UC Berkeley
Other Recent Work

Title
The Dynamics of Competition in the Internet Search Engine Market

Permalink
https://escholarship.org/uc/item/0h17g08v

Author
Gandal, Neil

Publication Date
2001-01-02

The Dynamics of Competition in the Internet Search Engine Market

Neil Gandal, Tel Aviv University, UC-Berkeley, and CEPR

January 26, 2001

Abstract

Search engines hold the key to helping consumers access the wealth of information on the web. In this paper I examine the evolution of and competition in the internet search engine market. The goal of my analysis is to examine whether early entrants benefit in the long-run from their first-mover position in internet markets.

I find that while early entrants (Yahoo, Lycos, Excite, Infoseek, and Altavista) still have an advantage, the pure "brand effect" advantage has been declining over time. Yahoo has maintained its leadership position by providing a superior product. The success of a wave of recent new entrants suggests that entry barriers are still quite low in the internet search engine market.

Keywords: Internet, Search Engines, Entry, Empirical Study.

JEL Classification Numbers: D43, L86.

0

1 Introduction

Many observers believe that the emergence of the internet has profoundly changed society. Perhaps the most important feature of the internet is the tremendous amount of publicly accessible information. One example is the recent decision of the Encyclopedia Brittanica, which was previously sold door-to-door, to put all its 32 volumes on the web free of charge. According to estimates by Lawrence and Giles (1998, 1999), the number of publicly indexable pages on the world wide web grew from 320 Million in December 1997 to 800 million in February 1999.

Search engines hold the key to helping consumers access and sort the wealth of information on the web. Media Metrix data show that many search engines are consistently among the top 25 most visited web sites. This suggests that internet users spend a significant amount of time using search engines to find relevant information. Despite the sophistication of the search engines, they are far from comprehensive.

Search engines are also more likely to index "popular" sites (i.e., those sites that have more links to them) and sites that are in the United States. Lawrence and Giles remark that "search engines can be compared to a phone book which is updated irregularly, is biased toward listing more popular information, and has most of the pages ripped out (http://wwwmetrics.com/)."[1] For further discussion, see two outstanding web sites: Daniel Sullivan's web site at http://www.searchenginewatch.com, and Greg Notess web site at http://www.notess.com/search.

In this paper I examine the evolution of and competition in the internet search engine market. The goal of my analysis is to examine whether early entrants benefit in the long-run from their first-mover position in internet markets. This is an important question since "first-mover" is often cited as a strong competitive advantage in internet firms' business

---

[1]Most internet search engines do not in general permit advertisers to pay to improve their position in search engine results. The exceptions are GoTo and Lycos. See http://www.searchenginewatch.com for more information.

plans. Does a first mover advantage in internet markets translate into market leadership in the long run? This paper will analyze the internet search engine market in order to provide some answers to these important questions.

To the best of my knowledge, Prusa and Schmitz (1991) is the only paper that empirically (econometrically) examines whether "first-movers" become market leaders. They find that new firms in the PC software industry have an advantage (over incumbents) in developing new software categories, while incumbents have a comparative advantage in product improvements in existing categories.

Despite the attention that the trade press devotes to "digital economy," there is little systematic empirical work about electronic commerce. Two exceptions include Brynjolffson and Smith (1999), which examines online vs. offline prices for books and Gallaugher and Downing (1999), which examines competition in the internet search engine market.

In their study, Gallaugher and Downing (1999) restrict attention to four early firms (Yahoo, Lycos, Excite, and Infoseek). Using market reach data, where market reach is the percent of consumers conducting searches with a particular search engine in a given month, they find that the fixed (brand) effects explain most of the variation in market reach. Market reach is similar to market share; the difference is that (the sum of) market reach can exceed one if consumers use more than a single search engine during a given month.

Both market share and market reach are problematic variables because they do not capture the tremendous growth in demand for searches. Media Metrix data on market reach (which are available at www.searchengine.com) show that market reach for Yahoo, Lycos, Excite, and Infoseek varies only slightly during the March 1997 - December 1998 period, the period examined by Gallaugher and Downing (1999). Yet the number of unique users of search engines grew from 43.1 Million in August 1997 (the earliest date for which I have data) to 96.1 Million in December 1998. Thus, it is not surprising that the fixed (brand) effects explain most of the variation in market reach. Their conclusion that pure brand effects are quite strong in this market is surprising and is likely due to the fact that they (i) analyze a

very early period in evolution of the industry, (ii) restrict attention to the four leading firms at the time, and (iii) use market search as the dependent variable.

Indeed, the trade press suggests that there is fierce competition in the search engine market. Barry Parr, the director of internet and e-commerce strategies at International Data Corp, believes that internet search is essentially a commodity service. According to industry analysts, most users "can't even differentiate between the major search engines." (See Weisul, Kimberly "Search Engines Chase Profit," *Interactive Online*, May 10, 1999 at http://www.zdnet.com/intweek/stories/news/0,4164,2255145,00.html.)

My analysis differs from Gallaugher and Downing in several important ways. I use the number of unique visitors to search engines to measure demand and include all of the competitors in the search engine market in my analysis. In addition, I estimate models that are typically employed in the analysis of oligopoly markets.

I find that while early entrants (Yahoo, Lycos, Excite, Infoseek, and Altavista) still have an advantage, the pure "brand name" advantage has been declining over time. I find that Yahoo has maintained its leadership position by providing a superior product.

The success of a recent wave of new entrants in the internet search engine market suggests that entry barriers are still quite low in the internet search engine market. This makes sense, given the fact that consumers pay no fee for the use of search engines and there are little (if any) consumer switching costs.

There is a large theoretical literature on the economics of emerging industries. A common theme in this literature is that both "learning by doing" (supply side) and "learning by using" (the demand) play a key role in the evolution of new industries. Since the internet search engine was truly a new product, the learning effects were particularly important in this case. Rob (1991) developed a theoretical model that shows that under uncertainty regarding the size of the market, entry will occur in waves. Vettas (1998) obtained similar results in an extension of the Rob (1991) model to a setting with uncertainty on both sides of the market. Empirically, this has been the case so far in the internet search engine market, with the first

wave consisting of Yahoo, Lycos, Excite, Infoseek, and Altavista. Indeed, there is empirical evidence that entry into new markets generally occurs in waves. See Geroski (1995) and the references cited within.

A large empirical literature has examined the post-entry performance of firms. These studies typically examine the entry and exit rates over time, the number of firms in the industry over time, the survival rate of new firms, and the evolution of firm size over time. Since this is a very large literature, I will not try to survey it. For a good summary of recent work see the Audretsch and Mata (1995) introduction to the special edition of the International Journal of Industrial Organization on the post-entry performance of firms.

In section 2, I briefly describe the search engine industry. Section 3 describes the data I employ; the analysis and discussion of the results are in section 4. Section 5 concludes.

2   The Search Engine Industry

Search engine competition began in 1994, with the entry of Yahoo and Lycos. By 1995, three additional search engines were competing in the market: Excite, Infoseek, and Altavista. Until, mid 1998, the market was dominated by these five firms and Yahoo was the industry leader. In August 1997, Yahoo led with 14.8 million unique users, while Infoseek and Excite respectively had 7.9 million and 7.6 million unique visitors during the month; Lycos and Altavista followed with 4.9 and 4.7 million unique users during the same period. Webcrawler, an early entrant that quickly lost significant market share had approximately 3.2 million unique users in August 1997. No other search engine had more than 1.7 million unique users. Ignoring a small competitive fringe, there were approximately 43.1 million unique visitors to search engines in August 1997.

As the market for searches grew, market structure began to change. The concentrated oligopoly market broke down in the middle of 1998 when several "late" entrants, obtained a fairly significant share of the market by offering high quality products. Table 1 shows that

4

by August 1999, five late entrants had each obtained market shares in the 5-6 percent range, nearly as much as the 7 percent share held by Altavista at that time.

Table 1 shows that the number of unique users of search engines increased by more than 100 percent during the August 1997 - August 1998 period and by nearly 50 percent during the August 1998 - August 1999 period. Consequently, despite the fact that Yahoo's number of unique users increased by 125 percent during the August 1997- August 1999 period, its market share fell from 34 to 25 percent during same period.

Table 1 also shows that in August 1997, the ratio of the number of unique visitors to search engines divided by the number of web users was approximately 1.0. This ratio rose steadily from 1997 to 1999 and stood at 1.6 in August 1998, and reached 2.1 in August 1999. Since nearly all web users visit at least one search engine per month, the data suggest that in 1997 each user employed a single search engine; by August 1999, consumers were on average using multiple search engines.

Why was there such an increase in the average number of search engines used by each consumer? Lawrence and Giles (1998) estimate that several search engines covered more than 20 percent of the indexable web in December 1997; in Lawrence and Giles (1999), they estimate that no search engine covered more than 16 percent of the indexable web in February 1999. The decline in coverage is due, in part, to the fact that the number of publicly indexable pages on the world wide web grew from 320 Million in December 1997 to 800 million in February 1999.

Similarly, Notess has consistently found that there is very little overlap among search engines' results. In an analysis, he undertook in September 1999, he found that five "very small" searches run on thirteen search engines yielded 140 unique pages. 66 of the 140 pages were found by a single search engine, while 30 were found by only two engines. Another interesting finding is that there is little overlap among the Inktomi-based databases, which include Snap, HotBot, and Yahoo from my sample. (See http://www.searchengineshowdown.com for details.)

Another reason for the increase in the use of search engines is due to the relative ease with which additional searches can be conducted. Several search engines now routinely offer a "second opinion" at the end of the search, that is a click of the mouse yields additional results for the search by another search engine.

3 Data Description

My goal empirically is to determine to what extent (i) being a "first-mover" in this industry was enough to establish a long-run leadership position, and (ii) inherent quality matters. In order to examine this question, I employ five months of data from the August 1998 - August 1999 period, using equally spaced intervals (August 1998, November 1998, February 1999, May 1999, and August 1999). I use one data point per quarter because the monthly changes are relatively small. I do not use data prior to August 1998, because (with the exception of August 1997 data) only market reach data are publicly available for the earlier period. As discussed above, such data do not capture the growth in the number of unique visitors to each search engine over time.

I include all search engines that had more than 3.6 million searches (or 2 percent of the market) in August 1999. Media Metrix, the source of my data, does not publicly report the data for firms with a smaller market share. Hence, there are 11 firms in my data set. Following PC Magazine and Media Metrix, I do not include America Online, Netscape, or Microsoft in the "search engine" category. While these firms provide search services, their web sites have significant traffic not associated with searching capabilities.

All of the firms are in the data set for the full period.[2] No firm exited during this period. The only significant firm to fall below the cutoff during the 1997-1999 period was Webcrawler, and its market size was already well below the cutoff in August 1998. Descriptive statistics

---

[2]Data on unique visitors is not publicly available for "About" for the August 1998 period. Hence there is one missing observation. Miningco changed its name to About in the spring of 1999. For ease of presentation, in the paper, I refer just to About.

on the variables in the study are in table 2. I now describe the variables in some detail.

- Monthly data on unique visitors to each search engine comes from Media Metrix, the leader in the provision of such data. The current data, which are often quoted by the trade press, are publicly available from Media Metrix at http://www.mediametrix.com. The variable denoted SEARCHES, measures (in millions) the number of unique visitors to each search engine. Media Metrix data are attractive because they measure web traffic by individual web sites rather than by web properties.[3] LSEARCHES is the natural log of this variable.

- The variable UNIVERSE is the number of people (in millions) who use the web. This variable is likely a function of the quality adjusted price of internet access. As the price of internet access fell and the quality of internet access increased, more consumers obtained access to the web. Table 1 shows that the number of internet users increased by 38 percent from August 1998-August 1999. The variable LUNIVERSE is the natural log of UNIVERSE.

- The variable AGE denotes the number of years that a search engine has been in the market. This variable is defined as 1999 less "the year that the firm entered the market." Nothing in the analysis changes if I define the variable so that I measure age in quarterly intervals.

- The dummy variable EARLY takes on the value 1 if the firm was among the five firms that dominated the market in its early years.

- Data on characteristics of search engines comes from two sources: PC Magazine and Lawrence and Giles (1999). PC Magazine analyses all of the 11 search engines in the

---

[3]The only exception regarding search engine data is in the case of the "Go Network." In January 1999, Infoseeek was acquired by Disney and Media Metrix lumped together three web sites "disney.go.com" , "infoseek.go.com", and "espn.go.com" into the go.com data. Hence the increase in Infoseek/Go Network's market share in 1999 is due to the aggregation of the data.

data set, while Lawrence and Giles (1999) examine seven search engines in my data set. Both of the PC Magazine and Lawrence and Giles analyses were conducted during the period I use in the analysis.

The PC magazine data contain ratings on five characteristics listed below. The ratings are poor (assigned a value of zero), fair (one), good (two), and excellent (three). PC Magazine sent 50 different requests to all of the search engines on various topics first using the basic search and then using refined searching methods. (For details, see http://www.zdnet.com/pcmag/stories/reviews/0,6755,2330316,00.html.) They report results in the following five areas:

- "Relevant hits from the initial search," denoted RELEVANTHITS, where the top ten hits from each query are evaluated for relevance by examining the content in the pages.

- "Eliminates Dead Links," denoted ELIMDEAD, counts the number of "dead" links in the top 10 hits from each query and then assigns a rating. They are careful not to count errors due to the fact that the server may not have been up at the time.

- "Eliminates Duplicate Links," denoted ELIMDUPL, counts the number of duplicate links in the top ten hits and then assigns a rating.

- "Customizes Search Effectively," denoted CUSTOMIZES, examines the tools and options that a search engine has available in order to refine a search. These tools are then rated by examining the quality of the hits from these advanced searched.

- "Effective Anticipatory Results," denoted ANTICIPATES, measures how well a search engine anticipates the goal of the search, and whether this results in more relevant results.

Two of these characteristics (ELIMDEAD, ELIMDUPL) are purely quantitative, while

8

RELEVANT HITS is also essentially quantitative, since it is easy to design fairly objective tests. (For example, does a search for Neil Gandal result in my home page appearing in the top ten hits?) The other two characteristics are clearly subjective; I discuss this issue further below.

The Lawrence and Giles (1999) article contains the following quantitative data on the performance of seven search engines (Yahoo, Go/Infoseek, Excite, Lycos, Altavista, Hotbot, and Snap) in my data set:

- "Percentage of Invalid Links," denoted PERDEAD, estimates the percentage of dead links using 1050 queries. They are careful not to count errors due to the fact that a particular server may not have been up at the time.

- "Coverage with Respect to Web Size," PERWEB, which measures the percentage of the web covered by each search engine. As mentioned above, no search engine in the data set covered more than 16 percent of the web in 1999.

- "Median Age of Matching Documents," MEDAGE, which measures the (median) time between when new documents were added to the search engines and when these documents were last modified. This variable may be problematic, since not all web pages list the date of the last update.

The one overlapping variable between the PC Magazine and Lawrence and Giles (1999) data relates to dead links. It is reassuring to know that there is a negative correlation (-0.48) between PERDEAD and ELIMDEAD for the seven search engines in my data set that were examined by PC Magazine and Lawrence and Giles (1999).

4 Analysis of the Data

Here I employ two models that have been used to examine competition in oligopoly markets. In section 4.1, I employ a variant of a growth model. Growth models are often employed to

examine industry dynamics. In section 4.2, I employ a model based on recent advances in estimating discrete-choice models of product differentiation.

4.1 A Growth Model

With a single characteristic on each product that does not change over time (x_j), it is possible to estimate a growth model of the form

$$log(SEARCHES_{j,t}) = \beta_o + \beta_1 log(SEARCHES_{j,(t-1)}) + \beta_2 x_j + \epsilon_{j,t}, \tag{1}$$

where the subscript j refers to the product and the subscript t refers to time; $\epsilon_{j,t}$ is an i.i.d. error term with mean zero and constant variance. From (1), $log(SEARCHES_j) = (\beta_0 + \beta_2 x_j)/(1 - \beta_1)$ in the long run. That is, differences in the product characteristics drive the long run market size differences among the products.

In the estimation, of course, I employ more than one characteristic, and some of these characteristics are industry specific (i.e., UNIVERSE), while other characteristics are product specific. The product specific characteristics do not change over time in my analysis.[4]

The results using the growth model (1) with product and industry characteristics are shown in table 3. The variable log(searches$_{-1}$) explains quite a bit of the variation in (1). The positive coefficient on EARLY and the negative coefficient on AGE show that other things being equal, the early entrants still have an advantage, but that this advantage has been declining over time.[5] The coefficient on LUNIVERSE measures the elasticity of searches with respect to the size of the internet universe. The positive sign on LUNIVERSE suggests that an increase in internet users has led to an increase in the number of searches. Although

---

[4]Greg Notess' search engine comparison site issues "Dead Link Reports" once every two to three months. The problem with these data are that they do not include all of the search engines in my sample. If these reports had included all of the search engines in the sample, this characteristic would have changed over time. Nevertheless, this probably does not affect the analysis that much, since Notess' reports show that firm performances are highly correlated over time. In other words, the relative ratings are quite similar over time.

[5]The coefficient of EARLY is statistically significant, while the coefficient on AGE is not statistically significant.

its effect on the number of searches is not statistically significant, this estimated elasticity (0.69) is fairly large in an economic sense.

Table 3 shows that providing relevant hits (RELEVANTHITS) is the characteristic that consumers value most; the coefficient on RELEVANTHITS is statistically significant. The coefficient on eliminating dead links (ELIMDEAD) and ANTICIPATES are also positive, but not statistically significant. The coefficient on eliminating duplicate links is negative and insignificant. The effect of CUSTOMIZES is negative and significant. Since CUSTOMIZES is a very subjective characteristic, the results suggest that the tastes of the reviewers may differ from the preferences of consumers at large. The results are virtually unchanged if the three data points of the GoNetwork are excluded.[6]

Adding a dummy variable for Yahoo in the estimation of the growth model (1) yields an associated coefficient of 0.26, with a t-statistic of 0.22. In other words, controlling for quality, the Yahoo brand name only adds marginal value to the product. Hence, the results suggest that the brand name Yahoo is much less important to consumers than the fact that the search engine provides relevant hits and does a good job in eliminating dead links.

4.2 A Discrete Choice Model of Product Differentiation

In order to examine the robustness of the results using the growth model, I now employ a model based on recent advances in estimating discrete-choice models of product differentiation. These techniques, developed by Berry (1994) and Berry, Levinsohn, and Pakes (BLP) (1995), enable structural estimation of the demand side of a differentiated product market.

---

[6]In the case of the PC Magazine data, no product received the lowest possible rating on any of the variables. Since for three of these variables, ELIMDEAD, ANTICIPATE, and ELIMDUPL, no product received the highest rating possible, these variables are essentially dummy variables (i.e, they have two values). Hence I constructed two dummy variables for both RELEVANTHITS and CUSTOMIZES. In the case of the growth model, the estimated coefficient for the dummy variable RELHITHIGH (which takes on the value 1 if the rating was the highest possible and 0 otherwise) was 0.42 (t=2.14), while the estimated coefficient for the dummy variable RELHITMIDDLE (which takes on the value 1 if the rating was the second highest possible and 0 otherwise) was 0.23 (t=1.19). The results are similar for CUSTOMIZES. This suggests that the continuous interpretation of these variables works well.

The utility of product j to consumer i, denoted u_{ij}, depends on both product and consumer characteristics. Following Berry (1994), I employ a random utility model of the form

$$u_{ij} = x_j \beta - \alpha p_j + \xi_j + \epsilon_{ij}, \tag{2}$$

where the first two terms are the mean valuations of product j's observed characteristics; x_j is a vector of observable product characteristics and p_j is the price. The parameters α and β represent the mean valuations of the observable characteristics. ξ_j represents the average value of product j's unobserved characteristics, while ϵ_{ij} is the deviation of buyer preferences around this mean.

The error term ϵ_{ij} introduces heterogeneity and determines the substitution patterns among products. The multinomial logit model assumes that there is no buyer heterogeneity: in particular, the logit assumes that ϵ_{ij} are identically and independently distributed across consumers and choices with the extreme value (Weibull) distribution function. Given the discrete choice set, under this assumption it can be shown that the probability of choosing product j, (the market share of product j) is

$$s_j = \frac{e^{\delta_j}}{\left(\sum_k e^{\delta_k}\right)}, \tag{3}$$

where

$$\delta_j = x_j \beta - \alpha p_j + \xi_j, \tag{4}$$

is the mean utility level from product j. Since there is little or no vertical differentiation among products and since income plays no role in consumer choice in the search engine market, the logit model seems appropriate in this case. In order to employ this model, it is necessary to assume the same demand function over time, but this seems reasonable since the data set is for a single year: August 1998 - August 1999. By inverting the market share equation (3), one obtains

$$log(s_j/s_0) = x_j\beta - \alpha p_j + \xi_j, \qquad (5)$$

where s_0 is the proportion of consumers who do not use search engines. I assume that all internet users access a search engine in a given month. Hence, s_0 in the percent of the U.S. population in millions (denoted POPULATION) without internet access. Hence, log (s_j/s_0) is log(SEARCHES/[POPULATION-UNIVERSE]).[7]

Once consumers have internet access, there is no cost to consumers of using a search engine. This is because most consumers in the U.S. pay a monthly internet access fee that allows unlimited use and local phone service in the U.S. is not metered. Further, there is no charge to consumers for the use of search engines. (This is true for all of the search engines in the dataset and generally is the case.)

Hence the price term drops out of (5). Since the product characteristics are exogenous in the short run, consistent estimates of the β parameters can be obtained by an OLS regression on (5).

Since the percentage change in the numerator of searches is much larger than the percentage change in [POPULATION-UNIVERSE], over the 12 month period of the analysis, the left hand side of the growth model (equation (1)) and the left hand side of the discrete choice model of product differentiation (equation (5)) essentially differ by a constant factor. The main difference between the two is that (1) captures the dynamics more explicitly than (5).

Hence it is not surprising that the results from the first discrete choice model of product differentiation (model I) in table 4 are qualitatively similar to the results from the growth model in table 3. The main empirical difference is that the t-statistics are smaller for the parameters in the growth model. This is because the variable log(searches$_{-1}$) explains quite

---

[7]Since some consumers use more than one search engine in a month, this model can be thought of as an approximation to the true choice model. Nevo (1998) similarly uses a discrete choice model of product differentiation to model demand in the market for breakfast cereals.

a bit of the variation in (1).

Table 4 presents the results from estimating the demand functions in (5). Similar to the growth model, the positive and significant coefficient on EARLY and the negative and significant coefficient on AGE in table 4 show the early entrants still have an advantage, but that this advantage has been declining over time. The positive and significant (elasticity) coefficient on LUNIVERSE suggests that an increase in the number of internet users has led to an increase in the number of searches.[8]

Similar to the growth model, relevant hits (RELEVANTHITS) and eliminating dead links (ELIMDEAD) are the characteristics that consumers value most. The estimated coefficients for these characteristics are positive and statistically significant in this case, while only relevant hits was statistically significant in the case of the growth model. Again the estimated coefficient on ELIMDUPL is insignificant, suggesting that duplication of results is not that bothersome to consumers.

The coefficients on ANTICIPATES and CUSTOMIZES are negative and significant; one possible explanation may be that consumers do not care about these advanced features. As noted earlier, these two characteristics are very subjective.[9]

In model II in table 4, I estimate (5) using the Lawrence and Giles data. They only examined seven search engines in the data set; hence there are just 35 observations. Since "relevant hits" appears to be a key characteristic, I also include this variable from the PC Magazine Ratings.

The results using the Lawrence and Giles data are fairly similar to the results from model II in table 4. The estimated coefficient for EARLY is positive and significant, while the estimated coefficient for AGE is negative although not quite statistically significant; the

---

[8]Note that the dependent variable is a (very non-linear) function of LUNIVERSE. Given the low correlation between LUNIVERSE and the dependent variable (0.25), this does not pose a serious problem.

[9]Again the results are also virtually unchanged if the three data points of the GoNetwork are excluded. In the case of the discrete choice model with an additional (dummy) variable for YAHOO, the coefficient on the Yahoo dummy variable is slightly larger than in the growth model, yet statistically insignificant (0.55, t=1.48). Again, this suggests that inherent quality is more important than brand name in this market.

coefficient on LUNIVERSE is positive and statistically significant.

The coefficient on RELEVANTHITS is again positive and significant. The positive and significant coefficient on the percentage of the web covered (PERWEB) suggests that consumers do care about coverage, at least up to a certain point. (Recall that no search engine covered more than 16 percent of the web.) The coefficient estimate for PERDEAD is negative as expected, but not quite statistically significant. The one surprise here is that the coefficient on median age (MEDAGE) is positive and significant. This may be because of the selection problem associated with this variable. (Recall that many of the web pages do not report the last date the page was modified.)

The discrete-choice model of product differentiation provides evidence that the results using the growth model are quite robust.

5   Conclusion

This paper examined the evolution of and competition in the emerging internet search engine market. Both a growth model and a discrete choice model of product differentiation suggest that the pure "first-mover" advantage of early entrants (Yahoo, Lycos, Excite, Infoseek, and Altavista) has declined over time. This suggests that the search engine market has relatively low barriers to entry and that competition is fierce. The results also suggest that consumers are primarily interested in search engines that provide relevant hits and, to a lesser extent, search engines that eliminate dead links. This make sense, since these two characteristics essentially measure how up-to-date is the search engine. The coverage that Notess' search engine comparison site gives to the issue of dead links provides independent support for the importance of this characteristic.

How are these results consistent with the fact that Yahoo has managed to hold its lead over time? The answer is that Yahoo has maintained its lead by continuing to innovate, that is, by providing a superior product. Indeed, Yahoo was one of only two search engines

(Hotbot was the other one) in the sample to receive an evaluation of either excellent or good in all five of the PC Magazine categories.

In this paper I examined the evolution of and competition in the internet search engine market. My analysis suggests that in internet markets with low switching costs and no obvious signs of network externalities, early entrants will benefit from their first-mover position in the long-run only to the extent that they continue to innovate and stay ahead in the quality dimensions. Pure brand rents in these markets will likely be short-lived.

6   References

Audretsch, D., and J. Mata, 1995, "The Post-entry Performance of Firms," *International Journal of Industrial Organization,* 13: 413-419.

Berry, S., "Estimating Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics,* 25: 334-347, 1994.

Berry, S., Levinsohn, J., and A. Pakes "Automobile Prices in Market Equilibrium," *Econometrica,* 63: 841-890, 1995.

Brynjolffson, E., and M. Smith (1999), "Frictionless Commerce? A Comparison of Internet and Conventional Retailers, MIT mimeo.

Gallaugher, J., and C. Downing, 1999, "Market Leadership and the web Portal Industry: An Empirical Investigation, mimeo.

Geroski, P., 1995, "What do we Know About Entry," *International Journal of Industrial Organization,* 13: 421-440.

Geroski, P., Machin, S., and C. Walters, 1997 " Corporate Growth and Profitability," *Journal of Industrial Economics,* 45: 171-89.

Lawrence, S., and L. Giles, 1998 "Searching the World Wide Web," *Science,* 280:98-100. (http://www.neci.nj.nec.com/ lawrence/science98.html).

Lawrence, S., and L. Giles, 1999 "Accessibility of Information on the Web," *Nature,* 400: 107-109.

Nevo, A., 1999, "Measuring Market Power in the Ready-To-Eat Cereal Industry, U.C. Berkeley mimeo.

Prusa,T., and J. Schmitz, 1991, "Are New Firms an Important Source of Innovation? Evidence from the PC Software Industry," *Economics-Letters*, 35: 339-342.

Rob, R., 1991, "Learning and Capacity Expansion Under Demand Uncertainty," *Review of Economic Studies,* 58: 655-675.

Sutton, J., 1997 "Gibrat's Legacy," *Journal of Economic Literature,* 35: 40-59.

Vettas, N., 1998, "Demand and Supply in New Markets: Diffusion with Bilateral Learning," *RAND Journal of Economics,* 29:215-233.

Product	Entry Date	Unique Users in Millions		
		August 1997	August 1998	August 1999
Yahoo	1994	14.8	26.0	33.0
Infoseek/Go Network	1995	7.9	12.5	18.5
Lycos	1994	4.9	12.0	14.9
Excite	1995	7.6	14.5	14.1
Altavista	1995	4.7	9.5	9.2
Webcrawler	1995	3.2		
About	1996		*	8.6
Looksmart	1996		3.2	8.5
Snap	1997		4.0	8.3
Hotbot	1996		5.5	7.2
GoTo	1998		2.6	7.1
Askjeeves	1997		0.4	4.0
Total # (in millions) of Unique Users		43.1	90.2	133.7
Total # (in millions) of Web Users		44.7	55.5	63.1

Table 1: The Search Engine Market August 1997 - August 1999. (*Data on unique visitors is not publicly available for "About" for the August 1998 period. In November 1998, About had approximately 4.2 million unique visitors.)

Variable	N	Mean	std. dev.	minimum	maximum
log(SEARCHES)	54	2.06	0.87	-0.86	3.51
log(SEARCHES/[POPULATION-UNIVERSE])	54	-3.30	0.87	-6.22	-1.83
EARLY	54	0.46	0.50	0.00	1.00
AGE	54	3.28	1.23	1.00	5.00
LUNIVERSE	54	4.08	0.05	4.02	4.15
RELEVANTHITS	54	1.91	0.68	1.00	3.00
ELIMDEAD	54	1.56	0.50	1.00	2.00
ELIMDUPL	54	1.46	0.50	1.00	2.00
ANTICIPATES	54	1.65	0.48	1.00	2.00
CUSTOMIZES	54	1.81	0.59	1.00	3.00
PERWEB	35	9.40	4.63	2.50	15.5
PERDEAD	35	5.26	3.95	2.20	14.00
MEDAGE	35	75.57	44.56	33.00	174.00

Table 2: Descriptive Statistics

	Growth Model	
Independent Variable	Coeff.	T-Stat
CONSTANT	-2.28	-0.76
Log(SEARCHES(-1))	0.72	9.20
EARLY	0.34	1.79
AGE	-0.051	-1.15
LUNIVERSE	0.69	0.93
RELEVANTHITS	0.21	2.17
ELIMDEAD	.073	0.94
ELIMDUPL	-0.10	-1.21
ANTICIPATES	0.071	0.88
CUSTOMIZES	-0.17	-2.31
Durbin Watson	2.08	
Number of Obs.	43	
Adjusted R^2	.95	

Table 3: Growth Model: Dependent Variable Log(SEARCHES)

	Model I PC Mag. Data		Model II L&G Data	
Independent Variable	Coeff.	T-Stat	Coeff.	T-Stat
CONSTANT	-22.14	-6.52	-14.51	-6.53
EARLY	3.40	9.93	2.05	5.43
AGE	-1.13	-6.54	-0.24	-1.43
LUNIVERSE	4.84	5.88	2.19	4.08
RELEVANTHITS	1.61	9.42	0.66	3.48
ELIMDEAD	0.17	1.71		
ELIMDUPL	-0.14	-1.23		
ANTICIPATES	-0.93	-4.67		
CUSTOMIZES	-0.21	-2.06		
PERWEB			0.051	3.57
PERDEAD			-.043	-1.48
MEDAGE			0.0064	4.45
Durbin Watson	1.51		1.95	
Number of Obs.	54		35	
Adjusted R^2	.89		.92	

Table 4: Discrete Choice Models of Product Differentiation: Dependent Variable: log(SEARCHES/[POPULATION-UNIVERSE]).