Engineering Management
Field Project

Applications of Outlier and Anomaly Detection in Sponsored Search Advertising Campaigns

By

Josh Peck

Fall Semester, 2011

An EMGT Field Project report submitted to the Engineering Management Program
and the Faculty of the Graduate School of The University of Kansas
in partial fulfillment of the requirements for the degree of
Master's of Science

Tom H. Bowlin
Tom Bowlin
Committee Chairperson

Ray Dick
Ray Dick
Committee Member

Terry Sullivan
Terry Sullivan
Committee Member

Date accepted: 12/1/11

Acknowledgements

I would like to thank the faculty and staff of the KU Engineering Management program for their support and tutelage.

Executive Summary

Organizations using sponsored search advertising rapidly find their staff overwhelmed with the amount of quantitative data available to them. One area that is often overlooked is management of dynamic change in the online marketplace. This research attempts to provide a predictive model to determine when an ad group is likely to decline in profitability.

Correlation analysis shows that, at the ad group level, the advertising metric most predictive of change in 7-day profit margin is revenue-per-click (RPC). Additionally, the likelihood of a negative change in RPC predicting a negative change in 7-day profit margin can be as high as 76% when applying these methods to ad groups that have a high number of impressions. The likelihood of false positives is low (3%-7%) when the number of impressions is high, so applying these methods would likely yield an improvement in profit over ad-hoc analysis. The anomaly detection methods show considerably less effectiveness when applied to ad groups with fewer impressions and as such should not be used in an unsupervised manner without further research.

TABLE OF CONTENTS

Acknowledgements i
TABLE OF CONTENTS iii
FIGURES iv
TABLES v
ABBREVIATIONS vi
GLOSSARY vii
Introduction 8
  Research Need 9
  Research Objectives 9
Literature Review 11
  Sponsored Search Overview 11
  Sponsored Search and Financial Markets 13
  Anomaly Detection Overview 16
  Specific Anomaly Detection Techniques 18
Research Approach 22
  Computational Methods 23
Results and Conclusions 24
  Exploratory Data Analysis 24
  Contrasting Results of Anomaly Detection Techniques 27
  Conclusions 32
Suggestions for Additional Research 34
REFERENCES 35
APPENDICES 36
  Appendix A: Histograms of Advertising Metrics 36
  Appendix B: Descriptive Statistics for Advertising Metrics 41
  Appendix C: Correlation Tables 42
  Appendix D: Anomaly Detection Technique Results 45

FIGURES

Figure 1: Mean Impressions by Day of Week 25
Figure 2: 7-Day Profit Margin Change Histogram 26
Figure 3: Change in RPC vs. 7-Day Profit Margin 27
Figure 4: Daily Change in Average Position Histogram 36
Figure 5: Daily Change in Clicks Histogram 36
Figure 6: Daily Change in Cost Histogram 37
Figure 7: Daily Change in CPC Histogram 37
Figure 8: Daily Change in CTR Histogram 38
Figure 9: Daily Change in Impressions Histogram 38
Figure 10: Daily Change in 7-Day Profit Margin Histogram 39
Figure 11: Daily Change in Revenue Histogram 39
Figure 12: Daily Change in RPC Histogram 40

TABLES

Table 1: Ad Group Data Fields 24
Table 2: Correlation for Changes in Daily Metrics 27
Table 3: Anomaly Detection Results -- All Ad Groups 29
Table 4: Anomaly Detection Results -- 4th QR Ad Groups 31
Table 5: Descriptive Statistics for Daily Metrics 41
Table 6: Descriptive Statistics for Changes in Daily Metrics 41
Table 7: Correlation for Changes in Daily Metrics 42
Table 8: Correlation for Daily Metrics 42
Table 9: Correlation for Changes in Daily Metrics (4th Quartile Subset) 43
Table 10: Correlation for Daily Metrics (4th Quartile Subset) 43
Table 11: Correlation for Changes in Daily Metrics (2nd-3rd Quartile Subset) 43
Table 12: Correlation for Daily Metrics (2nd-3rd Quartile Subset) 44
Table 13: Correlation for Changes in Daily Metrics (1st Quartile Subset) 44
Table 14: Correlation for Daily Metrics (1st Quartile Subset) 44
Table 15: Results of Anomaly Detection Techniques 46

ABBREVIATIONS

CPC	Cost-per-click, computed as $\frac{cost}{clicks}$
GFP	Generalized first-price auction
GSP	Generalized second-price auction
IQR	Interquartile range, computed as $Q_3 - Q_1$
MAD	Median absolute deviation
PPC	Pay-per-click
RPC	Revenue-per-click, computed as $\frac{revenue}{clicks}$
SERP	Search engine result page
UUID	Universally unique identifier

GLOSSARY

ad group. A group of keyword listings in a sponsored search advertising effort.

campaign. A group of ad groups in a sponsored search advertising effort.

landing page. The first webpage presented to a potential client in an online advertising campaign.

organic search results. Search results that a search engine provides for free.

paid search. Another term for "sponsored search."

sponsored search. A type of online advertising where search engines provide advertisements alongside search results.

sponsored search results. Search engine results provided by advertisers.

Introduction

Search engines are an important indicator of how many people use the Internet. comScore reports more than 17.1 billion explicit core searches were performed in July 2011. Over 11 billion (approximately 65%) of those searches were performed on Google (comScore 2011). Sponsored search marketing consists of providing relevant sponsored search engine results integrated with organic search results on the search engine results page (SERP). Advertisers are willing to make a significant investment in this form of advertising because of the large volume of Internet searches being performed and the contextual nature of the search results. Google reported $28 billion in advertising revenue in 2010, which is largely derived from their AdWords sponsored search advertising platform (comScore 2011; Google 2010).

A significant portion of the appeal of online advertising lies in the availability of metrics to determine the efficacy of an advertising effort. Historically, online marketers primarily sought to increase page views; however, they now desire more information about the online advertising campaign as it relates to profitability. One of the features of sponsored search that appeals to advertisers is the ability to have direct feedback on their advertising efforts (Jansen and Mullen 2008). Sponsored search advertising campaign managers are able to track, test, and target user activities in ways that are not possible in traditional media. Many advertising agencies take advantage of the improved feedback mechanisms of online advertising campaigns to improve advertising efficiency for their clients.

Research Need

There are numerous techniques being used to analyze sponsored search data and optimize the many factors that affect the performance of a sponsored search advertising campaign, ranging from statistical analysis to machine learning (Chandola, Banerjee, and Kumar 2009; Hodge and Austin 2004). Because many of these techniques are computationally intensive or require human intervention, it should be useful to apply computationally simple techniques to identify areas within the advertising campaign where additional attention or computation should be focused.

Research Objectives

The research and experiments performed will achieve the goals of 1) researching computational techniques, 2) identifying significant advertising parameters, 3) applying selected techniques, and 4) comparing results of each technique as applied to available data sets.

Numerous computational techniques are available for identifying anomalies in time-series data. The research portion of this project will identify strengths and weaknesses of different methods and select a subset of available techniques to apply to the available data sets based upon effectiveness in predicting a profitability change as well as computational efficiency.

Each of the selected methods for identifying anomalies will be applied to the data set in carefully designed experiments to determine which techniques best correspond with a change in the advertising marketplace. Because sponsored search advertising campaigns can be very complex, each technique will be applied to advertising metrics determined to be significant in predicting a change in the profitability of an advertising campaign. Then, results will be compared between the anomaly detection techniques to assess which techniques warrant further study and implementation into an advertising campaign monitoring system.

Literature Review

Discussion of the literature relevant to this research will include overviews of sponsored search and anomaly detection as well as in-depth discussion of specific anomaly detection techniques as may be applied to sponsored search markets.

Sponsored Search Overview

Web search engines are tools for effectively finding information on the Internet. These tools are widely used (Jansen and Mullen 2008). Until 1998, most search engines financed their operations by selling banner advertising on the search engine results page. Banner advertising was typically priced on a CPM (cost per mille/thousand), which provided incentive for the search engine provider to keep users on the site as long as possible. Since the role of a search engine is to connect users with other websites, this resulted in a dilemma for search engine providers. Search engines that provided the best search results were more popular with the users, yet they typically kept the searcher on the SERP for less time (Fain and Pedersen 2006).

In 1996, Yahoo! struck a deal with Proctor & Gamble to charge only when a user clicked on a banner advertisement, which was an early version of CPC (cost-per-click) pricing. In 1998, GoTo (later acquired by Overture, which was in turn acquired by Yahoo!) combined the core elements of what evolved into modern sponsored search (Fain and Pedersen 2006; Jansen and Mullen 2008). These elements include advertiser-provided content, advertiser-provided bids, a review process by the search engine, matching advertiser content to user queries, display of advertiser content, processes that gather data and meter clicks to charge advertisers (Fain and Pedersen 2006). In 1999, BeFirst (now MIVA) followed with a similar sponsored search product. In 2002, Google adopted the model and incorporated click feedback. In 2005, AskJeeves adopted the model and MSN search extended it to include behavioral targeting (Fain and Pedersen 2006; Jansen and Mullen 2008). The sponsored search model quickly became popular among search engine providers because it aligned the interests of the searcher (most relevant search results) with the interests of the search engine provider (increased revenue to finance operations) (Dhar and Ghose 2010). The model also became popular with advertisers because the direct accounting provided by sponsored search marketing offered more accountability than traditional forms of direct response marketing.

As sponsored search increased in popularity, search engine providers began modifying the algorithms used to determine which advertisements to show and how to price advertising allocation. GoTo.com's original algorithm was a generalized first-price (GFP) auction, in which the providers (advertisers) were ranked by bid price on a keyword (Jansen and Mullen 2008). Jansen and Mullen (2008) describes how the GFP auction is applied to sponsored search and the resulting bidding wars in terms of an example:

> **Example (first-price auction):** Suppose advertiser A will pay up to \$ 1 for the keyword 'coffee', while advertiser B values the same keyword at \$ 0.74. If B starts by bidding the lowest possible price, say \$ 0.10, then A would bid \$ 0.11 to win the first advertisement slot. Advertiser B would respond by bidding \$ 0.12, and so forth. Once A bids \$ 0.75, then B will not bid \$ 0.76 since it only values the keyword at \$ 0.74. To acquire the second advertisement slot, B simply has to bid \$ 0.10. Now, A only needs to bid \$ 0.11 to win the first slot, and so the cycle starts over (Jansen and Mullen 2008)

In February 2002, Google introduced a generalized second-price (GSP) auction to its advertising platforms to increase stability. The GSP auction permits the highest bidding advertiser to pay the bid price of the second-place advertiser plus a small delta.

Jansen and Mullen describe how the GSP auction is applied to sponsored search with another example:

> **Example (second-price auction):** Consider our previous example where advertiser A is willing to pay up to $ 1 for the keyword 'coffee', and advertiser B will pay up to $ 0.74. In a sealed-bid environment, recall that only the auctioneer knows the value of all bids. If A bids $ 1, and B bids $ 0.74, then in a second-price auction, A pays $ 0.74 + $ 0.1 (the minimum delta) while B pays the minimum bid amount of $ 0.10. Now suppose A and B bid strategically rather than honestly. If A bids $ 0.78 (i.e., less than what the advertisement is worth to A) and B bids $ 0.80 (i.e., more than what the advertisement is worth to B), then B wins with a price of $ 0.79. This is 5 cents more than the slot is worth to B. On the other hand, if A bids $ 0.70 and B bids $ 0.65 then A gets the slot for $ 0.66. However, B, with a bid of $ 0.74 could have won the slot for $ 0.71, and effectively loses 3 cents of value. However, when the bidders can see, or infer, their competitors' bids, clearly there are other strategic possibilities. For example, if A truthfully bids $ 1, then B can safely bid $ 0.99 instead of $ 0.74, forcing A to pay $1 for the advertisement slot instead of $ 0.75. (Jansen and Mullen 2008)

Additionally, Google made adjustments to the allocation algorithm to include factors such as click-through-rate (CTR), keyword relevancy, and landing page quality. These modifications were introduced to improve the user experience and increase the company's profits by penalizing advertisers who were using deceptive practices or who had exceptionally poor (non-credible or fraudulent) websites. Later in 2002, Yahoo! Search Marketing (then Overture) updated its pricing algorithms to include a quality index and utilize the GSP-style auction to address similar issues on its advertising platform (Jansen and Mullen 2008). Both Google and Yahoo! now run continuous auctions in which advertisers compete for placement (Fain and Pedersen 2006).

Sponsored Search and Financial Markets

As sponsored search advertising auctions matured and became more competitive, they began to take on properties of novel auctions (Jansen and Mullen 2008) and as such began to demonstrate similarities to financial markets. This evolution has caused some to study sponsored search markets in terms of economics and game theory (Dhar and Ghose 2010).

It is useful to discuss the mechanics of how sponsored search advertising campaigns operate to better contrast sponsored search markets with financial markets. Advertisers must submit their website via keyword listings to the search engine when they wish to market their product to consumers in the sponsored search marketplace. Bids are assigned to each keyword listing to assist the search engine in assigning a ranking to the listing (Ghose and Yang 2009). When a searcher enters a search term into a search engine, the search engine uses the search term to rank the organic search results contextually, and then uses the search term, keyword bid, and a number of other factors to operate a modified GSP auction to rank the sponsored search results. The sponsored search results are typically tailored web results including title, description, and URL for the advertising campaign being operated by the advertiser (Jansen and Mullen 2008). The advertiser may also specify additional rules for the advertising campaign to limit the geographic location in which the advertisement is shown (geo-targeting), the language in which the advertisement should be displayed (language targeting), to limit the time the advertisement is shown (scheduling or day-parting), to specify the algorithm used to match the search term (match type), and to specify the amount of money the advertiser is willing to spend on the advertising campaign (budget) (Google).

Financial exchanges and sponsored search auctions share common characteristics and utilize similar business models; both businesses rely on network effects and volume of transactions to generate revenue. In financial markets, revenue is generated by the volume of trades (as a percentage of each transaction), while in sponsored search auctions the volume of clicks on advertisements generates revenue. Both types of markets attempt to create network effects: financial exchanges by attracting buyers and sellers, and sponsored search markets by attracting search query volume and sponsored search advertisements. Price discovery and portfolio optimization occurs in both markets based upon the item being traded and careful selection of assets (Dhar and Ghose 2010).

There are differences between financial and sponsored search markets that must be understood to contrast the two. When a stock is sold on a financial market, no other buyer is permitted to own that specific equity; however, in a sponsored search market, as Dhar and Ghose (2010) state, "advertisers bid for rank and not just on 'winning' the auction by being the highest bidder (although being towards the top of the auction is considered desirable)." Both types of markets also differ in the level of transparency provided to the actors: in financial markets, the algorithm for winning is well established and "known to everyone." Sponsored search markets, however, are "relatively opaque"; price is one factor in ranking, but other factors such as historical performance, landing page quality, and CTR are also considered (Dhar and Ghose 2010). One reason for the differences between the markets is that both emerged with different levels of technological maturity. In the case of financial exchanges, running a continuous auction was historically computationally costly so the exchange benefited from limiting the number of auctions (and thus the number of stocks) traded on the exchange. Because sponsored search markets emerged at a time when computation was relatively cheap, the risk of running a continuous auction is greatly reduced, "so the universe of symbols in sponsored search markets is potentially much larger, with no regard to the inherent interest in them" (Dhar and Ghose 2010)

The similarity between sponsored search markets and financial markets has naturally led researchers to study methods that historically have been applied to financial markets. These methods may also apply to sponsored search markets as they mature (Dhar and Ghose 2010).

Anomaly Detection Overview

"*Anomaly detection* refers to the problem of finding patterns in data that do not conform to expected behavior" (Chandola, Banerjee, and Kumar 2009) and is used in a large number of subject domains including health care, intrusion detection, fault detection, military surveillance (Chandola, Banerjee, and Kumar 2009), fraud detection, loan application processing, and time-series monitoring (Hodge and Austin 2004). Hawkins describes the problem of outliers as "one of the oldest in statistics" (Hawkins 1980) that has been tackled through many approaches labelled as "outlier detection," "anomaly detection," or with names, depending on which author is describing the technique (Hodge and Austin 2004).

Aspects of anomaly detection include understanding the nature of the input data, understanding the types of anomalies, and understanding data labels for the data set. The nature of input data can be conceived as a collection of data instances with attributes that may be considered to be of a specified type (binary, categorical, continuous). Data instances may be univariate or multivariate and may have relationships with other data instances (sequential, spatial). Types of anomaly may include point anomalies (anomalous with respect to the data set), contextual anomalies (anomalous in a specific context), or collective anomalies (combination of data instances is anomalous). Data labels, if available, should mark data instances as *normal* or *anomalous*. Additionally, data labels can require great expense to generate, since it often requires a human expert to manually perform labeling. Data labels are prone to error because anomalous conditions are by definition difficult to predict and may arise dynamically (Chandola, Banerjee, and Kumar 2009). Anomaly detection systems typically output a score that is applied to each data instance. This score is a measure of the degree to which the data instance is considered an outlier or has an expected value. Essentially, the system assigns an "anomalous" or "normal" label to the data instance.

Several considerations must be addressed when selecting an appropriate technique for an outlier detection system. First, the method must scale to accommodate the number of data points to be processed. Second, the method must accurately model the data instances and accurately highlight outlying points. Finally, the method must be capable of selecting a neighborhood of interest in which anomalies are likely to be found (Hodge and Austin 2004). Data labeling also drives the selection of techniques for outlier detection systems.

Specific anomaly detection techniques can be loosely grouped into unsupervised methods (referred to as Type 1 by Hodge and Austin), supervised methods (referred to as Type 2 by Hodge and Austin), and semi-supervised methods (referred to as Type3 by Hodge and Austin) based upon the use of data labels (Chandola, Banerjee, and Kumar 2009; Hodge and Austin 2004). Unsupervised techniques are widely applicable since they require no training data. Semi-supervised techniques are moderately applicable since only normal data instance labels are required. Supervised classification requires both normal and anomalous data instance labels and is limited to specific subject domains where all anomalous conditions are knowable (Chandola, Banerjee, and Kumar 2009; Hodge and Austin 2004).

Specific Anomaly Detection Techniques

Anomaly detection techniques can be further grouped into categories based upon the underlying mathematical or computational techniques used to analyze the data instances.

Statistical techniques were among the earliest methods used to evaluate the presence of outliers and are generally useful for quantitative data or ordinal data that can be coerced into numeric form (Hodge and Austin 2004). The basic assumption of a statistical technique is that "normal data instances occur in high probability regions of the stochastic model, while anomalies occur in the low probability regions of the stochastic model" (Chandola, Banerjee, and Kumar 2009). Statistical models used for anomaly detection can be further classified into parametric and non-parametric models. Parametric techniques can be based upon existing statistical data distributions such as the Poisson distribution and the normal distribution or based upon a distribution created from a regression model. The low probability areas of the statistical model are well known for existing data distributions (Seo 2002). However, in the case of a distribution generated from a regression model, each data point must be compared with the model to score the magnitude of the data point's residual difference from the regression model (Chandola, Banerjee, and Kumar 2009).

Nonparametric statistical techniques make fewer assumptions about the data and do not require prior definition of the statistical model for normal data, but rather infer the statistical model from the given data (Chandola, Banerjee, and Kumar 2009). The histogram technique is an example of a nonparametric technique for anomaly detection in which the training data set is used to build a histogram, and then each testable data instance is checked to see if it falls into one of the existing bins. An anomaly is considered to be any data point that does not fall into one of the existing histogram bins (Chandola, Banerjee, and Kumar 2009).

Kernel functions may be used as a nonparametric technique (Chandola, Banerjee, and Kumar 2009) or semi-parametric technique (Hodge and Austin 2004), which Hodge and Austin describe as "[aiming] to combine the speed and complexity growth advantage of parametric methods with the model flexibility of non-parametric methods." Kernel-based techniques are able to do this by estimating the probability density function (PDF) of the normal instances, and then comparing test data instances with the low-density region(s) of the PDF (Chandola, Banerjee, and Kumar 2009; Hodge and Austin 2004). Computational complexity of statistical methods largely relies upon the nature of the underlying statistical model (Chandola, Banerjee, and Kumar 2009).

Nearest neighbor-based techniques, described as proximity-based by Hodge and Austin, are another prevalent area of research in the field of anomaly detection and typically rely upon the assumption that "normal data instances occur in dense neighborhoods, while anomalies occur far from their closest neighbors." Nearest neighbor techniques are typically grouped into methods that either measure the distance to nearby neighbors or measure the density of the neighborhood of each data instance (Chandola, Banerjee, and Kumar 2009). Proximity-based techniques are typically useful for both Type 1 (unsupervised) and Type 2 (supervised) outlier detection because they make no assumptions about the underlying data model and are relatively simple to implement (Hodge and Austin 2004). A common approach scores the magnitude of the anomaly by measuring the Euclidian or Gaussian distance to the k^{th} nearest neighbor (*k*-

NN). Much work on k-NN has centered on reducing the complexity from $O(n^2)$ or exponential computational growth with regard to the number of data points (Knorr and Ng 1998), because the traditional method requires a distance calculation between each data instance (Chandola, Banerjee, and Kumar 2009; Hodge and Austin 2004).

An alternative approach is to estimate the relative density of the neighborhood for each data instance. Density-based methods also suffer from high computational complexity, and perform poorly when the density of a given region varies.

Clustering-based techniques, where a clustering algorithm is used to group the data instances into clusters and then the distance from the data instance to the nearest cluster centroid is used to compute a score are conceptually similar to nearest-neighbor techniques. Clustering-based and proximity-based techniques share the ability to operate in an unsupervised mode. However, clustering-based techniques are typically less computationally complex once the data has been clustered, since significantly fewer distances must be calculated. The key drawbacks to clustering-type methods lie in the fact that clustering pre-processing is computationally difficult and the underlying clustering algorithm may have implications that require consideration when implementing an anomaly detection system (Chandola, Banerjee, and Kumar 2009). Hybrid techniques, which combine two or more of the previously discussed methods, have arisen recently to overcome weaknesses and exploit the strengths of other techniques. For example, a system may use multiple techniques (statistical, clustering, nearest-neighbor), then employ a meta-classifier to reconcile the results of each. The strength of utilizing multiple techniques is that they are able to overcome the weaknesses of any one technique, however they do so at increased computational cost (Hodge and Austin 2004).

Categorical data has been a challenge for most anomaly detection algorithms, since statistical methods and nearest-neighbor methods require real numbers as values. Machine learning anomaly detection techniques such as decision trees and rule-based systems, which process a series of antecedents before producing a conclusion, have been shown to operate with much lower levels of computational complexity; however, they are susceptible to problems with incomplete training data and often do not detect new types of anomalies well. For example, when utilizing a rule-based technique, all possible anomalous conditions must be known ahead of time. In many domains, it is prohibitively expensive to utilize data mining techniques or human experts to build an exhaustive set of possible anomalies. Similarly, decision trees suffer from over-selection, due to incompleteness, which requires human experts or other machine learning techniques to either prune or pre-select the records (Hodge and Austin 2004).

Research Approach

The goal of this research is to devise a predictive algorithm to operate at an early stage in a hybrid anomaly detection system. The algorithm should be run as the first statistical test on the data to obtain a quick score of potential anomalies that could affect future profits. In order to be successful, the algorithm must be effective at predicting changes in future profitability for components of the advertising campaign, have the ability to score each anomaly to assist with prioritization, and have a low rate of false-positives. Non-functional requirements of the algorithm include the need for minimal data storage overhead and low computational costs.

The first stage of the research is to perform exploratory data analysis with the goal of determining factors affecting future profitability of an advertising campaign. As the factors affecting future performance of the advertising campaign are better understood, anomaly detection techniques are applied to the data set to determine which techniques are best able to trigger an alert when a change in profitability is likely to occur in the future. Performance of the anomaly detection techniques is initially tested against the entire data set to better understand which techniques are most promising. The most promising techniques are then applied to the data set on a day-by-day basis to simulate the conditions in which the algorithm is expected to operate.

It is essential to establish an appropriate metric for detecting changes in the performance of the advertising campaign because the goal of this research is to detect immediate changes in the advertising marketplace. For this purpose, a 7-day forward moving average of profit margin for each ad group in the advertising campaign is utilized, which could be summarized as:

$$profit\_margin\_7day_{t_i} = \frac{\left(\sum_{t_i}^{t_{i+6}} profit\_margin_{t_i}\right)}{7}$$

Computational Methods

Two non-functional requirements of limited data storage and low computational costs give immediate guidance indicating it would be preferable to use univariate parametric statistical anomaly detection techniques. The techniques and analysis described by Seo provide the basis for algorithm selection and analysis. The methods applied include:

- Standard Deviation Method
- Tukey's Method
- MADe Method

Results and Conclusions

Exploratory Data Analysis

The data set is aggregated by the unique identifier (UUID) for the ad group (ad_group_hash) and by the date of the corresponding statistic (datestamp). The resulting data set consists of a series of data points with one data instance per day per ad group. The data set used for testing contains 45,695 observations across 703 ad groups for the 90-day period between March and June 2009 for metrics shown in Table 1.

Field Name	Type	Description
ad_group_hash	string	UUID to uniquely identify ad group
datestamp	string	date of recorded data instance
avg_pos	numeric	weighted average of shown ad positions
clicks	numeric	number of clicks
cost	numeric	daily cost in USD
cpc	numeric	cost per click
ctr	numeric	click-through rate
dow	string	day of week for data instance
impressions	numeric	number of times ad presented
profit	numeric	daily profit in USD
profit_margin	numeric	profit margin
profit_per_click	numeric	profit per click in USD
revenue	numeric	revenue in USD
rpc	numeric	revenue per click in USD

Table 1: Ad Group Data Fields

Upon initial observation, the variation of performance metrics across ad groups is significant (i.e. some ad groups have many clicks and high cost, others have few clicks and low cost), so it is useful to focus on per-click metrics such as CPC and RPC for comparison. Additionally, it appears useful to operate on differences ($datavalue[t_n] - datavalue[t_{n-1}]$), rather than raw numeric values, so that negative changes are negative values and positive changes are positive values. Further, it is apparent that observations from weekends are inherently anomalous, likely due to advertiser demand or overall search volume, from the bar chart in Figure 1: Mean Impressions by Day of Week.

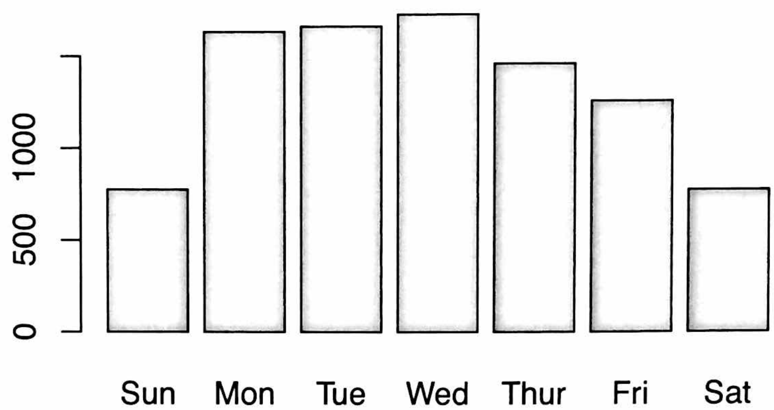

Figure 1: Mean Impressions by Day of Week

These weekend dates have been screened from the observations so that only data instances gathered on weekdays are presented for further analysis. It is obvious from the daily change in impression counts and from graphical representations of the daily impression counts that there is a wide range of search volume indicated in the data set. Ad groups have further been grouped by volume of impressions. Ad groups with impression counts in the 1st quartile, 2nd-3rd quartiles, and 4th quartile are grouped so that differences related to search volume can be further analyzed.

It is useful to observe whether the advertising metrics are stable over the period described in the data set. It is expected that in a stable system, the variability of advertising metrics would be centered strongly around zero and this is observed in the histogram for the 7-day change in profit margin as well as other advertising metrics available in Appendix A: Histograms of Advertising Metrics.

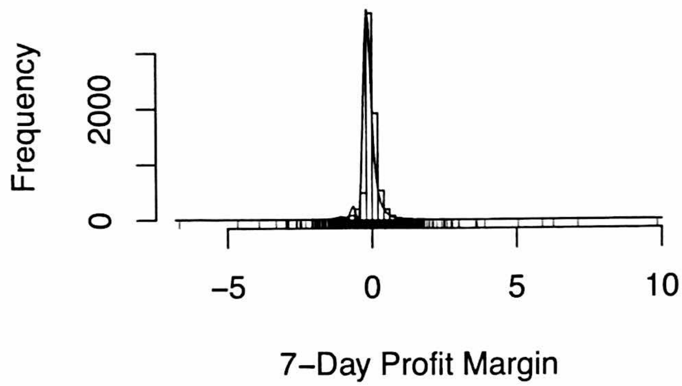

Figure 2: 7-Day Profit Margin Change Histogram

The next goal of this research is to determine which advertising metrics (or changes in them) are most likely to correspond with a change in profitability. For this purpose, the Pearson r correlation is calculated for each of the summarized metrics across the data set as a whole as well as within each of the impression-quartile subsets. The results of this computation can be found in Appendix C: Correlation Tables. Results for the data set as a whole can be seen in Table 2: Correlation for Changes in Daily Metrics. It should be noted that the $profit\_per\_click\_delta$ and $profit\_margin\_delta$ fields are closely coupled to the calculation used to determine 7-day profit margin and are discarded. The next best correlation is for the rpc_delta field with a value of approximately 0.41 and a visual inspection (Figure 3: Change in RPC vs. 7-Day Profit Margin) further reinforces that there is a relationship between 7-day profit margin and the daily change in RPC. Correlation between the daily change in RPC and the 7-day profit margin is confirmed for the subsets 4^{th} quartile and 2^{nd}-3^{rd} quartile subsets (approximately 0.40 and 0.48, respectively). However the 1^{st} quartile subset demonstrates reduced correlation at approximately 0.30.

cpc_delta	-0.008146597
cost_delta	-0.006351958
clicks_delta	-0.005072117
impressions_delta	-0.002193153
ctr_delta	0.008678448
avg_pos_delta	0.037747119
revenue_delta	0.095668933
profit_delta	0.127528134
rpc_delta	0.410483013
profit_per_click_delta	0.411382764
profit_margin_delta	0.459858648

Table 2: Correlation for Changes in Daily Metrics

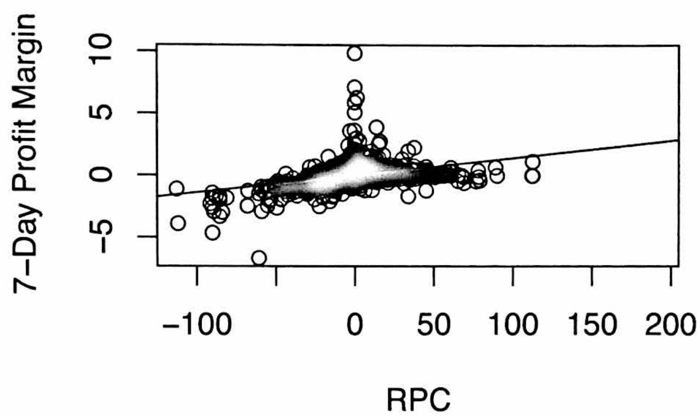

Figure 3: Change in RPC vs. 7-Day Profit Margin

Contrasting Results of Anomaly Detection Techniques

Each anomaly detection technique is incrementally tested against the entire data set as well as each of the quartile range subsets to determine if the technique applied to the daily change in the independent variable, RPC, is predictive of a change in the 7-day profit margin. Conversely, the likelihood that a positive change in 7-day profit margin, given that the technique has detected an anomaly, is measured. Only negative changes in

RPC and 7-day profit margin are tracked for this report, since those are the changes of most interest as a component of a system attempting to detect problems with an advertising campaign (typically a company is much less concerned if an advertising campaign is outperforming expectations).

In addition to applying each method to each date in each ad group in the data set, each method has been applied using multiple parameters. For example, when utilizing a technique that considers an anomaly to be n standard deviations from the mean, both 2.0 and 3.0 have been used as parameters and the resulting anomalies for each trial are recorded. The three methods have been selected from Seo (2002) based upon analysis provided and their applicability to the data set. The three methods selected are:

- Standard Deviation Method with 2σ and 3σ from mean
- Tukey's Method with 1.5 IQR and 3.0 IQR from quartile boundaries
- MADe Method with 2.0 MAD and 3.0 MAD from median

The results of each method / parameter combination for the full data set and each subset can be found in Table 15: Results of Anomaly Detection Techniques and the following table describe the results derived from iterating over the each ad group of the entire data set.

All Observations having Impressions						
Total Observations	20670					
Observations where pm_7d_delta negative	2762					
Observations where pm_7d_delta positive	2853					
	2.0 * SD	**3.0 * SD**	**Tukey 1.5**	**Tukey 3.0**	**2.0 * MAD**	**3.0 * MAD**
negative anomalies detected	207	44	569	365	792	540
p(negative anomaly detected)	0.020948234	0.005611998	0.055394291	0.035994194	0.077164973	0.053942912
p(pm_7d_delta < 0 | negative anomaly detected)	0.574879227	0.522727273	0.485061511	0.449315068	0.513888889	0.453703704
p(pm_7d_delta > 0 | negative anomaly detected)	0.014492754	0.022727273	0.049209139	0.04109589	0.055555556	0.048148148

Table 3: Anomaly Detection Results -- All Ad Groups

The two standard deviations method, one of the more simple methods applied, shows positive results and detects 207 anomalies for further inspection. The conditional probability that the 7-day future moving average of profit decreases when an anomaly is detected is 57.5%. Inversely, the probability of a positive change in profitability when an anomaly is detected (a false positive) is only 1.5%. As expected, the three standard deviations method detects fewer anomalies, only 44. For the three standard deviations method, the conditional probability that 7-day profit margin decreases when an anomaly is detected is 52.3%, and the risk of a false positive is 2.3%.

Tukey's 1.5 IQR method detects 569 anomalies for further inspection with a 48.5% probability that 7-day profit margin decreases when an anomaly is detected. Conversely, the probability of profit margin increasing when an anomaly is detected is about 5%. Tukey's 3.0 IQR method detects fewer anomalies, with 365 anomalies detected. For this method, the conditional probability that 7-day profit margin decreases given that an anomaly is detected is 45% with a 4.1% probability that 7-profit increases under the same conditions.

The two median absolute deviation method detects 792 anomalies, with a conditional probability that 7-day profit margin decreases when an anomaly is detected of 51.4%. Conversely, the likelihood that the 7-day profit margin increases when an anomaly is detected is 5.5%. When the 3.0 MADs method is applied, the number of anomalies detected drops to 540. The probability of 7-day profit margin increasing given that an anomaly is detected is 45.4% and the risk of a false positive is 4.8%.

A number of interesting results are discernable when the ad groups are grouped into subsets by the number of impressions. The ad groups in the fourth quartile range for impressions demonstrate significantly improved conditional probabilities that 7-day profit margin decreases given that an anomaly is detected. These probabilities typically approach 70% and the two standard deviation method appears to demonstrate the strongest results with 76% probability that 7-day profit margin decreases when an anomaly is detected. Likewise, the three standard deviations method demonstrates strong results with a 70% conditional probability of 7-day profit margin decrease when an anomaly is detected. The risk of false positives for the two and three standard deviation methods is not significantly higher at 3.1% and 0%, respectively.

Ad Groups in 4th QR for Impressions						
Total Observations	8710					
Observations where pm_7d_delta negative	1248					
Observations where pm_7d_delta positive	1294					
	2.0 * SD	**3.0 * SD**	**Tukey 1.5**	**Tukey 3.0**	**2.0 * MAD**	**3.0 * MAD**
negative anomalies detected	63	10	143	73	223	122
p(negative anomaly detected)	0.014695752	0.003329506	0.033295063	0.017566016	0.051320321	0.030080367
p(pm_7d_delta < 0 | negative anomaly detected)	0.761904762	0.7	0.72027972	0.684931507	0.713004484	0.704918033

p(pm_7d_delta > 0 | negative anomaly detected)	0.031746032	0	0.055944056	0.068493151	0.071748879	0.073770492

Table 4: Anomaly Detection Results -- 4th QR Ad Groups

Conversely, the ad groups in the first quartile rank for impressions demonstrate significantly reduced utility. The probability of the 7-day profit margin decreasing given an anomaly detected drops to a range from 0% to 12.5%. None of the methods appear to be very effective on the ad groups in this quartile. The two MADs method detected the most anomalies for this group, but it only detected 17 anomalies over the 90-day interval.

Ad Groups in 1st QR for Impressions						
Total Observations	1625					
Observations where pm_7d_delta negative	7					
Observations where pm_7d_delta positive	3					
	2.0 * SD	**3.0 * SD**	**Tukey 1.5**	**Tukey 3.0**	**2.0 * MAD**	**3.0 * MAD**
negative anomalies detected	8	1	12	9	17	16
p(negative anomaly detected)	0.0073846 15	0.001846154	0.013538462	0.009230769	0.019692308	0.017846154
p(pm_7d_delta < 0 | negative anomaly detected)	0.125	0	0.083333333	0.111111111	0.117647059	0.0625
p(pm_7d_delta > 0 | negative anomaly detected)	0	0	0	0	0	0

Conclusions

This research began with the idea that computational methods may be able to help predict when a change in profitability is likely to occur in a sponsored search advertising campaign. The discussed approaches appear to provide an improved level of information about the potential for a future decrease in profitability for an ad group. The benefit is especially apparent in the ad groups in the fourth quartile of impressions with 70% or more probability of decreasing 7-day profit margin when an anomaly is detected.

Before assessing which method is most effective, it is useful to consider the potential consumers of information generated by a predictive model of this type. If an employee is the consumer of information generated by this model, it seems reasonable to expect that a very low rate of false positives would be important and that the raw number of anomalies detected should be small. If another computer algorithm were the consumer of the information provided by these approaches, many more anomalies could be reviewed and more computationally intensive methods could be used.

The best performing algorithm overall was the two standard deviations method with a 57% probability of predicting a decrease in 7-day profit margin. However, it only detected 207 anomalies (or about 2% of the observations where the 7-day profit margin began decreasing). This method may be useful if an employee was the consumer of the information, since it has a low probability that the 7-day profit margin will increase and the employee would only have to inspect about two anomalies per day.

It may be useful to consider the two median absolute deviations method if another computer algorithm is to consume the output of this algorithm, since it detected 792 (or about 7%) of the observations where 7-day profit margin decreased with only a 5% probability that 7-day profit margin would increase.

Regardless of which algorithm is selected it appears that an improvement can be made with relatively little computation overhead. It is also notable that the statistical methods applied are more useful for ad groups with higher numbers of web searches. It should prove worthwhile to pursue utilizing these methods as one component of a system that monitors the health of sponsored-search advertising campaigns.

Suggestions for Additional Research

The number of observations for ad groups in the first quartile rank for impressions was relatively low in this data set. It may be useful to consider a data set with longer history or to utilize other methods to group the low volume ad groups and treat them as a whole, rather than measuring each individually. It should be noted that these ad groups were the least likely to demonstrate normal characteristics. As such, it is expected that statistical tests designed to address normal distributions would be less useful. Since the inter-day delay between conversions in these ad groups is so great, it may be useful to consider a Poisson or exponential probability distribution to more appropriately model the sparseness of activity in these ad groups.

The advertising campaign information utilized for this research only represents a single advertising vertical market. While it is likely that other advertising vertical markets would operate similarly, it could be informative to apply the methods described in this report to an additional data set before integrating them into a system.

This research operated on historical data in an unsupervised manner. Improved data labeling may provide additional information that would be useful. If an employee were assessing the output of this system on a regular basis, simple binary feedback ("Was this alert useful, yes or no?") may provide further validation and demonstrate other directions for further development.

REFERENCES

Chandola, V., A. Banerjee, and V. Kumar. "Anomaly Detection: A Survey." *ACM Computing Surveys (CSUR)* 41, no. 3 (2009): 1-58.

comScore, Inc., "Comscore Releases July 2011 U.S. Search Engine Rankings", comScore, Inc. http://www.comscore.com/Press_Events/Press_Releases/2011/8/comScore_Releases_July_2011_U.S._Search_Engine_Rankings.

Dhar, V., and A. Ghose. "Sponsored Search and Market Efficiency." *Information Systems Research* 21, no. 4 (2010): 760-772.

Fain, D.C., and J.O. Pedersen. "Sponsored Search: A Brief History." *Bulletin of the American Society for Information Science and Technology* 32, no. 2 (2006): 12-13.

Ghose, A., and S. Yang. "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets." *Management Science* 55, no. 10 (2009): 1605-1622.

Google, "Adwords Targeting and Placements", Google http://adwords.google.com/support/aw/bin/static.py?hl=en&guide=23292&page=guide.cs (accessed 7/1/2011 2011).

________, "Google Investor Relations: 2010 Financial Tables", Google http://investor.google.com/financial/2010/tables.html (accessed July 2011).

Hawkins, D.M. *Identification of Outliers*: Chapman & Hall, 1980.

Hodge, V., and J. Austin. "A Survey of Outlier Detection Methodologies." *Artificial Intelligence Review* 22, no. 2 (2004): 85-126.

Jansen, B.J., and T. Mullen. "Sponsored Search: An Overview of the Concept, History, and Technology." *International Journal of Electronic Business* 6, no. 2 (2008): 114-131.

Knorr, E.M., and R.T. Ng. "Algorithms for Mining Distance-Based Outliers in Large Datasets." 392-403: Citeseer, 1998.

Seo, S. "A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets." University of Pittsburgh, 2002.

APPENDICES

Appendix A: Histograms of Advertising Metrics

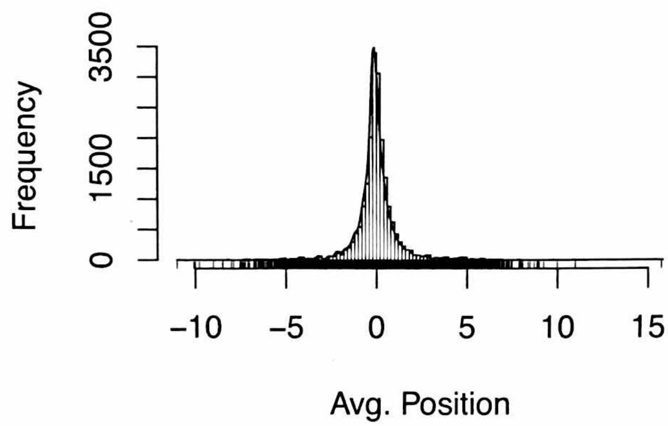

Figure 4: Daily Change in Average Position Histogram

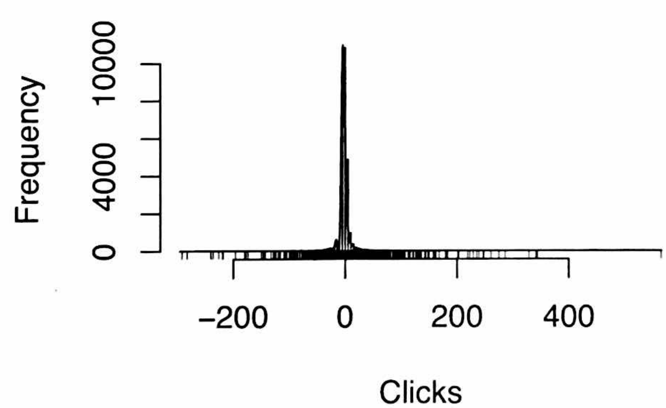

Figure 5: Daily Change in Clicks Histogram

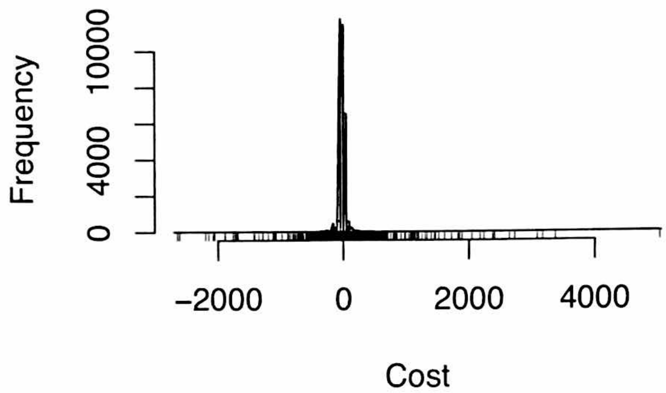

Figure 6: Daily Change in Cost Histogram

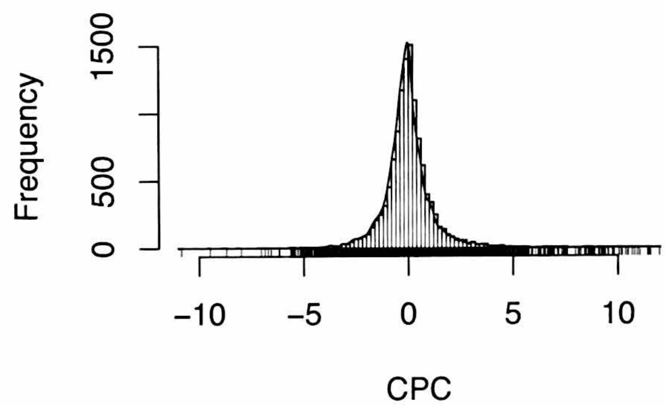

Figure 7: Daily Change in CPC Histogram

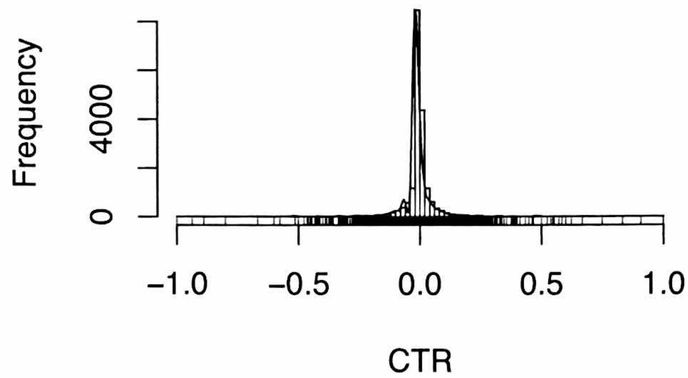

Figure 8: Daily Change in CTR Histogram

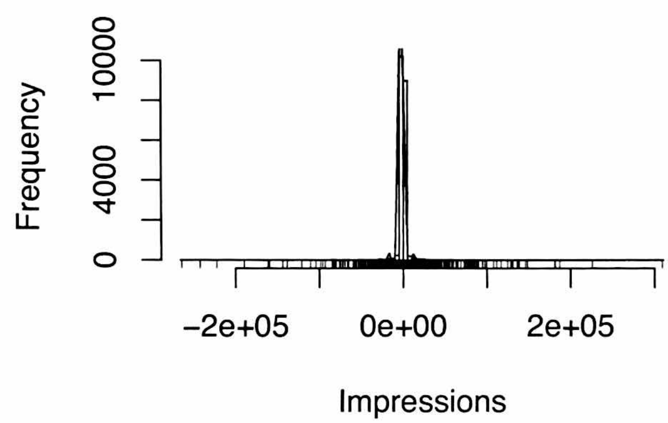

Figure 9: Daily Change in Impressions Histogram

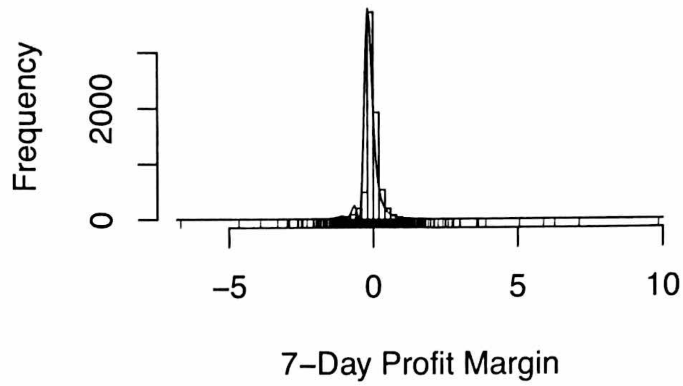

Figure 10: Daily Change in 7-Day Profit Margin Histogram

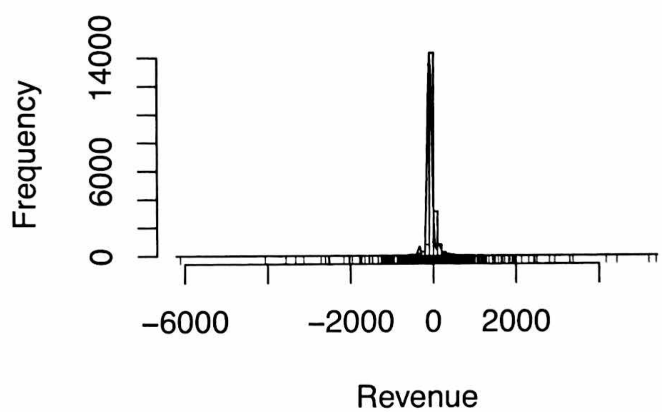

Figure 11: Daily Change in Revenue Histogram

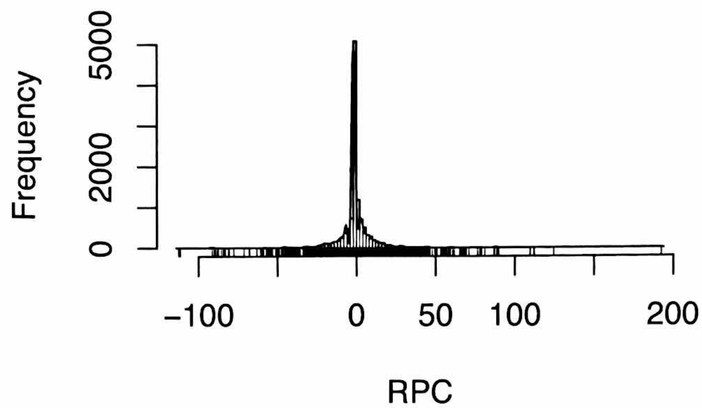

Figure 12: Daily Change in RPC Histogram

Appendix B: Descriptive Statistics for Advertising Metrics

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
avg_pos	0	2.141	3.571	3.481	4.635	15.74
clicks	0	0	2	12.66	8	638
cost	0	0	8.49	76.92	38.3	6013.11
cpc	0.01	2.99	4.581	4.887	6.442	19.59
ctr	0	0	0.01493	0.04738	0.06098	1
impressions	0	24	93	3412	554	343393
profit	-5967.92	-11.387	0	6.095	0	2784.86
profit_margin	-1	-1	-0.90266	-0.04216	0.18752	68.0184
profit_per_click	-19.59	-4.1043	-2.0761	0.1916	1.0486	187.32
revenue	0	0	0	83.02	29.7	7880.33
rpc	0	0	0.509	5.078	6.75	202.5

Table 5: Descriptive Statistics for Daily Metrics

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
avg_pos_delta	-11.00	-0.36	0.00	0.06	0.41	15.74
clicks_delta	-291.00	-1.00	0.00	0.25	1.00	565.00
cost_delta	-2650.66	-6.32	0.00	1.30	6.06	5032.59
cpc_delta	-10.85	-0.49	0.00	0.07	0.49	11.97
ctr_delta	-1.00	-0.01	0.00	0.00	0.01	1.00
impressions_delta	-263905.00	-34.00	0.00	85.98	31.00	309193.00
profit_delta	-6076.19	-7.65	0.00	0.14	8.00	5206.30
profit_margin_delta	-69.02	-0.37	0.00	0.01	0.33	69.02
profit_per_click_delta	-114.88	-2.22	-0.02	0.04	2.15	183.39
revenue_delta	-6110.70	0.00	0.00	1.44	0.00	5376.60
rpc_delta	-112.50	-1.73	0.00	0.11	1.81	192.60

Table 6: Descriptive Statistics for Changes in Daily Metrics

Appendix C: Correlation Tables

cpc_delta	-0.008146597
cost_delta	-0.006351958
clicks_delta	-0.005072117
impressions_delta	-0.002193153
ctr_delta	0.008678448
avg_pos_delta	0.037747119
revenue_delta	0.095668933
profit_delta	0.127528134
rpc_delta	0.410483013
profit_per_click_delta	0.411382764
profit_margin_delta	0.459858648

Table 7: Correlation for Changes in Daily Metrics

impressions	-0.26666349
avg_pos	-0.10744993
clicks	0.03040732
cost	0.03841286
revenue	0.07560169
profit	0.14233836
cpc	0.20082173
ctr	0.22782898
profit_per_click	0.34643324
rpc	0.38413068
profit_margin	0.4311065

Table 8: Correlation for Daily Metrics

clicks_delta	-0.012985079
cost_delta	-0.010947107
impressions_delta	-0.003827838
ctr_delta	-0.002698734
cpc_delta	0.015534015
revenue_delta	0.074053306
avg_pos_delta	0.074341175
profit_delta	0.105232235
rpc_delta	0.403550502
profit_per_click_delta	0.40400842
profit_margin_delta	0.455599688

Table 9: Correlation for Changes in Daily Metrics (4th Quartile Subset)

impressions	-0.3455036
avg_pos	0.01469402
cost	0.12196322
clicks	0.1293237
revenue	0.15141963
profit	0.15618673
ctr	0.29186823
cpc	0.35441803
profit_per_click	0.40158679
profit_margin	0.48145403
rpc	0.48532785

Table 10: Correlation for Daily Metrics (4th Quartile Subset)

cpc_delta	-0.017747073
cost_delta	-0.000316247
impressions_delta	0.00314513
avg_pos_delta	0.011670164
clicks_delta	0.014405514
ctr_delta	0.015326406
revenue_delta	0.24911565
profit_delta	0.274417413
rpc_delta	0.42025031
profit_per_click_delta	0.421499582
profit_margin_delta	0.462509208

Table 11: Correlation for Changes in Daily Metrics (2nd-3rd Quartile Subset)

avg_pos	-0.10660682
impressions	-0.07525128
cpc	-0.07133311
cost	-0.01875511
clicks	0.02078552
ctr	0.12897398
revenue	0.16860234
profit	0.24652943
rpc	0.31856913
profit_per_click	0.33610553
profit_margin	0.3880403

Table 12: Correlation for Daily Metrics (2nd-3rd Quartile Subset)

ctr_delta	-0.44915162
cpc_delta	-0.22009121
impressions_delta	-0.1161733
avg_pos_delta	-0.07567083
clicks_delta	-0.07259666
cost_delta	-0.05587011
revenue_delta	0.14633687
profit_delta	0.28577394
rpc_delta	0.29559959
profit_margin_delta	0.33791158
profit_per_click_delta	0.35774003

Table 13: Correlation for Changes in Daily Metrics (1st Quartile Subset)

clicks	-0.3900316
impressions	-0.34801561
cost	-0.33044677
rpc	-0.30166692
profit_margin	-0.29481654
revenue	-0.28672721
profit_per_click	-0.28013537
profit	-0.11853423
cpc	-0.05673278
avg_pos	-0.03106414
ctr	-0.01147717

Table 14: Correlation for Daily Metrics (1st Quartile Subset)

Appendix D: Anomaly Detection Technique Results

All Observations having Impressions						
Total Observations	20670					
Observations where pm_7d_delta negative	2762					
Observations where pm_7d_delta positive	2853					
	2.0 * SD	**3.0 * SD**	**Tukey 1.5**	**Tukey 3.0**	**2.0 * MAD**	**3.0 * MAD**
negative anomalies detected	207	44	569	365	792	540
p(negative anomaly detected)	0.020948234	0.005611998	0.055394291	0.035994194	0.077164973	0.053942912
p(pm_7d_delta < 0 | negative anomaly detected)	0.574879227	0.522727273	0.485061511	0.449315068	0.513888889	0.453703704
p(pm_7d_delta > 0 | negative anomaly detected)	0.014492754	0.022727273	0.049209139	0.04109589	0.055555556	0.048148148
Ad Groups in 4th QR for Impressions						
Total Observations	8710					
Observations where pm_7d_delta negative	1248					
Observations where pm_7d_delta positive	1294					
	2.0 * SD	**3.0 * SD**	**Tukey 1.5**	**Tukey 3.0**	**2.0 * MAD**	**3.0 * MAD**
negative anomalies detected	63	10	143	73	223	122
p(negative anomaly detected)	0.014695752	0.003329506	0.033295063	0.017566016	0.051320321	0.030080367
p(pm_7d_delta < 0 | negative anomaly detected)	0.761904762	0.7	0.72027972	0.684931507	0.713004484	0.704918033
p(pm_7d_delta > 0 | negative anomaly detected)	0.031746032	0	0.055944056	0.068493151	0.071748879	0.073770492
Ad Groups in 2nd-3rd QR for Impressions						
Total Observations	10335					
Observations where pm_7d_delta negative	1507					
Observations where pm_7d_delta positive	1556					
	2.0 * SD	**3.0 * SD**	**Tukey 1.5**	**Tukey 3.0**	**2.0 * MAD**	**3.0 * MAD**
negative anomalies detected	136	33	414	283	552	402
p(negative anomaly detected)	0.028350266	0.008127721	0.080599903	0.055732946	0.107982583	0.079729076
p(pm_7d_delta < 0 | negative anomaly detected)	0.514705882	0.484848485	0.415458937	0.399293286	0.445652174	0.393034826
p(pm_7d_delta > 0 | negative anomaly detected)	0.007352941	0.03030303	0.048309179	0.035335689	0.050724638	0.042288557
Ad Groups in 1st QR for Impressions						

Total Observations	1625					
Observations where pm_7d_delta negative	7					
Observations where pm_7d_delta positive	3					
	2.0 * SD	**3.0 * SD**	**Tukey 1.5**	**Tukey 3.0**	**2.0 * MAD**	**3.0 * MAD**
negative anomalies detected	8	1	12	9	17	16
p(negative anomaly detected)	0.007384615	0.001846154	0.013538462	0.009230769	0.019692308	0.017846154
p(pm_7d_delta < 0 | negative anomaly detected)	0.125	0	0.083333333	0.111111111	0.117647059	0.0625
p(pm_7d_delta > 0 | negative anomaly detected)	0	0	0	0	0	0

Table 15: Results of Anomaly Detection Techniques