

Institution:	UDur: University of Durham
Unit of Assessment:	Unit 10: Mathematical Sciences
Title of case study:	Statistical Modelling For Digital Marketing *** Redacted version with confidential information removed***
1. Summary of the impact	<p>This work revolutionizes the way in which pay-per-click (PPC, internet search) advertising is optimised. Recent estimates suggest that 80% of companies market online, yielding \$30 billion of revenue to companies managing the process. The research of the Durham team gives a collaborating company, Summit Media Ltd, a substantial lead in statistical tools to provide PPC optimisation. This lead is giving the company a more dominant position in the UK and is leading it to pursue international business more aggressively, as an industry leader. The company confirms that implementation of the research has generated new business worth millions of pounds and enabled the employment of additional staff.</p>
2. Underpinning research	<p>This work has been carried out by David Wooff (lead) with PDRAs Dr Jillian Anderson and Dr Amin Jamalzadeh. Funding for the research is two parallel EPSRC KTP awards 2009-2013 (£211,000) and 2012-2015 (£138,000) with industrial funding from Summit Media Ltd (SM), the initial end-user. A summary of staff engaged in the project, with dates, is:</p> <p>Prof David Wooff, Feb 2010 –</p> <p>Dr Jillian Anderson, Feb 2010 – Sep 2011, Jul 2012–Jun 2013 (now a full-time company employee)</p> <p>Dr Amin Jamalzadeh, Sep 2011 – Jul 2012 (now a full-time company employee)</p> <p>The research concerns exploitation of vast amounts of information routinely collected as part of SM's digital marketing activities. SM manages these activities for clients such as major UK online retailers -Argos, John Lewis, Homebase, and so forth. The purpose of the research is to optimize Pay-Per-Click (PPC) positioning and pricing in order to maximise customer traffic and revenue subject to constraints, for example to maximize expected revenue subject to a stated budget and cost-of-sale target. Traditionally this has been thought almost impossibly large scale. A typical client account may require optimization over hundreds of thousands of keywords (search terms), each of which has a multivariate data history going back months or years.</p> <p>The major advances in this research are:</p> <ul style="list-style-type: none"> • a logical structuring of the problem and its decomposition into conditionally independent components; • statistical modelling of each component using classical and Bayesian statistical methods; • development of methods to carry out in reasonable time millions of regression models as required; • predictive simulation of revenues, generated from the underlying models; • optimization of the allocation of budget to keywords, together with their positions; • multi-layered diagnostic checking to compare observed to predicted outcomes.

The most recent (Jan 2012) relevant academic publication in this area suggests a methodology which is appropriate for up to three keywords. By contrast, the Durham group's recent test-bed dataset contains 115000 keywords and optimization procedures which take around one hour on a desktop. The Durham methodology is implemented in bespoke software, written by the company, which calls code written by us in the statistical language R. This software went live in November 2012.

The methodology is explained in references [2-6] cited in section 3 below, which build on the preliminary work published in [1]. It is internationally leading and in an area, digital commerce, flagged as an EPSRC theme.

3. References to the research

[1] Wooff, D. A. & Jamalzadeh, A. (2013), *Robust and scale-free effect sizes for non-Normal two-sample comparisons, with applications in e-commerce*, Journal of Applied Statistics 40, 2495-2515, doi: 10.1080/02664763.2013.818625, <http://dx.doi.org/10.1080/02664763.2013.818625>.

[2] Wooff, D. A. & Anderson, J. (2013), *Time-weighted multi-touch attribution and channel relevance in the customer journey to online purchase*, Journal of Statistical Theory and Practice, doi: 10.1080/15598608.2013.862753, <http://dx.doi.org/10.1080/15598608.2013.862753>.

[3] Wooff, D. A. & Anderson, J. (2013), *Time-weighted attribution of revenue to multiple e-commerce marketing channels in the customer journey*, <http://dro.dur.ac.uk/11458>.

[4] Wooff, D. A. & Anderson, J. (2013), *Inferring marketing channel relevance in the customer journey to online purchase*, <http://dro.dur.ac.uk/11459>.

[5] Wooff, D.A. (2013), *Optimization of pay-per-click bidding for search engines*. Invited conference presentation, GDRR 2013: Third Symposium on Games and Decisions in Reliability and Risk, Kinsale, County Cork, Ireland, July 8th - 10th, 2013.

[6] Anderson, J. (2013), *Weighted Attribution of Revenue to Marketing Channels within a Customer Purchase Path*, and Jamalzadeh, A. (2013), *Measuring PPC Brand Keywords Incrementality Using Geo Experiments*, talks at the Young Statisticians' Meeting YSM2013, Imperial College London, July 4th-5th 2013.

References [2-6] explain work undertaken through the parallel KTP awards described in section 2 above. The first of these awards, which has now finished, was given the highest grade of 'Outstanding' by the KTP grading panel for its achievement in meeting KTP's Objectives.

4. Details of the impact

Suppose you want to buy a laptop. What you might do is go to a search engine and type *laptop* (known as the *keyword*) in the search field. If you are using Google, you will see sponsored links (adverts), as well as the results of natural search. The adverts appear because the keyword you typed is one that a company has paid Google to display whenever someone searches for it. Broadly what happens is that if you then click on the advert, the sponsoring company pays Google a small amount. This is called Pay Per Click (PPC). Total paid-for search spend was £3.1 billion in the UK in 2012, growing at about 14% per year¹, with Google's market share about 88% in December 2012².

Impact case study (REF3b)

The amount a company is prepared to pay per click determines the *position* of the sponsor's advert in the list. A count is made of the number of times the advert appears - these are called *impressions*. Once you click on the link, you are taken to the sponsor's website. This is a *visit*. Once within that website, you might visit several pages and might buy something from the sponsoring company. Your browsing history is recorded using *cookies* stored on your computer, and is usually aggregated with the browsing histories of thousands of customers visiting the sponsor's website from various sources. This results in an enormous volume of data. For example, for every keyword for a given sponsor, there is a daily summary which reports the total number of adverts shown, the resulting number of visits, the total amount paid by the sponsoring company for the clicks, the total sales revenue generated by the clicks, and so forth.

One of Summit Media's tasks is to maximise the flow of traffic to a client. Traditionally they manage the keywords and PPC bids using their expertise. However, there were many relationships which were poorly understood. For example, what is the relationship between number of impressions and PPC price, and does this lead to increased sales revenue for the sponsor. Further, how does this vary by keyword and how does it vary by sponsor?

The Durham research has addressed these issues by modelling the complex interrelationships, simulating from the constructed models, and then providing individual predictions of expected revenue (together with uncertainties) for every advert position for every keyword in the client's portfolio. The portfolio is then optimized through an algorithm developed by the Durham team. Several different feasible solutions are available, depending on the risk attitude of the client – for example higher returns might be associated with greater uncertainty. Many of the issues posed challenges which the group has needed to overcome: for example, the massive scale of the problem requires special statistical modelling.

This has a very major impact for the collaborating company in several important ways. Firstly, far greater precision and certainty in their operations, as previous practice focused only on default pricing and rough heuristics, together with sharper human judgements for a tiny fraction of keywords. Secondly, a much reduced need for human intervention and so a significant saving in personnel costs. Thirdly, the ability to obtain much better revenue streams for clients because of understanding the uncertainties involved and through optimization of budget. Fourthly, the ability to attract more clients because it will be able to offer both a cheaper service and a more precisely quantified expected revenue.

Summit Media, based in Hull, approached Durham University while it was seeking a higher education research partner to enhance the services it offered to clients. During the resulting KTP period, from 2010-13, the company... **[text removed for publication]**.

Summit Media listed the main achievements of the KTPs as (i) the “development of a market leading solution (Forecaster) which has revolutionised the way pay PPC advertising is optimised”; (ii) the creation of a new department (Insight, now headed by one of the KTP associates as a full time employee); (iii) tighter and automated control over key business drivers; (iv) client retention and (v) The acquisition of new clients... **[text removed for publication]**.

In the KTP final report³, Summit Media says that it “definitely wouldn't be making similar progress” without the Durham KTPs and that “the main outcome from the project, Forecaster, is seen as the single most important piece of innovation that Summit has developed over the past few years... “ **[text removed for publication]**.

5. Sources to corroborate the impact

1. Internet Advertising Bureau UK 2012 Full Year Digital Adspend Factsheet, available from <http://www.iabuk.net/research-subjects/online-adspend> (copy accessed 17/10/2013).

Impact case study (REF3b)

2. Google market share data from Experian Hitwise, available at <http://www.experian.co.uk/marketing-services/news-december-2012-search.html> (copy accessed 17/10/2013).
3. Technology Strategy Board: Knowledge Transfer Partnership KTP007499 Partners Final Report.
4. [text removed for publication].
5. [text removed for publication].
6. [text removed for publication].
7. [text removed for publication].