

University of Texas/Chitika Study:

In 2006, The Top 50K Blogs Generated \$500M in Ad Revenue

Vijay Mookerjee, Ph.D, and Milind Dawande, Ph.D
University of Texas, Dallas, TX.

*in collaboration with
Chitika, Inc.*

Study Objective: The Blogosphere has significantly increased in popularity within the past few years. This popular Internet resource has quickly become a viable advertising channel for top brands and media planners. To date, there have been little to no studies on the dollar amount spent on advertising in the blogosphere. Using revenue trends and statistics from a representative sample for the 12000+ Chitika publisher network, we estimated the advertising revenue earned by bloggers in 2006.

Methodology: Within Chitika's 12,000+ network there are blogs from all across the spectrum. A stratified sampling technique was used to ensure that blogs across different levels of popularity were represented. The popularity of a blog was measured by its rank as reported in the Technorati Rank service. Blogs were sampled from within the top 100 rank as well as from medium to low ranks. The highest and lowest ranks in the final sample were 48 and 117,360 and there were a total of 332 data points in the sample.

- Using Chitika revenue numbers from 2006 and assuming that bloggers will, on average, diversify across 3 revenue sources, and assuming that they will typically try out other sources that will generate similar revenue streams, it is safe to estimate that the total estimated revenue for their blog will be at least 3 times the revenue attained through Chitika.
- With the sample data set, (BlogID, Rank, 2006 Revenue) standard statistical predictive modeling techniques (in this case, the specific regression model used was a variant of the Pareto model) were used to estimate the advertising revenue earned by blogs across all ranks.

Model: A double exponential model was used to estimate the dollar size of the blog advertising market.

Revenue = e^y

$y = e^{\alpha(\text{rnk})+\beta}$, where rnk is the popularity rank based on the Technorati index.

α , β are coefficients that need to be estimated.

Methodology: Taking natural Logarithms (twice) of both sides in the above model, we get the revised (or linearized) model that was fitted to estimate the coefficients α and β

$$\text{Ln}(\text{Ln}(\text{Revenue})) = \alpha + \beta(\text{rnk}) + \varepsilon$$

ε is a normally distributed error term with zero mean

Results: The revised model is linear and can be estimated using standard least squares statistical techniques. Using the Chitika Data Set, the least squares estimates for the coefficients were found to be

$$a \text{ (estimate for } \alpha) = 2.324792$$

$$b \text{ (estimate for } \beta) = -0.00000365$$

The fit of the revised model (i.e., the linearized form), was $R^2 = 0.3015$. This fit is modest, but acceptable given the exploratory nature of the study. The important point to note here is that the fit for a single exponential model for the same data is significantly poorer. Thus a double exponential model better explains the blogging phenomenon than the single exponential form typically used to model power-law-type phenomena.

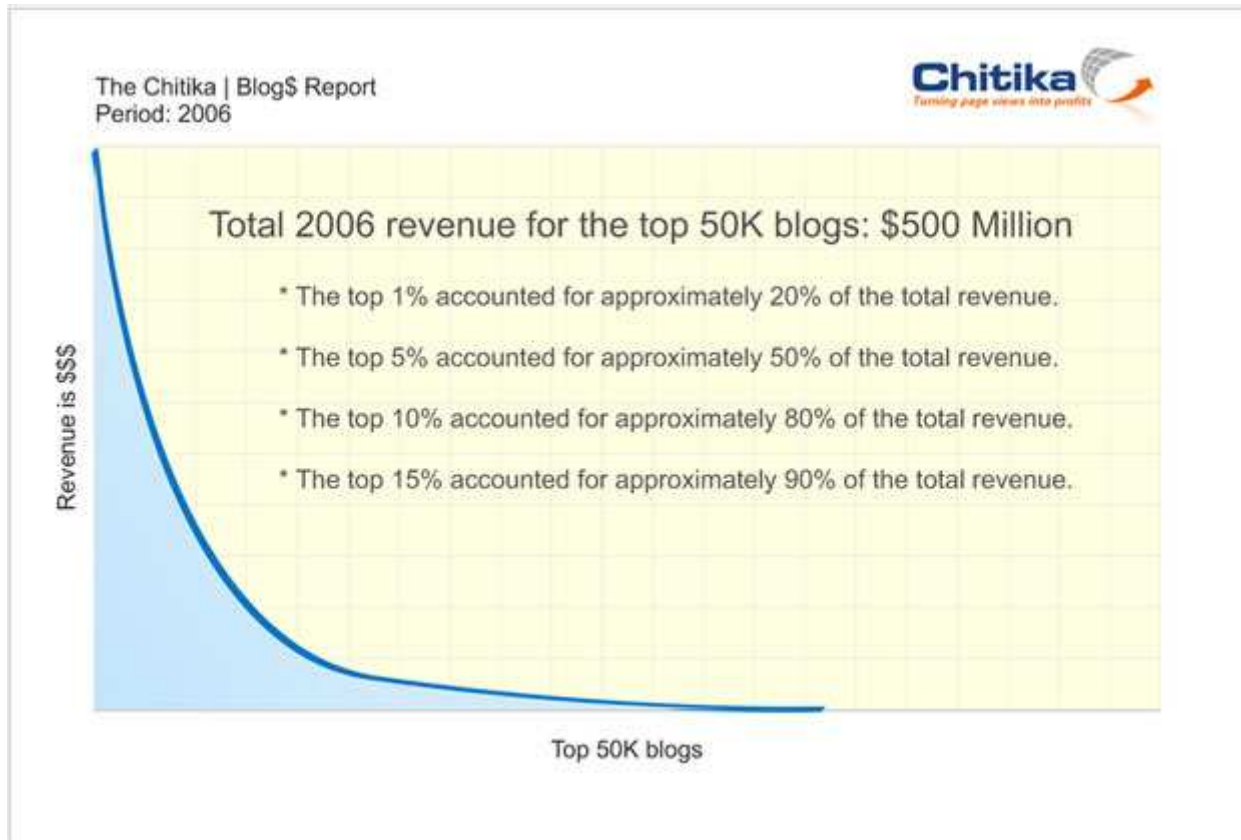
Using the above model, we are able to estimate the total size of the market (for the first 50,000 blogs) as below

$$\text{Total Size} = \sum_{i=1}^{50000} e^y, \text{ where } y = e^{2.324792 - 0.00000365(i)}$$

Size Correction: Upon closer observation, it was found that model tends to overestimate the revenue at less popular blogs, but underestimates it at more popular ones. Because, there are relatively fewer very popular blogs, the initial estimate of the model was lowered. However, this correction does not affect the basic conclusion of the finding (i.e., concerning the double exponential form), but only the numerical estimate of the total size of the market.

Summary of Findings: Ad revenue in a blog is more sensitive to the rank of the blog than what one would expect in a typical Zipf Law 80/20 curve situation. One reason for this may be the social value of advertising in a blog. If online advertising is like advertising in a mall, advertising in the blogosphere is like advertising in a country club.

- The top 1% accounted for approximately 20% of the total revenue.
- The top 5% accounted for approximately 50% of the total revenue.
- The top 10% accounted for approximately 80% of the total revenue.
- The top 15% accounted for approximately 90% of the total revenue.



Implications: We speculate that Blogging is a generation ahead of the Internet in terms of the “networking” effects that govern this phenomenon. Advertisers need to consider this aspect, namely, the importance of the popularity of the Blog, similar to the way they need to consider the importance of targeting opinion leaders or social connectors, in a viral marketing campaign.

This study was done as an initial attempt in establishing the advertising revenue of blogs. Suggestions for alternate methods of estimating these numbers as well as ideas for future studies are welcome and encouraged: blogdollarreport@chitika.com

~~~~~