Managing Web Sites For Profitability: Balancing Content and Advertising

Rajiv M. Dewan  
Simon School of Business, University of Rochester

Marshall L. Freimer  
Simon School of Business, University of Rochester

Jie Zhang  
Simon School of Business, University of Rochester

This paper can be downloaded from the Social Science Research Network Electronic Paper Collection:  
http://papers.ssrn.com/abstract=283008
MANAGING WEB SITES FOR PROFITABILITY: BALANCING CONTENT AND ADVERTISING

RAJIV M. DEWAN, MARSHALL L. FREIMER, AND JIE ZHANG

ABSTRACT. Balancing the amount of advertising and content affects the profitability of the web site and its attractiveness to potential visitors. This tradeoff is modeled as a control problem for a web site manager who is maximizing the net present value of cash flows by controlling the amount of advertising and content displayed on the web site over its life. The model is calibrated and justified using web site advertisement and audience data. A new measure of web site traffic is developed and it is statistically significant in explaining the market value of firms that own advertising supported web sites.

1. Introduction

As the initial burst of enthusiasm for electronic commerce (e-commerce) and web related services wanes, profitability, if not profit, is the yardstick used to measure the success of ecommerce ventures. The primary means of managing profitability in most commerce settings is by adjusting the price of goods and services delivered. This tool is not available to most content ecommerce ventures as they do not charge for all or most of their content. While a few web sites, such as Financial Times and Britannica have begun to charge for some content, the majority of content on the web is free. Most of these content web sites are supported by advertisements.

Date: July 6, 2001.

We thank Nielsen Netratings for providing us audience data at a reduced rate.
An advertisement impression, a click through, or an eventual purchase of an advertiser’s offering may generate revenue for the web site that shows the advertisement. Even though prominent web sites are attempting to expand the revenue stream, advertisements remain the primary source of revenue. Consider the Yahoo family of sites. Apart from advertisements, it has created a number of new ways of charging for services. It offers annual membership for premium services at an annual fee, charges for listing in auctions, and for a fee and a transaction charge it provides a platform for over 13,000 retailers. All of these revenues totaled $1.1 Billion in the fiscal year 2000. Of this total, 90%, or $1.0 billion was from advertisements [3].

Advertisements are not viewed benignly by site visitors. Advertisements clutter up the page, reduce the amount of space available for user valued content, grab attention using garish colors or animation, and in general reduce the usability of the web site. Further, they add to the time it takes to download the page. Some sites such as CNET stage the downloading of page resources to make sure that advertisements are downloaded before other content. In any case, a complex formatted page may not fully render in readable form until all the content is downloaded. This delay in page rendering imposes delay costs on users. While the visitor may find a particular advertisement of interest, advertisements are, more likely than not, impose a cost on the visitor and reduce the attractiveness of the web site. The users would respond
to additional advertisements by visiting the site less often and/or visiting competing sites. In this sense the number of advertisements is analogous to price. This is examined further in the next section.

Setting the advertising policy for a site presents the web site owner with a tradeoff between revenue per customer and the number of customers who choose to visit the site. This decision directly affects the profitability of the site in each period. The number of visitors in any one period depends on the number of visitors in the prior period with some gain from growth of the number of users and loss from attrition. Consequently, the advertising policy cannot be made independently in each period but has to be managed over the life time of the site. We model this problem as an infinite horizon planning model with discounting. The objective is to maximize the market value of the firm which is the net present value or discounted sum of advertisement revenues and expenditures for site content. In section 3 we describe the model in greater detail.

Finally, we close the loop between the model and reality by a multistep estimation procedure. Using time series data on visitors to a web site and results of the model, we estimate the number of visitors in the future. This is used to estimate the market value of the site. The relationship between the estimated market value and the current market value obtained from market data is examined to validate the model.
2. ARE ADVERTISEMENTS LIKE PRICE CHARGED?

We would like to make the case that the number of advertisements on a advertisement supported web site acts like price in other markets. Advertisements, on an ex-ante basis, are a cost to the users. They clutter up the page, reduce the amount of valuable content visible and increase download time. Advertisements provide revenue to the web site owner on a per impression basis. To make this analogy complete, we conducted an empirical examination to see if the number of advertisements is like price from an economics perspective. A well known fact about markets is that the price for goods is lower in more competitive markets. Does this hold for web advertisements?

To examine whether the number of advertisements is like price from an economics perspective, we looked at newspaper sites. *American Journalism Review*, a trade press publication for the journalism industry, classifies general news dailies as national or regional on a number of factors [1]. We examined all the national newspaper sites and the top regional newspaper sites listed at the site. Since the advertisements come in many different shapes, we used the percentage of screen devoted to advertisements as a measure of amount of advertisements. We used the median screen size of 800x600 to evaluate the pages. This data is shown in Table 1.

We expect the market for viewers for web sites of national news dailies to be more competitive than that of regional news dailies. So we hypothesize:
<table>
<thead>
<tr>
<th>National</th>
<th></th>
<th>Top Regional</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Christian Science Monitor</td>
<td>5.25%</td>
<td>Boston Globe</td>
<td>18.03%</td>
</tr>
<tr>
<td>New York Times</td>
<td>4.43%</td>
<td>Chicago Tribune</td>
<td>12.70%</td>
</tr>
<tr>
<td>USA Today</td>
<td>2.93%</td>
<td>Dallas Morning News</td>
<td>12.51%</td>
</tr>
<tr>
<td>Washington Times</td>
<td>4.50%</td>
<td>Chicago Sun Times</td>
<td>15.29%</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td>2.93%</td>
<td>New York Post</td>
<td>12.35%</td>
</tr>
<tr>
<td>Nando Times</td>
<td>7.35%</td>
<td>New Orleans T. Picayune</td>
<td>10.80%</td>
</tr>
<tr>
<td>Washington Post</td>
<td>5.85%</td>
<td>Miami Herald</td>
<td>8.80%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The Kansas City Star</td>
<td>15.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The News &amp; Observer</td>
<td>5.63%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Salt Lake Tribune</td>
<td>5.80%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pittsburgh Post Gazette</td>
<td>6.18%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sun Sentinel</td>
<td>13.68%</td>
</tr>
</tbody>
</table>

Table 1. Percentage of 800x600 screen devoted to advertisements

\( H_a \): The average advertisement level at web sites of national dailies is different than at web sites of regional dailies.

\( H_0 \): There is no difference in mean amount of advertisements on web sites of national and regional dailies.

We tested these hypothesis on the data above. On web sites of national dailies, 4.7% of the screen was devoted to advertisements on average. This average for regional dailies was 9.6%. We did an Analysis of Variance (Single factor ANOVA) test for equality of means and this
null hypothesis was rejected with probability of 98.5%. The details of the analysis are shown in Table 2.

To explore the differences further, define a dummy variable:

\[ x_i = \begin{cases} 
1, & \text{if site } i \text{ is regional;} \\
0, & \text{otherwise.}
\end{cases} \]

We regressed the percentage of advertisements against the dummy variable \( x_i \). The results are shown in Table 3. Note that the one-tail probability of the slope of the dummy variable being zero is 0.007. Hence, with great confidence we can conclude that web sites of regional dailies have a greater percentage of advertisements.

### 3. Modeling the Web Site Management Problem

Consider a web site manager who is managing an advertising supported web site that offers content of value to visitors. There are many
design factors that affect the attractiveness of the page to a potential viewer. More well organized, frequently updated content attracts more viewers but is also expensive to acquire, format and place on the web on an ongoing basis. The web site owner uses a greedy strategy for acquiring content — he picks the cheaper ones (on a marginal basis) first. This results in an increasing cost of content as the content amount is increased. We model this cost as $kq_t^2$ where $q_t$ is the amount of content in period $t$.

In contrast, advertising reduces the attractiveness of the page — it adds clutter, increases download time and leaves less room on the page for the content that attracts a visitor. The web site owner, however, gets revenue from advertisements. More impressions (an instance of an advertisement viewed by a visitor) result in greater revenue. The relationship between impressions and revenue depends on the type of contract between the advertiser and the web site. The simplest kinds of contracts are written on “cost per thousand impressions” or CPM basis. More complex contracts may include click-throughs or actual sale performance. Even these can be simplified to a per impression rate. In this model we assume that the revenue in period $t$ is $cp_tu_t$ where $c$ is the revenue per impression, $p_t$ and $u_t$ are the number of advertisements shown to each user and the number of visitors in period $t$, respectively.

Putting the revenue together, the web site owner’s profit in period $t$ is $cp_tu_t - kq_t^2$. In each period the manager has the option of varying the amount of content and advertisements. These are the variables he controls. We take the objective to be maximize the net present value
of the firm as this maximizes shareholder wealth. Firms may have operated with different objectives in the short run, but economics and finance dictate that the long term object has to be the maximization of net present value of future cash flow streams. Let $\beta$ be the rate of continuous discounting. The manager’s objective function is to:

$$\max_{p_t, q_t} \int_0^\infty (c p_t u_t - k q_t^2) e^{-\beta t} dt$$

The formulation of the problem is not yet complete for we have not considered the effect of the manager’s decisions on the potential viewers. Let $u_t$ denote the number of viewers in time $t$. Starting at time 0 with $u_0$ viewers, we want to estimate the changes in viewership from period to period. Let $\dot{u}_t$ denote the change in number of viewers in period $t$.

Content valued by the visitors has the opposite effect of advertisements — it attracts visitors. Further growth in number of visitors to a web site may come from the growth in the number of users the world wide web on a global basis. There may be some attrition as some of the existing users may stop using the web site. Putting all of these factors together, we posit that:

$$\dot{u}_t = a - b p_t + e q_t - f u_t$$

where $\dot{u}_t$ is the change in number of visitors in period $t$, $a$ represents the contribution from natural growth in web population, $p_t$ and $q_t$ are amounts of advertisements and content in period $t$, respectively. The last term is the attrition that is proportional to the number of current visitors.
Putting this together with the objective (3) we get the following control problem for the manager:

$$\max_{p_t, q_t} \int_0^\infty (c p_t u_t - k q_t^2) e^{-\beta t} dt$$

Subject to:

$$\dot{u}_t = a - b p_t + e q_t - f u_t$$

$$p_t + q_t \leq 1$$

$$p_t, q_t \geq 0, \ u_0 \text{ given}$$

This is a control problem with $p$ and $q$ as controls and $u$ as the state variable. The constraint $p_t + q_t \leq 1$ represents the limit on page size (perhaps based on the median screen size.) In the rest of the paper we consider the case when this constraint is binding\(^1\). Set $p = 1 - q$.

The Hamiltonian of the control problem is:

$$H(u, q, \lambda) = e^{-\beta t} (c(1 - q)u - kq^2) + \lambda (a - b + bq + eq - fu)$$

The optimality condition is:

$$H_q = -e^{-\beta t}(cu + 2kq) + \lambda(b + e) = 0 \quad (1)$$

The auxiliary or adjoint equation is:

$$\dot{\lambda} = -H_u = -e^{-\beta t} c(1 - q) + f \lambda \quad (2)$$

\(^1\)We have solved the case when the page composition constraint is not binding. In this case we get $u$ to be a constant. This case does not seem empirically relevant as we rarely see pages that fit well within a 800 by 600 pixel screen, a screen of median size, such that the web site owner can increase the advertisements and have no impact on the amount of content.
Using (3) and (1) to eliminate \( q, \lambda \) and \( \dot{\lambda} \) from (2), we get a second order differential equation shown below.

\[
 ce^2 + b(ce - 2k(f + \beta)) + a(c(b + e) + 2k(f + \beta)) \\
+ (-2fk(f + \beta) - bc(2f + \beta))u(t) - 2k\beta u'(t) + 2ku''(t)
\]

A general solution for \( u(t) \) that fits this system is

\[
u(t) = A_1 + A_2e^{rt} + A_3e^{-rt}
\]

with constants \( A_1, A_2, A_3, \) and

\[
r = \sqrt{(bc + ce + f k + \beta/2)(f + b/2) - \beta/2}
\]

The constants \( A_1, \ldots, A_3 \) and \( r \) describe the market for content and one may expect them to be the same within an industry/genre of web sites. Using the boundary condition of \( u(0) = u_0 \) and \( \lim_{t\to\infty} u'(t) = 0 \), we obtain the constants. The specific solution obtained is shown below:

\[
(3) \quad u(t) = u_0e^{-rt} + (1 - e^{-rt})u_a
\]

where

\[
u_a = \frac{c(a + e)(b + e) + 2k(a - b)(f + \beta)}{(2f + \beta)(bc + ce + f k) + f k\beta}
\]

Note that \( u(t) \) moves asymptotically from its initial value \( u_0 \) to its asymptotic value \( u_a \) at \( t \) increases. This is shown in the figure 1 below for \( \beta = 0.02, a = 2,000, b = 1,000, c = 20, e = 1,000, f = 0.2, k = 10,000,000, \) and \( u_0 = 20,000 \) thousand page views.

Using the specific solution (3) and equations (3) to (2), we obtain asymptotic value \( q_a \) for \( q(t) \) as:

\[
 q_a = \frac{(f + \beta)(-a + b + fu_0) + f(b + e + fu_0) + 2f^2ku_0(f + \beta)/(b + e)}{(2f + \beta)(bc + ce + 2fk) - 2f^2k}
\]
Using the same parameters as in Figure 1, the proportion of content, \( q(t) \) is plotted below. The proportion of advertisements is \( 1 - q(t) \). We note that the firms make an initial investment in the site by having extra content and fewer advertisements. The proportion of content falls as the web site matures. This fits with the notion of investing in visitors.

4. Market value of firms

In this section we close the loop between the model and reality by looking at the association between the model predicted market value of the firm the stock price based market capitalization. We do it in three steps:
1. Using monthly audience data obtained from Nielsen Netratings, we estimate the asymptotic traffic, $u_a$, and the growth rate, $r$ at a number of content sites.

2. Define and compute a special statistic of traffic at a web site that we call the discounted total traffic (DTT) — the discounted sum of predicted future traffic at a site.

3. Examine the association between the discounted total traffic and market capitalization that is based on stock prices.

4.1. **Estimating $u_a$ and $r$.** The solution to the control problem predicts that the number of visitors will increase asymptotically from a starting value to an asymptotic value. The final value and the rate of change depend on the firm - the attractiveness of the content, degree
of competition and response of customers to advertising. We do not have time series data on proportion of content and advertising, but we do have time series data on number of visitors. We obtained this data from Nielsen Netratings. They estimate the global viewership of top 500 sites using panels. Using this data we estimated $u_a$ and $r$ for a number of sites that are primarily advertising supported. In particular, for site $i$, let

$$u_i(t) = u_{i0}e^{-r_it} + (1 - e^{-r_it})u_{ia} + \epsilon$$

where $u_{ia}$ and $r_i$ are the asymptotic traffic and growth rate for the site.

The first order conditions for minimizing the sum of square errors are:

$$\sum_{t=0}^{n} (u_i(t) - u_{ia} + (u_{ia} - u_{i0})e^{-r_it})(1 - e^{-r_it}) = 0$$

$$\sum_{t=0}^{n} (u_i(t) - u_{ia} + (u_{ia} - u_{i0})e^{-r_it})(u_{ia} - u_{i0})e^{-r_it}t = 0$$

We wrote out the second order conditions and found the sum of squared errors to be convex in the estimators. While the first order conditions are necessary and sufficient, they are hard to solve analytically. Consequently, we used the numerical Newton-Raphson method to obtain the estimates. These estimates and the quality of fit for a set of advertisement supported content sites are exhibited in Table 4. We used monthly audience measurement data for these sites from February 1999 to May 2001.

4.2. The Discounted Total Traffic at a site. In Table 4 note that there is a wide variation in growth rates. This reflects firms at different
<table>
<thead>
<tr>
<th>Firm</th>
<th>Market Cap (US$ Millions)</th>
<th>Asy. Traffic (Millions $u_a$)</th>
<th>Growth Rate Per Month ($r$)</th>
<th>Fit ($R^2$)</th>
<th>DTT (Millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yahoo</td>
<td>8,871.5</td>
<td>11.44</td>
<td>0.0307</td>
<td>0.95</td>
<td>2,743.1</td>
</tr>
<tr>
<td>Terra</td>
<td>3,606.2</td>
<td>1.08</td>
<td>0.2071</td>
<td>0.51</td>
<td>269.7</td>
</tr>
<tr>
<td>HomeStore</td>
<td>3,005.5</td>
<td>42.50</td>
<td>0.00023</td>
<td>0.79</td>
<td>619.4</td>
</tr>
<tr>
<td>CNET</td>
<td>1,360.1</td>
<td>9,572.84</td>
<td>0.86E-6</td>
<td>0.82</td>
<td>585.1</td>
</tr>
<tr>
<td>Goto</td>
<td>1,047.4</td>
<td>9,376.94</td>
<td>0.2E-6</td>
<td>0.82</td>
<td>157.8</td>
</tr>
<tr>
<td>Infospace</td>
<td>984.5</td>
<td>2,783.15</td>
<td>1.24E-6</td>
<td>0.36</td>
<td>268.0</td>
</tr>
<tr>
<td>MP3</td>
<td>333.7</td>
<td>0.04</td>
<td>0.0708</td>
<td>0.79</td>
<td>9.7</td>
</tr>
<tr>
<td>HotJobs</td>
<td>313.2</td>
<td>1,273.28</td>
<td>2.0E-6</td>
<td>0.84</td>
<td>138.2</td>
</tr>
<tr>
<td>Starmedia</td>
<td>138.7</td>
<td>1,646.07</td>
<td>1.56E-6</td>
<td>0.88</td>
<td>173.4</td>
</tr>
<tr>
<td>NetZero</td>
<td>121.7</td>
<td>6,109.47</td>
<td>0.73E-6</td>
<td>0.70</td>
<td>319.5</td>
</tr>
<tr>
<td>Switchboard</td>
<td>113.6</td>
<td>0.04</td>
<td>0.0485</td>
<td>0.71</td>
<td>9.5</td>
</tr>
<tr>
<td>LookSmart</td>
<td>105.6</td>
<td>23.74</td>
<td>0.00025</td>
<td>0.41</td>
<td>370.8</td>
</tr>
<tr>
<td>SportsLine</td>
<td>75.9</td>
<td>17.93</td>
<td>0.00039</td>
<td>0.21</td>
<td>438.3</td>
</tr>
<tr>
<td>Ask Jeeves</td>
<td>72.9</td>
<td>0.29</td>
<td>0.0442</td>
<td>0.85</td>
<td>69.5</td>
</tr>
<tr>
<td>iVillage</td>
<td>41.5</td>
<td>0.17</td>
<td>0.0447</td>
<td>0.82</td>
<td>41.5</td>
</tr>
<tr>
<td>ArtistDirect</td>
<td>27.2</td>
<td>7.95</td>
<td>0.000227</td>
<td>0.16</td>
<td>113.7</td>
</tr>
<tr>
<td>Women</td>
<td>19.4</td>
<td>0.19</td>
<td>0.0638</td>
<td>0.62</td>
<td>46.2</td>
</tr>
<tr>
<td>InfoNautincs</td>
<td>10.8</td>
<td>1.67</td>
<td>0.00019</td>
<td>0.07</td>
<td>21.4</td>
</tr>
<tr>
<td>LaunchMedia</td>
<td>9.5</td>
<td>6.04</td>
<td>0.00023</td>
<td>0.28</td>
<td>88.3</td>
</tr>
<tr>
<td>Salon</td>
<td>3.6</td>
<td>4.07</td>
<td>0.00026</td>
<td>0.28</td>
<td>64.3</td>
</tr>
</tbody>
</table>

**Table 4.** Estimates of asymptotic traffic and growth rate levels of maturity. It is hard to interpret and compare firms with such large variations in asymptotic traffic and growth rates. Consider the following simplifications of the problem:

1. The growth rate tends to zero as the Internet matures ($a = 0$).
2. Content and advertisements have equal but opposite effect ($b = e$).
3. Attrition rate, $f$, is small compared to $b$ or $e$.
4. The discount rate, $\beta$, is taken to be 0.4%.

With these simplifications, the discounted net present value of the firm turn out to be a constant plus a discounted measure of total traffic.
Multiple $R^2$ 0.84  
Adjusted $R^2$ 0.83  
$F$ 95.1  

<table>
<thead>
<tr>
<th>Coeff.</th>
<th>t-Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($A$)</td>
<td>-45.2</td>
<td>-0.204</td>
</tr>
<tr>
<td>Slope ($B$)</td>
<td>3.2</td>
<td>9.8</td>
</tr>
</tbody>
</table>

Table 5. Regression of Market Capitalization Vs Discounted Total Traffic rate

Multiple $R^2$ 0.56  
Adjusted $R^2$ 0.27  
$F$ 7.6  

<table>
<thead>
<tr>
<th>Coeff.</th>
<th>t-Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($A$)</td>
<td>-9.4</td>
<td>-0.03</td>
</tr>
<tr>
<td>Slope ($B$)</td>
<td>3.0</td>
<td>2.76</td>
</tr>
</tbody>
</table>

Table 6. Regression of Market Capitalization Vs Discounted Total Traffic rate for smaller firms

over the lifetime of the web:

\[
\text{Discounted Total Traffic (DTT)} = \frac{u_{i0}}{\beta} + \frac{u_{i0} - u_{i0}}{r_i + \beta}
\]

4.3. Market Capitalization. We obtained the market capitalization of web site owning firms classified into search and portal category by Wall Street Research Net, [2], and estimated the coefficients in the following structural equation:

(5) \quad \text{Market Capitalization}_i = A + B \cdot \text{DTT}_i + \epsilon

The result of the linear regression is shown in Table 5.
We note that the adjusted $R^2$ is 84%, the F statistic is 95.1, and the slope is almost surely not zero. This is very re-assuring. However, examining the data more carefully, we note that Yahoo is much larger than others and has a strong influence on the fit. So we re-estimated the coefficients for all but Yahoo. The results are shown in Table 6. The adjusted $R^2$ drops to 26.9%, the F statistic drops to 7.6 (still a very good fit) but the slope remains almost the same and statistically significant.

REFERENCES


WILLIAM E. SIMON SCHOOL OF BUSINESS ADMINISTRATION, UNIVERSITY OF ROCHESTER, ROCHESTER, NEW YORK 14627

E-mail address, R. Dewan: dewan@simon.rochester.edu
URL: http://rdewan.simon.rochester.edu