# The Correlation between Visits and Product Sales on Three Business-to-Customer Internet Web Sites

**Robert B. Mellor**
Faculty of Computing, Information Systems & Mathematics
Kingston University
London KT1 2EE

**Keywords**
Internet statistics, web site hits, product turnover, marketing.

## Abstract

Short-term analysis of 3 web sites selling B2C products revealed a chaotic picture where absolute visit rates, absolute numbers of products sold, and number of visits per product sold, varied wildly with type of product and with time. However long-term analysis, from Jan. 1997 to Jan. 2003, surprisingly revealed that on average one product was sold every 10229 visits, with little significant variance (max. 4.28%) between web site annual averages. Both automated submissions (keyword submission) to search engines, and bulk e-mailing, could increase visit rates, but their effect on product sales was negligible. The explanation proposed is that the majority of visits recorded are machine-made, and not caused by humans.

## Abbreviations and Definitions

B2C: Business to Customer. Customer: A human successfully completing a financial transaction. HTML: Hyper Text Markup Language. HTTP: Hyper Text Transfer Protocol. SME: Small and Medium-Sized Enterprise. UNIX: A common server platform. URL: Uniform Resource Locator (the mechanism for addressing resources on the Internet). Visitor: A resolvable HTTP request.

## Introduction

Even used indiscriminately, Internet-mediated marketing methods such as automated submissions, bulk e-mailings and banner advertising, may result in increased visitor rates to the relevant web sites. However increased visit rates are not synonymous with increased interest from customers, and therefore do not automatically result in higher levels of product sales. During the late 1990s it was common for companies pay a fee to place banner advertisements on e.g. search engine pages. This banner was "exposed" to the hypothetical end-customer a certain number of times and, if the banner was sufficiently interesting, then the customer could click on the banner and thus visit the advertisers web site. In recent years this unfocussed "exposures" approach has been partly replaced by "pay per click" approaches, where the advertiser pays only for a successful click-through (for review see Prasad *et al*, 2003).

However the essential paradox remains that the company placing the advertising pays for increased visits (the volume of which can be measured by analysing web log files), but still does not know what value a click has. This is of prime importance because, in their classic paper, Novak & Hoffman (2000) compared costs for various forms of Internet marketing with the costs of comparable advertising channels and found that Internet advertising costs around twice as much as e.g. newspaper advertising.

This report analyses three Internet web sites from companies that offer widely different products in three different EU countries. Product sales are correlated with Internet visits and two methods of increasing visits are investigated. These are automated keyword submissions to search engines, and bulk e-mailing of offers, where the e-mail includes a web site URL.

## Materials and Methods

**Data sources.**
Twenty-five disparate companies distributed throughout all the EU countries were asked if they wished to participate in this survey. Although these twenty-five companies were active in quite different branches, they had in common that they offer B2C Internet retailing of products in the price class from around 100 Euros to around 1000 Euros, and had done so for at least three years. Data from the fourteen companies agreeing to participate was screened and found to be sufficiently complete in only three cases.

An overview of the three on-line companies is given in Table 1. To protect confidentiality the three companies involved have been designated by letters. The three companies involved have given full access to their Internet statistics and to their Internet-related sales statistics, for the periods shown.

| Company | Type/ employees | Area of Business | Start of Internet presence | Average product price range (Euros) | Period during which statistics have been made available for this study |
|---|---|---|---|---|---|
| A | SME /120 | Travel Agency | 1994 | 700 -1000 | 1997-2000 |
| B | Micro-Business /1 | Astrology & Partner Matching | 1995 | 100-350 | 1997-2000 |
| C | Private College /40 | Adult Education | 2000 | 200-800 | 2000-2003 |

***Table 1:*** *Business overview over the three data sources.*

All 3 web sites were HTML on UNIX platforms. This means that the log file raw data was collected by the same method in all cases. None of the web sites distributed cookies. All 3 companies report a server up time of over 99.5%. An overview of their Internet presence is given in Table 2.

| **Company Identification** | **Top Level Domain** | **Approx. maximum daily hit rate during the period studied** | **Marketing and business strategy.** |
|---|---|---|---|
| A | .DK | 19000 | "Clicks and Mortar" |
| B | .DE | 350 | Pure Internet |
| C | .ORG.UK | 1000 | e-learning |

***Table 2:*** *Key factors in the Internet presence of the three data sources.*

During the period covered in this study, Company A accepted on-line orders but did not accept on-line payments directly, whereas Companies B and C employed third-party mechanisms for accepting on-line credit card payments. All three companies accepted off-line business.

**Geographical, seasonal and language considerations**
Company A services the whole of Denmark and southern Sweden. The web site is mostly in Danish with some English. Company A reports that annual selling curves exhibit a marked polarity, with high activity in the winter months ("winter break" travel, bookings for summer holidays) and low activity in the summer months, when the customers are away on holiday.

Company B sells services on a regular year-round basis in countries representing the developed world. The web site is in English and German.

Company C admits students immediately upon registration, but still reports a faint seasonality with peaks in February and August, the traditional semester starting months. Company C's marketing is focussed on emerging countries, especially the Middle East. The web site is mostly in English, with some Arabic.

**Internet and sales statistics**
Internet statistics were collected by UNIX log-file analysis. Further details about such analyses can be found in Wu & Chen (2002).

Company A used the freeware www_stats (ftp.ics.uci.edu/pub/websoft/wwwstat), whilst Companies B and C both used licensed versions of Webalizer (www.mrunix.net/webalizer). Because all three web sites run on UNIX platforms, there were no differences in the format of the raw data. Compatibility was checked by taking a sample of the log files from Company A, analysing it with a demo version of Webaliser, and comparing these results with the results obtained from Company C's own www_stats program. No significant differences were found.

Both of these programs, www_stats and Webalizer, reset "visits" to zero at 00.01 o'clock on the first of each month.

Customer buying was recorded as sales of product units, i.e. the number of invoices sent. Thus e.g. returning customers are not counted as one individual. Because of Company A's lack of on-line payment system, customers ordering at the end of a month may receive an invoice dated the following month. Therefore, in the case of Company A, sales months are calculated from the 5th of a month to the 4th of the following month.

**Other methods**
Automated submissions (keyword submission) to search engines were achieved in all cases using Trellian SubmitWolf (www.trellian.com). E-mails were sent using GroupWise (Company A) or Eudora (Company C) in plain ASCII text. Successful e-mails refer to the number sent minus those returning a mailer daemon error.

## Results

**The effect of automated keyword submission on visit rates.**
The effect of automated keyword submission was illustrated both by positive and negative examples. Firstly, Company A hired a submissions service in autumn 1998. The results are shown in Figure 1. Assuming that the inclination of the basic curve was due to expansion in general Internet use, and thus would have continued anyway during the experimental period, then the submission service was responsible for an approximately 19% increase in visits during the immediate post-submission two months.

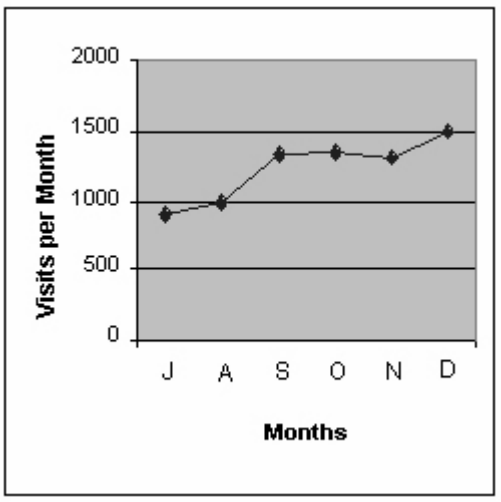


***Figure 1:*** *Company A, effect of automated submissions on visit rates during the second half of 1998. Submissions were effected on the 30th and 31st of August.*

Secondly, Company C routinely used submission upon posting new or updated files on their server, but stopped this practice during autumn 2002. Figure 2 shows a

corresponding slump in visit rates during the months when no submissions were made.

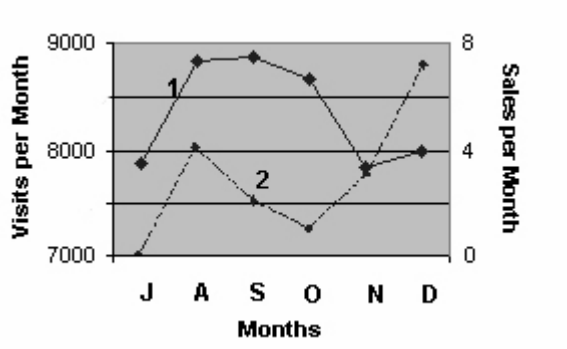


***Figure 2:*** *Company C, Curve 1: Effect of neglecting automated submissions on visit rates during the second half of 2002. Submissions stopped on the 1st of September. Curve 2: Sales during the same period.*

**The effect of bulk e-mail on visit rates.**
Both Company A and Company C make use of bulk e-mailing as a marketing technique. No useful statistics exist for Company A that correlates bulk e-mailing with web site visits. However Company A has a business pipeline dedicated to e-mail marketing based on an e-mail address database containing around 12000 e-mail addresses at any one time. About 400 new additions to the database occur per month via a HTML form on a dedicated HTML "subscription" page. Bulk e-mailing takes place quarterly and results in typically approx. 1000 mailer daemon error messages, reporting that customers have changed their e-mail address, mis-spelt their e-mail address, post box full, and other errors. This corresponds to a customer "loss" rate of approx. 3% per month. On an annual average (over 3 years), Company A reports the sale of one product per 21870 successful e-mails sent.

Company C sends out one e-mail message in bulk in January each year (2001 and 2002). The message was sent twice, with 14 days between the two. Firstly it was sent to those who have requested information, but had not bought a product. The second mailing consists of exactly the same message, but sent to those who had studied (or were still studying) at the College. Figure 3 shows that both mails elicited a significant response in visits to the web site in 2001 (the curve for 2002 is not shown, but the tendency was very similar). In both 2001 and 2002 an estimated approximately 25% of successful e-mail recipients re-visited the web site.

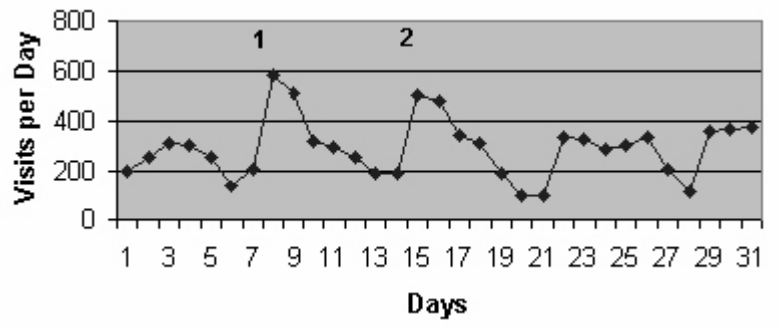


***Figure 3:*** *Company C, effect of e-mailings on visit rates during the days of January 2001. At point 1, Sunday the 7th, 1233 successful e-mails were sent to those who had requested more information. At point 2, Sunday the 14th, 799 successful e-mails were sent to existing customers.*

**The correlation between Internet visits and product sales**

Figure 4 shows the number of visits and number of products sold per month for each company's web site from January 1998 to December 2002. For brevity, data for 1997 (Company A and Company B) is excluded from Figure 4, but included in Table 3.

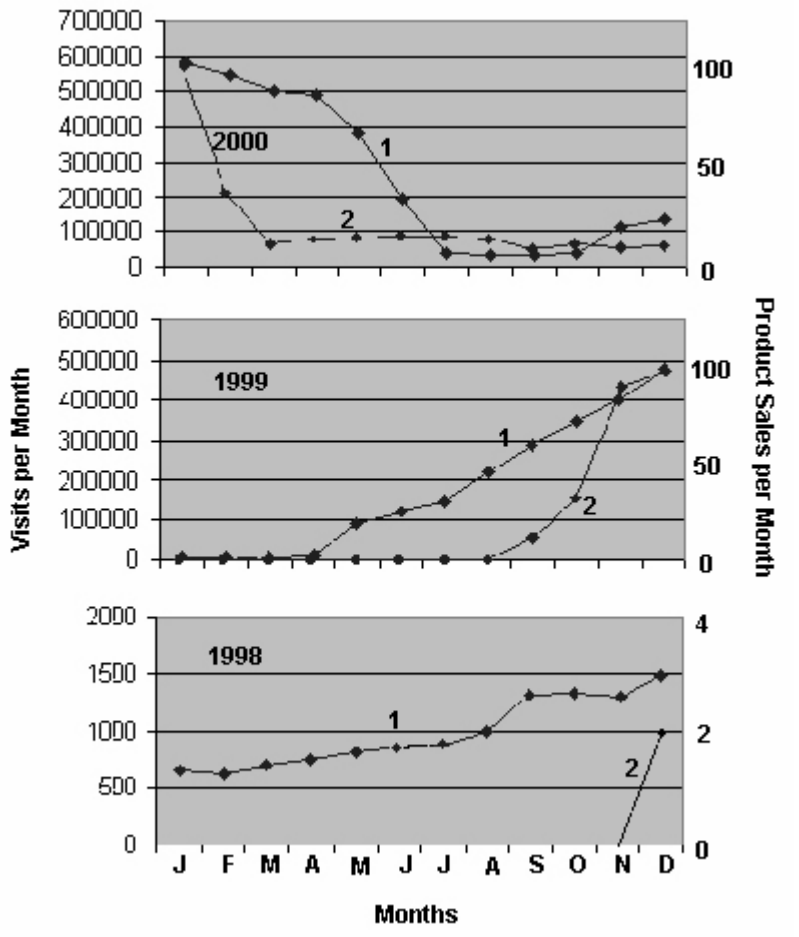


Fig 4A

Fig 4B

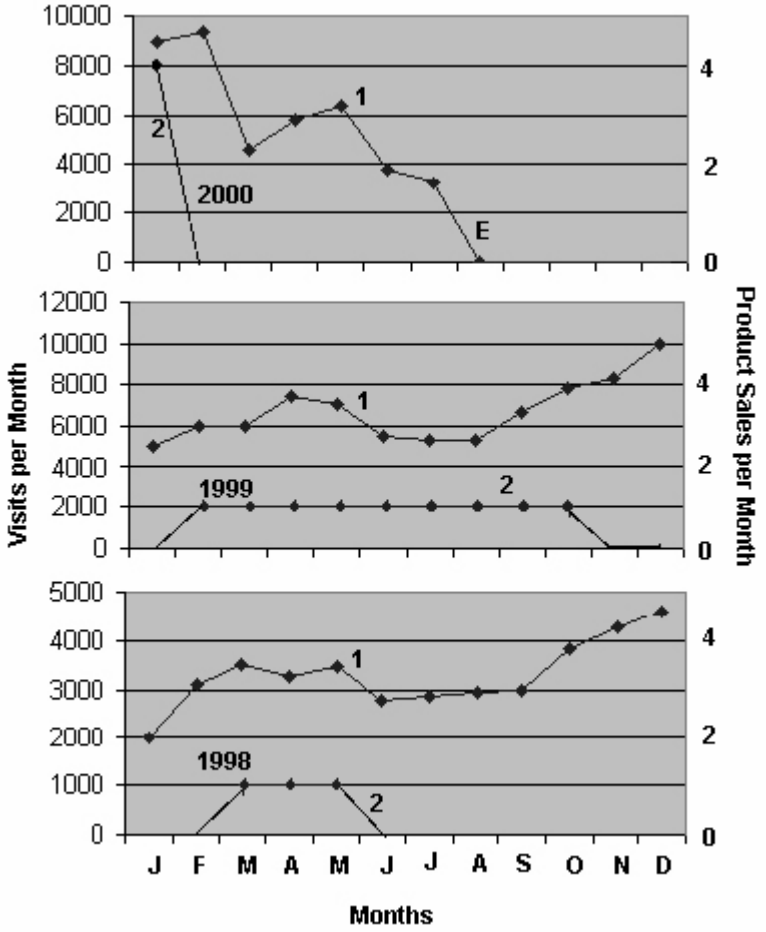
10000
8000
6000
4000
2000
0
4
2
0
1
2
2000
E
12000
10000
8000
6000
4000
2000
0
1
1999
2
Visits per Month
Product Sales per Month
5000
4000
3000
2000
1000
0
1
1998
2
J F M A M J J A S O N D
Months

Fig 4C

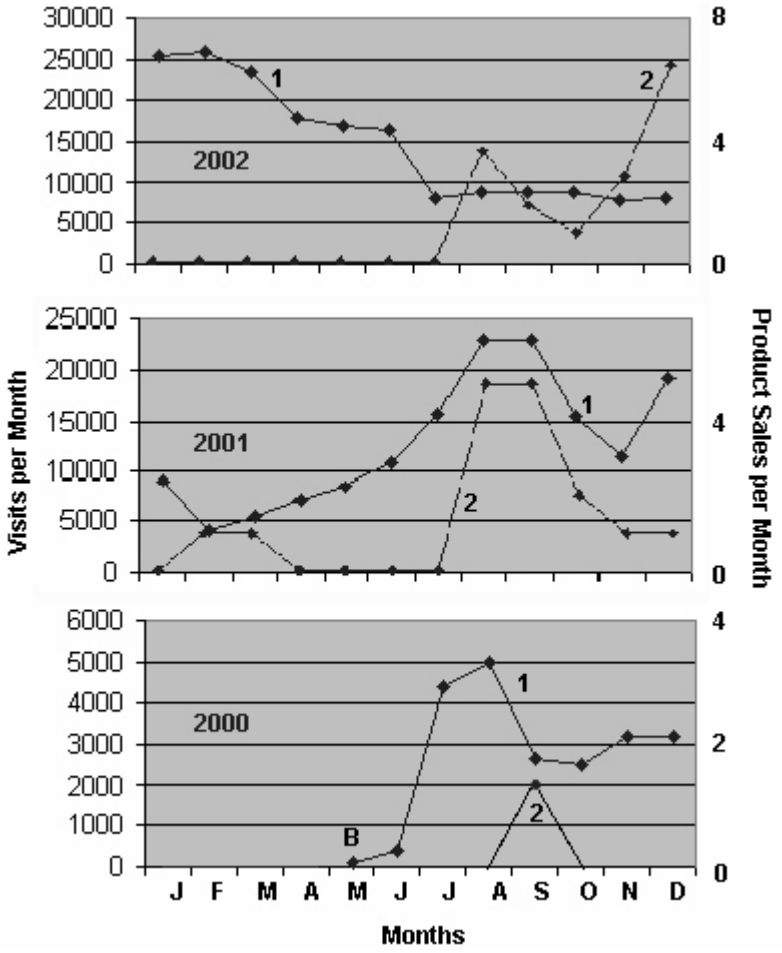


***Figure 4A, B & C:*** *Monthly visits and monthly sales for Company A (figure 4A), Company B (figure 4B) and Company C (figure 4C). In all cases curve 1 refers to visits whilst curve 2 refers to sales during the same period. B; beginning of operations. E; end of operations.*

The data from Figure 4 is summarised below in Table 3.

| Company | Year | Visits | Sales | Visits/Sale/ Year | Visits/Sale (av.) |
|---|---|---|---|---|---|
| A | 1997 | 9758 | 1 | 9758 | 9791 |
| | 1998 | 11712 | 2 | 5856 | |
| | 1999 | 2116207 | 240 | 8818 | |
| | 2000 | 3081136 | 290 | 10625 | |
| B | 1997 | 7359 | 0 | not applicable | 10595 |
| | 1998 | 39637 | 3 | 13212 | |
| | 1999 | 80531 | 9 | 8948 | |
| | 2000 | 41994 | 4 | 10499 (6 months) | |
| C | 2000 | 21400 | 1 | 21400 (8 months) | 10301 |
| | 2001 | 152868 | 16 | 9554 | |
| | 2002 | 175967 | 17 | 10351 | |

***Table 3:*** *Overview over annual visits and sales - (N.B.: Company B started operations in June 1997 and ceased operations in August 2000. Company C started operations in May 2000.).*

## Discussion

The Internet statistics software resets "visits" to zero at 00.01 o'clock on the first of each month. This is a clear source of systematic error because e.g. a visitor visiting a web site on the 27th and 28th of a month is counted as one, whereas the same person visiting e.g. on the 31st and 01st of the following month is counted twice. This factor is however acceptable because:

1. It is the *de facto* standard
2. It is the same in all 3 web sites studied
3. Sales statistics follow the same principle, i.e. returning customers are counted as plural sales.

A further minor source of error is the possibility of multiple customers counting as one product unit sale. This is possible in the case of Company A where one sale may represent one individual or more individuals (group travel) and in the case of Company B, where one sale may represent one individual or two individuals (partner matching). This was because the only statistics available concern the number of invoices sent.

**The effect of automated keyword submission on visit rates and product sales.**
As shown in Figure 1, submissions had a measurable effect on Company A's visit rates, and that this effect stopped after approximately 2 months. During this time Company A's sales remained constant (data not shown), i.e. the increased visit rates had no obvious sales effect. Clearly it could be argued that customers became aware of Company A's products during the submissions period (which was a seasonal "low period") and may later return to buy, but the short-lived nature of the submissions effect can equally well argue against this point of view.

Figure 2 showed that when Company C stopped using submissions software, visits slumped. However, within the same period, product sales rose to record levels, perhaps due to that the season from which the data is taken is also a "high period", being a traditional College entrance season.

The most likely conclusion from the above data is that, when provoked by automated keyword submissions, increased visit rates are due to HTTP requests from search engine spiders, crawlers and similar cataloguing software systems. There is no evidence from this study that such submissions-provoked increases resulted in higher levels of product sales.

**The effect of bulk e-mail on visit rates and product sales.**
Company C sends annual e-mail messages detailing "best student of the year", policy changes, new courses etc, to those who have expressed interest, and to past/present students. Company C's e-mailing database contains under 4000 addresses, thus the two annual bulk e-mailing actions (in 2001 and 2002) can be estimated to have resulted in between 6600 and 8000 successful mails. Despite the relatively high rate of interest expressed, measured as post-mailing web site visits, careful comparison with the e-mailing database with new customers reveals that no e-mail recipient has ever subsequently bought a product. The most likely explanation is that those customers who are on the past/present students list, have simply had their needs fulfilled. That those who had expressed interest (and even returned to the web site) still failed to buy, is probably best expressed as "if they were not convinced enough to purchase upon their first visit, then it will be unlikely that they will be sufficiently convinced upon their second visit".

Figure 3 (data from Company C) also illustrates the "week wave" phenomenon of visitor behaviour, with marked troughs in visit rates at weekends. This common phenomenon is often attributed to visitors preferring to use the Internet from their place of work, instead of using their domestic line (if any). This interpretation is supported by statistics from Company A, where visit rates peak during lunch break times, 12.00 to 14.00 (data not shown).

Company A sends highly crafted e-mail messages on a quarterly basis to those who have expressly subscribed to the service. The mails contain 1-2 paragraphs of details of special- and last minute offers, together with a direct link to the relevant HTML page and the e-mail and telephone number of the appropriate employee. Despite this high degree of precision, Company A reveals that the sales response is only one per 21870 successful e-mails. This apparent extremely low sales efficiency may, however, be slightly misleading. Jayawardhena, *et al* (2003) reported that "the outcomes of purchase intentions did not necessarily correlate with consumer segmentation according to purchase orientations", and thus it is possible that e-mail recipients were stimulated to buy quite different products. Such purchases would not have been reported in this business channel.

World-wide, "spam" accounted for an estimated 10-15% of all e-mail messages sent in 1999 (Wood, 1999), an estimate which has risen to about 90% of all e-mail

messages sent in 2003 (Mellor, 2003). Spam often contains unfocussed marketing material that the recipient has not requested and normally finds irrelevant. Often, spam originates with unscrupulous persons who typically use a type of software called a "ripper". This software crawls the Internet in a random fashion and extracts any e-mail addresses found in HTML files. Lists of e-mail addresses found are subsequently compiled and then sold to the naive and unwary, who use it to send their marketing offers. The largest of such lists may contain up to 5000000 e-mail addresses. How successful is spam? The answer can be no more than a guess; if Company A uses highly focussed and expressly requested material and finds only one customer in approx. 22000 successful e-mails, then the sales rate connected to spam-advertising may be 100 times smaller, roughly one sale per 2.2 million e-mails. The data presented here may thus explain why spam is so prevalent; the sales rate resulting from spam-advertising is so tiny that the lists have to be used again and again in order to get any return at all on outlay. Indeed, with an optimistic profit of $10 per sale, spam senders must therefore send 7.3 million e-mails per day to make $1000 a month before tax.

**The correlation between Internet visits and product sales**

This study uses long-term (annual) statistics because monthly visit rates vary considerably with market vagrancy. Such fluctuations are exemplified here by e.g. Company A, who without further use of automatic submission software, experienced a 100-fold increase in visits between spring 1999 and spring 2000. There are probably various reasons for this, and the similar fluctuations experienced by Companies B and C. Most obvious are local factors, including the various marketing campaigns carried out by these companies at diverse points in time. However these also occurred against a backdrop of external factors, including the general increased popularity of the Internet at this time (Sexton, *et al*, 2002). The subsequent decline during 2000 may have been connected to the NASDAQ crash of 2000 (as hypothesised by the UCLA Center for Communications Policy, 2001). This manifests itself here in a failure of visit rates to maintain the same growth rate after Company A's seasonal summer low (despite both visits and sales being somewhat higher in 2000 than in 1999). Similar factors may have played a part in Company B stopping trading in 2000, and in Company C's slow start in 2000.

Not only visits, but also sales may likewise be subject to uncontrollable external factors, e.g. Company C experienced a total lack of student enrolments from the Middle East for 10 months following the September 11$^{th}$ incident in 2001.

**Results from short-term analysis**

Figure 4 shows that more visits will not automatically result in proportionally higher revenues. For example; if short-term (two-month) timeframes are taken, then, in 1999, Company B experienced constant sales for increasing visit rates. In 2002 Company C experienced low sales for high visit rates (Jan.- March), but increasing sales for decreasing visit rates in Oct.- Dec.

In Aug. - Dec. 1999, Company A experienced increasing sales for increasing visit rates but decreasing sales for high visit rates in spring 2000, ending the year with constant sales rates for decreasing, constant and, lastly, slightly increasing visit rates.

Thus at any one time, wildly differing revenues can result from increasing, stable or declining visit rates, and e.g. stable revenues from wildly differing visit rates. Indeed, the only constant appears to be that zero hits on a web site will result in no revenues for the owner.

**Results from long-term analysis**

However Table 3 shows that on long-term average there are approximately (but consistently) around 10000 times more visits than customers. From Table 3, column 6 (Visits/Sale (av.)), the overall average can be calculated to be 10229, with the maximum variance being 4.28%. In the raw data (Table 3, column 5 - Visits/Sale/Year), the single largest variances are Company A in 1998 (5856, which corresponds to -42.75%, but the two sales in 1998 are perhaps not representative) and Company C in 2000 (21400, which corresponds to +209%, but one sale in seven months is perhaps not representative). Otherwise the range of variance in the raw data is quite narrow, falling between -13.8% and +29.2%.

This high degree of correlation points to the existence of an underlying unifying factor, because the customers are otherwise so diverse in nationality, interests etc. between the three web sites, and that the companies involved use different marketing methods and strategies at different times, have different seasonal selling curves, etc. The most reasonable explanation is that customers are few, and that their contribution to total visits is "swamped" by a "background" rate, which, on average, is approx. 10000 times greater.

Visits (all HTTP requests) can be divided into 4 general categories:

1. Those originating from customers (i.e. those purchasing a product),
2. Those originating from potential customers (e.g. those clicking on a link to the web site, even if they immediately regret this and click on the browser "back" button).
3. Directed requests from machines (e.g. cataloguing software from indexing engines, often called "spiders", "crawlers" or "bots", software checking for broken links, monitoring software etc.).
4. General machine communication (e.g. the routine background requests which enable the Internet to function).

Automatic machine requests (category 4) are responsible for maintaining inter-server communication and the structure of the Internet. This type of traffic accounts for around 50% of HTTP traffic on the Internet (Mellor, 2003). Such machine requests, as well as category 3 cataloguing software, and other software which e.g. checks for broken links, may well follow links contained in banner advertising, thus giving a misleading impression that banners are more popular than they in reality are. Banner hosts that are crawled often by cataloguing software may thus appear to be better sources of visitors than they, in reality, are.

It may be possible to produce software that can distinguish between the first two categories and the latter two. However until such software is available, the amount of visits attributable to each category remains a matter of interpretation.

Thus Figure 4 and Table 3 are interpreted as showing that visits from the first category (customers) are "swamped" by a very large "background" due to visits from the latter three categories. Indeed, because of short-term market vagrancy, this "background" can be considered to be variable and random (or otherwise influenced by uncontrollable external forces). Because the "background" is so very much higher than the customer visit rate, then it hardly changes the significance if the background at any particular point in time is e.g. only 9000 times higher, or e.g. 11000 times higher.

Figures 1, 2 & 3 show that automated submissions simply temporarily increase the "background". This is because automated submissions attract machine-made HTTP requests (category 3 requests), who's sales potential are zero. The success of bulk e-mailing in attracting customers seems to be surprisingly low, even highly focussed e-mailings result in only one sale in approx. 22000 e-mails, so bulk e-mailing appears to attract few category 1 and many category 2 requests. So although one may be tempted to say that in the long-term, there occurs one sale per approx. 10000 visits, using these tools to e.g. double visits, will by no means necessarily double sales.

Obviously the value of a click for advertisers varies with market differentiation (e.g. societal, demographic, seasonal and other) factors, as well as product pricing and other, e.g. "added value", aspects of their service. However the data presented here suggests a general rule-of-thumb, namely, that in the long-term, the value of a click (or click-through) should not exceed the average profit on an average product divided by 10000. Clearly, mere "exposures" will have a lower value.

## References

Jayawardhena, C., Wright, L. T. & Masterson, R. (2003). An investigation of online consumer purchasing. Qualitative Marketing Research, 6, 58-65.

Mellor, R. B. (2003). The Web Managers Handbook, Copenhagen, Globe. 265 pages.

Novak, T. P. & Hoffman, D. L. (2000). Advertising pricing models for the world wide web. In D. Hurley, B. Kahin & H. Varian (Eds.), Internet publishing and beyond: The economics of digital information and intellectual property. MIT Press, Cambridge. 1-21.

Prasad, A., Mahajan, V. & Bronnenberg, B. (2003). Advertising versus pay-per-view in electronic media. International Journal of Research in Marketing, 20, 13-30.

Sexton, R. S., Johnson, R. A. & Hignite, M. A. (2002). Predicting Internet/e-commerce use. Internet Research, 12, 402-410.

UCLA Center for Communications Policy (2001). Surveying the digital future year 2. Available as http://ccp.ucla.edu/pdf/ucla-internet-report-2001.pdf

*Mellor - The value of an Internet click.*

Wood, D. (1999). Programming internet email. O'Reilly, Sebastopol, 362 pages

Wu, Y-H & Chen, A. L. P. (2002). Prediction of web page accesses by proxy server log. World Wide Web, 5, 67-88