have a click through rate (CTR) 9.9% lower than that of a long tail (LT) keyword.

The p-value = 0.000 indicates there is a significant difference in the means of the samples, and, therefore, the null hypothesis (H_{20}) should be rejected in favour of the alternative hypothesis (H_{2A}). Therefore, the data supports the following: $H_{2A}: CTR_{HE} - CTR_{LT} < 0$. Which means the click through rate (CTR) of long tail (LT) keywords is significantly higher than head (HE) keywords.

5.4. Targeted and generic advert text: Hypothesis 3 and 4

The first null hypothesis under Objective 3 states that head (HE) keywords without targeted text adverts (WOTA) will yield similar click-through rates (CTRH) than head (HE) keywords with targeted text adverts (WTA). The alternative hypothesis states that head (HE) keywords with targeted adverts (WTA) will yield higher click-through rates (CTRH) than head (HE) keywords without targeted adverts (WOTA).

$$H_{3_0}: CTRH_{WOTA} - CTRH_{WTA} = 0$$

$$H_{3_A}: CTRH_{WOTA} - CTRH_{WTA} < 0$$

The second null hypothesis under Objective 3 states that long tail (LT) keywords without targeted text adverts (WOTA) will yield similar click-through rates (CTRL) than long tail (LT) keywords with targeted text adverts (WTA). The alternative hypothesis states that long tail (LT) keywords with targeted adverts (WTA) will yield higher click-through rates (CTRL) than long tail keywords without targeted adverts (WOTA).

$$H_{4_0}: CTRL_{WOTA} - CTRL_{WTA} = 0$$

$$H_{4_A}: CTRL_{WOTA} - CTRL_{WTA} < 0$$

The dataset to test these two hypothesis was 7,160 keywords, of which 200 were head (HE) keywords and 321 long tail (LT) keywords with generic text ads, and 895 head (HE) keywords and 5,744 long tail (LT) keywords with targeted text ads. The data was distributed as show in Figure 8 and Figure 9.

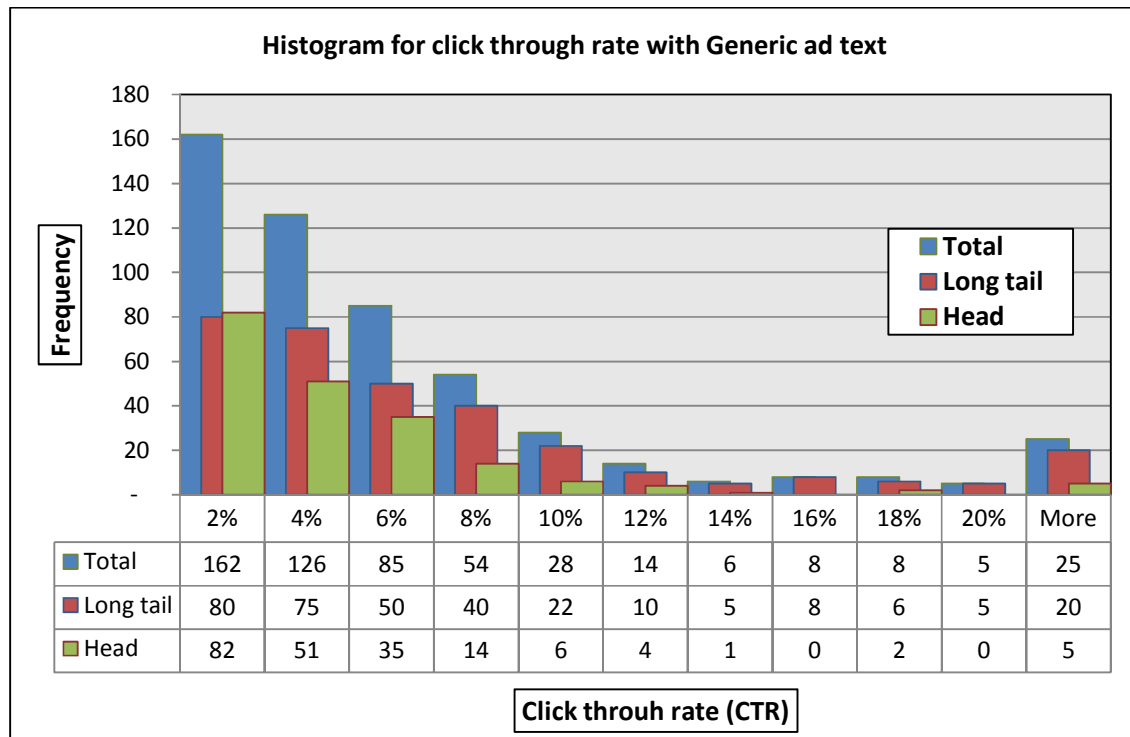


Figure 8: Histogram for click-through rate (CTR) distribution for generic text ad

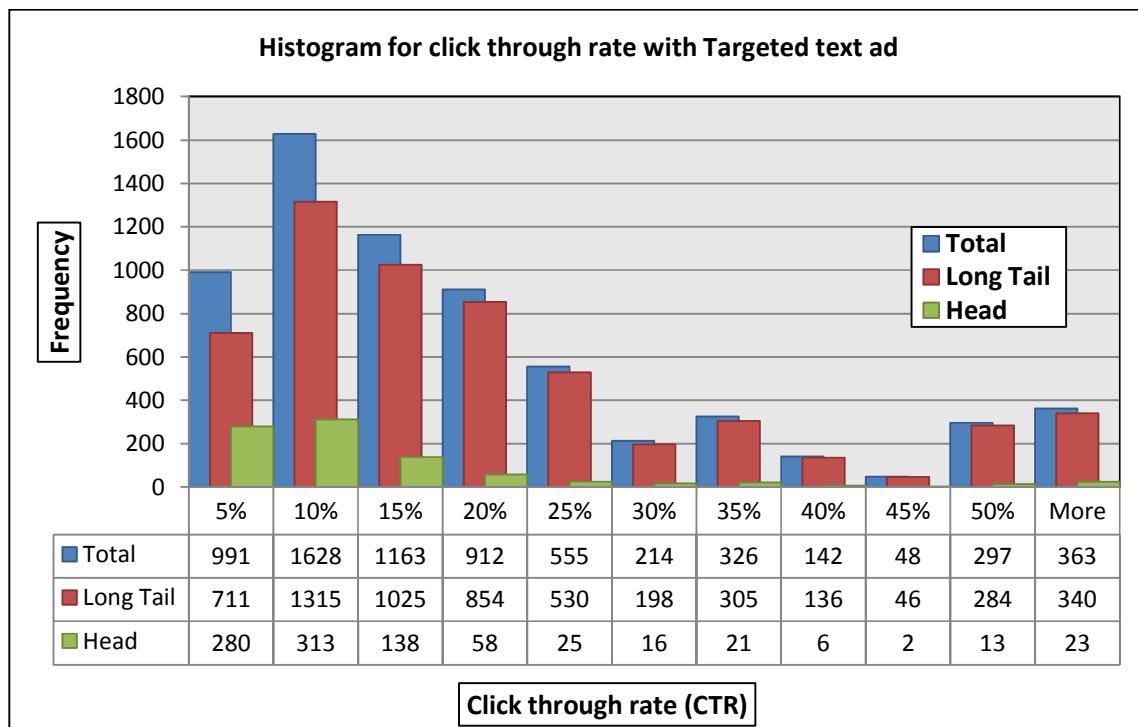


Figure 9: Histogram for click-through rate (CTR) distribution for targeted text ad

The mean click-through rate (CTR) of the head (HE) keywords with generic advert text was 3.95%, with a standard deviation of 4.89%. For long tail (LT) keywords with generic text ads, the mean of click through rate (CTR) was 7.66%, with a standard deviation of 13.04%. Head (HE) keywords with targeted text ads had a mean of 11.87% with a standard deviation of 15.86%, while long tail (LT) keywords with targeted text ads had a mean of 21.12% and a standard deviation of 21.88%. The complete descriptive statistics are in Table 5.

<i>CTR Targeted/Generic</i>				
	Head Targeted	Head Generic	Long Tail Targeted	Long Tail Generic
Mean	11.87%	3.95%	21.12%	7.66%
Standard Error	0.53%	0.35%	0.29%	0.73%
Median	7.69%	2.61%	14.29%	4.30%
Mode	33.33%	4.55%	50.00%	25.00%
Standard Deviation	15.86%	4.89%	21.88%	13.04%
Sample Variance	2.52%	0.24%	4.79%	1.70%
Kurtosis	16.981	16.739	7.347	30.976
Skewness	3.825	3.664	2.491	5.086
Range	100%	33%	200%	100%
Minimum	0%	0%	0%	0%
Maximum	100%	33%	200%	100%
Sum	106	8	1213	25
Count	895	200	5744	321

Table 5: Descriptive statistic for click-through rate (CTR) for targeted and generic

Hypotheses 3 and 4 were combined to form four data sets, each with a mean. The method needed to compare the means of more than two data sets was an analysis of variance (ANOVA) (Zikmund, 2003, p529; Albright *et al.*, 2006, p537). The result of the ANOVA is shown in Table 6.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	15.776	3	5.259	124.012	.000
Within Groups	303.452	7156	.042		
Total	319.229	7159			

Table 6: ANOVA for equality of means in click through rate (CTR) for targeted and generic, head and long tail samples

The $p = 0.000$ allows us to reject equal means hypothesis, indicating that there is a significant difference in variance in the means between groups. In order to explore these differences and examine hypotheses 3 and 4, a post hoc test was performed. Sheffe (post hoc test) performs simultaneous joint pairwise comparisons for all possible pairwise combinations; it is used to examine all possible linear combinations of group means (SPSS, 2006). The results from the post hoc test are show below in Table 7.

(I) HvsLTG	(J) HvsLTG	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Head Targeted	Head Generic	.07921 [*]	.01611	.000	.0342	.1242
	Long tail Targeted	-.09248 [*]	.00740	.000	-.1132	-.0718
	Long tail Generic	.04212 [*]	.01340	.020	.0047	.0796
Head Generic	Head Targeted	-.07921 [*]	.01611	.000	-.1242	-.0342
	Long tail Targeted	-.17169 [*]	.01481	.000	-.2131	-.1303
	Long tail Generic	-.03709	.01855	.262	-.0890	.0148
Long tail Targeted	Head Targeted	.09248 [*]	.00740	.000	.0718	.1132
	Head Generic	.17169 [*]	.01481	.000	.1303	.2131
	Long tail Generic	.13460 [*]	.01181	.000	.1016	.1676
Long tail Generic	Head Targeted	-.04212 [*]	.01340	.020	-.0796	-.0047
	Head Generic	.03709	.01855	.262	-.0148	.0890
	Long tail Targeted	-.13460 [*]	.01181	.000	-.1676	-.1016

Table 7: ANOVA post hoc test, Sheffe for equality of means in click-through rate (CTR) for targeted and generic, head and long tail samples

The p-value = 0.000 for the difference in means between head (HE) targeted (WTA) and head generic or without targeted ad text (WOTA), indicated the null hypothesis can be rejected in favour of the alternative hypothesis. Therefore, the data supports the following: $H_{3A}: CTR_{WOTA} - CTR_{WTA} < 0$. This indicates that head (HE) keywords with targeted ad text achieve a higher click-through rate (CTR) than head (HE) keywords without ad text.

The p-value = 0.000 for the difference in means between long tail (LT) targeted (WTA) and long tail (LT) generic or without targeted ad text (WOTA), indicated the null hypothesis can be rejected in favour of the alternative hypothesis. Therefore, the data supports the following: $H_{4A}: CRTL_{WOTA} - CRTL_{WTA} < 0$. This indicates that long tail (LT) keywords with targeted ad text (WTA) achieve a higher click-through rate (CTR) than long tail (LT) keywords without ad text (WOTA).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^b
Corrected Model	15.776 ^a	3	5.259	124.012	.000	.049	372.036	1.000
Intercept	21.134	1	21.134	498.376	.000	.065	498.376	1.000
Add_Type	4.860	1	4.860	114.605	.000	.016	114.605	1.000
HeadLongtail	1.785	1	1.785	42.090	.000	.006	42.090	1.000
Add_Type * HeadLongtail	.326	1	.326	7.691	.006	.001	7.691	.792
Error	303.452	7156	.042					
Total	574.369	7160						
Corrected Total	319.229	7159						

Table 8: Test of between subject effects

The p-value = 0.006 on Table 8 indicate that there is significant interaction effect. An interaction effect occurs when the effect of an independent variable (keyword selection) is different for different levels or categories of another independent variable (advert type) (Malhotra, 2007, p520; Hair, Anderson, Tatham and Black, 1998, p408). In other words the

sum of the effect of targeted advert text and a long tail keyword selection is greater than the individual effects.

It is interesting to note that p-value = 0.262 for the difference in means between long tail (LT) generic and head (HE) generic, indicating that there is not a significant difference in the means between these two samples.

5.5. Pages viewed: Hypothesis 5

The first null hypothesis under Objective 4 states that the pages viewed (PV) by users from head (HE) keywords is equal to those from long tail (LT) keywords. The alternative hypothesis states that pages viewed (PV) by users from head (HE) keywords is not equal to those from long tail (LT) keywords.

$$H_{5_0}: PV_{HE} - PV_{LT} = 0$$

$$H_{5_A}: PV_{HE} - PV_{LT} \neq 0$$

The dataset to test this hypothesis was 7,160 keywords, of which 1,095 were head keywords (HE) and 6,065 were long tail (LT) keywords. The data was distributed as shown in Figure 10. The mean of the head (HE) keywords page views was 10.6, with a standard deviation of 7.97 pages per visit. For long tail (LT) keywords, the page views mean was 6.4, with a standard deviation of 8.66. The complete descriptive statistics are presented in Table 9.

Hypothesis 5 needed to compare the means of two independent data sets; the most appropriate statistical technique to do this was a t-Test (Zikmund, 2003, p524; Albright *et al.*, 2006, p496). The results of the t-Test are shown below in Table 10.

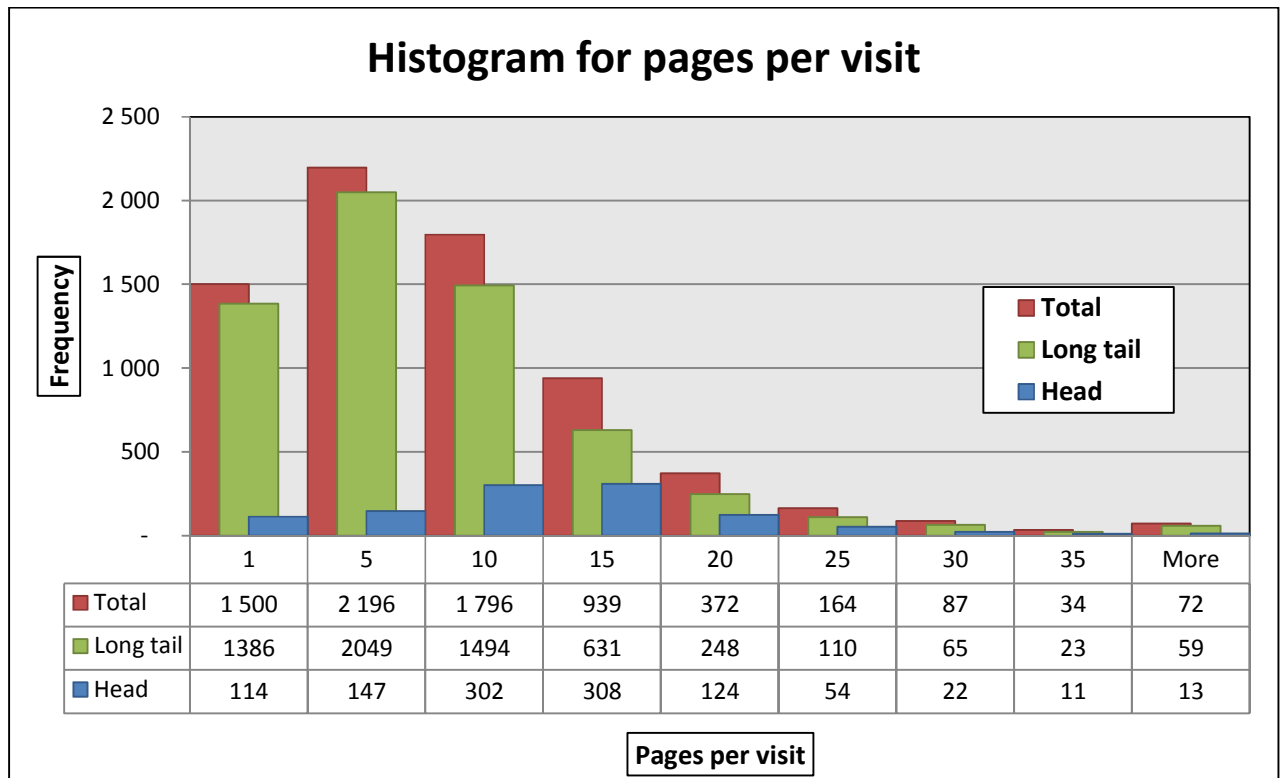


Figure 10: Histogram for pages per visit

<i>Page views per visit</i>		
	Head	Long tail
Mean	10.68	6.40
Standard Error	0.24	0.11
Median	9.8	4.33
Mode	0	0
Standard Deviation	7.97	8.66
Sample Variance	63.56	75.05
Kurtosis	7.89	223.48
Skewness	1.90	9.46
Range	65	291
Minimum	0	0
Maximum	65	291
Sum	11690	38818
Count	1095	6065

Table 9: Descriptive statistic for page views per visit

Pages per Visit	Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal Variances Assumed	9.61	.002	15.21	7158	.000	4.275	.28111	3.72462	4.8267
Equal Variances Not Assumed			16.11	1597	.000	4.275	.26537	3.75517	4.7961

Table 10: t-Test for equality of means in page views per visit

Levene's test shows a significance of 0.002. Since it is less than 0.1, equal variance cannot be assumed. With no assumption of equal variance the t-value = 16.11, with degree of freedom = 1,597. The mean difference = 4.275, indicating that the average head (HE) keyword generates 4.275 more page views (PV) per session compared to a long tail (LT) keyword.

The p-value = 0.000 indicates there is a significant difference in the means of the samples, and, therefore, the null hypothesis (H_{5_0}) should be rejected in favour of the alternative hypothesis (H_{5_A}). Therefore, the data supports the following: $H_{5_A}: PV_{HE} - PV_{LT} \neq 0$. This means the pages viewed (PV) per session for head (HE) keywords are significantly higher those of long tail (LT) keywords.

5.6. Time spent: Hypothesis 6

The second null hypothesis under Objective 4 states that the time spent (TS) by users from head (HE) keywords is equal to those of long tail (LT) keywords. The alternative hypothesis

states that time spent (TS) by users from head (HE) keywords is not equal to those of long tail (LT) keywords.

$$H_{6_0}: TS_{HE} - TS_{LT} = 0$$

$$H_{6_A}: TS_{HE} - TS_{LT} \neq 0$$

The dataset to test this hypothesis was 7,160 keywords, of which 1,095 were head keywords (HE) and 6,065 were long tail (LT) keywords. The data was distributed as show in Figure 11. The mean of the head keywords was 534.8s, with a standard deviation of 459s per visit. For long tail (LT) keywords the mean time was 377.96s, with a standard deviation of 539s. The complete descriptive statistics are presented in Table 11.

Hypothesis 6 needed to compare the means of two independent data sets; the most appropriate statistical technique to do this was a t-Test (Zikmund, 2003, p524; Albright *et al.*, 2006, p496). The results of the t-Test are shown below in Table 12.

Levene's test shows a significance of 0.16. Since it is more than 0.1, equal variance can be assumed. With the assumption of equal variance the t-value = 9.05, with degrees of freedom = 7,158. The mean difference = 156.8s, indicating that the average head (HE) keyword users spends 156.84 seconds longer per session compared to a long tail (LT) keyword.

The p-value = 0.000 indicates there is a significant difference in the means of the samples, and, therefore, the null hypothesis (H_{6_0}) should be rejected in favour of the alternative hypothesis (H_{6_A}). Therefore, the data supports the following: $H_{6_A}: TS_{HE} - TS_{LT} \neq 0$. This means the time spent (TS) per session for head (HE) keywords is significantly higher that of long tail (LT) keywords.

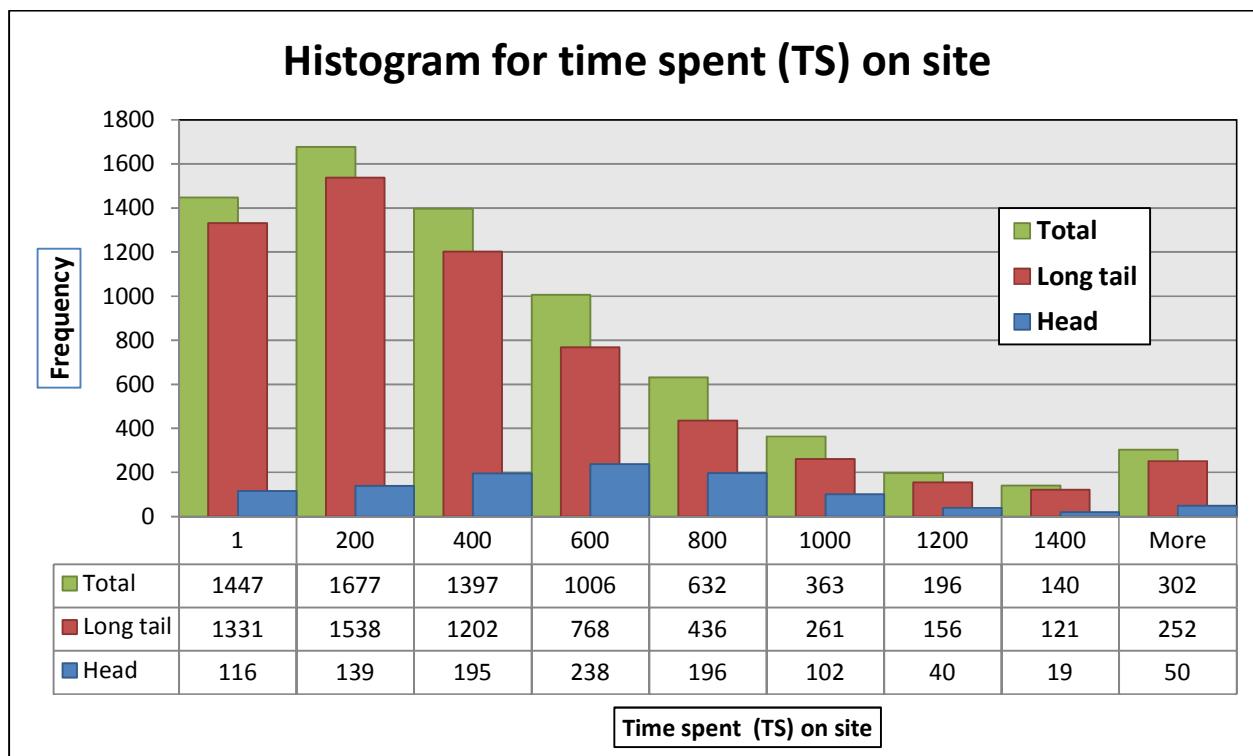


Figure 11: Histogram for time spent (TS) on site

<i>Time spent on site</i>		
	Head	Long tail
Mean	534.80	377.96
Standard Error	13.87	6.93
Median	491.24	221.94
Mode	0	0
Standard Deviation	459	539
Sample Variance	210682	290918
Kurtosis	9.041	42.770
Skewness	2.066	4.481
Range	4545	10691
Minimum	0	0
Maximum	4545	10691
Sum	585609	2292351
Count	1095	6065

Table 11: Descriptive statistic for time spent (TS) on site

Ave Time on Site	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	1.977	.160	9.05	7158	.000	156.84	17.333	122.86	190.82
Equal variances not assumed			10.12	1689	.000	156.84	15.504	126.43	187.25

Table 12: t-Test for equality of means in time spent (TS) on site

5.7. Bounce rate: Hypothesis 7

The third null hypothesis under Objective 4 states that the bounce rate (BR) of users from head (HE) keywords will be equal to those from long tail (LT) keywords. The alternative hypothesis states that the bounce rate (BR) for head (HE) keywords will not be equal that of long tail (LT) keywords.

$$H_{7_0}: BR_{HE} - BR_{LT} = 0$$

$$H_{7_A}: BR_{HE} - BR_{LT} \neq 0$$

Bounce rate is the percentage of users that view only one page of a website, before they leave the website. The dataset to test this hypothesis was 7,160 keywords, of which 1,095 were head (HE) keywords and 6,065 were long tail (LT) keywords. The data was distributed as show in Figure 12. The mean of the head (HE) keywords bounce rate (BR) was 15.68%, with a standard deviation of 20.09% pages per visit. For long tail (LT) keywords the bounce rate (BR) mean was 21.46%, with a standard deviation of 30.6%. The complete descriptive statistics are presented in Table 13 below.

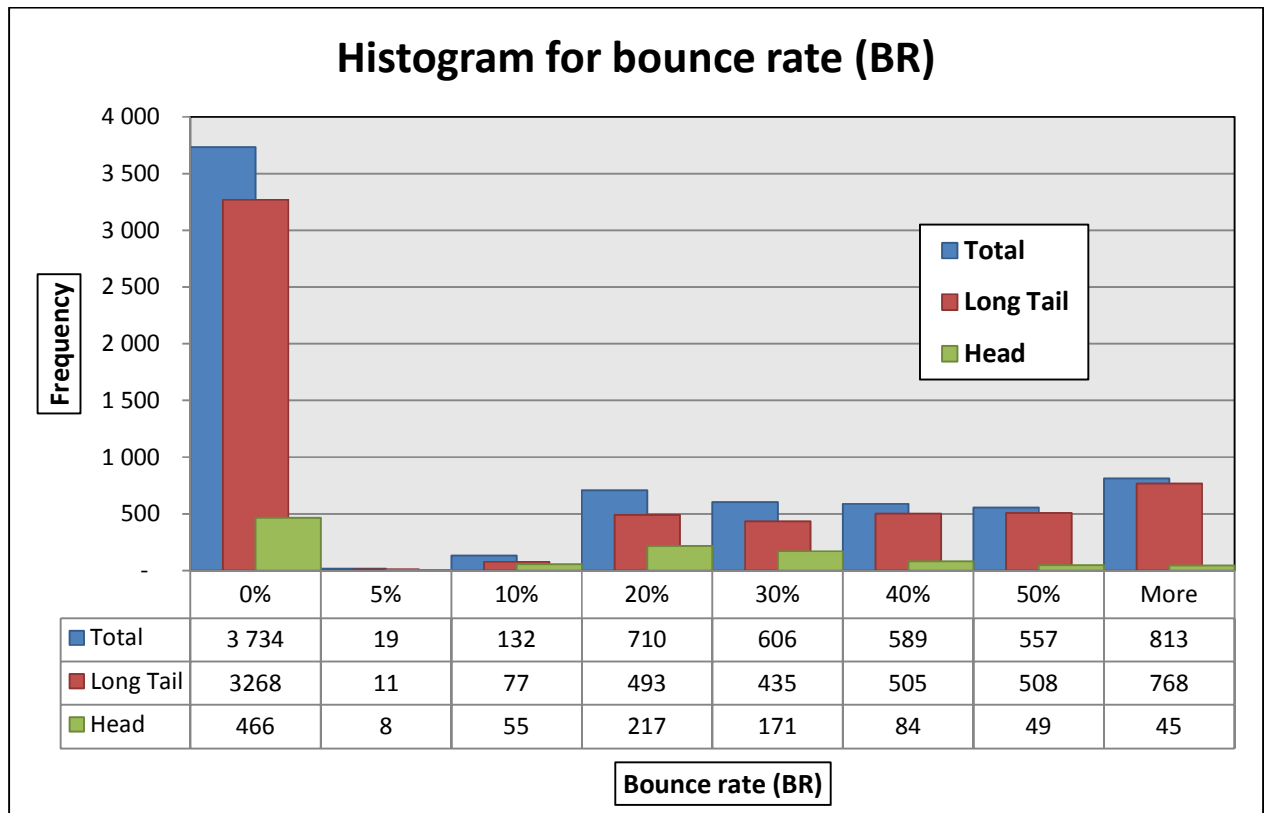


Figure 12: Histogram for bounce rate (BR)

<i>Bounce Rate (BR)</i>		
	Head	Long tail
Mean	15.68%	21.46%
Standard Error	0.61%	0.39%
Median	11.11%	0.00%
Mode	0.00%	0.00%
Standard Deviation	20.09%	30.60%
Sample Variance	4.04%	9.36%
Kurtosis	5.0193	1.0284
Skewness	1.9753	1.4341
Range	100%	100%
Minimum	0%	0%
Maximum	100%	100%
Sum	172	1301
Count	1095	6065

Table 13: Descriptive statistic for bounce rate (BR)

Hypothesis 7 needed to compare the means of two independent data sets; the most appropriate statistical technique to do this was a t-Test (Zikmund, 2003, p524; Albright *et al.*, 2006, p496). The results of the t-Test are shown below in Table 14.

Bounce Rate	Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal Variances Assumed	263.157	.000	-6.02	7158	.000	-.0578	.0096	-.0766	-.0390
Equal Variances Not Assumed			-7.99	2135	.000	-.0578	.0072	-.0720	-.0436

Table 14: t-Test for equality of means in bounce rate (BR)

Levene's test shows a significance of 0.00. Since it is less than 0.1 equal variance cannot be assumed. Without the assumption of equal variance the t-value = -7.99, with degrees of freedom = 2,135. The mean difference = -5.78%, indicating that the average head (HE) keyword user has a bounce rate (BR) 5.78% less than the average long tail (LT) keyword.

The p-value = 0.000 indicates there is a significant difference in the means of the samples, and, therefore, the hypothesis (H_{70}) should be rejected in favour of the alternative hypothesis (H_{7A}). Therefore, the data supports the following: $H_{7A}: BR_{HE} - BR_{LT} \neq 0$. This means the bounce rate (BR) for head (HE) keywords is significantly lower than that of long tail (LT) keywords.

5.8. The long tail

The dataset of 7,160 keywords had been purposefully categorised by the company in head (HE) and long tail (LT) keywords. But with a little investigation the data strongly resembled a power law or long tail (LT) distribution. See Table 15:

% of keywords	% of clicks	Mean CTR	Mean CPC
1%	46.4%	8.0%	R 1.045
5%	64.1%	10.8%	R 0.923
10%	72.6%	11.6%	R 0.850
20%	82.0%	12.5%	R 0.805
50%	94.1%	14.3%	R 0.747
75%	97.9%	16.3%	R 0.746
100%	100.00%	18.9%	R 0.738

Table 15: Data distribution on clicks per keywords

Table shows clearly that the data is a 'classical' long tail (LT) distribution. Amazingly, only 1% of the keywords got close to 47% of the total number of clicks. The data also seems to confirm the Pareto principle, or 80/20 rule, with 82% of the total clicks coming from 20% of the keywords.

The steady increase in the click-through rate (CTR) and steady decrease in the cost-per-click (CPC) in Table 15 shows further evidence of the influence of the long tail (LT). The mean click-through rate (CTR) increases from 8% to more than 18%, while the cost-per-click (CPC) mean falls from R1.04 to below R0.74. The distribution is further illustrated in Figure 13.

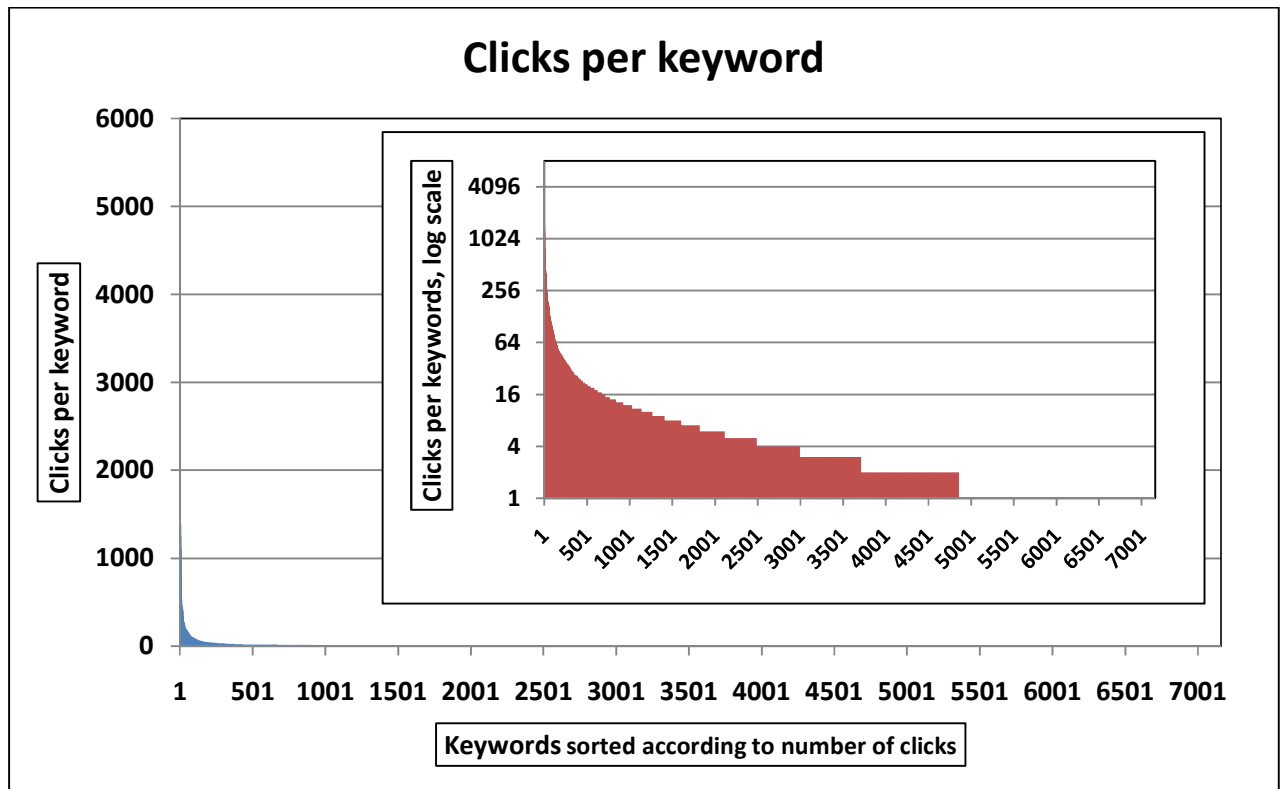


Figure 13: Clicks per keywords, sorted according to number of clicks. On normal and log scale on vertical axis.

With a normal scale on the vertical axis it is difficult to clearly picture the graph. With a log scale the long tail (LT) is very evident, showing that there are more than 2,000 keywords that only got a single click, and that there are less than 1,000 keywords that generated more than 16 clicks.

5.9. Total clicks and cost per keyword

The dataset also included further information regarding the total clicks achieved and the total cost per keyword. The total dataset for this information was 7,160 keywords, with 1,095 head (HE) and 6,065 long tail (LT) keywords. The descriptive statistics for this data are shown in Table 16.

	<i>Total Clicks</i>		<i>Cost per Keyword</i>	
	Head	Long tail	Head	Long tail
Mean	49.05	5.19	R 50.74	R 3.45
Standard Error	7.81	0.11	R 8.04	R 0.07
Median	6	2	R 6.10	R 1.66
Mode	1	1	R 0.84	R 0.35
Standard Deviation	258	9	R 266.07	R 5.65
Sample Variance	66719	74	70794	32
Kurtosis	193	72	253.86	56.72
Skewness	12.77	6.34	14.36	5.80
Range	4799	184	R 5 583	R 108
Minimum	1	1	R 0.20	R 0.07
Maximum	4800	185	R 5 583	R 108
Sum	53706	31457	R 55 563	R 20 916
Count	1095	6065	1095	6065

Table 16: Total clicks and total cost per keyword

The mean for total cost per keyword for head (HE) keywords = R50.74 and R3.35 for long tail (LT) keywords. The total money spent on head (HE) keywords = R 55 564, while only R20 916 was invested in long tail (LT) keywords. For this investment, head (HE) keywords generated 53,706 clicks, while long tail (LT) keywords resulted in 31,457 clicks.

This highlights the high return from long tail (LT) keywords. The total clicks seem to confirm that head (HE) keywords generate a lot more clicks, with a mean of 49.05 for head (HE) and 5.19 for long tail (LT). To confirm if this difference is significant a t-test was performed. The results are displayed in Table 17.

Levene's test shows a significance of 0.00. Since it is less than 0.1, equal variance cannot be assumed. Without the assumption of equal variance the t-value = 5.62, with degrees of freedom = 1,094. The mean difference = 43.86, indicating that the average head (HE) keyword generates 43.86 more clicks than the average long tail (LT) keyword.

Clicks	Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal Variances Assumed	414.7	.000	13.19	7158	.000	43.86	3.33	37.34	50.37
Equal Variances Not Assumed			5.62	1094	.000	43.86	7.81	28.542	59.177

Table 17: t-Test for equality in means of total clicks

5.10. Cost-per-click (CPC) for targeted and generic ad text

In comparing the performance of the targeted advert text to that of generic advert text, the cost-per-click (CPC) also tells a very interesting story. The dataset was 7,160 keywords, of which 200 were head (HE) keywords and 321 long tail (LT) keywords with generic text ads, and 895 head (HE) keywords and 5,744 long tail (LT) keywords with targeted text ads.

The mean cost-per-clicks (CPC) of the head (HE) keywords with generic advert text was R1.67, with a standard deviation of R0.38. For long tail (LT) keywords with generic text ads the mean of cost-per-click (CPC) was R1.32, with a standard deviation of R0.57. Head (HE) keywords with targeted text ads had a mean of R0.9, with a standard deviation of R0.39, while long tail (LT) keywords with targeted text ads had a mean of R0.65, and a standard deviation of R0.21. The complete descriptive statistics are shown below in Table 18.

CPC Targeted Generic				
	Head Targeted	Head Generic	Long Tail Targeted	Long Tail Generic
Mean	R 0.90	R 1.67	R 0.65	R 1.32
Standard Error	R 0.01	R 0.03	R 0.00	R 0.03
Median	R 0.82	R 1.73	R 0.64	R 1.36
Mode	R 0.35	R 2.10	R 0.54	R 2.10
Standard Deviation	R 0.39	R 0.38	R 0.21	R 0.57
Sample Variance	0.154	0.141	0.046	0.322
Kurtosis	0.381	-0.702	-0.272	-0.980
Skewness	0.842	-0.208	0.284	0.004
Range	R 2.31	R 1.75	R 1.13	R 2.21
Minimum	R 0.15	R 0.72	R 0.07	R 0.29
Maximum	R 2.47	R 2.47	R 1.20	R 2.50
Sum	808	335	3718	424
Count	895	200	5744	321

Table 18: Cost-per-click (CPC) for targeted and generic advert text

This data was combined to form four data sets, each with a mean. The method needed to compare the means of more than two data sets was an analysis of variance (ANOVA) (Zikmund, 2003, p529; Albright *et al.*, 2006, p537). The result of the ANOVA is shown in Table 19.

CPC	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	355.941	3	118.647	1599	.000
Within Groups	530.831	7156	.074		
Total	886.772	7159			

Table 19: ANOVA for cost-per-click (CPC) for targeted and generic advert text

The $p = 0.000$ allows us to reject equal means hypothesis, indicating that there is a significant difference in variance in the means between groups. In order to explore these differences, a post hoc test was performed. Sheffe (post hoc test) performs simultaneous

joint pairwise comparisons for all possible pairwise combinations; it is used to examine all possible linear combinations of group means (SPSS, 2006). The results from the post hoc test are shown below in Table 20:

(I) HvsLTG	(J) HvsLTG	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Head Targeted	Head Generic	R -0.77	R 0.021	.000	-0.831	-0.712
	Long tail Targeted	R 0.26	R 0.010	.000	0.228	0.283
	Long tail Generic	R -0.42	R 0.018	.000	-0.468	-0.369
Head Generic	Head Targeted	R 0.77	R 0.021	.000	0.712	0.831
	Long tail Targeted	R 1.03	R 0.020	.000	0.972	1.081
	Long tail Generic	R 0.35	R 0.025	.000	0.284	0.421
Long tail Targeted	Head Targeted	R -0.26	R 0.010	.000	-0.283	-0.228
	Head Generic	R -1.03	R 0.020	.000	-1.081	-0.972
	Long tail Generic	R -0.67	R 0.016	.000	-0.718	-0.630
Long tail Generic	Head Targeted	R 0.42	R 0.018	.000	0.369	0.468
	Head Generic	R -0.35	R 0.025	.000	-0.421	-0.284
	Long tail Targeted	R 0.67	R 0.016	.000	0.630	0.718

Table 20: ANOVA post hoc test, Sheffe for equality of means in cost-per-click (CPC) for targeted and generic, head and long tail samples

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^b
Corrected Model	355.941 ^a	3	118.647	1599.5	.000	.401	4798.361	1.000
Intercept	2196.437	1	2196.437	29609.6	.000	.805	29609.641	1.000
Add_Type	221.995	1	221.995	2992.7	.000	.295	2992.659	1.000
HeadLongtail	39.342	1	39.342	530.4	.000	.069	530.354	1.000
Add_Type * HeadLongtail	1.008	1	1.008	13.6	.000	.002	13.584	.958
Error	530.831	7156	.074					
Total	4788.293	7160						
Corrected Total	886.772	7159						

Table 21: Test of between subject effects

The p-value's for the all the combination of the groups are 0.000, indicating that there are significant differences between all the subsets of the data. The p-value = 0.000 in Table 21

indicated that there is significant interaction effect. An interaction effect occurs when the effect of an independent variable (keyword selection) is different for different levels or categories of another independent variable (advert type) (Malhotra, 2007, p520; Hair, Anderson, Tatham and Black, 1998, p408). In other words the sum of the effect of targeted advert text and a long tail keyword selection is greater than the individual effects.

5.11. Sales and rental keywords

The dataset also provided the opportunity to examine the relationship between rental and sales keywords. Of the total 7,160 keywords, there were 517 head (HE) sales keywords, 578 rental head (HE) keywords, 4,390 long tail (LT) sales keywords and 1,675 long tail (LT) rental keywords. The data was distributed as shown in Figure 14 and Figure 15.

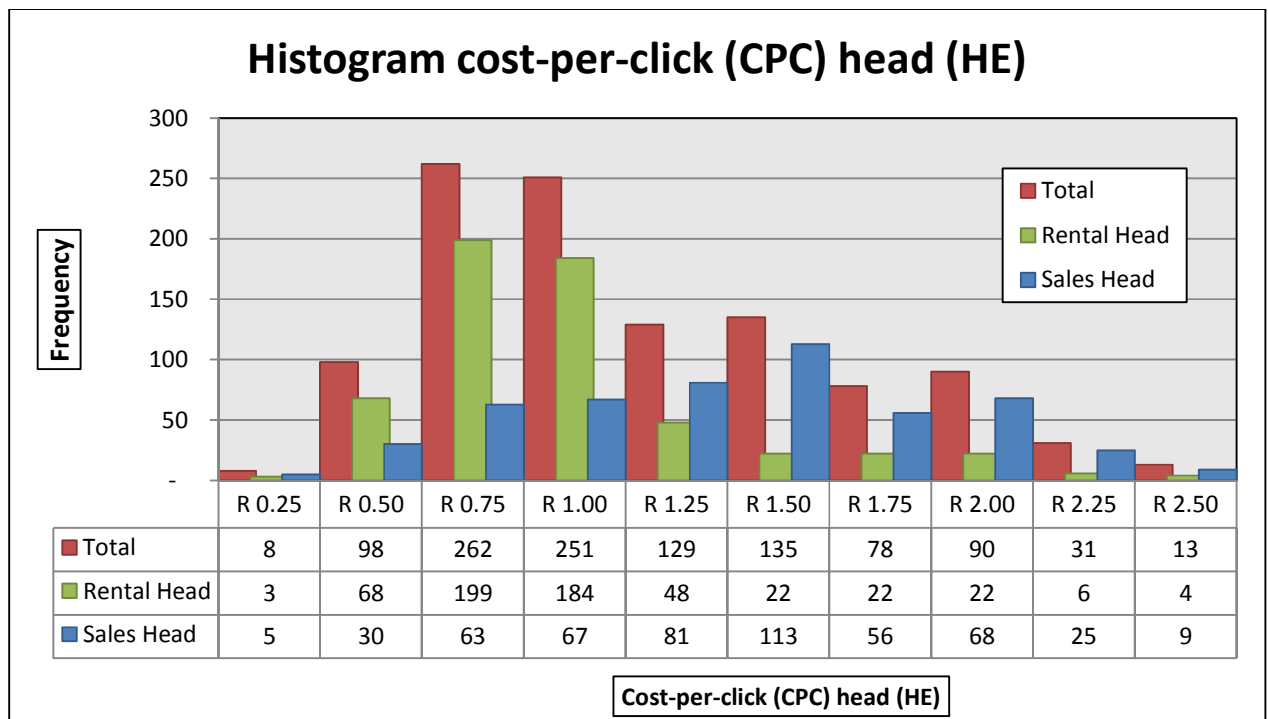


Figure 14: Histogram for cost-per-click (CPC) for head (HE) sales and rental keywords

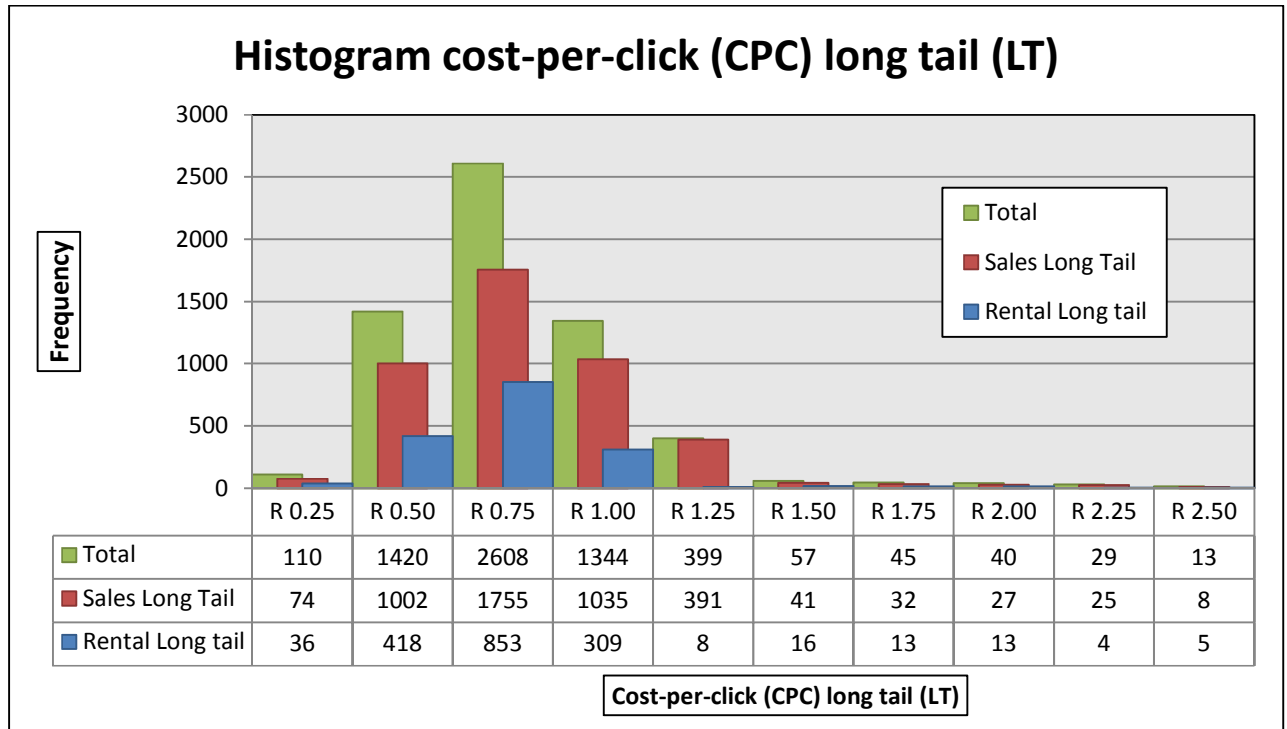


Figure 15: Histogram for cost-per-click (CPC) for long tail sales (LT) and rental keywords

The mean of the sales head (HE) keywords' cost-per-click (CPC) was R1.25, with a standard deviation of R0.50. Rental head (HE) keywords had a mean cost-per-click (CPC) of R0.86, with a standard deviation of R0.40. Long tail (LT) sale keywords had a mean of R0.70, with a standard of R0.29, while long tail (LT) rental keywords had a mean of R0.64, with a standard deviation of R0.27. The complete descriptive statistics are shown in Table 22.

The method needed to compare the means of more than two data sets was an analysis of variance (ANOVA) (Zikmund, 2003, p529; Albright *et al.*, 2006, p537). The result of the ANOVA is shown in Table 23.

	Cost-per-click (CPC)			
	Sales Head	Rental Head	Sales Long tail	Rental Long tail
Mean	R 1.25	R 0.86	R 0.70	R 0.64
Standard Error	R 0.02	R 0.02	R 0.00	R 0.01
Median	R 1.28	R 0.77	R 0.66	R 0.62
Mode	R 0.35	R 0.83	R 0.35	R 0.54
Standard Deviation	R 0.50	R 0.40	R 0.29	R 0.27
Sample Variance	0.2499	0.1567	0.0862	0.0719
Kurtosis	-0.6981	2.2770	5.6138	11.2987
Skewness	0.0212	1.4885	1.6711	2.4751
Range	R 2.31	R 2.25	R 2.43	R 2.42
Minimum	R 0.15	R 0.22	R 0.07	R 0.07
Maximum	R 2.47	R 2.47	R 2.50	R 2.49
Sum	648	494	3072	1071
Count	517	578	4390	1675

Table 22: Descriptive statistics cost-per-click (CPC) for sales and rental, head (HE) and long tail (LT) keywords

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	168.522	3	56.174	559.67	.000
Within Groups	718.250	7156	.100		
Total	886.772	7159			

Table 23: ANOVA for equality of means in cost-per-click (CPC) for rental and sales, head (HE) and long tail (LT) samples

The $p = 0.000$ allows us to reject equal means hypothesis, indicating that there is a significant difference in variance in the means between groups. In order to explore these differences, a post hoc test was performed. Sheffe (post hoc test) performs simultaneous joint pairwise comparisons for all possible pairwise combinations; it is used to examine all possible linear combinations of group means (SPSS, 2006). The results from the post hoc test are show below in Table 24. The p-values are 0.000 for all the combination of groups, indicating the differences in means are significant.

(I) Category	(J) Category	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Sales Head	Rental Head	R 0.40	R 0.02	.000	R 0.35	R 0.45
	Sales Long Tail	R 0.55	R 0.01	.000	R 0.51	R 0.60
	Rental Long Tail	R 0.62	R 0.02	.000	R 0.57	R 0.66
Rental Head	Sales Head	R -0.40	R 0.02	.000	R -0.45	R -0.35
	Sales Long Tail	R 0.16	R 0.01	.000	R 0.12	R 0.19
	Rental Long Tail	R 0.22	R 0.02	.000	R 0.17	R 0.26
Sales Long Tail	Sales Head	R -0.55	R 0.01	.000	R -0.60	R -0.51
	Rental Head	R -0.16	R 0.01	.000	R -0.19	R -0.12
	Rental Long Tail	R 0.06	R 0.01	.000	R 0.04	R 0.09
Rental Long Tail	Sales Head	R -0.62	R 0.02	.000	R -0.66	R -0.57
	Rental Head	R -0.22	R 0.02	.000	R -0.26	R -0.17
	Sales Long Tail	R -0.06	R 0.01	.000	R -0.09	R -0.04

Table 24: ANOVA post hoc test, Sheffe for equality of means in cost-per-click (CPC) for rental and sales, head (HE) and long tail (LT) samples

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	168.522 ^a	3	56.174	559.666	.000	.190
Intercept	2649.081	1	2649.081	26393.027	.000	.787
SalesRental	47.017	1	47.017	468.431	.000	.061
HeadLongTail	132.329	1	132.329	1318.409	.000	.156
SalesRental * HeadLongTail	25.462	1	25.462	253.678	.000	.034
Error	718.251	7156	.100			
Total	4788.295	7160				
Corrected Total	886.773	7159				

Table 25: Test of between subject effects

The p-value = 0.039 in Table 25 indicated that there is significant interaction effect. An interaction effect occurs when the effect of an independent variable (keyword selection) is different for different levels or categories of another independent variable (rental/sales) (Malhotra, 2007, p520; Hair, Anderson, Tatham and Black, 1998, p408). In other words the sum of the effect of rental or sales keywords and a long tail keyword is greater than the individual effects.

The dataset also provides insight into the click-through rates (CTRs) of sales and rental keywords. The mean for sales head (HE) keywords' click through rate (CTR) was 7.1%, with a standard deviation of 10.6%. Rental head (HE) keywords had a mean click-through rate (CTR) of 13.4%, with a standard deviation of 17.2%. Long tail (LT) sales keywords had a mean of 17.9%, with a standard deviation of 20.3%, while long tail (LT) rental keywords had a mean of 27%, with a standard deviation of 23.7%. The descriptive statistics are shown in Table 26.

	Click-through rate (CTR)			
	Sales Head	Rental Head	Sales Long tail	Rental Long tail
Mean	7.1%	13.4%	17.9%	27.0%
Standard Error	0.5%	0.7%	0.3%	0.6%
Median	4.3%	8.4%	11.1%	20.0%
Mode	33.3%	33.3%	50.0%	50.0%
Standard Deviation	10.6%	17.2%	20.3%	23.7%
Sample Variance	0.0112	0.0297	0.0414	0.0562
Kurtosis	36.3894	14.0389	8.7367	6.1496
Skewness	5.1338	3.5530	2.7610	2.2082
Range	99.8%	99.9%	199.9%	199.9%
Minimum	0.2%	0.1%	0.1%	0.1%
Maximum	100.0%	100.0%	200.0%	200.0%
Sum	37	77	785	453
Count	517	578	4390	1675

Table 26: Descriptive statistics' click-through rate for sales and rental, head (HE) and long tail (LT) keywords

The method needed to compare the means of more than two data sets was an analysis of variance (ANOVA) (Zikmund, 2003, p529; Albright *et al.*, 2006, p537). The result of the ANOVA is shown in Table 27.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	20.474	3	6.825	163.469	.000
Within Groups	298.755	7156	.042		
Total	319.229	7159			

Table 27: ANOVA for equality of means in cost-per-click (CPC) for rental and sales, head (HE) and long tail (LT) samples

The $p = 0.000$ allows us to reject equal means hypothesis, indicating that there is a significant difference in variance in the means between groups. In order to explore these differences, a post hoc test was performed. Sheffe (post hoc test) performs simultaneous joint pairwise comparisons for all possible pairwise combinations; it is used to examine all possible linear combinations of group means (SPSS, 2006). The results from the post hoc test are show below in Table 28.

(I) Category	(J) Category	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Sales Head	Rental Head	-6.33%	1.24%	.000	-9.78%	-2.87%
	Sales Long tail	-10.80%	0.95%	.000	-13.45%	-8.14%
	Rental Long tail	-19.94%	1.03%	.000	-22.82%	-17.07%
Rental Head	Sales Head	6.33%	1.24%	.000	2.87%	9.78%
	Sales Long tail	-4.47%	0.90%	.000	-7.00%	-1.94%
	Rental Long tail	-13.62%	0.99%	.000	-16.37%	-10.86%
Sales Long tail	Sales Head	10.80%	0.95%	.000	8.14%	13.45%
	Rental Head	4.47%	0.90%	.000	1.94%	7.00%
	Rental Long tail	-9.15%	0.59%	.000	-10.79%	-7.50%
Rental Long tail	Sales Head	19.94%	1.03%	.000	17.07%	22.82%
	Rental Head	13.62%	0.99%	.000	10.86%	16.37%
	Sales Long tail	9.15%	0.59%	.000	7.50%	10.79%

Table 28: ANOVA post hoc test, Sheffe for equality of means in click-through rate (CTR) for rental and sales, head (HE) and long tail (LT) samples

The p-values are 0.000 for all the combination of groups, indicating the differences in means are significant.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	20.474 ^a	3	6.825	163.469	.000	.064
Intercept	95.252	1	95.252	2281.537	.000	.242
SalesRental	5.331	1	5.331	127.699	.000	.018
HeadLongTail	13.274	1	13.274	317.952	.000	.043
SalesRental * HeadLongTail	.177	1	.177	4.242	.039	.001
Error	298.755	7156	.042			
Total	574.369	7160				
Corrected Total	319.229	7159				

Table 29: Test of between subject effects

The p-value = 0.039 in Table 29 indicated that there is significant interaction effect. An interaction effect occurs when the effect of an independent variable (keyword selection) is different for different levels or categories of another independent variable (rental/sales) (Malhotra, 2007, p520; Hair, Anderson, Tatham and Black, 1998, p408). In other words the sum of the effect of rental or sales keywords and a long tail keyword is greater than the individual effects.

5.12. Conclusion

The results of the hypotheses testing are listed in Table 30. All seven of the hypotheses were rejected in favour of the alternative hypotheses. These results will help meet the objectives of the research and will be discussed in Chapter 6.

The data showed strong evidence of being a classical long tail (LT) distribution, with 1% of the keywords providing 46.4% of the clicks. The data also provided valuable insight in the

cost-per-click (CPC) for targeted and generic advert text, with targeted advert text showing significantly lower cost-per-click (CPC) for head (HE), with a mean difference of R0.77 compared to generic advert text. Long tail (LT) targeted advert text was also significantly higher than generic, with a mean difference of R0.67. Nevertheless, head (HE) keywords did generate more page views, head (HE) users spent more time, and had a lower bounce rate.

Hypothesis	Result
H1 ₀	Reject H1 ₀
H2 ₀	Reject H2 ₀
H3 ₀	Reject H3 ₀
H4 ₀	Reject H4 ₀
H5 ₀	Reject H5 ₀
H6 ₀	Reject H6 ₀
H7 ₀	Reject H7 ₀

Table 30: Summary of hypotheses testing

Finally, it was interesting to explore the dynamics between head (HE)/long tail (LT) and sales/rental keywords. The data indicated that there is a significant difference in the means of sales head (HE), sales long tail (LT), rental head (HE) and rental long tail (LT) keywords.

6. DISCUSSION OF RESULTS

6.1. Introduction

The fundamental question this research aims to answer is: *“Can a long tail distribution in keyword selection help improve the effectiveness of a sponsored search online advertising campaign?”* In order to answer this question chapter 6 is structured according to the research objectives:

- Objective 1: to determine if a long tail (LT) approach to keyword selection can improve the return on investment (ROI) of the campaign by bringing the overall average cost-per-click (CPC) down significantly.
- Objective 2: to explore if the long tail (LT) keywords will yield significantly higher click-through rates (CTRs) than head (HE) keywords.
- Objective 3: to see if targeted ad text, related to the keyword used, will significantly increase the click-through rate (CTR) for either head (HE) or long tail (LT) keywords.
- Objective 4: to compare the behaviour of users that click on head (HE) and long tail (LT) keywords by comparing commonly used measures like pages viewed (PV), bounce rate (BR) and time spent (TS).

Malhotra (2007, p89) urged marketing managers to use causal research to examine the presumed casual relationship between variables. The data suggests that there is a causal relationship between a long tail (LT) keyword selection and return on investment (ROI). This gives the marketing manager a tool to improve the sponsored search advertising, which is based on empirical evidence.

6.2. Cost-per-click: Objective 1

The first objective of this research was to determine if a long tail (LT) approach to keyword selection can improve the return on investment (ROI) of the campaign by bringing the overall average cost-per-click (CPC) down significantly. The selection of keywords is one of the four basic decisions an advertiser needs to make when setting up a sponsored search campaign (Rutz *et al.*, 2007). However, very little of the recent literature has focussed on keywords and none on strategies for keyword selection (Kumar, 2008; Rutz *et al.*, 2007; Rutz *et al.*, 2008; Ghose *et al.*, 2007).

The research investigated a sponsored search campaign to apply Anderson's (2004) long tail (LT) concept. Long tail (LT) is just a more popular name for what statisticians call the power law distribution, where the 'head' (HE) is very high but short, and the 'tail' (LT) low but long (McDonald, 2008). The data set provided a virtual textbook example, to such an extent that a graph to show the long tail (LT) needs to be placed on a log scale to identify the trend, illustrated again in Figure 16.

Such results provide evidence that the data supports Anderson's (2004) premise that the improved search technology and the almost limitless inventory (potential keywords in this case) will increase the size and contribution of the long tail (in terms of clicks), while making the relative contribution of the head (HE) smaller and smaller. This is in contrast to Elberse's (2008) view, which suggested that the long tail would be very difficult to profit from.

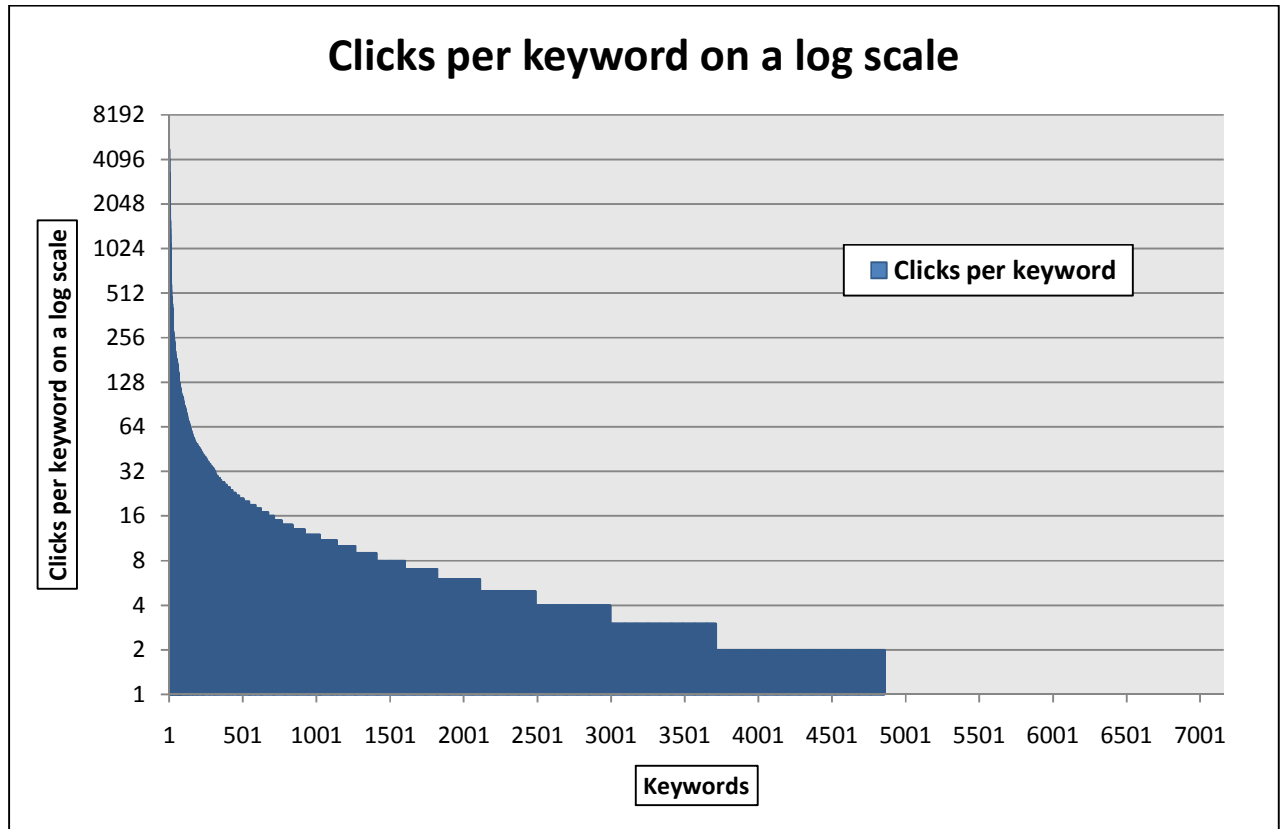


Figure 16: Clicks per keyword, sorted according to number of clicks on a log scale

If traffic metric are viewed as more important, as Ilfeld *et al.* (2002) suggested than online firms should focus their online advertising on attracting users to their website. Since the company under study is an online portal, return on investment (ROI) was defined as the maximum number of clicks (traffic to its website) that can be generated for every rand spent. Return on investment (ROI) is, therefore, driven by the cost-per-click (CPC) on an inverse relationship; the lower the cost-per-click (CPC) the higher the return on investment (ROI).

In order to see if the long tail (LT) could achieve a lower cost-per-click (CPC) than head (HE) keywords, null hypothesis was: $H_{10}: CPC_{HE} - CPC_{LT} = 0$. The null hypothesis, however, was rejected in the t-Test in favour of the alternative Hypothesis: $H_{1A}: CPC_{HE} - CPC_{LT} > 0$. In

fact, with a mean difference of R0.36 and a mean of R0.68 for long tail (LT) keywords, the implications is that head (HE) keyword are, on average, more than 50% more expensive than long tail (LT) keywords.

The implication for return on investment (ROI) is significant, for every rand spent on a head (HE) keyword the company got on average 0.95 clicks, while for every rand spend on a long tail (LT) keyword the company got on average 1.46 clicks. For a sponsored search campaign currently spending R10 000 per month, that could mean increasing the traffic resulting from the campaign from 9,578 to 14,641 if only long tail (LT) keywords are included, which will be a significant improvement. Thus long tail keywords are both cheaper and more effective.

The data provides additional evidence of the improved return on investment (ROI) of long tail (LT) keywords by considering the total cost and clicks achieved by long tail (LT) and head (HE) keywords. Although head (HE) keywords, as suggested by Anderson's (2004) long tail (LT) theory, achieve significant higher clicks on average, 49 compared to long tail (LT) keywords' five, the total clicks of long tail (LT) keywords appear to be significant, with 31,457 compared to the 53,706 generated by head (HE) keywords. The total cost really brings the improved return on investment (ROI) to light; with R55 563 invested in head (HE) keywords this is a 50% worse return on investment (ROI) compared to the R20 916 investment in long tail (LT) keywords. Thus the long tail is an exceedingly good lens to improve return on marketing investment.

Edelman *et al.* (2007) showed that the 'generalized second-price' (GSP) auction used by search engines aims to maximise revenue for search engines. However, the auction pricing assigns an individual price to each of the millions of keywords (Edelman *et al.*, 2006). The

data suggests this individual price setting also allows firms to employ strategies like the long tail (LT) to improve their return on investment (ROI) by looking for keywords where the equilibrium price (cost-per-click) is lower.

A lower equilibrium price for long tail (LT) keywords is one possible reason for the lower cost-per-click (CPC) achieved. However, since, search engines use a quality score combined with the maximum bid to rank (Edelman *et al.*, 2007; Jansen *et al.*, 2008), and since quality score is highly influenced by the historical click-through rate (CTR) (Edelman *et al.*, 2007; Jansen *et al.*, 2007; Jansen *et al.*, 2008; Li *et al.*, 2009; Pedersen, 2008), the lower cost-per-click (CPC) could be as a result of the higher click-through rate (CTR). Objective 2 seeks to explore this argument.

When we turn to the other aspect of sponsored search, this too should add to the campaign's effectiveness. Szymanski *et al.* (2006) suggested that sponsored search provides a very high return on investment (ROI) compared to other marketing methods; the data suggests that the long tail (LT) approach to keyword selection could improve this relative advantage to sponsored search even further. The interaction effect between the independent variables keyword selection and advert text was significant and will be explored in objective 3.

6.3. Click-through rate: Objective 2

In order to try and help explore some of the possible reasons why there is a significant difference in the cost-per-click (CPC) of head (HE) and long tail (LT) keywords, Objective 2 aimed to investigate if long tail (LT) keywords would yield significantly higher click-through

rates (CTRs) than head (HE) keywords. A solution to low current click-through rates is critical as display advertising click-through rates (CTRs) have been steadily declining, falling from around 7% in the 1990s to levels as low as 0.7% in 2002, 0.2% in 2006 and 0.1% in 2008 (Hoffmann and Novak, 2000c; Chatterjee, Hoffman and Novak, 2003; Rutz *et al.*, 2007; Anderson, 2009; Fulgoni and Morn, 2009). The data suggests that the click-through rates (CTRs) for both head (HE) and long tail (LT) keywords are significantly higher than those achieved by banner advertising. This possibly provides further evidence of Szymanski *et al.*'s (2006) suggestion that sponsored search provides a very high return on investment (ROI) compared to other marketing methods.

This is critically important since rank is determined by a combination of the advertiser's maximum bid and the quality score of the keywords (Edelman *et al.*, 2007; Jansen *et al.*, 2008). And since historical click-through rate (CTR) is one of the key elements in determining the quality score, a significantly higher click-through rate (CTR) could, in part, explain the lower cost-per-click (CPC) achieved by long tail (LT) keywords (Edelman *et al.*, 2007; Jansen *et al.*, 2007; Jansen *et al.*, 2008; Li *et al.*, 2009; Pedersen, 2008). A prominent position or placement (for example, the slots listed at the top or highlighted in special colour determined by the rank) is commonly believed to be desirable, exactly because of higher click-through rate (CTR) (Ansari *et al.*, 2003; Ghose *et al.*, 2007).

With the aim of meeting the second objective - whether long tail (LT) keywords achieved a higher click-through rate (CTR) – the null hypothesis was: $H_{20}: CTR_{HE} - CTR_{LT} = 0$. The null hypothesis was rejected in the t-Test in favour of the alternative hypothesis: $H_{2A}: CTR_{HE} - CTR_{LT} < 0$. The mean difference of 9.98% indicated that the click-through rate (CTR) of long tail (LT) keywords was almost double that of head (HE) keywords.

The mean of 20.4% achieved by long tail (LT) keywords is particularly impressive, translating into the average long tail (LT) keyword generating a click for every five times it is displayed. Even head (HE) keywords had a mean click-through rate (CTR) of 10.42%, which is a click for every 10 impressions. Although sponsored search results have been shown to be just as relevant to a user search query as organic results, research has previously suggested the browsers have a negative bias towards sponsored search results (Jansen *et al.*, 2006; Jansen *et al.*, 2007). The click-through rates achieved in this research suggest that users are starting to relax the negative bias towards sponsored search results.

Jansen *et al.* (2007) argued that keyword selection is a dynamic form of metatagging, by making firms associate search terms with specific pages on the firm's website. Since the landing page also influence the quality score (Edelman *et al.*, 2007; Jansen *et al.*, 2007; Jansen *et al.*, 2008; Li and Jhang-Li, 2009; Pedersen, 2008). The long tail (LT) provides the advertiser the opportunity to make the landing or destination page the user reaches relevant to the user's search term, thereby possibly increasing the quality score, rank and click-through rate (CTR), thus reducing the cost-per-click (CPC).

More and more people are using long tail (LT) type search terms, looking for information related to their search instead of looking for a specific homepage of a website (Jones, 2008). If this trend continues, long tail (LT) terms would yield more and more clicks in the coming years. This also takes Jansen *et al.*'s (2007) idea of keyword selection being a form of metatagging further by allowing the advertisers to be able to take users from the search engine to the specific page that is relevant to them, without going through the home page. This is a very important point; it is similar to being able to enter any office in a building with

tens of thousands of offices, directly without going through the front door, using elevators or other tools that takes time and effort to get to the destination. It can, therefore, make the user's experience more enjoyable and help him/her find what he/she is looking for much easier. This direct entry might have other consequences; for example the branding effect might be lessened as a result of less home page views.

This significant improvement in click-through rate (CTR) achieved by long tail (LT) keywords can, at least partially, be explained by the ability of long tail (LT) keywords to be closer to the user's search point on the continuum suggested by Bhatnagar *et al.* (2001). Since the long tail (LT) keywords were themed around geographical locations, mostly suburb locations combined with property type, the high click through rates (CTRs) could indicate that users found the advertising more relevant. This approach should work for most services that are geographical in nature, but also points to using the user's context, as search users might search for context before principle.

It stands to reason that these keywords were more relevant to the users' intention, behaviour and context of looking for a specific property type in a specific area. This supports Ansari *et al.*'s (2003) suggestion that the Internet has the ability to be highly addressable, allowing customisation and delivering the right content to the right person at the right time.

It also provides evidence for Wilber *et al.*'s (2009) argument that sponsored search allows for individual level targeting (one-on-one) as the ultimate means of targeting behaviour. It could be argued that targeting behaviour in this research achieved the same as segmenting according to demographics, since people searching for specific property types in specific areas would arguably share very similar demographic traits. But targeting the behaviour

seems a much better proactive approach to reach the audience compared to a more after-the-fact approach of segmenting. The data provides some evidence that one-on-one marketing allows firms to communicate with a much broader potential customer base on an individual basis, it will be interesting to see if this, as predicted by Ferguson *et al.* (2006), will lead to the end of mass marketing.

The improved targeting achieved by employing long tail (LT) keywords has reduced cost-per-click (CPC) and improved the click-through rate (CTR), and by a combination of these factors increased the return on investment (ROI), providing support to Chandra's (2009) claim that improved targeting will improve advertising return.

There is, however, another element to the targeting, which is the use of targeted advert text, which is subject of Objective 3.

6.4. Targeted advert text: Objective 3

The third objective of this research was to determine if targeted advert text, related to the keyword used, will significantly increase the click-through rate (CTR) for either head (HE) or long tail (LT) keywords. The design of the advert text is one of the four basic decisions an advertiser needs to make when setting up a sponsored search campaign (Rutz *et al.* 2007). The targeting of this advert text seems logical in light of the overwhelming evidence from the literature that shows improved results from targeting (Rossi *et al.*, 1996; Allenby *et al.*, 1999; Iyer *et al.*, 2005; Chandra, 2009).

The company under study had specifically set up the ad groups to allow for targeting advert text around the keywords in the ad groups. This included reference to buying or renting, suburbs and property type. This is evidence of the two factors motioned by Iyer *et al.* (2005) that are leading to increased targeting of advertising;

- Better information about customers: which keyword they are searching for and then matching the content of the advert to speak to it; and
- Fragmentation of media: or in this example, the further fragmentation on one medium - search engine - by employing the long tail (LT), or purposefully looking for more fragmented niches (long tail (LT) keywords).

The aim of the third objective of this research was to examine the effect of targeted advert text on both long tail (LT) and head (HE) keywords. The third null hypothesis looked specifically at head (HE) keywords, therefore: $H_{3_0}: CTR_{H_{WOTA}} - CTR_{H_{WTA}} = 0$; while the fourth null hypothesis looked at long tail (LT) keywords: $H_{4_0}: CTR_{L_{WOTA}} - CTR_{L_{WTA}} = 0$. Running an analysis of variance (ANOVA) confirmed that the means were not equal, and the post hoc test confirmed the rejection of both hypotheses.

The mean difference in click-through rate (CTR) for head (HE) keywords was a remarkable 7.9%, with the mean for generic advert text only 3.95% and targeted advert text 11.87%. The improvement for head (HE) keywords was a staggering 200%. The mean difference in click-through rate (CTR) for long tail (LT) was also impressive at 13.46%, with a mean of 21.12% for targeted advert text and 7.66% for generic advert text. The improvement for long tail (LT) keywords was an equally astounding 175%. The combined effect, comparing head (HE) generic to long tail (LT) targeted, is a spectacular increase of 430%.

The cost-per-click (CPC) for generic and targeted advert text also tells a very interesting story: an analysis of variance (ANOVA) confirmed the means were not equal and a post hoc test again revealed the relationship between groups.

The mean difference in cost-per-click (CPC) for head (HE) keywords was R0.77, with the mean for generic advert text being R1.67 and targeted advert text R0.90. The decrease was an astonishing 46% in the average cost-per-click (CPC) of head (HE) keywords. The mean difference in cost-per-click (CPC) for long tail (LT) keywords was R0.67, with the mean for generic advert text being R1.32 and targeted advert text R0.65. Incredibly, the decrease was more than 50%. The combined effect, comparing head (HE) generic to long tail (LT) targeted, was a overall reduction 61%.

Chandra (2009) suggested that the ability to target advertising is essential to maximise return on investment (ROI) in advertising. Both the click-through rate (CTR) and cost-per-click (CPC) data about targeted advert text provide very substantial evidence to support this argument. Kim *et al.* (2005) argued that one of the major challenges in targeted advertising is finding the customers most likely to be interested in the product or service. The combination of targeted advert text and long tail (LT) keywords appear to help solve this for sponsored search advertising.

Both the selection of keywords and the design of the text advert play a substantial part in the targeting of sponsored search campaigns, and, as noted by Rutz *et al.* (2007), they are under the complete control of the advertiser. This is a noticeable departure from traditional media where the targeting was mostly done around choosing a specific publication, radio

station or television channel according to the demographics of their audience. The media owner influenced the composition of this audience by the content that they offered, therefore the advertiser had little or no influence over the targeting other than the choice of publication, radio station or television channel.

With sponsored search the channel is almost a given, with only a very limited amount of search engines available, but the means of targeting is completely controlled by the advertiser, not the media owner. The data indicates that increased targeting can improve return on investment (ROI), and since the means of targeting is controlled by the advertiser, this could become a significant competitive advantage for a firm.

Chandra (2009) noted that in traditional media more targeted advertising rewards the media owner by fetching higher prices. This data suggests that, for sponsored search, the effect was the opposite. This might be as a result of the pricing model. In traditional media, the cost-per-impression pricing ensures that the media owner gets paid when the advertising is placed or displayed.

In sponsored search, the media owner only gets paid if the user acts on the advertising, therefore, because the data also suggests a higher click-through-rate (CTR) for targeted advertising, it is better for the search engine to earn less more often. This might ensure more revenue than earning more, less often. It is probably one of the motivating factors behind the search engines introducing quality scores to promote more relevant advertising. However, since the branding effect is lessened as a result of long tail keywords users not entering through the home page, some advertisers might choose to pay a premium to have this

added benefit. Therefore, the keyword selection might be subject to the specific objective of the advertiser.

The data clearly shows the improvement on return on investment (ROI) from both using a long tail (LT) distribution in keyword selection and targeted advert text in design, the interaction effect is also significant. The combination of long tail keywords and targeted advert text has a bigger impact that they would have independently. It is also important to compare the quality of the users that are directed towards the website by sponsored search. Objective 4 seeks to examine this.

6.5. User comparison: Objective 4

Ilfeld *et al.* (2002) noted that for online firms increased Internet traffic would lead to increased brand awareness, the others suggested online firms ignore the branding effect of online display advertising. The branding impact of display advertising was however strongly supported in research by Briggs *et al.* (1997), Dreze *et al.* (2003), Manchanda *et al.* (2006), Yoo (2008) and Fulgoni *et al.* (2009). The data in this research suggests that long tail (LT) distribution in keywords selection increased the return on investment (ROI) for sponsored search advertising. Before the complete endorsement of this finding can be given, the quality, or more accurately, the behaviour of the users should be examined since, in order for increased traffic to benefit the company by increased brand awareness (among other benefits), the resulting users should interact with the website once they have landed on it.

Hoffman *et al.* (2000b, 2000c) proposed several measures as proxies for interactivity of users, including time (TS) spend on site, bounce rate (BR) and number of pages viewed (PV). This

extension is one of the key advantages of the measurability of online advertising (Hoffman *et al.*, 2000b, 2000c). Objective 4 is aimed at comparing the behaviour of users that click on long tail (LT) and head (HE) keywords by comparing the time they spend on the website, the amount of pages they visited as well as the percentage that leave the website without looking at any pages other than the landing page (bounce rate (BR)).

This first measurable examined was the pages viewed (PV). In order to see if there is a difference in the mean of pages viewed between head (HE) and long tail (LT) keywords, The fifth null hypothesis was formulated as: $H5_0: PV_{HE} - PV_{LT} = 0$. This hypothesis was rejected in the t-Test, however, in favour of the alternative hypothesis: $H5_A: PV_{HE} - PV_{LT} \neq 0$. The mean difference was 4.27 pages viewed (PV). Head (HE) keywords had a mean of 10.68 and long tail (LT) keywords had a mean of 6.40.

This seems to indicate that the average head (HE) keyword users will view 66% more pages when compared to the average long tail (LT) user. This significant difference in the pages viewed (PV) could mean that the quality of users generated by long tail (LT) is inferior to those of head (HE) keywords. Before exploring this further, let's consider the other results.

The second measurable examined was time spent (TS) on the site. To examine the relationship between head (HE) and long tail (LT) users' time spent on the website, the sixth null hypothesis was formulated as: $H6_0: TS_{HE} - TS_{LT} = 0$. Hypothesis 6 was rejected in the t-Test, however, in favour of the alternative hypothesis: $H6_A: TS_{HE} - TS_{LT} \neq 0$. The mean difference was 156.84s, indicating that the average head (HE) user spent 156.8s more on the website when compared to the average long tail (LT) user. The mean for head (HE) keywords was 534.8s, and long tail (LT) keywords were 377.96s.

This seems to indicate that the average head (HE) keywords users will spend 41% more time on the website compared to long tail (LT) keyword users. This significant difference in time spent (TS) is another indication that the quality of long tail (LT) users might be inferior to the head (HE) users. There is one last measure to compare before discussing possible explanations.

The third measurable examined was the bounce rate (BR), that is the percentage of users that leave the website without looking at any pages other than the page they landed on. To examine the relationship between head (HE) and long tail (LT) bounce rate (BR), the seventh null hypothesis was formulated as: $H7_0: BR_{HE} - BR_{LT} = 0$. Hypothesis 7 was rejected, however, in favour of the alternative hypothesis: $H7_A: BR_{HE} - BR_{LT} \neq 0$. The mean difference was 5.78%. The mean for head (HE) keywords was 15.68%, and long tail (LT) keywords mean was 21.46%.

This indicates that the bounce rate (BR) for long tail (LT) keywords is on average 38% higher than that of head (HE) keywords. Consequently the average long tail (HE) user is 38% more likely to leave after viewing only a single page. This significant difference in time spent (TS) is another indication that the quality of long tail (LT) users might be inferior to head (HE) users. Comparing these findings to the website average provides further insight, as shown in Table 31.

Although no statistical tests were done to examine the significance of the site data, it appears on the surface that head (HE) keyword users view more pages, stay longer and have a lower bounce rate (BR) than the average user. A long tail (LT) user appears on the

surface to view less pages, spends less time but has a lower bounce rate (BR) compared to the average user.

Behaviour	Site Average July 2009	Head Keywords	Long tail Keywords
Pages Per Visit	9.09	10.68	6.4
Average Time	520	535	378
Bounce Rate	33.98%	15.68%	21.46%

Table 31: Comparing users' behaviours to site average.

Although this provides possible further evidence that long tail (LT) users might be of an inferior quality, there is another possible explanation: Jansen *et al.* (2007) suggested that advertisers become part of the search process by associating keywords to pages buried deep within their website. This makes the landing page more relevant to the user and improves the user's experience (Jansen *et al.*, 2007). However, this could have a significant impact on the user's behaviour, which was used to measure quality of clicks.

The office building explanation is very relevant again. If you could enter any office in an office building directly, would you spend more or less time in the building? Therefore, it is possible that because users go directly to the page with content relevant to their search query, it could result in them viewing fewer pages and spending less time. This could also result in them leaving quicker as they are able to identify immediately that the content is not relevant to them while, if they landed on the homepage, they might view one or two more pages before they come to the same conclusion.

Another alternative explanation is that head (HE) and long tail (LT) users behave differently because they have different objectives. Rogers and Sheldon suggested in Chipp and Ismail (2008, p67) that there are four types of online behaviour, information accumulation, communication, surfing and shopping. Long tail (LT) users might be more goal orientated falling in the shopping category while and head (HE) users might fall in the browsing category. Long tail (LT) would then be more tuned towards sales, whilst head (HE) might be more open towards brand building attempts.

It is also possible that the quality of long tail (LT) users is indeed inferior to head (HE) keyword users. This could influence the application of the long tail (LT) theory as a strategic tool for keyword selection in order to improve return on investment (ROI). However, even with these concerns, the evidence overwhelmingly indicates that long tail (LT) will be a very useful tool for account managers to improve the overall return on investment (ROI).

The reports turns now to interesting results that came out of the data and which did not form part of the stated research objectives.

6.6. Sales and rental keywords

Since the company under study had separated the account in to sales head (HE), sales long tail (LT), rental head (HE) and rental long tail (LT) campaigns, an analysis of means for the cost-per-click (CPC) and click-through rates (CTRs) could be done for these groups. The results were very interesting. The sales head (HE) cost-per-click (CPC) had a mean of R1.25, while rental head's (HE) cost-per-click (CPC) mean was R0.86, with a mean difference of R0.4. Sales long tail (LT) cost-per-click (CPC) mean was R0.70, while rental



long tail (LT) cost-per-click (CPC) mean was R0.64, with a mean difference of R0.06. All the differences were significant.

The results for click-through rates (CTRs) show similar results. The sales head (HE) click-through rate (CTR) mean was 7.1%, while rental head's (HE) click-through rate (CTR) mean was 13.4%, with a mean difference of 6.33%. Sales long tail (LT) click-through rate (CTR) mean was 17.9%, while rental long tail (LT) was 27%, with a mean difference of 9.15%. All the differences were significant. The data strongly suggests that rental keywords get significantly higher click-through rates (CTRs) and significantly lower cost-per-click (CPC), which would result in a much higher return on investment (ROI). There could be many factors that contributed to this. The interaction effect for the keywords selection and rental/sales type keyword was significant, this suggest that the combination of rental and long tail keywords have bigger effect than just the two factors individually.

One possible explanation is that there is less competition for rental keywords, so since the search engines, as noted by Edelman *et al.* (2007), use the auction model to price the keywords individually, the cost is lower. If there were fewer adverts displayed this also increased the probability of the user clicking on the advert, which will raise the historical click-through rate (CTR) which, in turn, will increase the quality score and then lower the cost-per-click (CPC). This could have created a positive reinforcing cycle that conspires to improve the advertiser's return on investment (ROI).

This again suggests the individual pricing results from the auction model allows the firm to successfully employ strategies that identify keywords with lower equilibrium prices that can be added to the campaign and tailored to products. Although the rentals and sales

categories are unique to a specific industry, the result can be generalised by suggesting firms look for categories or product niches in their offering that might offer similar advantages.

6.7. Research question

The fundamental question this research aims to answer is: *“Can a long tail distribution in keyword selection help improve the effectiveness of a sponsored search online advertising campaign?”*

The overwhelming answer is: YES. The evidence strongly supports this answer, despite some of the concerns raised by the results of Hypotheses 5, 6 and 7. However, it is unlikely that the account manager will forego head (HE) keywords since the evidence still indicates that they will be responsible for a large volume of clicks. Moreover, different strategies could be designed for different browser goals. It will be most prudent to employ the strategic bidding as suggested by Xu *et al.* (2009) to be placed second or third, therefore maximising the return for head (HE) keywords.

Account managers can then grow the keyword in the account by adding long tail (LT) keywords. It is possible that long tail (LT) keywords could match the volume of head (HE) keywords in the future, but they should immediately bring the overall cost-per-click (CPC) down and improve the click-through rates (CTRs) significantly. It is also beneficial to look for specific categories or products where better performance can be achieved, similar to the rental example in this research.

7. CONCLUSION

7.1. Introduction

This final chapter will highlight the main findings of the research, summarising them with a main set of results. Some recommendations will be provided for companies and account managers. The limitations of the research will be discussed, as well as some recommendations made for future research.

7.2. Findings

The research aimed to provide the following:

1. Add to the body of knowledge about sponsored search on a keyword level.
2. Provided evidence that the combination of the granular approach resulting from applying the long tail (LT) to keyword selection and targeting advert copy will increase the click-through rate (CTR).
3. Provide a theoretical approach to keyword selection that can deliver sustainable results, giving the marketer tools to increase return on investment (ROI).

The research contributed to the body of knowledge on a keyword level by showing that a long tail (LT) distribution strategy improved the return on investment (ROI) achieved by the campaign by reducing the cost-per-click (CPC). The long tail (LT) keyword had an average cost-per-click (CPC) 50% lower than head (HE) keywords, therefore producing a 50% higher return on investment (ROI). The long tail (LT) keywords also contributed a substantial amount of clicks.

The individualising achieved by the granular approach of selecting long tail (LT) keywords, combined with the increased targeting of the advert text in order to relate directly to the keywords selected, impacted the click-through rate (CTR) significantly. The data shows a 95% increase in click-through rate (CTR) for long tail (LT) keywords compared to head (HE) keywords. While targeted advert text improves the performance of head (HE) keywords by 200% and long tail (LT) keywords by 175%, the combined improvement from head (HE) keywords with generic advert text to long tail (LT) keywords with targeted advert text is an amazing 430%.

The long tail (LT), therefore, not only improves the click-through rate (CTR) by selecting more targeted keywords, it also allows for an unrivalled ability to target the advert text to the users search query. The combined effect of the long tail (LT) keyword with targeted advert text improves the cost-per-click (CPC) from R1.67 for head (HE) keywords with generic advert text, to R0.65 for long tail (LT) keywords with targeted advert text. This combined benefit is an improvement of more than 61%.

If, as Anderson (2004) predicted, the long tail (LT) continues to get longer and fatter, this strategy could not only be sustainable but become ever more effective in the future. The long tail (LT) might become one of the marketers' key tools in improving the return on investment (ROI) for sponsored search campaigns. Firms that can build the capabilities in this area might well find a strategic advantage in the significantly lower advertising costs that this strategy can offer.

7.3. Recommendations

The recommendation will be separated into recommendations for the academia, the firm and the account manager. Firstly, the recommendations for academia are reviewed.

Anderson's (2004) long tail (LT) has been applied to many different online and offline settings. This research has added the application of the long tail (LT) to sponsored search. There are however a need for more strategies that can be used for keywords selection to enable better return on marketing investment.

Chandra (2009) showed that targeted advertising increased advertisers' willingness to pay, and allowed media owner to charge higher rates. This research provided an example where more targeted advertising was in fact employed to lower the cost of advertising. Other examples of this phenomenon need to be explored.

Dong *et al.* (2009) stressed that the firm's strategic behaviour influenced the effect of individual level targeting. This research suggest that strategic behaviour plays a critical role in evaluating the effectiveness of the application of long tail (LT) to sponsored search, if the firm did not aim to maximise traffic to its website, the perceived effectiveness could be deferent and needs to be explored.

Briggs *et al.* (1997), Dreze *et al.* (2003), Manchanda *et al.* (2006), Yoo (2008) and Fulgoni *et al.* (2009) all strongly supported the branding effects of online display advertising. This research hinted that there might be some branding benefit to sponsored search advertising,

especially for head keywords. Some research on sponsored search and branding would add great value to the literature.

Secondly, the recommendations for the firm are discussed. Sponsored search has already redefined the way advertising is done on the Internet; the next step is redefining advertising in its entirety. In developing markets the full effect of sponsored search is yet to be felt, but in developed markets it is already on a path that might soon make it the number one advertising medium. Therefore, no firm can afford not to be part of this revolution.

Firms should make sure they have a well planned strategy to develop the skills needed to turn their sponsored search campaign into a strategic advantage. This is particularly applicable to online firms, but it might also have important implication for traditional retailers, manufacturers and other sectors.

Unlike traditional media where targeting is mostly driven by media owners to achieve higher cost per impression rates, sponsored search increases targeting, and personalising leads to lower rates for the media owner and should, therefore, be driven by the advertiser. Some companies use an agency for their strategic advertising planning. These companies need to ensure that the agency not only has competencies in traditional media, but has specific skill sets necessary to employ strategies like the long tail (LT) to sponsored search campaigns in order to maximise the return on investment (ROI).

Firms should start as soon as possible. Historical click-through rates (CTRs) form part of the quality score (Edelman *et al.*, 2007; Jansen *et al.*, 2007; Jansen *et al.*, 2008; Li *et al.*, 2009; Pedersen, 2008), therefore there is no time to waste. Firms that start early and build a high

click-through rate (CTR) history will have a key advantage. The only way new entrants can even try to catch up is to spend a significant amount of money to try and rank higher, but they might never do so.

Lastly, some recommendations for the account/marketing manager are presented. Account and marketing managers should not abandon head keywords; they will make up the bulk of clicks for the foreseeable future. Account and marketing managers should rather ensure that adverts rank in order to maximise their return on investment (ROI). Xu *et al.* (2009) suggested account managers aim to be placed second or third, as this is the ideal placement; it should be tested to confirm it works for the specific account. It might even be possible to get a better return on investment (ROI) with a lower rank.

Managers should progressively build a campaign to include long tail (LT) keywords, but campaigns must be designed in such a way as to allow for very targeted advert text. This can be done by grouping long tail (LT) keywords around a theme that is then incorporated into the advert text.

Landing pages must be relevant to the keywords selected, to ensure maximum effectiveness. Landing pages also influence the quality score (Edelman *et al.*, 2007; Jansen *et al.*, 2007; Jansen *et al.*, 2008; Li *et al.*, 2009; Pedersen, 2008). A great indicator is how the page ranks in the organic results; the ideal situation is to have organic results on the first page to increase the quality score to 10/10. Competitor holding a quality score of 3/10 for the same keyword will have to bid almost four times the maximum bid to show higher.

It is also advisable to look for certain categories of products where the equilibrium prices might be lower than the average and/or there is less competition. This could result in fewer adverts being displayed which could increase click-through rates (CTRs), lowering cost-per-click (CPC) and improving return on investment (ROI).

Lastly, ensure a precise measure of the return on investment (ROI) expected, and then manage the campaign to attain that. Google Adwords (2009) provides great tracking and analytical tools to assist in this.

7.4. Limitations

This research has the following limitations:

- The company aims to attract users to its portal and, therefore, defines return on investment (ROI) as generating more users per rand spent; results might not be applicable where return on investment (ROI) is defined differently.
- An internal validity that might have influenced the results is the history. History refers to external event that are external to the experiment event but occurred simultaneously to the experiment (Malhotra, 2007, p226). The economic conditions during the experiment could have influenced the rental/sales results. With more consumers looking to rent as result of the depressed market conditions and the lack of liquidity in home loan finance.
- Only a single company in a single industry is used, therefore the results might not be relevant to other companies or industries.



- The research was only conducted on one sponsored search provider; results on others might be different.
- The targeted and generic advert text campaign was not run at the same time, nor for the same period of time. Results might have been different if they were run at the same time and for the same period of time.
- The increased click-through rate (CTR) achieved by long tail (LT) and targeted advert text could be solely as a result of less ads being shown. As the market develops, the influence these strategies have on click-through rates (CTRs) might change.
- The proxies used to measure user behaviour and compare users of long tail (LT) and head (HE) keywords could not take into account the subtle differences in the levels of the users entering the website. Long tail (LT) users might, in fact, be of a better quality than head (HE) users.

7.5. Suggestions for future research

It has now been 11 years since Goto.com first introduced the concept of sponsored search (Fain and Pederson 2006; Goldfarb and Tucker 2007; Jansen and Spink 2007; Jansen *et al.*, 2008; Wilbur *et al.*, 2009). The body of academic knowledge is growing fast, but there still a huge need for more research on the subject. Here are a few suggestions for future research:

- The effectiveness of the long tail (LT) strategy suggested in this research over time should be examined. Will it become more effective or less?
- Strategies for keyword design that can help tailor the campaign more to the users context needs to be developed.

- Comparisons should be made between the effectiveness of sponsored search and other online and off-line advertising by incorporating all benefits, including direct action, branding and others.
- User satisfaction of long tail (LT) and head (HE) users that click through to a website needs to be explored. Are long tail (LT) users more satisfied since they are taken directly to the relevant page?
- The branding effect of sponsored search advertising should be researched. Does it increase the brand equity of firms?
- Examine the influence of brand on click-through rate (CTR) and cost-per-click (CPC). Are users more likely to click on the sponsored search advert if it is a well known brand?
- How are users' search journeys influenced by long tail (LT) terms? Do they start with generic then more to long tail (LT) terms when they cannot find what they are looking for, or do they start with long tail (LT) terms?



REFERENCES

- Adams, R. (2004) Intelligent advertising. *AI & Society*, 18 (1) p68-81.
- Albright, S.C., Winston, W.L. and Zappe, C. (2006) *Data Analysis & Decision-making with Microsoft Excel*. United States: Thomson South-Western.
- Allenby, G.M., Leone, R.P. and Jen, L. (1999) A Dynamic Model of Purchase Timing with Application to Direct Marketing, *Journal of the American Statistical Association*, 94 (466) p365-374.
- Anderson, C. (2004) The Long Tail. *Wired Magazine*, October 2004, p170–177.
- Anderson, C. (2009) *Free: The future of a Radical Price*, p238. New York: Hyperion.
- Animesh, A., Ramachandran, V. and Viswanathan, S. (2005) *Quality uncertainty and adverse selection in sponsored search markets*. Maryland, NET Institute, Working Paper No.05-27. Available from <http://www.netinst.org/Siva.pdf> [Accessed on 01/05/2009].
- Ansari, A. and Mela, C.F. (2003) E-Customization, *Journal of Marketing Research*, 40 (2) p131-145.
- Bhatnagar, A. and Papatla, P. (2001) Identifying Locations for Targeted Advertising on the Internet, *International Journal of Electronic Commerce*, 5 (3) p23-44.
- Briggs, R. and Hollis, N. (1997) Advertising on the Web: Is There Response before Click-through? *Journal of Advertising Research*, 37 (2) p33-45.
- Brynjolfsson, E., Hu, Y.J. and Smith, M.D. (2006) From Niches to Riches: Anatomy of the Long Tail, *MIT Sloan Management Review*, 47 (4) p66-71.
- Brynjolfsson, E., Hu Y. and Simester, D. (2007) *Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales*. USA. SSRN. Available from: <http://ssrn.com/abstract=953587> [Accessed 26/02/2009].
- Chandra, A. (2009) Targeted Advertising: The role of subscriber characteristics in media markets, *The Journal of Industrial Economics*, 57 (1) p58-84.
- Chatterjee, P., Hoffman, D. L. and Novak, T. P. (2003) Modelling the Clickstream: Implications for Web-based Advertising Efforts, *Marketing Science*, 22(4) p520-541.
- Chen, Y. and He, C. (2006) *Paid Placement: Advertising and Search on the Internet*. NET Institute, Working Paper No. 06-02. Available at SSRN: <http://ssrn.com/abstract=936472> [Accessed 28/10/2009].
- Chickering, D.M. and Heckerman, D. (2003) Targeted Advertising on the Web with Inventory Management, *Interface*, 33 (5) p71-77.
- Chipp, K and Ismail, Z. (2008) *e-Commerce: a southern African perspective*. Pretoria: Van Schaik Publishers
- Cho, C. and Cheon, H.J. (2004) Why do people avoid advertising on the Internet? *Journal of Advertising*, 33 (4) p89-97.
- Danaher, P.J. (2007) Modeling Page Views Across Multiple Websites with an Application to Internet Reach and Frequency Prediction, *Marketing Science*, 26 (3) p 422-437.
- Dawes, J. (2009) Brand loyalty in the UK sportswear market, *International Journal of Market Research*, 51 (4) p449-463.
- Dreze, X. and Hussherr, F. (2003) Internet advertising: is anybody watching? *Journal of Interactive Marketing*, 17 (4) p8-23.

- Dreze, X. and Zufryden, F. (1998) Is Internet Advertising Ready for Prime Time? *Journal of Advertising Research*, 38 (3) p7-18.
- Dong, X., Manchanda, P and Chintagunta, P.K. (2009) Quantifying the Benefits of Individual-level Targeting in the Presence of Firm Strategic Behavior, *Journal of Marketing Research*. 46 (2) p207-221.
- Economist (2004) Virtual fun. *The Economist*,. 371 (8375) p15-16.
- Economist (2007) Who's afraid of Google? *The Economist*. 384 (8544) p9-9.
- Edelman, B. and Ostrovsky, M (2006) Strategic bidder behaviour in sponsored search auctions, *Decision Support Systems*, 43 (1) p192-198.
- Edelman B., Ostrovsky, M. and Schwarz, M (2007) Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keyword, *American Economic Review*. 97 (1) p 242-259.
- Elberse, A. (2008) Should You Invest in the Long Tail? *Harvard Business Review*, 86 (7/8) p88-96.
- Enders, A., Hungenberg, H., Denker, H. and Mauch, S. (2008) The long tail of social networking: Revenue models of social networking sites, *European Management Journal*, 26 (3) p199-211.
- Fain, D.C. and Pedersen, J.O. (2006) Sponsored Search: A Brief History, *Bulletin of the American Society for Information Science and Technology*, 32 (2) p12-13.
- Feng, J., Bhargava, H.K. and Pennock, D.M. (2007) Implementing sponsored search in Web search engines: Computational evaluation of alternative mechanisms, *INFORMS Journal on Computing*, 19 (1) p137-148.
- Ferguson, R. and Hlavinka, K. (2006) The long tail of loyalty: how personalized dialogue and customized rewards will change marketing forever, *The Journal of Consumer Marketing*, 23 (6) p 357-361.
- Fulgoni, G.M. and Morn, M.P. (2009) Whither the Click? How Online Advertising Works, *Journal of Advertising Research*, 49 (2) p134-142.
- Ghose, A. and Yang, S. (2007) *An Empirical Analysis of Search Engine Advertising: Sponsored Search and Cross-selling in Electronic Markets*. USA, NET Institute Working Paper No.07-35, Available from http://www.netinst.org/Ghose-Yang_07-35.pdf [Accessed on 01/05/2009].
- Ghose, A. and Yang, S. (2009) *An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets*, NET Institute Working Paper. Available at SSRN: <http://ssrn.com/abstract=1022467> [Accessed 30/10/2009].
- Goldfarb, A. and Tucker, C.E. (2007) *Search engine advertising: Pricing ads to context*. USA, NET Institute Working Paper No.07-23, Available from: <http://ssrn.com/abstract=1021451> [Accessed on 01/05/2009].
- Google (2009) *Google annual report*, California, USA. Google. http://investor.google.com/documents/2008_google_annual_report.html [Accessed 25/02/2009].
- Google Adwords (2009) *Welcome to AdWords*, USA, Google Inc. <http://adwords.google.com> [accessed 06/11/2009]
- Hair F.J. Anderson, E.A, Tatham, R.L. and Black, W.C. (1998) *Multivariate Data Analysis*. New Jersey: Prentice-Hall
- Hoffman, D.L. and Novak, T.P. (2000a) How to acquire customers on the Web, *Harvard Business Review*, 78 (3) p179-188.
- Hoffman, D.L. and Novak, T.P. (2000b) *When exposure-based Web advertising stops making sense (and what CDNOW did about it)*. Nashville, eLab. Available from: <http://www.weima.org/download/when%20exposure-based%20web%20advertising%20stops%20making%20sense.pdf> [Accessed on 03/05/2009].

- Hoffman, D.L. and Novak, T.P. (2000c) *Advertising Pricing Models for the World Wide Web*. Cambridge: MIT Press. Available from <http://www.weima.org/download/advertising%20pricing%20models.pdf> [Accessed on 03/05/2009].
- Hoffman, D.L., Novak, T.P. and Venkatesh, A. (2004) Has the Internet become indispensable? *Communications of the ACM*, 47 (7) p37-42.
- Hollis, N. (2005) Ten Years of Learning on How Online Advertising Build Brands, *Journal of Advertising Research*, 45 (2) p255-2686.
- Hulland, J., Chow, Y.H. and Lam, S. (1996) Use of causal models in marketing research: A review, *International Journal of Research in Marketing*, 13 (2) p181-197.
- IDC (2008) *IDC – Press Release*, Framingham, Massachusetts, USA: IDC Available from: <http://www.idc.com/getdoc.jsp?containerId=prUS21304208> [Accessed 01/05/2009].
- Ilfeld, J.S and Winer, R.S.. (2002) Generating Website Traffic, *Journal of Advertising Research*, 42 (5) p49-61.
- Iyer, G., Soberman, D. and Villas-Boas, J.M. (2005) The Targeting of Advertising, *Marketing Science*, 24 (3) p 461-476.
- Jansen, B.J. (2007a) The Comparative Effectiveness of Sponsored and Non-sponsored Links for Web E-Commerce Queries, *ACM Transactions on the Web*, 1 (1) Article 3 p1-25.
- Jansen, B.J. (2007b) Click Fraud, *Computer*, 40 (7) p85-86.
- Jansen, B.J. and Molina, P.R. (2006) The effectiveness of Web search engines for retrieving relevant ecommerce links, *Information Processing and Management*, 42 (4) p 1075-1098.
- Jansen, B.J. and Mullen, T. (2008) Sponsored Search: an overview of the concept, history and technology, *International Journal Electronic Business*, 6 (2) p114- 131.
- Jansen, B.J. and Resnick, M. (2006) An Examination of Searcher's Perceptions of Non-sponsored and Sponsored Links During Ecommerce Web Searching, *Journal of the American Society for Information Science and Technology*, 57 (14) p 1949-1961.
- Jansen, B.J. and Spink, A. (2007) Sponsored Search: Is Money a Motivator for Providing Relevant Results? *Computer*. 40 (8) p52-57.
- Jones, J. (2008) The Long Tail of Search Marketing, *NZ Marketing Magazine*, 27 (3) p24-25.
- Kazienko, P. and Adamski, M. (2007) AdROSA- Adaptive personalization of web advertising. *Information Sciences*, 177 (11) p2269-2295.
- Kim, Y., Street, W.N., Russell, G.J. and Menczer, F. (2005) Customer Targeting: A Neural Network Approach Guided by Genetic Algorithms, *Management Science*, 51 (2) p264-276.
- Kitts, B., LeBlanc, B., Meech, R. and Laxminarayan, P. (2006) Click Fraud, *Bulletin of the American Society for Information Science and Technology*, 32 (2) p20-23.
- Kumar, P. (2008) *Theory and Evidence...Search Advertising in Electronic Markets: A Study on the Impact of Keyword Wordographics*, Bachelor of Science, Leonard N. Stern School of Business, New York University.
- Laffey, D. (2007) Paid Search: The Innovation that changed the Web, *Business Horizons*, 50 (3) p211-218
- Le Rey, J. (2008) *Turning the tail: from Long Tail to Big Tail*, Rome, Blogspot.com, Available from <http://adscriptum.blogspot.com/2008/06/turning-tail-from-long-tail-to-big-tail.html> [Accessed on 04/05/2009].
- Li, Y. and Jhang-Li, J. (2009) Pricing display ads and contextual ads: Competition, acquisition and investment, *Electronic Commerce Research and Application*, 8 (1) p16-27.

- Malhotra, N.K. (2007) *Marketing research: An Applied Orientation, Fifth Edition*. New Jersey: Pearson Education Inc.
- Manchanda, P., Dube, J., Goh, K.Y. and Chintagunta, P. K. (2006) The Effect of Banner Advertising on Internet Purchasing, *Journal of Marketing Research*, 43 (1) p98-108.
- Mangani, A. (2004) Online Advertising: Pay-per-view versus pay-per-click, *Journal of Revenue and Pricing Management*, (2) 4 p295-302.
- McCubbin, K. (2009) *UK Affiliate Marketing Blog – Kristy's Affiliate Marketing Guide*. UK. Affiliate Stuff. Available from: <http://www.affiliatestuff.co.uk/> [accessed 11/11/2009]
- McDonald, S. (2008) The Long Tail and its Implications for Media Audience Measurement, *Journal of Advertising Research*, 48 (3) p313-319.
- Pedersen, J.O. (2008) Introduction to the Special Issue: Click Fraud, *International Journal of Electronic Commerce*, 13 (2) p5-7.
- Poiesz, T.B. and Robben, H.S. (1994) Individual Reactions to Advertising: Theoretical and Methodological Developments, *International Journal of Advertising*, 13(1) p25-53.
- PricewaterhouseCoopers (2009) *IAB Internet Advertising Revenue Report*, New York, IAB. Available from http://www.iab.net/media/file/IAB_PwC_2008_full_year.pdf [Accessed 01/05/2009].
- Rosen, S. (1981) The Economics of Superstars, *The American Economic Review*, 71 (5) p845-858.
- Rossi, P.E., McCulloch, R.E. and Allenby, G. M. (1996) The Value of Purchase History Data in Target Marketing, *Marketing Science*, 15 (4) 1996 p321-340.
- Rust, R.T. and Varki, S. (1996) Rising from the Ashes of Advertising, *Journal of Business Research*, 37(3) p173-181.
- Rutz, O.J. and Bucklin, R.E. (2007) *A Model of Individual Keyword Performance in Paid Search Advertising*. USA. Available from <http://ssrn.com/abstract=1024765> [Accessed on 01/05/2009].
- Rutz, O.J. and Bucklin, R.E. (2008) *From generic to branded: A model of spillover in paid search advertising*. USA. Available from <http://ssrn.com/abstract=1024766> [Accessed on 27/10/2009].
- Sherman, L. and Deighton, J. (2001) Banner advertising: Measuring Effectiveness and Optimizing Placement, *Journal of Interactive Marketing*, 15 (2) p60-64.
- Sorensen, H. (2008) Long Tail Media in the Store, *Journal of Advertising Research*, 48 (3) p329-338.
- SPSS, (2006) *SPSS 14.0 for Windows Integrated Student Version help function*. Pearson Prentice Hall
- Szymanski, B.K. and Lee, J-S. (2006) 'Impact of ROI on bidding and revenue in sponsored search advertisement auctions', *Second Workshop on Sponsored Search Auctions*, Ann Arbor, MI.
- Vise, D.A. (2005) *The Google Story: Inside the Hottest Business, Media and Technology Success of Our Time*. New York: Bantam Dell Publishing Group, p1.
- Wilbur, K.C. and Zhu, Y. (2009) Click Fraud, *Marketing Science*, 28 (2) p293-308.
- Xu, L., Chen, J. and Whinston A.B. (2009) *To Place Better or Price Cheaper? Bidding and Pricing under Keyword Advertising*. Austin, USA: University of Texas. Available from: <http://ssrn.com/abstract=1350510> [Accessed 01/05/2009].
- Yoo, C.Y (2008) Unconscious processing of Web advertising: effects on implicit memory, attitude towards the brand and consideration set, *Journal of Interactive Marketing*, 22 (2) p2-18.



Zikmund, W.G. (2003) *Business Research Methods*. United States: Thomson South-Western.