

Applying Quantitative Marketing Techniques to the Internet

by

Alan L. Montgomery
Carnegie Mellon University
Graduate School of Industrial Administration
5000 Forbes Ave., Posner Hall 255A
Pittsburgh, PA 15213-3890

email: alan.montgomery@cmu.edu

July 2000

Abstract:

Quantitative models have proved valuable in predicting consumer behavior in the offline world. These same techniques can be adapted to predict online actions. The use of diffusion models provides a firm foundation to implement and forecast viral marketing strategies. Choice models can predict purchases at online stores and shopbots. Hierarchical Bayesian models provide a framework to implement versioning and price segmentation strategies. Bayesian updating is a natural tool for profiling users with clickstream data. I illustrate these four modeling techniques and discuss their potential for solving Internet marketing problems.

Keywords: Choice Models, Clickstream Data, Hierarchical Bayesian Models, Internet Marketing, Versioning, Viral Marketing, World Wide Web

The novelty of the Internet has led many to believe that new techniques and methods must be developed to take advantage of this technology. This is not the case. Many methods used in promotion, pricing, forecasting, and control in marketing research for the past 30 years have been or could be applied to the Internet. My treatment is not meant to be a comprehensive survey, but to illustrate how quantitative techniques can be applied successfully to Internet marketing problems.

The introduction of many personalized selling and customer relationship methods on the Internet are really reintroductions of concepts that fell out of favor over the past 50 years as merchandisers focused on mass marketing to exploit the economies of scale of national and regional operations. Blattberg and Deighton [1991] laid out a good framework for Internet marketing in their article on interactive marketing, which they wrote several years before the popularization of the Internet. A key notion of interactive marketing is that consumers are addressable. Instead of treating consumers as homogeneous masses, managers can identify and communicate with each individual.

Advances in both technology and manufacturing make interactive marketing possible. Manufacturers can now create custom products en masse [Wind and Rangaswamy 1999]. For example, Cannondale manufactures bicycles with over 8 million different frame and color variations. Customers can order bicycles that are built to their specifications. Previously, it would not have been profitable to serve a niche containing an individual customer. The challenge to the producer is to take over functions traditionally handled by channel members, such as retailers and wholesalers, to satisfy a consumer's needs.

To implement a successful interactive marketing strategy, a firm must use information effectively. Many of the notions of interactive marketing have been refined in recent years by researchers studying direct marketing [Allenby, Leone, and Jen 1999; Bitran and Mondschein 1996; Bult and Wansbeek 1995; DeSarbo and Ramaswamy 1994; Gonul and Shi 1998]. Information is critical to identifying, differentiating, and interacting with customers. The problem is

extracting insights from huge databases. Quantitative models can transform this information into profitable decisions.

Promotion using Viral Marketing

The introduction of Hotmail is a classic internet start-up success story. In 1996, Sabeer Bhatia and Jack Smith pioneered a new product category—free Web-based email. Originally they were interested in developing database software called JavaSoft. The two were working for another company at the time and were reluctant to communicate through their employer’s e-mail system, fearing that others might discover their business plans. They came up with the idea of creating an off-site mail server that could be accessed through a web browser with email text displayed in HTML-formatted documents. Since email is basically a database, they could adapt the computer code already written for JavaSoft. Eventually they formed Hotmail and offered basic email services for free. Hotmail generated revenue by selling banner advertisements and offering advanced services for a fee.

A compelling aspect of the Hotmail story is its phenomenal growth rate. Bhatia and Smith launched the company in July 1996 and by the end of the month, it had more than 20,000 subscribers. By September 1996, it had more than 100,000 subscribers. It passed the one-million-subscriber mark in January 1997. One and a half years after it went online, Hotmail had signed up over 12 million subscribers (Figure 1).

How did Hotmail accomplish this phenomenal growth rate? The traditional answer would be lots of advertising. Yet practically all the initial seed money from the venture capitalists was spent on hardware and personnel, leaving very little for traditional advertising. An alternative promotional campaign was suggested by one of Hotmail’s venture capitalists, who advocated placing an advertising message, “Get your free email at Hotmail”, with a hyper-link back to Hotmail at the bottom of every out-going email. The recipient would know that the email message came via Hotmail and its use would be an implicit endorsement of the service. The more

email messages a Hotmail subscriber sent, the more advertising would be distributed. Every recipient would become a potential new subscriber. Initially the founders were against this type of promotion, thinking that their users would be repelled by any advertising, but they had few other advertising options.

Hotmail's subscriber base grew faster than that of any other on-line company, and Hotmail is now the largest email provider in the world. A traditional print publication would be considered extremely successful if it could garner a subscriber base of 100,000 in several years. AT&T's Universal Mastercard is held up as a triumphant example of targeted marketing, because it recruited more than a million new customers in a year. Many Internet companies spend 25% of revenues on traditional advertising (for example, television, radio, and magazines). Even Juno, which offered an Internet email service, spent \$20 million on advertising to recruit a third as many subscribers as Hotmail. Hotmail spent only \$500,000 on advertising over two years and got 12 million subscribers during this time.

The key to Hotmail's success was viral marketing, a term coined in 1997 by Steve Jurvetson and Tim Draper, Hotmail's venture capitalists, to denote a type of marketing that infects its customers with an advertising message, which passes from one customer to the next like a rampant flu virus. Some Internet marketers now hope to replicate Hotmail's success by letting their customers perform their marketing functions for them.

Many think viral marketing is a creation of the Internet, but it isn't. Viral marketing has been described in the marketing literature for more than 30 years but termed word-of-mouth or the diffusion of innovation. Bass [1969] used these types of models to describe the sales of televisions, clothes dryers, dishwashers, refrigerators, and other consumer durables. Other researchers used them to describe the sales of a wide variety of products, including pharmaceuticals, hi-tech products, agricultural fertilizers, and medical equipment. Mahajan, Muller, and Bass [1991, 1995] provide excellent reviews of diffusion models and their applications to marketing. Predating the use of diffusion models in marketing was research on the transmission of news and epi-

demics by Taga and Isii [1959]. Bartholomew [1982] provides an extensive treatment of these types of models in the social sciences.

The Bass model [1969] predicts the number of users that will adopt an innovation. It assumes that the total population has m individuals that will eventually adopt the product or service, and defines the number of individuals who have adopted at time t as $n(t)$. This means at week t , $m-n(t)$ users have not adopted but may adopt. Bass assumes that a constant proportion (α) of these potential adopters are converted because of the effects of advertising. Also, there are $n(t)(m-n(t))$ potential pairs of adopters to nonadopters, and Bass assumes a constant proportion (β) of them will convert. Formally, this yields the following differential equation:

$$n'(t) = \alpha(m - n(t)) + \beta n(t)(m - n(t)) \quad (1)$$

The solution to this differential equation is:

$$n(t) = \frac{m(e^{\alpha t + \beta m t} - 1)}{e^{\alpha t + \beta m t} + \frac{\beta m}{\alpha}} \quad (2)$$

This model expresses the total number of adopters at any time as a function of time, total market potential (m), and parameters that measure the effects of advertising (α) and word-of-mouth (β). An even better model would incorporate the growth of the Internet population, the opening of more than one Hotmail account, and subscribers who no longer using their Hotmail account.

Diffusion models can fit the growth of Hotmail's subscriber base very well. The model given in equation (2) is fitted to Hotmail's actual user data for the first year. The fitted subscriber predictions are denoted by a solid line (Figure 1). The parameters m , α , and β are estimated using statistical methods. I obtained estimates of $\alpha=.0012$, $\beta=.008$, and $m=9,670,000$ with t measured in weeks. The coefficient which measures word-of-mouth effects (β) is much larger than that which measures advertising (α). There is a remarkable agreement between the actual subscriber base and the model's predictions.

There are a couple of problems that often arise when estimating diffusion models. First, other quadratic and exponential time trends can also fit the data very well, although the predictions of future growth from these other models can be quite different. Consider that an exponential time trend would predict the subscriber base will increase forever. The predictions for Hotmail's diffusion model indicate rapid growth to continue for another year before starting to level off. These predictions form a characteristic "S"-shaped curve found in diffusion models. Second, parameter estimates can be unstable if they are based on only a few observations. This makes forecasting unreliable without additional information about the diffusion process. Otherwise, the forecasts will not be accurate until a good portion of the process has occurred. These problems have been addressed extensively by researchers (for a review see Mahajan, Muller, and Bass [1991, 1995]).

Several conditions are necessary to get diffusion processes to work. First, a population that communicates freely is needed. Email provides a direct channel for individuals to communicate regardless of where they live. Verbal communication may require physical contact. If people are geographically dispersed and individuals do not travel often, diffusion marketing may not work in reaching the whole population quickly. For example, a diffusion process that starts in India may diffuse quickly in India, but may not affect the American market for some time, because of the physical separation between these markets.

The speed of the diffusion process is related to how rapidly and often users communicate with non-users. The success of Hotmail's diffusion process was due to the high frequency of email use. The Internet provides many other vectors besides email for rapid communication, such as Usenet, chat, and web discussion groups. Each can be an effective diffusion tool. Small specialized scientific communities which communicate only through physical contact at conferences a few times per year might be poor candidates for viral marketing (i.e., small values of β). In contrast, highly integrated groups of consumers, such as teenagers at the same high school, are excellent candidates (i.e., large values of β). Napster, an online community that distributes musi-

cal recordings, has exploited diffusion effects in its marketing program by using the social networks of college students. Viral marketing will work in scientific communities, but the diffusion process will be much slower. To build a substantial user base quickly in a small and scattered scientific community, traditional promotion and advertising techniques may be preferable.

Second, users must convey the fact that they are now users. In the case of a flu virus, this is automatic, and for many luxury products, such as BMW automobile owners, may happily volunteer that they are users. That Hotmail subscribers automatically identified themselves contributed to Hotmail's success with viral marketing. If the Hotmail hypertext message were not at the bottom of every email, recipients would have had no automatic means of connecting to Hotmail.

Third, the value of the product must be fairly high relative to its cost to encourage consumers to adopt. The email service Hotmail offered was free but valuable to consumers, creating an economic incentive to subscribe. Hotmail would have taken much longer to diffuse if it had cost \$20 per year. Onsale, an online retailer and auctioneer of surplus computer merchandise, is another example of a firm that used diffusion effects to rapidly expand its subscriber base, although it did not provide an integrated communication channel as did Hotmail. The value Onsale offered to users was the excitement of auctions and potential bargains which created positive word-of-mouth effects.

These three conditions make diffusion effects work online and offline. Online environments are attractive because these conditions are easier to establish. Consumers are more willing to talk to each other in online chat rooms than in a retail store at the shopping mall. However, practitioners should remember that diffusion effects can also have negative consequences. When Intel released the Pentium chip and later discovered that it had a mathematical error, the word-of-mouth effects were strongly negative and included Pentium jokes told at Intel's expense. They were neutralized only after Intel replaced the defective chips.

An extensive body of research in diffusion modeling exists that researchers can draw upon to explain viral marketing. Complex formulations of diffusion models take into account the

growth of the population, that some users influence nonusers more than others, and the reversion of users to nonusers. Many advances have been made in recent years to improve estimation techniques [Van den Bulte and Lilien 1997], to incorporate marketing-mix variables, such as price and advertising, into the diffusion process [Horsky 1990], and to use past products to predict diffusion patterns of similar products [Sultan, Farley, and Lehmann 1990]. However, the Internet provides some unique challenges for diffusion modeling that may yield new models [Rangaswamy and Gupta 1999]. For example, the Internet allows the collection of individual-level information, such as social networks (e.g., who does a customer know), interest in a product or service, and past purchases that could be used to forecast individual level usage. In comparison, the Bass model uses only aggregate level information.

Measuring the Value of Product Components

Typically marketers think of products being composed of components or attributes. An important attribute is brand name. Advertising is usually intended to increase consumers' affinity for a specific brand. All portals use their brand names to try to extend users' positive feelings for the brand to other products and to companies that are partners. For example, Yahoo! has Yahoo! Shopping, Yahoo! Auctions, Yahoo! Careers, and Yahoo! Photos. Partnering, which is widespread on the Internet, is related to umbrella branding and brand extensions [Aaker and Joachimsthaler 2000].

If a firm is to use branding effectively, it must measure the value of its brand, which marketers call brand equity. Brand equity can be measured directly through survey devices or indirectly through choice models and changes in stock prices. A popular technique that enables a firm to measure brand equity and the value of each of the components that make up a product is conjoint analysis [Green and Wind 1975], also known as trade-off analysis. Using conjoint analysis, an analyst can infer respondents' valuations of attributes from the preferences they express for different products [Green and Srinivasan 1990]. Wind et al. [1989] describe its use in devel-

oping Courtyard by Marriott, Dreze and Zufryden [1997] use it for web site design, and Dahan and Srinivasan [1998] discuss its use for testing product concepts on the Internet.

In conjoint analysis, the researcher presents consumers with alternative products that differ in the levels of their attributes and asks them to rate the products. The researcher does not ask consumers to rate the attributes directly, because the most natural task is evaluating the product as a bundle of attributes. The analyst then uses statistical techniques to infer the value of each product component from the overall ratings. Frequently in conjoint analysis a user is asked to choose their most favored item from a set of choices. Statistically these choices are predicted using a multinomial logit model [McFadden 1980].

Imagine trying to determine the value attached to the Amazon brand in a book-selection process. How much will it increase the value of a book beyond the value it would gain from a lesser known brand, such as 1Bookstreet? Brynjolfsson and Smith [2000] consider just such a conjoint task in their study of consumers' book purchases through DealPilot, which has been re-named EvenBetter. DealPilot is a shopbot or a computer agent that searches over 30 different stores for a given book, and lists the prices at the various stores in ascending order. The shopbot also retrieves other information such as delivery time and costs, tax, and shipping method. The range of prices for an average DealPilot search is \$12, and no bookseller consistently offers the lowest price. Clearly, searching multiple vendors using a shopbot should benefit a consumer.

A multinomial logit model can predict which book or alternative will be selected. It assumes that the utility for the i th alternative has a deterministic and random component:

$$u_i = \sum_{j=1}^M \beta_j x_{ij} + \epsilon_i, \quad (3)$$

The deterministic component is a weighted sum of the M attributes (x_{ij}) which comprise a product and a random error (ϵ_i). The cumulative distribution function of these errors is assumed to be an extreme value distribution, $\Pr[\epsilon_i < \epsilon] = \exp\{-e^{-\epsilon}\}$. If a consumer maximizes utility, the probabil-

ity of choosing the i th alternative is:

$$Pr\{i \text{ chosen}\} = \frac{\exp\left\{\sum_{j=1}^M \beta_j x_{ij}\right\}}{\sum_{k=1}^N \exp\left\{\sum_{j=1}^M \beta_j x_{kj}\right\}} \quad (4)$$

In this shopbot example, the attributes of each book are item price, shipping cost, tax, expected number of days until delivery, and an indicator for brand if it is sold by one of the three large booksellers (Amazon, Borders, and Barnes & Noble). Brynjolfsson and Smith [2000] use actual purchase data from a panel from over 20,000 unique visitors during late 1999. These researchers estimated weights using maximum likelihood estimation (Table 1). These estimates imply that consumers are more than twice as sensitive to a dollar spent on taxes as they are to a dollar spent on the item. Also, the value of Amazon's brand name can be imputed from the model parameters to be \$2.46 more ($=.477/-.194$) than the value of an unbranded bookstore. All three major booksellers have substantial brand equity.

The attributes, utility of the alternatives, and choice probabilities which correspond with a search are listed in Table 2. The choice probabilities are derived from the attributes and their associated weights; essentially the probability that a customer chooses a product is proportional to its attractiveness (or utility) compared to the sum of attractiveness over all alternatives in the set. For example, the exponential of utility for the first alternative is $\exp\{-.194 \cdot 29.56 - .369 \cdot 2.50 - .019 \cdot 14\} = \exp\{-7.36\} = .000639$. To compute the probability that this alternative is chosen, take the ratio of this value to the sum of the exponential utility of all alternatives offered. In this case the probability of choosing the first alternative is 26 percent ($=.000639/.002503$). The most likely choice would be alternative #3 from Barnesandnoble.com at a price of \$37.19. The probability of purchasing this book is 48 percent.

The marketing literature includes a great deal of research concerning choice modeling. A

seminal paper in marketing by Guadagni and Little [1983] considered the problem of predicting purchase choice for an individual shopper at a supermarket. They estimated the choice probabilities using previous purchases at the store. Researchers have made many advances in choice modeling over the years, particularly in estimating the variation in differences across consumers [Kamakura and Russell 1989; Chintagunta, Jain, and Vilcassim 1991; Rossi and Allenby 1993]. Telang, Mukhopadhyay and Wilcox [1999] describe an application of choice modeling to search engine selection on the Internet. Degeratu, Rangaswamy, and Wu [1999] consider the use of choice modeling of online grocery purchases.

Pricing using Versioning

Pricing is a key component in any marketing strategy, since it determines how much of the value that the company creates for the customer it retains and how much the customer enjoys. Versioning refers to developing many products that are related through a common design (for example a line of VCRs or an encyclopedia with print, CD, and web versions) but they are sold at different prices. Each of the individual products has unique features that appeal to certain users who may be willing to pay extra for them. Shapiro and Varian [1998] provide a good discussion of the application of versioning to information goods. Conventionally researchers in economics and marketing refer to versioning as price segmentation or price discrimination [Nagle 1984]. In the long theoretical history of price discrimination, versioning is the latest twist as applied to information products. However, the extreme properties of information products (that is high fixed costs and low variable costs) can make versioning the difference between a profitable marketing strategy and a failure.

The explosion of computerized scanners at the point-of-sale (POS) in retail environments has resulted in the development of many tools for price segmentation. New models allow retailers to track prices paid by an individual consumer and respond with customized coupons and promotional offers. Rossi, McCulloch, and Allenby [1996] consider the use of a multinomial

probit model to determine the price sensitivity of an individual based upon previous purchases from a single product category. Formally, the model is defined as:

$$u_{ijt} = \sum_{k=1}^M \beta_{ijk} x_{ijk} + \epsilon_{ijt}, \quad (5)$$

The u_{ijt} defines the utility, which is not observed, of the i th person during time t for alternative j . In total there are N choices from which the consumer can choose. The M covariates (x_{ijk}) provide information about the attributes of each choice such as price or brand name. The errors are assumed to follow a multivariate normal model, i.e., $[\epsilon_{i1} \dots \epsilon_{iN}]' \sim N(0, \Sigma)$. The analyst observes one choice for the i th person at time t , and assumes that the choice corresponds to the alternative with the maximum utility, i.e., $\text{argmax}(u_{i1} \dots u_{iN})$.

The multinomial probit model is similar in construction to the multinomial logit model. The primary difference is that the errors of equation (5) are assumed to follow a multivariate normal distribution instead of an extreme-value distribution as in the multinomial logit model. Another important feature of the model proposed by Rossi, McCulloch, and Allenby [1996] is that the parameters are allowed to vary across consumers. The parameter differences across consumers are related to a matrix of demographic characteristics for individual i (Z_i):

$$\beta_{ij} \sim N(Z_i \gamma_j, V_{\beta}), \text{ where } \beta_{ij} = [\beta_{ij1} \dots \beta_{ijM}]' \quad (6)$$

This incorporates idiosyncratic behavior, since the coefficients for each individual varies around the mean. If $V_{\beta} \rightarrow 0$ then the coefficients across the individuals are equal. This variation allows for some users to be more price sensitive, while others may have a stronger preference for a particular type of product (e.g., a can of tuna packed in water as opposed to oil). This technique forms a hierarchy of models, and is usually referred to as a hierarchical Bayesian model or a random coefficient model. For another application of hierarchical Bayesian models in a regression context see Montgomery [1997].

Rossi, McCulloch, and Allenby [1996] consider the use of this model to determine

whether to offer a coupon and its value to an individual at check-out. In a blanket coupon drop, all consumers receive a coupon of the same value, say 10 cents, valid on their next purchase. A coupon drop using information solely based on demographics increased profits over a blanket coupon drop by 10 percent. Determining a coupon drop using only information from the last purchase increased the relative profit gain to 60 percent. Using the entire purchase history resulted in a 90 percent increase. The use of all information from the purchase history and demographics resulted in a 160 percent increase in profits. In this scenario coupons range in value up to 30 cents, with 1/6 of the consumers getting no coupon at all. These results show that the largest profit gains come, not from the knowledge of demographic characteristics about a household, but instead from knowledge about previous purchases. Managers need tools, such as probit models, to fully realize the value of the information contained in these purchase histories.

These techniques can be applied directly to the Internet. For example, Barnesandnoble.com could decide to offer a consumer no coupon, free shipping, \$5 off, or \$10 off on their next purchase based upon their previous purchase history. However, instead of blindly making the same offer to all consumers (like a blanket coupon drop through email), it could decide to selectively make this offer to consumers who are least likely to make purchases and who are price sensitive. It can infer price sensitivity by analyzing previous purchases and eliminate discounts to consumers who would purchase with or without a discount.

Another technique for price segmentation that has been suggested for information products is bundling [Bakos and Brynjolfsson 1999]. Bundling focuses on the characteristics of the products not those of the consumers. The idea behind bundling is to reduce the variation in the amount that individuals are willing to pay for an individual product by combining it with another product. The bundle will have less variation than its components. For example, consider a firm selling two consulting reports, one on retail sales and the other on unemployment. A firm in the retail sector places a higher value on the retail report (\$1,000) than on the unemployment report (\$600). However, the preference of a firm in the industrial sector is reversed. If the firm prices

each report at \$600, it obtains revenues of \$1,200 per report since both firms purchase it. Pricing each report at \$1,000 yields only \$1,000 in revenues per report because only one firm will buy each report. Therefore, \$600 appears to be the best price for each report. However, if the consulting firm bundles the two reports and prices the combined report at \$1,600, both firms will buy it. The consulting firm obtains total revenues of \$3,200, as opposed to the \$2,400 it would obtain by pricing them separately.

The Quantitative Challenges from Clickstream Data

A common thread through all techniques discussed is the need for data. Fortunately, a natural byproduct of users accessing WWW pages is a dataset that contains the sequence of URLs they visited, how long they viewed them, and at what time. This dataset is called the clickstream. To maximize its potential, managers can merge the clickstream with demographic and purchase information.

Three potential sources exist for collecting clickstream data: (1) The host server (the computer of the site being visited) keeps a record of visits, usually called a server log. As a user requests a page, the server records identifying information (IP address, previous URL visited, and browser type) in the server log. (2) A third party can capture information about web requests. For example, if a user contacts an Internet Service Provider (ISP) or Commercial On-line Service (COS), such as AOL, it can record any requests the user makes as it relays them to the requested server. Because many ISPs and COSs cache their users' requested pages, they do not pass all requests on to the server; instead, they serve many pages from local cache archives to speed up responses to user requests. Unfortunately, this means that server logs contain only a subset of the viewings that occur. Dreze and Zufryden [1998] discuss some of the challenges of using server log data to measure advertising effectiveness.

(3) A final—and perhaps the most reliable—source of clickstream data is a program installed on the computer where the browser program is running that can “watch” the user and re-

cord the URLs of each page viewed in the browser window as well as other application programs that the user is running. It records the actual pages viewed, and thus avoids the problem of cached requests. Such a program can also record how long windows are active. The drawback is that the analyst must choose the individuals and obtain their consent to participate in such a panel. Generally web users are randomly sampled to construct a representative sample. The information from this sample can be projected to the national population using statistical inference. The largest provider of such information is Media Metrix [Coffey 1999].

The clickstream of an actual user session as collected by Media Metrix (Table 3) shows that the user frequently views the same page repeatedly and sometimes pauses to do other tasks between page views (for example run other applications or watch television). Only five of the 12 viewings the user requested could generate a “hit” to the server. This illustrates the advantage of collecting data at a user's machine and not from a host site since it includes all requests, eliminating a potential source of bias.

Information about where and how frequently users access web sites is used for various tasks. Marketers use such information to target banner ads. For example, users who often visit business sites may receive targeted banner ads for financial services even while browsing at non-business sites. Web managers may use this information to understand consumer behavior at their site. Additionally, it can be used to compare competing web sites. Members of the financial community use such information to value dot com companies. Analysts use clickstream information to track trends in a particular site or within a general community. Financial analysts find this type of intelligence useful for assessing the values of companies because many traditional accounting and finance measures can be poor predictors of firms' values.

Another use of clickstream data is to profile visitors to a web site. Identifying characteristics about visitors to a site is an important precept of personalization. One way to find out characteristics of visitors is to ask them to fill out a survey. However, not everyone is willing to fill them out, creating what is known in marketing research as a self-selection bias. The information

may be inaccurate as well, for example visitors may give invalid mailing addresses to protect their privacy or inaccurately report incomes to inflate their egos. Also, completing a survey takes time, and the effort required may severely skew the type of individuals that complete it and the results.

An alternative way to profile users is with clickstream data. The demographic profiles of sites reported by companies like Media Metrix can be used to determine what type of individuals visit a site (Table 4). For example, Media Metrix reports that 66 percent of visitors to *ivillage.com* are female. Even without knowing anything about a user except that they visit *ivillage.com*, the odds are two to one that a visitor is female. This is quite reasonable because *ivillage.com* offers content geared toward issues of primary concern to women. Some gaming sites appeal primarily to teenage boys, and sports sites may draw predominately adult men. On the other hand, such portals as Yahoo! and Excite draw audiences that are fairly representative of the web as a whole.

Media Metrix can identify demographic characteristics of visitors using information provided to them by panelists. However, simply a knowledge of the web sites visited by a user and profiles of these web sites (that is, the demographic characteristics of a sample of users) is enough to make a good prediction about a visitor's demographics. For example, suppose we wish to predict whether a user is a woman. In general, about 45 percent of web users are female. Therefore without knowing what sites a person visited one would guess that there is a 45 percent probability of being female and a 55 percent probability of being male. If forced to choose, one would guess the user to be male, but this would be an inaccurate guess since the odds are almost equal.

However, if one knows that this individual visited *ivillage.com*, whose visitors are 66 percent women, the hypothesis that this user is female can be improved. This is a Bayesian hypothesis updating problem, and an analyst could apply Bayes formula to recompute the probability that the user was female using this new information:

$$\hat{p} = \frac{\hat{p} \cdot p}{\hat{p} \cdot p + (1 - \hat{p})(1 - p)} = \frac{.45 \cdot .66}{.45 \cdot .66 + .55 \cdot .34} = .62 \quad (7)$$

The original probability of being female is denoted by $p=.45$, and the new information we have is $p=.66$. The updated probability or posterior probability of our hypothesis is denoted by $\hat{p}=.62$. In other words, the probability this is a female user has increased from 45 percent to 62 percent. While most of the web sites visited by this individual indicate the user is most likely a female, some of the sites (aol.com, eplay.com, halcyon.com, lycos.com, and netradio.net) visited might point to the individual being male. However, based on information from all 22 sites the probability that the user is female is 99.97 percent (derived by updating the probability of being female for all sites listed in Table 4).

To assess the accuracy of this technique I applied it to actual usage information from a sample of 19,000 representative web users with one month of usage and known gender. There is a great deal of variation in users in this sample, such as some users visit only one site, while others may visit hundreds. If the model predicts more than an 80 percent probability that the user is male then the user is predicted to be male. Similar predictions are made for female users. There was enough information to classify 60 percent of users as either male or female. 81 percent of users classified as male are male, and 96 percent of users classified as female are female. The agreement between the predictions and actual accuracy validates the accuracy of these techniques. More advanced techniques that appropriately account for statistical dependence between web site visits can increase the accuracy of these predictions.

In this example, all of the web sites visited by a user were known, but this is not typical. However, many web advertising agencies, such as DoubleClick and Flycast, have such information for their advertising partners since they serve all their partners' banner ads. These partners give them wide coverage of the web. User profiling techniques allows them to accurately predict characteristics of their users without ever having to ask a user to fill out a survey.

Predicting the genders of an individual seems innocuous, but the same techniques can be applied to other more sensitive demographic variables, such as income (for example, does a user make more than \$75,000). Just as there are sites that provide valuable information about gender, there are sites that provide information about income (for example, millionaire.com, wsj.com, businessweek.com). While this example illustrates that these techniques can be used successfully, questions about privacy need to also be considered before using these techniques in practice.

The collection, processing, and use of clickstream data is not cost free, as some would claim. A major portal can collect 30 to 50 gigabytes of information from web accesses each day in a server log. Even a small site will generate several megabytes. To analyze these large datasets, managers need data mining and statistical techniques. These techniques generally require skill and their inaccurate applications can result in poor or misleading predictions about visitors and what they want to see. If, for example, a visitor is wrongly classified as male, the visitor could be shown messages that would not appeal to a female user, or even worse cause her to leave the site and never return. Therefore, to apply these techniques appropriately, one should take into account the value of personalization and the costs of misclassifying a visitor.

Conclusions

Marketing researchers have learned a great deal about interactive marketing over the last 30 years. The newness and excitement over the Internet leads many to overlook the fact that the ideas behind apparently novel web marketing applications have been considered previously. It is natural for a new communication medium like the Internet to develop its own language. These new terms are needed to convey a new application or differentiate it from old ones. For example, Iconocast (Dec 16, 1998) named 'viral marketing' the Internet marketing buzzword of the year. But some new terms can also lead to the misunderstanding that the ideas behind them are not well researched or understood. Viral marketing has a firm foundation in diffusion models. In

fact, many “new” Internet marketing concepts have a strong foundation in past academic marketing research.

The models discussed in this paper are not an exhaustive set of techniques nor do they cover all areas of marketing. They were chosen to illustrate how quantitative techniques can be used to solve practical Internet marketing problems. A wide variety of research in marketing has potentially strong applications to the Internet. Techniques that have not been mentioned in this paper include:

Autoregressive tobit models [Montgomery 1999] can use past usage patterns to forecast future usage. These models can handle periods of no usage, a common feature in this data.

Collaborative filtering matches consumers to others with similar interests using clustering techniques. These techniques are used in personalizing web sites. For example, Ungar and Foster [1998] describe a collaborative filtering system used by CDNow, an online music retailer, to recommend new artists to their customers through email using their previous purchases.

Computer agents [Chavez and Maes 1996] can be used to create marketplaces for buying and selling goods. Pazgal [1999] considers the impact of these software agents on Internet marketing.

Hazard models are an alternative to tobit models for modeling censored observations. In these models [Vilcassim and Jain 1991], the variable of interest is the time elapsed between observations instead of modeling observations using a fixed time interval (for example seconds or days).

Markov models have been applied to web-browsing behavior to consider the path visitors travel through a web site [Pirolli and Pitkow 1999]. One marketing application would be to relate path to price sensitivity (for example, if someone came to a page using a specified route should he or she receive a promotional offer?).

NBD models have been found useful for modeling repeat purchases [Gupta and Morrison 1991]. A potential application to the Internet would be modeling visitor frequencies at a web

site.

Learning models consider the time users spend online and assume that the rate of learning how to navigate a web site follows a power law. Johnson, Lohse, and Bellman [2000] fit these models to learning at many web sites. These models can provide an estimate of “stickiness”.

Trial-and-repeat models are used by consumer packaged goods manufacturers to decompose sales into initial trial and subsequent repeat purchases. Fader and Hardie [1999] forecast sales using a trial-and-repeat model with data from CDNow.

This list is incomplete and will expand as other success stories about quantitative marketing techniques are disclosed. There is a great knowledge base from which to draw in solving Internet marketing problems using quantitative models. Practitioners can apply and adapt this knowledge about quantitative modeling to Internet marketing problems.

Implementing these techniques can be challenging. First, the volume of data collected from the Internet can be huge. Second, information may not be in a quantitative form. For example, clickstream data often requires text processing techniques to translate qualitative information into quantitative forms that can be used in the models presented. Third, implementing these models online may require fast response time. Predictions from the models outlined in this paper are generally quick and efficient from a computational standpoint. However, estimating these models while a user is online may not be feasible due to the intense computation requirements. Finally, the techniques discussed require compatible privacy policies. Privacy is outside the scope of this discussion, but these policies can substantially impact the application and use of these models. In conclusion, successfully implementing quantitative techniques requires skill on the part of the practitioners, but also promises a big reward to those who can do it effectively.

Acknowledgments

I thank the editors and reviewers for their extensive suggestions and comments. They have been instrumental in shaping this paper. Also, I thank Media Metrix for their generous contribution of

data, although any opinions, findings, or recommendations are mine and do not necessarily reflect the views of Media Metrix.

URL's (alphabetized per underlined phrases in main text)

- 1Bookstreet (www.1bookstreet.com)
- Amazon (www.amazon.com)
- Barnesandnoble.com (www.barnesandnoble.com)
- Cannondale (www.cannondale.com)
- CDNow (www.cdnow.com)
- DoubleClick (www.doubleclick.com)
- EvenBetter (www.evenbetter.com)
- Flycast (www.flycast.com)
- Hotmail (www.hotmail.com)
- Iconocast (www.iconocast.com/issue/19981216.html)
- Juno (www.juno.com)
- Media Metrix (www.mediametrix.com)
- Napster (www.napster.com)
- Onsale (www.onsale.com)
- Pentium jokes (www-pcd.stanford.edu/cousins/pentium.html)
- Yahoo! (www.yahoo.com)

References

- Aaker, David A. and Joachimsthaler, Erich 2000, *Brand Leadership: Building Assets in the Information Society*, Free Press, New York.
- Allenby, Greg M. and Rossi, Peter E. 1993, "A Bayesian approach to estimating household parameters," *Journal of Marketing Research*, Vol. 30, pp. 171-182.

- Allenby, Greg M.; Leone, Robert P.; and Jen, L. 1999, "A Dynamic model of purchase timing with application to direct marketing," *Journal of the American Statistical Association*, Vol. 94, No. 446, pp. 365-374.
- Bakos, Yannis and Brynjolfsson, Erik 1999, "Bundling information goods: Pricing, profits, and efficiency," *Management Science*, Vol. 45, No. 12, pp. 1613-1630.
- Bass, Frank M. 1969, "A New product growth model for consumer durables," *Management Science*, Vol. 15, No. 1, pp. 215-227.
- Bartholomew, D. J. 1982, *Stochastic Models for Social Processes*, Third edition, John Wiley & Sons, New York.
- Bitran, G. and Mondschein, S. 1996, "Mailing decisions in the catalog sales industry," *Management Science*, Vol. 42, No. 9, pp. 1364-81.
- Blattberg, Robert C. and Deighton, John 1991, "Interactive marketing: Exploiting the age of addressability," *Sloan Management Review*, Vol. 33, No. 1, pp. 5-14.
- Brynjolfsson, Erik and Smith, Michael D. 2000, "The Great equalizer? The role of shopbots in electronic markets," Working Paper, Sloan School, MIT.
- Bult, J.R. and Wansbeek, T. 1995, "Optimal selection for direct mail," *Marketing Science*, Vol. 14, No. 4, pp. 378-394.
- Chavez, Anthony and Maes, Patti 1996, "Kasbah: An Agent marketplace for buying and selling goods," Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology, London, UK, April.
- Chintagunta, Pradeep; Jain, Dipak; and Vilcassim, N. 1991, "Investigating heterogeneity in brand preferences in logit models for panel data," *Journal of Marketing Research*, Vol. 28 (November), pp. 417-428.
- Coffey, Steve 1999, "Media Metrix methodology: Prepared for the advertising research foundation," Media Metrix White Paper, Winter 1999/2000. Retrieved on June 12, 2000 from http://www.mediametrix.com/usa/products/us__methodology_long.pdf

- Dahan, Ely and Srinivasan, V. 1998, "The Predictive Power of Internet-Based Product Concept Testing Using Visual Depiction and Animation," MIT, Sloan School, Working Paper. Retrieved on June 12, 2000 from <http://web.mit.edu/edahan/www/WorkingPaperonInternetConceptTestingbyDahanandSrinivasan.PDF>.
- Degeratu, Alexandru; Rangaswamy, Arvind; and Wu, Jianan 1999, "Consumer choice behavior in online and traditional Supermarkets: The Effects of brand name, price, and other search attributes," eBusiness Research Center Working Paper, 03-1999, Pennsylvania State University. Retrieved on June 12, 2000 from <http://www.ebrc.psu.edu/papers/pdf/03-1999.pdf>.
- DeSarbo, W. S. and Ramaswamy, V. 1994, "CRISP: Customer response based iterative segmentation procedures for response modeling in direct marketing," *Journal of Direct Marketing*, Vol. 8, No. 3, pp. 7-20.
- Dreze, Xavier and Zufryden, Fred 1997, "Testing web site design and promotional content," *Journal of Advertising Research*, Vol. 37, No. 2, pp. 77-91.
- Dreze, Xavier and Zufryden, Fred 1998, "Is Internet advertising ready for prime time?," *Journal of Advertising Research*, Vol. 38, No. 3, pp. 7-18.
- Fader, Peter S. and Hardie, Bruce G.S. 1999, "Forecasting repeat sales at CDNow: A Case study," Working Paper, The Wharton School, University of Pennsylvania. Retrieved on June 12, 2000 from <http://fourps.wharton.upenn.edu/ideas/pdf/fader-cdnow.pdf>
- Gonul, F. F. and Shi, M. 1998, "Optimal mailing of catalogs: A New methodology using estimable structural dynamic programming models," *Management Science*, Vol. 44, No. 9, pp. 1249-1262.
- Green, Paul E. and Srinivasan, V. 1990, "Conjoint analysis in marketing: New developments with applications for research and practice," *Journal of Marketing*, Vol. 54, No. 4, pp. 3-19.

- Green, Paul E. and Wind, Yoram 1975, "New way to measure consumers' judgments," *Harvard Business Review*, Vol. 53, No. 4, pp. 107-117.
- Guadagni, P.M. and Little, J. 1983, "A Logit model of brand choice calibrated on scanner Data," *Marketing Science*, Vol. 2 (Summer), pp. 203-38.
- Gupta, Sunil and Morrison, Donald G. 1991, "Estimating heterogeneity in consumers' purchase rates," *Marketing Science*, Vol. 10, No. 3, pp. 264-69.
- Horsky, Dan 1990, "A Diffusion model incorporating product benefits, price, income and information," *Marketing Science*, Vol. 9, No. 4, pp. 342-365.
- Johnson, Eric J.; Jerry Lohse; and Steve Bellman 2000, "Cognitive lock-in," Working Paper, Columbia Business School, Columbia University.
- Kamakura, Wagner and Russell, Gary 1989, "A Probabilistic choice model for market segmentation and elasticity structure," *Journal of Marketing Research*, Vol. 26, No. 4 (November), pp. 379-390.
- Mahajan, Vijay; Muller, Eitan; and Bass, Frank M. 1991, "New product diffusion models in marketing: A Review and directions for research," *Journal of Marketing*, Vol. 54, No. 1, pp. 1-26.
- Mahajan, Vijay; Muller, Eitan; and Bass, Frank M. 1995, "Diffusion of new products: Empirical generalizations and managerial uses," *Marketing Science*, Vol. 14, No. 3 (Part 2), pp. G79-88.
- McFadden, D. 1980, "Econometric models of probabilistic choice among products," *Journal of Business*, Vol. 53, pp. 513-29.
- Montgomery, Alan L. 1997, "Creating micro-marketing pricing strategies using supermarket scanner data," *Marketing Science*, Vol. 16, No. 4, pp. 315-337.
- Montgomery, Alan L. 1999, "Using clickstream data to predict WWW usage," Working Paper, Graduate School of Industrial Administration, Carnegie Mellon University.
- Nagle, Thomas T. 1984, "Economic foundations for pricing," *Journal of Business*, Vol. 57, No.

- 1, Part 2 (January), pp. S3-26.
- Pazgal, Amit 1999, "Software agents: The Future of marketing on the Internet?," Proceedings of the Fourth Conference on Information Systems and Technology, Cincinnati, Ohio.
- Pirolli, Peter and Pitkow, James E. 1999, "Distributions of surfers' paths through the World Wide Web: Empirical characterizations," *World Wide Web*, Vol. 2, No. 1-2, pp. 29-45.
- Rangaswamy, Arvind and Gupta, Sunil 2000, "Innovation Adoption and Diffusion in the Digital Environment: Some Research Opportunities" in *New Product Diffusion Models* (International Series in Quantitative Marketing, 11), eds. Vijay Mahajan, Eitan Muller, and Jerry Wind, Kluwer Academic Publishers, Hingham, Massachusetts.
- Roberts, Michael J. and Mahesh, Shripriya 1999, "Hotmail," HBS Case 899-165, Harvard Business School Publishing, Harvard University, Cambridge, Massachusetts.
- Rossi, Peter E. and Allenby, Greg M. 1993, "A Bayesian approach to estimating household parameters," *Journal of Marketing Research*, Vol. 30, No. 2 (May), pp. 171-182.
- Rossi, Peter E.; McCulloch, Robert E.; and Allenby, Greg M. 1996, "The value of purchase history data in target marketing," *Marketing Science*, Vol. 15, No. 4, pp. 321-340.
- Shapiro, Carl and Varian, Hal R. 1998, *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business School Press, Cambridge, Massachusetts.
- Sultan, Fareena; Farley, John U.; and Lehmann, Donald R. 1990, "A meta-analysis of applications of diffusion models," *Journal of Marketing Research*, Vol. 27, No. 1 (February), pp. 70-77.
- Taga, Y. and Isii, K. 1959, "On a stochastic-model concerning the pattern of communication - diffusion of news in a social group," *Annals of the Institute of Statistical Mathematics*, Vol. 11, No. 1, pp. 25-43.
- Telang, Rahul; Mukhopadhyay, Tridas; and Wilcox, Ronald T. 1999, "An empirical analysis of the antecedents of Internet search engine choice," Working Paper, Graduate School of Industrial Administration, Carnegie Mellon University. Retrieved on June 12, 2000 from

- <http://homenet.hcii.cs.cmu.edu/progress/SearchEngine.html>.
- Ungar, Lyle H. and Foster, Dean P. 1998, "Clustering methods for collaborative filtering", Working Paper, University of Pennsylvania. Retrieved June 12, 2000 from <http://www.cis.upenn.edu/~ungar/papers.html>.
- Van den Bulte, Christophe; and Lilien, Gary L. 1997, "Bias and systematic change in the parameter estimates of macro-level diffusion models," *Marketing Science*, Vol. 16, No. 4, pp. 338-353.
- Vilcassim, N. J. and Jain, D.C. 1991, "Modeling purchase-timing and brand-switching behavior incorporating explanatory variables and unobserved heterogeneity," *Journal of Marketing Research*, Vol. 28, No. 1, pp. 29-41.
- Wind, Jerry; Green, Paul E.; Shifflet, D.; and Scarbrough, M. 1989, "Courtyard by Marriott: Designing a hotel facility with consumer-based marketing models," *Interfaces*, Vol. 19 (January-February), pp. 25-47.
- Wind, Jerry and Rangaswamy, Arvind 1999, "Customerization: The Second revolution in mass customization," eBusiness Research Center Working Paper, 06-1999, Pennsylvania State University. Retrieved on June 12, 2000 from <http://www.ebrc.psu.edu/papers/pdf/06-1999.pdf>.

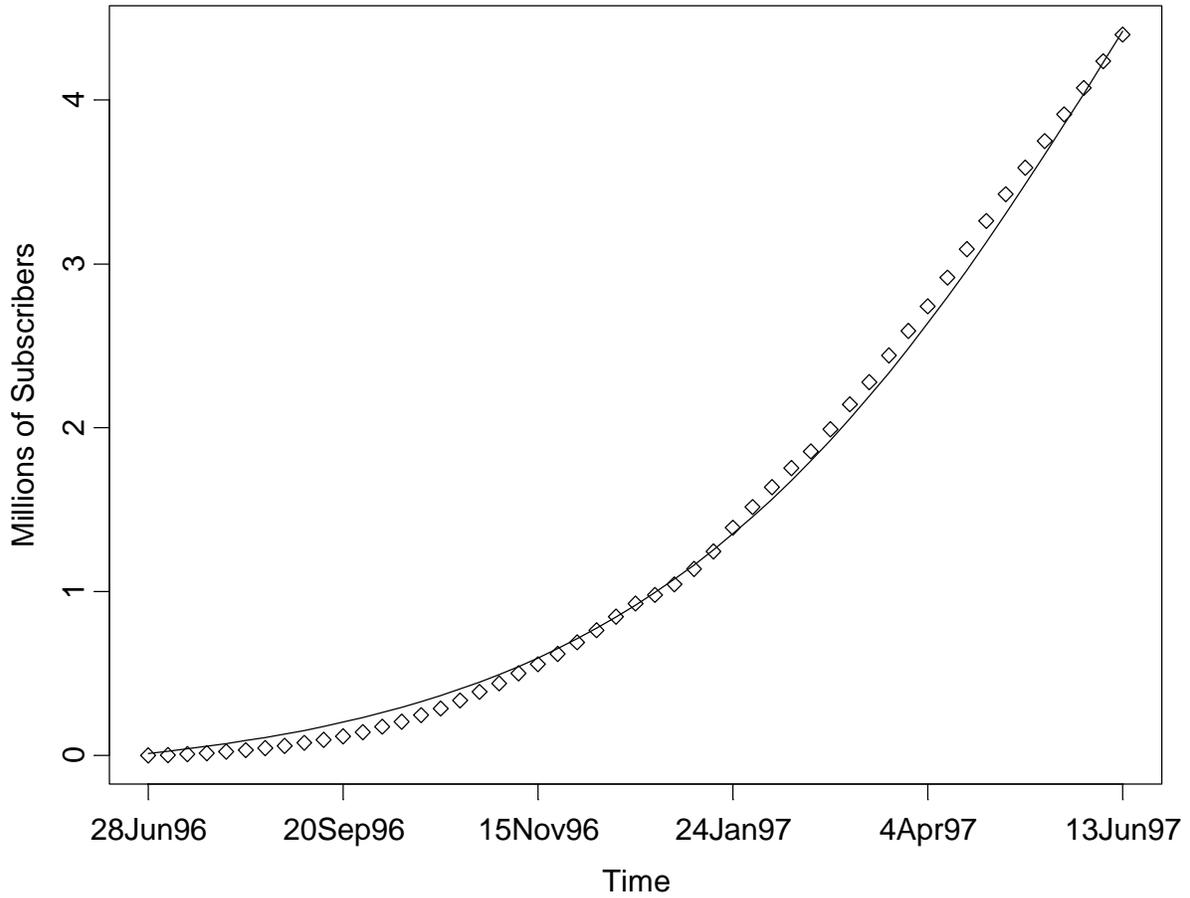


Figure 1: The cumulative number of Hotmail subscribers in millions (diamonds) at each weekly period closely matches the fitted predictions of the estimated Bass diffusion model (the line). The data is taken from Roberts and Mahesh [1999].

Parameters	Estimates
Item Price	-.194 (.001)
Shipping Price	-.369 (.002)
Sales Tax	-.433 (.014)
Delivery Average	-.019 (.001)
Amazon.com	.477 (.020)
Barnesandnoble.com	.174 (.023)
Borders.com	.269 (.020)

Table 1: Parameter estimates from Brynjolfsson and Smith [2000] of a multinomial logit choice model for consumer purchases at Dealpilot. The standard errors of the estimates are given in parentheses.

Alternative #	Store Brand	Item Price	Tax	Shipping	Delivery Time in Days	Utility	exp(Utility)	Choice Probability
1	Kingbooks	\$29.56	0	2.50	14.0	-7.36	.000639	26%
2	1Bookstreet	\$35.96	0	0	11.5	-7.63	.000487	19%
3	Barnesandnoble	\$31.96	1.28	3.95	4.5	-6.72	.001204	48%
4	Borders	\$39.95	0	3.90	5.0	-9.45	.000079	3%
5	Amazon	\$39.95	0	3.95	5.0	-9.26	.000095	4%
Total:							.002503	

Table 2: The product attributes or explanatory model variables from the alternatives retrieved by a shopbot are given in the first five columns. The attractiveness and choice probabilities derived from the multinomial logit model are given in the final three columns. The most likely choice is a book from Barnes & Noble which is \$2.40 more than the cheapest alternative.

Date and Time of Access	Seconds Viewed	URL
18JUL99:18:55:57	47	<i>http://www.voicenet.com/</i>
18JUL99:18:56:44	37	<i>http://www.weather.com/weather/us/cities/HI_Lahaina.html</i>
18JUL99:18:57:25	105	<i>http://www.weather.com/weather/us/cities/MI_Traverse_City.html</i>
18JUL99:19:03:00	7	http://www.weather.com/weather/us/cities/MI_Traverse_City.html
18JUL99:19:03:56	2	http://www.weather.com/weather/us/cities/MI_Traverse_City.html
18JUL99:19:03:58	6	http://www.weather.com/weather/us/cities/HI_Lahaina.html
18JUL99:19:04:58	2	http://www.weather.com/weather/us/cities/HI_Lahaina.html
18JUL99:19:05:00	1	http://www.weather.com/weather/us/cities/MI_Traverse_City.html
18JUL99:19:15:24	39	http://www.weather.com/weather/us/cities/MI_Traverse_City.html
18JUL99:19:17:00	7	http://www.weather.com/weather/us/cities/MI_Traverse_City.html
18JUL99:19:17:07	13	<i>http://www.realastrology.com/</i>
18JUL99:19:17:20	44	<i>http://www.realastrology.com/libra.html</i>

Table 3: In this sample of clickstream data from a single session of one user, italicized URLs denote those viewings that would likely have generated a hit that would be recorded in the server logs of those sites. The remaining viewings would probably be generated from the local browser cache and would not generate a hit on the corresponding server.

<i>Web Site Visited</i>	<i>Percentage of Female Viewers</i>
aol.com	48%
astronet.com	64%
avon.com	75%
blue-planet.com	52%
cartoonnetwork.com	56%
cbs.com	54%
country-lane.com	76%
eplay.com	47%
halcyon.com	41%
homearts.com	70%
ivillage.com	66%

<i>Web Site Visited</i>	<i>Percentage of Female Viewers</i>
libertynet.org	63%
lycos.com	39%
netradio.net	27%
nick.com	57%
onhealth.com	59%
on-linepsych.com	83%
simplenet.com	44%
thriveonline.com	76%
valupage.com	59%
virtualgarden.com	71%
womenswire.com	66%

Table 4: A list of Web sites one woman visited during July 1999 along with the percentage of female viewers as estimated by Media Metrix. Sites with fewer than ten viewings during the month are omitted from this list.