Software Marketing on the Internet: the Use of Samples and Repositories

by

Alexia Gaudeul
Dept. of Economics and ESRC Centre for Competition Policy
University of East Anglia

CCP Working Paper 08-23

The support of the Economic and Social Research Council is gratefully acknowledged.

ISSN 1745-9648

Software marketing on the Internet: the use of samples and repositories*

Alexia Gaudeul†

June 2, 2008

Abstract

This paper examines one of the most important marketing strategies by software producers on the Internet. That is whether to offer free samples and if so, whether to list the samples on shareware repositories. I show that firms with higher value products have a greater incentive to offer free samples but are more reluctant to do so if they are well known, and even when they do are less likely to be listed on shareware repositories. I then proceed to use four types of Probit-based models to corroborate the findings from the theoretical model.

JEL Classifications: D42, D43, D82, D83, L13, L15, L81, L86.

Keywords: Shareware; Software; Internet; Distribution; Intermediation; Directory; Repository; Advertising; Brand; Reputation; Asymmetric Information; Search; Sample.

This paper examines strategies in the marketing of experience goods over the internet. I want to explain the variety of strategies that are observed on the Internet when it comes to offering samples of one's product or not. I focus specifically on how offering samples gives access to

*Thanks go to Bruno Jullien for advising me on a previous version of this paper, to Tina Chang and Peter Moffatt for their advice on the empirical section, and to Bruce Lyons and anonymous referees for suggestions. Support from the ESRC Centre for Competition Policy is gratefully acknowledged.

†School of Economics, University of East Anglia and ESRC Centre for Competition Policy, email: a.gaudeul@uea.ac.uk, website: http://agaudeul.free.fr/

better distribution opportunities on the Internet. The paper uses software marketing as an example, but the model and its reasoning applies to any situation where firms can offer free samples of their products at little costs to themselves. For example, an online newspaper such as Economist.com may offer the first paragraphs of an article for free, a book seller such as Amazon.com may provide access to the first few pages of a book, a dating site such as Match.com may provide basic features free, and so on. Alongside those websites, there are intermediaries that provide access to those samples and then direct consumers back to the websites where products are sold. For example my.yahoo.com provides a news aggregation service for information services such as Reuter, AFP and others. books.google.co.uk offers access to extended excerpts of books and provides links to book retailers who sell the book. Download.com lists software for which trial versions are available. JSTOR gives access to the abstract and sometimes the first few pages of the papers in journals to which one is not subscribed, and so on. Bourreau and Lethiais (2007) give statistics on the extent to which marketing strategies based on the offering of free products are used on the Internet.

I focus in this paper on software distribution and marketing. Software may be offered as shareware, free on a trial basis. Shareware may have a built-in expiration date (after 30 days, the user can no longer get access to the program), or it may have some features disabled. After the trial period or in order to have access to the missing features, the consumer must buy the complete version of the program. Software publishers can offer shareware on their own website, or they can list their sample on a shareware repository. Savvy consumers will find software via a search engine or a shareware repository. Software publishers that are well-known are likely to come out on top in search results, but if the search is specific, then search results will also include specialized software publishers even if they are not very well known. Search results will also include shareware repositories. If the consumer chooses to consult a repository, then she will find out about a great number of software responding to her need and will be able to easily compare them by popularity, feedback from other consumers and of course by the features they provide. Consumers without access to the Internet, consumers who distrust online retailing and newcomers

to the Internet are more likely to know of and buy only from well-known publisher, such as those publishers that have a wide product line and that do brand advertising, are established players in the industry or have established a dominant position in one or many development areas.

Basic listing on the repository is usually free, though firms may choose to pay a fee for preferential placement.[1] Software is only checked for viruses, and spyware (software that collects private information) is normally excluded. Other than those cursory checkups, any software can be listed. Repositories attract clients and thus advertising revenues by offering an easily searchable central place with a wide diversity of software. Repositories do not usually handle transactions between the software publisher and the firm: once a consumer has sampled one firm's product and decided to buy it, she is directed to that firm's website. Repositories do not also usually play a role as a certifier, but consumers can rank results by the number of times one software was downloaded (Duan, Gu, and Whinston (2005) and Duan, Gu, and Whinston (2006) examine whether consumers will respond to that information). Repositories may however sell some add-on services (registering customers, processing payment, etc.). Some repositories are highly visible on the Internet and attract a great number of consumers (table in appendix B). While offering samples on the repositories reaches many consumers, it puts one in direct competition with existing alternatives. There is little other possible outlet however for those firms that find it the most difficult to find consumers (new entrants, niche products). This paper will contrast the strategies of the better known software publishers (Microsoft, Adobe, Apple) with that of second tier software publishers (Intuit, Symantec, Avanquest) and that of relatively obscure software publishers, such as Corel, Nova Development or Pantone. Better known software publishers will be shown to prefer not to offer samples of their product, while second tier publishers will offer samples on their website but not list their product on the repositories, thus avoiding competition. Lower tier publisher will have no other recourse than listing their product on the repositories. Firms' strategies will thus have two dimensions: offer samples or not, and if so, list the sample on the repository or only on their website.

[1] See Bhargava and Feng (2006) and Bhargava and Feng (2007) for an analysis of possible fee induced bias in search results.

Literature This paper is positioned among a large group of literature addressing the incentives for sample provision on the Internet. Its main contribution lies in the framework, both the theoretical model and the empirical testing, that links not only pricing, but marketing strategies with firm identities. The paper provides an insight into two significant factors influencing the sample provision decisions, i.e. what information delivery channel to use (own website or software repository) and which consumers to target (informed vs. uninformed). The following literature review considers other factors.

Milgrom and Roberts (1986) and Grossman (1981) consider that offering samples is one way to reduce consumer uncertainty when buying experience goods. Firms that offer samples of their products may decide to do so to increase willingness to pay but also to prevent being mixed up with low value software. Competition between firms faced with an imperfectly informed consumer leads them to disclose information about their product. In the same setting, Bourreau and Lethiais (2007) consider the cost of offering a portion of one's product for free and show that high value sellers will signal their value by offering samples. Boom (2004) reaches similar conclusions. Similar to this paper, Waldfogel and Chen (2006) consider how providing information via third parties may undermine the role of brand and reputation. However, internet intermediaries such as price comparison websites face their own reputational issues (Leggatt (2007)), so much so that consumers who used them in the past now tend to prefer going directly to well-known brands' websites.

Haruvy and Prasad (2005) consider that samples may be offered to build up a user base when network effects are important. In the special case of software such as Adobe Acrobat, a document preparation system, a firm may offer a free version of its software that only allows reading, listening or using documents, music or other types of files that were created using the full version of the software. This increases the audience for the users of the full version. In that case the free version does not only prove the value of the output generated using the full version, but also increases the value of the product to those who pay for the full version. The software publishers take into account two-sided market effects, attracting one side (non-paying

content consumers) to better attract paying content producers (Jullien (2005)). The same dynamic arguments can be made from reading Crémer (1984) and Villas-Boas (2006): samples may be offered to build up an informational advantage: a consumer knows more about a product she has tried than about a product she has not tried, so she is likely to become a regular customer after trying. Bergemann and Välimäki (2006) show that niche players will price low at the beginning so as to attract buyers while mass market players can price high and still get customers. Samples may also be offered to generate positive word of mouth; some lead consumers may be provided with free advance versions of a soon-to-be-launched product.[2] Finally, offering samples may be a way to impose or promote a standard in an industry: there is strong competition to become the dominant platform for music downloads so competitors offer free basic versions of their own music reader with their own standards (Apple's AAC/FairPlay vs. Microsoft's Windows Media Audio).

Lethiais (2001) considers that samples and other free content may be offered to generate advertising revenues. This is the case of Opera, a web browser, that offers three versions of its product: a basic one that is free, a more advanced one that is free but displays advertising, and finally, a full version that one has to pay for but that is free from advertising. Livejournal, a blog aggregator, also offers such a scheme. More worryingly, such 'adware' can turn into 'spyware', software that collects information about the user and reports it back to the software publisher who generates money by reselling that information to interested third parties, usually marketers. Offering samples may thus allow better rent extraction. Consumers with low willingness to pay have to endure advertising, consumers with high willingness to pay buy their way out of this by getting the full version of the product. Varian (2000) gives general theoretical arguments for versioning, Shapiro and Varian (1998) apply them to the business context and Ghose and Sundararajan (2005) undertake an empirical study of software versioning. Note that software publishers may expect those consumers with low willingness to pay to turn into consumers with high willingness to pay later on, who then buy software they became locked into through the use of its free version.

[2] See Chevalier and Mayzlin (2006) for an application to the online context.

I present a table in appendix C that classifies Internet retailing strategies that consist of offering free samples, free products or reduced prices. Examples of firms following those strategies are given, along with the explanation for their choice of strategy according to the literature exposed above. The contribution of this paper is to complement the 'value signaling' explanation for offering sample with an 'enhanced distribution' explanation. I examine in the model presented below the implications of assuming that offering samples gives access to enhanced distribution potentials. I then test the predictions from the model from data collected on the Internet. The paper concludes with an assessment of the accuracy of the model's predictions, some alternative explanations for the empirical findings, and some questions for future research.

1 The model

Consider a firm i selling a product of value θ_i to a unit mass of consumers. θ_i is known to the firm i but not to the consumers. Consumers expect value θ to be distributed according to the uniform probability density function over $[0, \overline{\theta}]$, bounded. The firm can choose to offer a sample or not. The firm incurs no cost in doing so, the sample is free and the consumer incurs no cost in trying the product. If firm i offers a sample, then the consumer learns its value θ_i. Portion π_i of consumers are aware of the existence of the firm i ('brand awareness' or 'visibility') and of no others. This variable π_i is a measure of the prominence of the 'virtual location' of the website (Häring (2003)). Portion μ of consumers independently visit a repository where a sample for a competing firm with a product of net value $a > 0$ ('alternative') is offered. The firm can list its product on the repository only if it offers a sample. If it lists on the repository, then those consumers who visit the repository learn of the product of firm i along with that of the alternative. The firm must offer the same price to all consumers. I will assume that consumers' expected value for firm i's product when that firm does not offer a sample corresponds to the expected value of the product of such a firm in equilibrium (this is the rational expectations hypothesis).

Suppose the firm does not offer samples. Denote θ_{NS} the expected value of the product of a firm that does not offer a sample. Portion $\pi_i(1 - \mu)$ of consumers visit the firm but not the

repository, learn there is no sample available from i and buy s.t. $\theta_{NS} - p \geq 0$. Portion $\pi_i \mu$ of consumers visit the firm and the repository, learn there is no sample available from i but that there is an alternative of net value a and buy product i s.t. $\theta_{NS} - p \geq a$. Firm i then has two choices. It can choose strategy NS under which it sells at $p = \theta_{NS}$ and makes profit:

(1) $\Pi_{NS} = \pi_i(1 - \mu)\theta_{NS}$

or it can choose strategy (NS,a) under which it sells at $p = \theta_{NS} - a$ and makes profit

(2) $\Pi_{NS,a} = \pi_i(\theta_{NS} - a)$

The firm will prefer the latter to the former s.t. $\theta_{NS} > \frac{a}{\mu}$, i.e. if consumers are relatively optimistic about the value of the firm's good.

Note that the strategy that consists of offering no sample and listing on the repository is not feasible as repositories require a sample to be listed.

Suppose now the firm offers a sample. It will list on the repository since doing so potentially allows it to reach a wider portion of consumers. The firm will sell from its own website s.t. $\theta_i - p_i > 0$ and from the repository s.t. $\theta_i - p_i > a$. If the firm wishes to sell on the repository, then it must set price $p_i = \theta_i - a$. Note that this means the firm will never list on the repository if $\theta_i < a$. Denote the strategy consisting of offering a sample and selling on the repository as 'S,R'. Profit is then:

(3) $\Pi_{S,R} = (\pi_i(1 - \mu) + \mu)(\theta_i - a)$

Denote the strategy consisting of offering a sample but not selling on the repository as 'S,NR'. The firm then sets price $p_i = \theta_i$ and makes profit

(4) $\Pi_{S,NR} = \pi_i(1 - \mu)\theta_i$

In the following remark, I note that samples will be offered by at least some firms. This is because if no sample is offered, then consumers buy only if the expected value of software is high enough. Under the rational expectations hypothesis, that expected value is the average value of the software of those firms that do not offer sample. Then, firms that are above that average will want to prove this by offering samples.

Remark 1 *At least some firms will offer samples, and those will have products of higher value than firms not offering samples.*

Proof. Suppose no firms offer samples. Then, in a rational expectations equilibrium, $\theta_{NS} = \frac{\overline{\theta}}{2}$ (consumers' expectations correspond to the expected value of the product of a firm that does not offer samples). But then firms with higher than average value will prefer (S,NR) to (NS), and (S,R) to (NS,a) as in both case they can sell to more consumers at a higher price. Therefore, I cannot have all firms not offering samples. ■

The following remark follows from the previous one and looks at the expression of profit in the case where strategy (NS,a) is chosen. It shows that if the alternative on the repository is too good (better than the average value of software in the industry), then firms will not want to adopt strategy (NS,a), and will sell only to those consumers who are not aware of the repository.

Remark 2 *If $a > \frac{\overline{\theta}}{2}$ then firm will never choose strategy (NS,a).*

Proof. Note that $\frac{\overline{\theta}}{2}$ is the best *a-priori* a firm that does not offer samples may have in a rational expectations equilibrium (this is the *a-priori* when no firm offer samples. In all other cases, from remark 1, it is higher value firms that offer samples so the *a-priori* on firms that do not offer samples decreases). This means that if $a > \frac{\overline{\theta}}{2}$ then the firm would make negative profits with strategy (NS,a) while its profits would be positive with strategy (NS) for example. ■

The following remark notes that firms with high value and low brand awareness prefer strategy (S,R) to strategy (S,NR). This is because firms with high value products can compete effectively against the alternative and will do so if the additional sales on the repository are important to them, which is true of those firms with lower brand awareness.

Remark 3 *Among firms that offer samples, firms with high value products and/or firms with low brand awareness will make sales via the repository while the others will sell only via their website.*

Proof. A firm that offers a sample will sell on the repository s.t. $\Pi_{S,R} > \Pi_{S,NR}$, that is, if $\theta_i > a(1 + \frac{1-\mu}{\mu}\pi_i)$. ■

The following proposition establishes when firms will not offer samples. There is unraveling of expectations. Consumers assume the value of those firms that do not offer samples is low and this leads firms with higher value software to offer samples. It is not possible to sustain an optimistic equilibrium whereby consumers would assume the value of those firms that do not offer samples is high so firms with high value software do not offer sample.

Proposition 1 *There is a unique rational expectations equilibrium where only firms with value* $\theta_i = 0$ *do not offer samples.*

Proof. A) Suppose $a < \frac{\overline{\theta}}{2}$ so strategy (NS,a) may be used (see remark 2).

Suppose $\theta_{NS} < \frac{a}{\mu}$. Then firm i will choose strategy (NS) rather than (NS,a). It prefers (NS) to (S,NR) s.t. $\theta_i < \theta_{NS}$ and (NS) vs. (S,R) s.t. $\theta_i < \frac{\pi_i(1-\mu)}{\pi_i(1-\mu)+\mu}\theta_{NS} + a$. Strategy (NS) is thus preferred for any $\theta_i < \min[\theta_{NS}, \frac{\pi_i(1-\mu)}{\pi_i(1-\mu)+\mu}\theta_{NS} + a]$. This means that for $\theta_{NS} < \frac{a}{\mu}(\pi_i(1-\mu)+\mu)$ then the firm chooses strategy (NS) whenever $\theta_i < \theta_{NS}$, while for $\theta_{NS} \in [\frac{a}{\mu}(\pi_i(1-\mu)+\mu), \frac{a}{\mu}]$ then it chooses strategy (NS) whenever $\theta_i < \frac{\pi_i(1-\mu)}{\pi_i(1-\mu)+\mu}\theta_{NS} + a$.

Suppose now $\theta_{NS} > \frac{a}{\mu}$. Then firm i will choose strategy (NS,a) rather than (NS). It prefers strategy (NS,a) vs. (S,NR) s.t. $\theta_i < \frac{1}{1-\mu}(\theta_{NS} - a)$ and (NS,a) vs. (S,R) s.t. $\theta_i < \frac{\pi_i}{\pi_i(1-\mu)+\mu}(\theta_{NS} - a) + a$. Strategy (NS,a) is thus preferred for any $\theta_i < \min[\frac{1}{1-\mu}(\theta_{NS} - a), \frac{\pi_i}{\pi_i(1-\mu)+\mu}(\theta_{NS} - a) + a]$. However, $\frac{1}{1-\mu}(\theta_{NS} - a) > \frac{\pi_i}{\pi_i(1-\mu)+\mu}(\theta_{NS} - a) + a$ for any $\theta_{NS} > (\mu + (1-\mu)(\pi_i(1-\mu)+\mu))\frac{a}{\mu}$, which is less than $\frac{a}{\mu}$ and is therefore true in the definition interval. This means that for any $\theta_{NS} > \frac{a}{\mu}$ then the firm does not offer samples for any $\theta_i < \frac{\pi_i}{\pi_i(1-\mu)+\mu}(\theta_{NS} - a) + a$.

Now, from the above, for any $\theta_{NS} < (\pi_i(1-\mu)+\mu)\frac{a}{\mu}$, then the firm does not offer samples for any $\theta_i < \theta_{NS}$. In a rational expectations equilibrium, I must have $\theta_{NS} = E(\theta_i | \theta_i < \theta_{NS}) = \frac{1}{2}\theta_{NS}$.

Solving this for θ_{NS} gives out $\theta_{NS} = 0$. I can check that this solution is within its domain of definition, so this constitutes a rational expectations equilibrium.

Suppose now $(\pi_i(1-\mu)+\mu)\frac{a}{\mu} < \theta_{NS} < \frac{a}{\mu}$, then the firm does not offer samples for any $\theta_i < \frac{\pi_i(1-\mu)}{\pi_i(1-\mu)+\mu}\theta_{NS} + a$. In a rational expectations equilibrium, I must have $\theta_{NS} = E(\theta_i|\theta_i < \frac{\pi_i(1-\mu)}{\pi_i(1-\mu)+\mu}\theta_{NS}+a) = \frac{1}{2}(\frac{\pi_i(1-\mu)}{\pi_i(1-\mu)+\mu}\theta_{NS}+a)$. Solving this for θ_{NS} gives out $\theta_{NS} = a\frac{\pi_i(1-\mu)+\mu}{\pi_i(1-\mu)+2\mu}$. I can check that this solution is not within its domain of definition as $a\frac{\pi_i(1-\mu)+\mu}{\pi_i(1-\mu)+2\mu} < (\pi_i(1-\mu)+\mu)\frac{a}{\mu}$, so this does not constitute a rational expectations equilibrium.

Suppose now $\theta_{NS} > \frac{a}{\mu}$. Then as seen above, I can check that the firm will not offer samples for any $\theta_i < \frac{\pi_i}{\pi_i(1-\mu)+\mu}(\theta_{NS}-a)+a$. In a rational expectations equilibrium, I must have $\theta_{NS} = E(\theta_i|\theta_i < \frac{\pi_i}{\pi_i(1-\mu)+\mu}(\theta_{NS}-a)+a) = \frac{1}{2}\frac{\pi_i}{\pi_i(1-\mu)+\mu}(\theta_{NS}-a)+a$. Solving this for θ_{NS} gives out $\theta_{NS} = \frac{(1-\pi_i)\mu}{2\mu(1-\pi_i)+\pi_i}a$, which is less than $\frac{a}{\mu}$ so there is no rational expectations equilibrium in that interval.

B) Suppose now $a > \frac{\overline{\theta}}{2}$ so strategy (NS,a) may not be used (see remark 2). Then again the only equilibrium is such that firm of value $\theta_i = 0$ do not offer samples, and other firms choose strategy (S,R) s.t. $\theta_i > a + a\frac{1-\mu}{\mu}\pi_i$, and (S,NR) otherwise. ■

This proposition allows me to define several zones in the reputation/value space where firms will choose strategies (S,R), (S,NR), (NS) or (NS,a) depending on the nature of the equilibrium and their characteristics (θ_i, π_i): firms will never choose strategy (NS,a) and only firms with a product of value 0 do not offer samples. The rest of the firms choose their strategy according to remark 3. In the following, I will show the result of internalizing the value of a as the maximum net utility gained from any firms on the repository. This choice of definition for a is based on the assumption that consumers are able to accurately identify the best deal from all firms listed on the repository. Intuitively, both value and visibility influence the decision to sell on the repository; the firms with the highest value will not fear competition from other firms, and will be all the more motivated to list on the repository as they are less visible on the Internet. This means that the higher the value and the lower the visibility, the more likely the firm will be to list on the repository.

Proposition 2 *Only firms with the highest ratio* $\frac{\theta_j}{1+\frac{1-\mu}{\mu}\pi_j}$ *in the universe of firms may sell via the repository.*

Proof. Consider two firms i and j and suppose one of them, j, is listed on the repository. Take as the value of a the net utility from buying from that firm, so $a = \theta_j - p_j$. Suppose firm i considers whether to sell via the repository as well, in which case it must set its price $p_i = \theta_i - a$, otherwise it would make no sales via the repository. I thus have $p_i = \theta_i - \theta_j + p_j$. From remark 3, firm i will list on the repository s.t. $\theta_i > (\theta_j - p_j)(1 + \frac{1-\mu}{\mu}\pi_i)$. Firm j has two options: either set its price such that firm i does not enter the repository, so it has to set its price $p_j = \theta_j - \frac{\theta_i}{1+\frac{1-\mu}{\mu}\pi_i}$ and makes profit $(\pi_j(1-\mu)+\mu)(\theta_j - \frac{\theta_i}{1+\frac{1-\mu}{\mu}\pi_i})$, or allow firm i in, in which case it makes no sales via the repository (only one firm makes sales on the repository, and firm i will enter only if it makes sales on the repository). In that latter case, firm j is reduced to sell via its own website at price $p_j = \theta_j$ and makes profit $\pi_j(1-\mu)\theta_j$. Firm j will want to stay on the repository only if $(\pi_j(1-\mu)+\mu)(\theta_j - \frac{\theta_i}{1+\frac{1-\mu}{\mu}\pi_i}) > \pi_j(1-\mu)\theta_j$, which translates into $\frac{\theta_j}{1+\frac{1-\mu}{\mu}\pi_j} > \frac{\theta_i}{1+\frac{1-\mu}{\mu}\pi_i}$. It will be indifferent between staying on the repository or leaving if $\frac{\theta_j}{1+\frac{1-\mu}{\mu}\pi_j} = \frac{\theta_i}{1+\frac{1-\mu}{\mu}\pi_i}$.

This means that only firms with the highest ratio $\frac{\theta_j}{1+\frac{1-\mu}{\mu}\pi_j}$ may list on the repository, and none of the others will. That value $\frac{\theta_j}{1+\frac{1-\mu}{\mu}\pi_j}$ is what was referred to as 'a' previously. ■

In a universe with an infinity of firms spanning the whole range of possible brand recognition and value, then, among those firms that offer samples, only the firm with the highest value ($\theta = \overline{\theta}$) and the lowest reputation ($\pi = 0$) will sell via the repository. The following corrolary examines the consequences of having a bound h on the value of the ratio $\frac{\theta_j}{1+\frac{1-\mu}{\mu}\pi_j}$ in the above proposition. In that case, there will be a range of firms that may be listing on the repository:

Corollary 1 *Suppose there is a bound* $h \leq \overline{\theta}$ *to the value of the ratio* $\frac{\theta_j}{1+\frac{1-\mu}{\mu}\pi_j}$ *over all firms. Consider the set of firms with ratio equal to* h*. Then only those firms may sell via the repository. They will set their prices as an increasing function of their visibility* π_j*, such that* $p_j = h\frac{1-\mu}{\mu}\pi_j$*.*

Proof. Suppose there is a bound $h \leq \overline{\theta}$ to the ratio's value, then all firms with $\frac{\theta_j}{1+\frac{1-\mu}{\mu}\pi_j} = h$ may list on the repository and will price at $p_j = \theta_j - h$. There will thus be a relationship such that $p_j = h\frac{1-\mu}{\mu}\pi_j$, so that price and visibility will be positively correlated. ■

The following graph (Figure 1) illustrates proposition 1 and its corollary 1. π ranges from 0 to 1 and θ ranges from 0 to $\overline{\theta}$. The hashed area is devoid of firms. On the right of the graph, bordering the limit of the hashed area, are those firms that use strategy (S,R). On the left are those firms with the lowest value products, that offer no samples. In between the two lines are all the firms using strategy (S,NR).

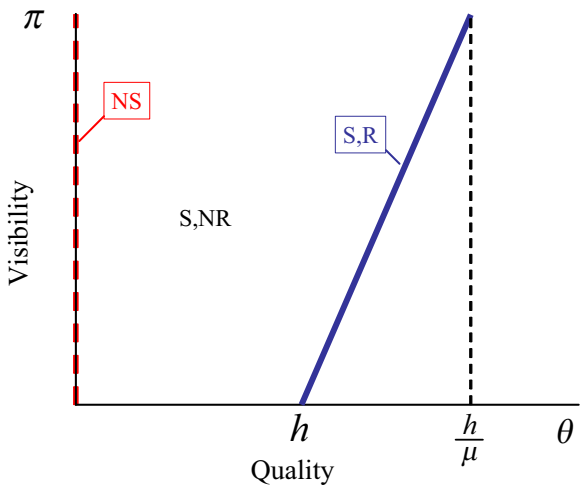

Figure 1: Graphical representation of strategies in the (θ, π) space.

I can now summarize the results from the model into the following hypotheses, which I will then test in the empirical part:

H1: Firms will use three types of strategies, either not offering samples or offering samples, and when offering samples, either selling via the repository or not (remark 3 and proposition 1).

H2: Everything else being equal, publishers with lower value products are less likely to offer samples (proposition 1).

H3: Among those firms that offer samples, those with higher value products are more likely to be selling via the repository (remark 3).

H4: Among those firms that offer samples, those with better visibility are less likely to be selling via the repository than less well known publishers (remark 3).

H5: Among those firms that offer samples and sell via the repository, there will be a positive

correlation between price and visibility. There will be no such correlation among other firms (corollary 1 to proposition 2).

2 Empirical analysis

I selected software in May 2006 from two databases, CNET download.com which is the dominant shareware repository (see table in appendix B), and amazon.com, a dominant Internet retailer. Software in each database is classified by category, and both data sources classify software broadly along the same scheme. Software that was patently misclassified was excluded. Six categories were selected for further data collection: databases, word processors, image editing, personal finance, software and programming. Software listed on CNET download.com was ranked by the number of times consumers downloaded the software. The top ten software were selected.[3] Similarly, the top ten software by sales in each category were selected from amazon.com. I checked through the Google search engine that this sampling did not miss any important software in each category. Software that appeared among the top ten results from Google when searching for software in a category was also included if it did not appear already, which happened only once. This sample selection strategy, where Google search results complement data from internet retailers and repositories, is a combination of techniques employed in Häring (2003) or Clay, Krishnan, and Wolff (2001) for example. There was very little overlay between software sampled on download.com and on amazon.com as only two software was listed among the top ten on both repositories. The total sample size was 116. The sample was distributed

[3] Since listing on the repository is free, all software that offers samples will potentially list on the repository even if its price is too high to compete. This is why I selected only the top ten software on the repository. Software that was listed among the top ten downloads on the repository was included not only because more downloads presumably means more sales, but also because consumers tend to choose relative popularity as their main way to arrange software listing when searching a repository (Duan, Gu, and Whinston (2006)) and will tend to ignore any software that does not appear on the first page of results.

One could also have selected software by the total number of downloads on the repository. However, the length of time the product was available differs greatly firm by firm, so a product that appears to have been downloaded often may have been so because it has been listed for a long time. As an alternative, taking the daily average of downloads does not either take into account the difference between older product that were present when download.com was starting, and newer products that benefit from the present popularity of this site. Moreover, recently listed products often display high daily download numbers because of novelty, and this number then tapers off in a way that is difficult to predict. Using the 'number of downloads last week' suffers from the same novelty bias.

among categories and strategies as follows (Table 1):

	Strategy			
Category	S,R	S,NR	NS	Total
Backup	10	1	9	20
Database	10	5	2	17
Illustration	10	6	4	20
Personal finance	10	3	7	20
Programming	10	5	4	19
Word processors	11	6	3	20
Total	**61**	**26**	**29**	**116**

Table 1: Sample composition, by category and strategy

Information was collected on the price of the software, the availability of samples on the software publisher's website, the terms under which samples were distributed, the number of backlinks to the internet address of the software publisher, the ratings of the software by users at amazon.com and download.com and the rating by the editors at download.com.[4]

- The number of websites linking to a publisher's website ('backlinks') was used as a proxy for visibility π and was collected through Google. Typing link:www.websitename.com in the Google search query field returns the number n_i of 'backlinks' to website i. This is a measure of visibility on the Internet that Google, the dominant search engine, uses to determine the ranking of that website in response to a specific query: a website that is linked to more often will be listed before a website that is linked to less often if both websites, through automated content analysis, seem to fit the consumer's query equally adequately (see Weiss (2005) for a discussion of the robustness of this measure of visibility). The number of backlinks will be converted in logarithmic scale. This is because the higher the number of links the lower is the contribution of one additional link in terms of likelihood to be found.[5]

[4]Other data on CNET download.com was collected as well, including the size of the downloads, the number of dowloads made since listing, 'last week's downloads' and the number of ratings made.

[5]Consider for example, two sites 1 and 2, where n_1 sites link to site 1 and n_2 sites link to site 2. Suppose there are $N = n_1 + n_2$ sites and all sites link either to site 1 or 2. In a random search the probability one finds a site that links to site 1 is $\pi_1 = \frac{n_1}{n_1+n_2}$. The logarithm of n_1 approximates the shape of this function.

- Price will be used as a proxy for value θ. The alternative to using price as a measure of value would be very data intensive and would involve designing comparable measures of value across software categories.[6] The model's prediction on the relation between the price of the software, its visibility and the publisher's choice of strategy are analogous to the predictions linking value, visibility and the choice of strategy. Indeed, whatever the strategy, price will be increasing with value. This choice of proxy is based on the assumption that consumers are not systematically wrong in their evaluation of offered alternatives. This choice also assumes away such things as lock-in effects, brand premium from advertising, network effects favoring the established players, etc... Denote p_i is the price of software i, j the category to which software i belongs. Two estimates of price were used, either from the price quoted at the repository or on the software publisher's website.

- I experimented with another variable, r_i, the rating of the software, as an alternative for the use of price as a proxy for value. There were three such measures, user and editors rating at download.com and user rating at amazon.com. Both of those measures were on a scale of 1 to 5 stars. Amazon and Download both provide guidelines as to the meaning of each star category (appendix A). Ratings were available for 89 of the 116 software in the sample.[7] In order to account for possible discrepancy between ratings at both websites, they were estimated in relative terms vs. the average rating of software on the same repository.[8] Software that receives lower rating by consumers may be less likely to be available as shareware so as to hide its type. Alternatively, low rated publishers may wish to offer samples to reassure the consumers that its product is underrated.

- I also used another variables as an alternative for 'backlinks', which measures visibility.

[6] There are some examples of the use of price as a measure of value in the literature on cross industry analysis of intra-industry trade (Greenaway, Hine, and Milner (1995)).

[7] Ratings based on less than 5 opinions were ignored unless, in the case of Amazon.com, a sufficient number of consumers signaled the reviews were useful (indeed, a review can be so 'definitive' that no reviewers would submit additional reviews).

[8] User ratings do not necessarily express the absolute value level of the software, or even its value in relative terms vis-à-vis the competition, but its value vs. what the consumer expected from the product. Products that are good value for money or exceed expectations will therefore tend to be highly rated, rather than products that are of higher value.

w_i, is an indicator categorizing publishers, taking value 1 if more than 4 software in the sample belonged to the publisher of the software, and value 0 else. This will be a proxy for 'brand recognition'. Diversified publishers, whose products take value 1, are usually either generalists like Microsoft and Apple, or publishers who have gained a position in one software category (Symantec for anti-virus software, Avanquest for database software, Adobe for illustration software, Intuit for financial software). Software benefits from the reputation of their publisher; consumers may have used previous versions and will trust the publishers. This makes offering samples or being listed on the repository less necessary.

Analysis Firms used the full set of strategies hypothesized in the model (**H1**). There was a significant relationship between price, visibility and the choice of strategy as seen in the following scatter plot that relates visibility ($\ln n_i$) and price with strategy (Figure 2):

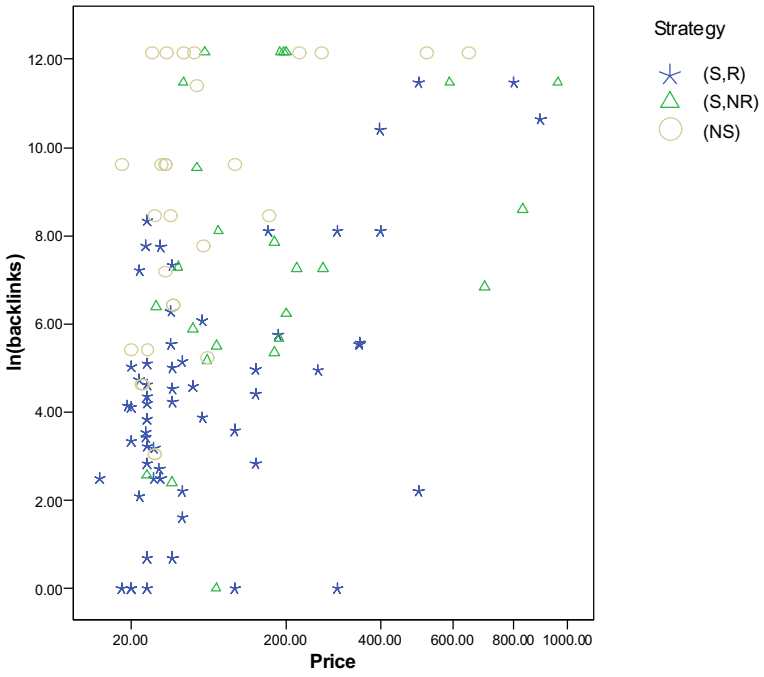

Figure 2: Price and number of backlinks, indexed by strategy.

As seen above, there are very few firms on the right of a diagonal starting from the origin to the top right corner of the graph. This supports corollary 1. Firms that do not offer samples

are on the left hand side of the graph, as hypothesized in **H2**. The prices of those firms that adopt strategy (S,NR) do not seem to differ in a systematic way from the prices of those firms that adopt strategy (S,R), which goes against hypothesis **H3**. Firms that offer samples and are listed among the top ten downloads on the repository are primarily among the less well-known, as hypothesized in **H4**. We will see there is a significant correlation between price and visibility among those firms (as hypothesized in **H5**), though this correlation also exists for those firms using strategy (NS). Those observations on the empirical distribution of strategies in the sample fit well with the model as seen in Figure 1 and are globally in support of the hypotheses from the model. In the following, I will test the statistical significance of those observations.

Four probit based regressions were performed. The first model (Probit 1) is a simple probit model that predicts the decision to offer samples or not based on the variables collected. The second model (Probit 2) considers the sub-sample of those firms that do offer samples, and uses a simple probit model to predict the decision to list on the repository. I limit the regression to this sub-sample because listing on the repository is only open to those firms that do offer samples. A third model (Heckman Probit) uses a Heckman probit selection model (see Heckman (1979)), thus using all the data available and taking into account the self-selection effect (firms not offering samples exclude themselves from listing on the repository). The decision of firms is thus supposed to be as follows: First choose whether to offer a sample or not, and if offering a sample is chosen, then choose whether to sell via the repository or not. The decision to offer a sample or not is therefore the selection variable, while the decision to list on the repository is the outcome variable.[9] A fourth model (Sartori) uses Sartori's estimator which is better fit for this special case in which identical explanatory factors influence selection and the subsequent outcome of interest (Sartori (2003)).

[9]One could also assume that firms first choose whether to list on the repository. If they do so they have to offer a sample. If they do not do so they can still choose to offer a sample on their website. However, one cannot observe variations in the choice whether to offer samples or not among those selected firms that choose to list on the repository. This is why one first has to evaluate the decision whether to offer a sample or not, and then the decision whether to list on the repository.

	Probit 1	Probit 1	Probit 2	Probit 2	Heckman Probit	Sartori estimator	Sartori estimator
	(1)	(2)	(1)	(2)		(1)	(2)
Dependent variable = **1 if available as sample, 0 else.**							
ln(backlinks)	-0.239***	-0.152**			-0.245***	-0.223***	-0.153**
	(-4.74)	(-2.30)			(-4.92)	(-4.84)	(-2.45)
publisher (w)		-0.821*					-0.649
		(-1.94)					(-1.62)
price	0.003***	0.003***			0.003***	0.003***	0.003***
	(2.96)	(2.91)			(3.11)	(2.89)	(2.75)
constant	1.861***	1.568***			1.883***	1.750***	1.520***
	(5.40)	(4.31)			(5.58)	(5.38)	(4.39)
Dependent variable = **1 if available on the repository, 0 else**							
ln(backlinks)			-0.202***	-0.157**		-0.256***	-0.180***
			(-3.45)	(-2.23)		(-5.48)	(-3.09)
publisher (w)				-0.542	-0.765**		-0.817**
				(-1.14)	(-2.21)		(-2.05)
price			0.000	0.000		0.001*	0.001
			(0.33)	(0.30)		(1.67)	(1.49)
constant			1.651***		0.931***	1.451***	1.211***
			(4.86)		(5.89)	(5.07)	(4.02)
athrho constant					-3.477		
					(-0.04)		
Pseudo R-square	0.225	0.254	0.165	0.178			
chi2	29.840	33.66	17.397	18.71	4.876	24.502	26.80
p	0.000	0.000	0.000	0.000	0.027	0.000	0.000
N	116	116	86	86	116	116	116

Significance levels: *<10% **<5% ***<1%, t-values in parenthesis.

Table 2: Estimation results

From table 2, the 'Probit 1(1)' model shows that the websites of firms that offer samples have higher prices (as hypothesized in **H2**) and they are significantly less linked to than others. This is easily understood in the context of the model, as firms with lower visibility have the most to gain from offering samples as it opens access to listing on the repository. Whether the firm was part of a diversified software publisher ($w_i = 1$), an alternative measure of visibility, also made it less likely to offer samples (Probit 1(2)), but r_i (rating by consumers, an alternative measure of value) was not a significant predictors of the decision of the firm (results not shown).

As a matter of interpretation of the results, take as the base case the average firm with price $p_i = \$140$ and visibility $\ln(n_i) = 6$. Statistically, there is 80% probability that this average firm will be observed to offer a sample. If its price was instead one standard deviation higher at $\$340$, everything else remaining the same, then this probability would be higher at 94%, while if its price was $\$0$, then this probability would be lower at 65%. If visibility was instead one standard deviation higher at 9.5, then it would be observed to offer a sample with lower probability 63%, while if visibility was instead 2.5, then this probability would be higher at 92%. As can be seen, the effect of changing price and visibility by one standard deviation is about the same; both factors have similar importance from a statistical point of view. This will also be the case in all results exposed below.

The 'Probit 2' model considers the sub-sample of firms that offer samples and examines their decision to list on the repository. As hypothesized in **H4**, firms that are listed on the repository are significantly less linked to than firms that are not listed there. No variables in addition to n_i turned out to be significant predictors, though w_i performed almost as well as n_i when replacing it as a predictor (result not shown). This is because the two are correlated as multi-product firms gain from brand visibility established among several products (The coefficient of correlation was 0.71, significant at the 1% level).

The 'Probit 1' and 'Probit 2' models thus means that n_i influences both the decision to offer a sample and to list on the repository, while from the model it should **only** influence the decision whether to list on the repository. p_i influences only the decision to offer a sample, while according to the model it should **also** influence the decision whether to list on the repository. w_i was shown to also influence the decision whether to offer a sample or not so that both the proxy for visibility and the proxy for brand recognition are a factor in the decision to offer a sample.

The 'Heckman' model is the full model taking into account the selection effect, i.e. the fact firms that can list on the repository are a sub-sample of all firms. I cannot use the same variables in both the outcome and selection equation in a Heckman selection model, so I will use w_i in the outcome equation and n_i in the selection equation. The results confirm that firms that offer

samples have higher price (**H2**) and are less well known than other firms, and among those firms that offer samples, those that sell via the repository are less well-known publishers (**H4**).

The 'Sartori' model uses Sartori's estimator (Sartori (2003)), which allows me to use the same explanatory variable for both the selection and outcome variable. This estimator is well-adapted to a model in which selection among alternatives is non-random and quasi-simultaneous, and identical explanatory variables influence the decision. The results are essentially the same in the two equations as those from the two separate probit models and considerably improve the predictions compared to the Heckman probit model. w_i, used jointly with n_i turns out to be a significantly positive predictor of whether the firm lists its sample on the repository or not (Sartori 2). As usual, r_i does not turn out to be a significant variable. There is thus both a 'brand' (or 'product line' or 'diversification') effect at work in addition to plain Internet visibility in the decision whether to offer samples or not.

Finally, I also explored correlations patterns between price and visibility. Overall, there was significant correlation between price p_i and visibility $\ln(n_i)$ (0.405, significant at the 1% level). The correlation was not significant among those firms that use strategy (S,NR) (0.330, not significant), as expected from **H5**. The highest correlation, as expected from hypothesis **H5**, was for firms that use strategy (S,R) (0.559, significant at the 1% level). However, correlation was also high, though less significant, among firms that use strategy (NS) (0.470, significant at the 5% level). This later correlation is not explained within the model and is not consistent with hypothesis **H5.**

In conclusion to this part, I find strong support for hypotheses **H1, H2, H4** and for the first part of hypothesis **H5**. Hypothesis **H3** is rejected (firms with higher value products are not more likely to be selling via the repository). I also find two additional effects not predicted in the model: firms that are more visible (high n_i) are less likely to offer samples, and firms that are diversified are less likely to list on the repository. The discussion that will follow will throw some light on those additional findings, but I first discuss additional possible factors:

I explored other factors that might influence the publisher's strategic decisions. The category the software was in did not influence the decision. I can thus reject the idea that offering samples is significantly more difficult or easier in some software categories than in others, or that conventions may have led consumers to expect to receive samples in some specific product categories and not in others. The identity of the software publisher, for those software publishers with many software in the dataset[10] did not improve the predictions from the model as well. Offering samples or not does not therefore seem to be a global strategy set at the publisher's level for all its software. Expressing visibility in relative terms vs. other software in the same category did not change the results. However, there was little variation in average number of backlinks by category, so it is not possible to test whether firms consider visibility relative to the repository or relative to their competitors to decide on whether to list on the repository or not (the model assumes they compare their visibility relative to the repository and then decide whether to list there or not). Expressing price in relative terms did change the results and make them less significant. This is consistent with the model if I assume that the distribution of types is the same across categories. I also looked for systematic differences depending on whether network effects or standards were particularly important in specific development areas, but no significant effect was observed.

I was mindful when collecting data to note whether the try-out version was 'adware', i.e. advertising financed, or whether the software offered was merely a reader-only version for documents created by another program. There were however only very few of those types of try-out versions in the sample, as adware and reader-only versions are usually freeware and not shareware. There was little evidence of competition based on try-out terms, which goes against the thesis exposed in Bourreau and Lethiais (2007): Data on the terms of the try-out version was collected for 84 of the 87 software that offered samples. Fifty-nine software put a limit on the length of use of the try-out version, of which 40 chose to set a limit at 30 days of free trial and 13 offered 15 days. The maximum try-out period length was 90 days, the minimum 10. The other 24 software put various limits on the try-out version: 16 offered only a sub-sample of features, and

[10] Twelve from Microsoft, six from Avanquest, and five from Adobe and from Symantec.

the rest either limited the number of uses that could be made of the software or crippled output from the software sample by for example putting watermarks on the documents processed by the software. Three software did not seem to put any limitation on the use of the try-out version.

Table 1 shows that software publishers may take into account the likely reduction in willingness to pay after sampling; samples were less likely to be available for software that responds to punctual needs: only one backup software and three personal finance software chose strategy S,NR. In that later category, samples were not available for that software that responds to tax filing needs. However, the non-availability of samples in those categories may be due to the low average price of software in those categories, as proposed in the model, rather than to any specific category effect. Indeed, 'backup' and 'personal finance' were the categories with the lowest average price ($50) compared with the other categories where average price was around $200 (Table 3).

	price (in $)		ln(backlinks)	
Category	Mean	Std Dev.	Mean	Std Dev.
Backup	52.40	48.69	6.13	3.55
Database	155.24	137.67	5.48	3.16
Illustration	173.40	231.75	6.84	2.85
Personal finance	47.05	46.70	4.71	3.77
Programming	250.68	298.05	7.65	3.43
Word processors	174.00	212.39	6.06	3.93
Total	**140.85**	**197.25**	**6.15**	**3.53**

Table 3: Average and standard deviation of price and visibility of software by category.

There were two software that offered a sample on the repository but did not seem to be offering samples on their website, which is not predicted by the model, and goes against **H1**. However, this is only a very small portion of the sample. Still, there were instances in which firms made it relatively difficult to obtain a sample, by for example requiring registration or sending samples only by post rather than offering to download the try-out version on their website. As hypothesized in the model, there was little use of price discrimination, as prices at the shareware repository almost always the same as on the publisher's website. A few discrepancies in prices were due to new versions of the software being offered on the website and/or to a degree of

latency in price menu changes or in ranking changes: new products take some time to reach the top of sales or downloads ranking.[11] There was however some evidence of discrimination in price between amazon.com and publishers' websites. Prices on amazon.com always were lower or equal to those on the publisher's website. This was often due to amazon.com listing an older version of the software that was no longer available on the website and had been replaced by a newer, pricier version.

3 Discussion

This paper explained that some firms, those that are better known but whose product does not compare too favorably with the existing less well-known competition, may want to avoid being put into direct competition with lesser-known products, and thus decide to sell at higher prices to consumers who are less aware of the alternatives. There may however be other explanations why better known firms may decide not to list on repositories, or, if they list, for them not being listed among the top selling firms on a shareware repository.

One explanation may be that consumers at the repository are in search of a bargain or of a basic product, and would not anyway buy better known, higher value products. Following on from this, a stigma may have become attached to listing on shareware repositories; it would signal one is targeting the low end of the market. This would explain our finding that diversified software publishers are less likely to list on the repository: the stigma associated with listing is more important to firms that have a product line reputation to defend than to others. That stigma would be self-fulfilling, as consumers in search of better value products would know not to go to the repository. It may also be that consumers expect firms that do not offer samples to be of good value because the mere fact the firm exists means other consumers trusted it enough to buy its product without sampling. Firms that do offer samples would not be bought from because offering a sample would be interpreted as a desperate measure to attract consumers to

[11] In the case of download.com, publishers are able to add downloads of older versions of their products to the downloads of their newer versions, thus keeping the ranking of the older version.

what is a lower value product. It might also be that such prejudice prevents established software publishers from realizing the importance of such distribution channels. They might also not want to encourage the emergence of those repositories by listing their products there. Consumers who visit the repository may do so as a last resort when established software did not fulfill their needs. They would be looking for alternatives to software they already know, and that already known software will usually be from one of the established software publishers. This means there would be little point for established software publishers in listing on repositories that are geared toward those consumers that are dissatisfied with their offering. Again, this would explain our additional finding that diversified software publishers are less likely to list on the repository.

Consumers may be more confident in the value of products offered by highly visible firms, since those firms will probably have attained high visibility by establishing their reputation over time in their development area. Consumers' uncertainty about their taste for highly visible product will be low, since highly visible products would normally have come to be so through advertising campaigns, those campaigns allowing the consumer to evaluate the product in an alternative way to sampling. Better-known products would then not need to offer samples because that would not bring much additional information to the consumer. This would explain our additional finding that firms that are more visible (high n_i) are less likely to offer samples

Finally, better known firms may be able to call upon more resources than other firms, and thus be able to produce better value or more complete software, which would explain the correlation between price and visibility among those firms that do not offer samples. This would be an explanation for the correlation between price and visibility observed in the data among those firms that do not offer samples.

Consumers' risk aversion might also play a role: higher priced, higher value products are more risky than lower priced ones so that publishers of those products would be more likely to offer samples in order to reduce uncertainty. It may also be that higher priced products are more complex, and thus more difficult to evaluate from technical documentation alone, which means offering a sample may be the most expedient option. Consumers would want to sample more

complex products that require higher learning costs before committing themselves to buying them. This would be an additional explanation for higher priced firms being more likely to offer samples in the data.

Compound sampling strategies may arise and play a role if consumers are uncertain about whether their need for the product is genuine or not. Those consumers would then use the products available on the repository as a way to test products in a product category. Products on the repository would serve as a stepping stone into the use of those products that are not listed there. An externality would thus be generated whereby trying a product reveals one's own taste for the competitor's product. This can be likened to Crémer (1984) who examines how a monopoly that offers coupons for the next purchase of its good discriminates between first time and second time buyers, offering low prices in a first period to expand the number of experienced buyers in the second period. In the context of this paper, the firm that does not offer samples benefits from the increase in the number of experienced buyers that is brought about by the existence of free or cheap alternatives on the repository.[12] This would explain why, from the data, firms that are highly visible on the Internet are less likely to offer samples; they do not need to do so as consumers try other products before buying their own.

4 Conclusion

This paper shows a significant relationship between the choice of a strategy and the prominence of a software publisher's website. Samples of lower value software from better-known publishers are usually not available, while samples of higher value software from less well-known publishers are often available. Less well-known software publishers tend to offer samples and be listed among the top downloads on the repository.

This paper examined one aspect of the strategies followed by firms on the Internet and provides a better understanding of the Internet marketing environment. The decisions by firms to

[12] Another common strategy is to start with a low-range and cheap product and then gradually learn through experience what one's needs are so as to later on buy a more sophisticated and more adequate product.

offer samples was shown to depend on the proportion of consumers who are aware of shareware repositories (are repositories well established or is their role limited?), on the consumers' ability to evaluate their need for a product (are consumers confident and knowledgeable, or do they lack sufficient information on products sold via the Internet?), and on the extent to which search engines and other Internet tools make the Internet a level playing field where opportunities to sell depend only on the product's own merit rather than on its visibility. The paper's empirical finding, as explained in the context of the model that is presented in this paper, confirms that not all consumers know of the existence of shareware repositories, so firms can profitably exploit less knowledgeable consumers even at the expense of sales on the repositories. Firms were also shown to be able to build on their strong position on the Web; seen in another light, this means the Internet does not make reputation or location irrelevant. Search technology is not used by enough consumers or is not efficient enough to identify all available alternatives. This means firms are not motivated to offer samples and provide consumers with all the information they need to perfectly rank alternatives. Whether such a state of imperfect competition will persist depends on whether firms keep on being able to advertise their product successfully enough to bypass repositories and avoid having to offer samples of their products, or if on the contrary consumers learn to better use Internet tools so firms are motivated to offer them more information about their product.

References

BERGEMANN, D., AND J. VÄLIMÄKI (2006): "Dynamic Pricing of New Experience Goods," *Journal of Political Economy*, 114(4), 827–858.

BHARGAVA, H. K., AND J. FENG (2006): "Pay for Play: Sponsored Recommendations in Information Gatekeepers," Discussion paper, UC Davis and University of Florida.

——— (2007): "Sponsored Results in Intelligent Recommenders: Impact on Quality Signaling and Consumer Welfare," Discussion paper, UC Davis and University of Florida.

BOOM, A. (2004): "'Download for Free' - When do Providers of Digital Goods Offer Free Samples," Discussion paper, Freie Universität Berlin, 2004/28.

BOURREAU, M., AND V. LETHIAIS (2007): "Pricing Information Goods: Free vs. Pay Content," in *Internet and Digital Economics*, ed. by E. Brousseau, and N. Curien, pp. 345–367. Cambridge University Press: Cambridge, UK.

CHEVALIER, J. A., AND D. MAYZLIN (2006): "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43(3), 345–354.

CLAY, K., R. KRISHNAN, AND E. WOLFF (2001): "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry," *Journal of Industrial Economics*, 49, 521–539.

CRÉMER, J. (1984): "On the Economics of Repeat Buying," *Rand Journal of Economics*, 15(3), 396–403.

DUAN, W., B. GU, AND A. B. WHINSTON (2005): "Do Online Reviews Matter? - An Empirical Investigation of Panel Data," Discussion paper, University of Texas at Austin.

——— (2006): "Herd Behavior and Software Adoption on the Internet: An Empirical Investigation," Discussion paper, University of Texas at Austin.

GHOSE, A., AND A. SUNDARARAJAN (2005): "Versioning and Quality Distortion in Software? Evidence from E-commerce Panel Data," Discussion Paper 05-14, NET Institute.

GREENAWAY, D., R. HINE, AND C. MILNER (1995): "Vertical and Horizontal Intra-Industry Trade: A Cross Industry Analysis for the United Kingdom," *The Economic Journal*, 105(433), 1505–1518.

GROSSMAN, S. J. (1981): "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law & Economics*, 24(3), 461–83.

HÄRING, J. (2003): "Different Prices for Identical Products? Market Efficiency and the Virtual Location in B2C E-Commerce," Discussion Paper 03-68, ZEW.

HARUVY, E., AND A. PRASAD (2005): "Freeware as a Competitive Deterrent," *Information Economics and Policy*, 17, 513–534.

HECKMAN, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–162.

JULLIEN, B. (2005): "Two-sided Markets and Electronic Intermediaries," *CESifo Economic Studies*, 51(2-3), 233–260.

LEGGATT, H. (2007): "British consumers wary of price comparison websites," Bizreport, http://www.bizreport.com, August 27.

LETHIAIS, V. (2001): "La Tarification des Services sur Internet: Intégration Verticale et Gratuité," *Revue Economique*, 52(7), 39–56.

MILGROM, P., AND J. ROBERTS (1986): "Relying on the Information of Interested Parties," *RAND Journal of Economics*, 17(1), 18–32.

SARTORI, A. E. (2003): "An Estimator for Some Binary-Outcome Selection Models without Exclusion Restrictions," *Political Analysis*, 11(2), 111–138.

SHAPIRO, C., AND H. VARIAN (1998): "Versioning: The Smart Way to Sell Information," *Harvard Business Review*, pp. 106–114.

VARIAN, H. R. (2000): "Versioning Information Goods," in *Internet Publishing and Beyond: The Economics of Digital Information and Intellectual Property*, ed. by B. Kahin, and H. R. Varian, pp. 190–202. MIT Press: Cambridge, MA.

VILLAS-BOAS, J. M. (2006): "Dynamic Competition with Experience Goods," *Journal of Economics & Management Strategy*, 15(1), 37–66.

WALDFOGEL, J., AND L. CHEN (2006): "Does Information Undermine Brand? Information Intermediary Use and Preference for Branded Web Retailers," *The Journal of Industrial Economics*, 54(4), 425–449.

WEISS, D. (2005): "Measuring Success of Open Source Projects Using Web Search Engines," in *Proceedings of The First International Conference on Open Source Systems (OSS 2005)*, ed. by M. Scotto, and G. Succi, pp. 93–99.

A The meaning of user ratings

The meaning of user ratings as specified at www.download.com:

> **1 star:** Though it may have some useful functions, this product is unstable, riddled with spyware, or otherwise difficult to use. Download at your own risk.
> **2 stars:** This program has major flaws, though it is not completely without merit. I cannot recommend this download unless it is the only one of its kind.
> **3 stars:** This is a decent, solid program. It has no major flaws, but it didn't knock my socks off. There may be similar programs that work better, have more features, or are easier to use.
> **4 stars:** Despite a few minor flaws, this program is generally excellent. I recommend this download.
> **5 stars:** I can't recommend this product highly enough.

The meaning of user ratings as specified at www.amazon.com:

> **1 star:** I hate it.
> **2 star:** I don't like it.
> **3 star:** It's OK.
> **4 star:** I like it.
> **5 star:** I love it!

B Shareware repositories

Name	Website	Backlinks
CNET Download	http://www.download.com	604,000
Softpedia	http://www.softpedia.com	26,200
Tucows	http://www.tucows.com	25,800
Softpicks	http://www.softpicks.net	12,500
Top Shareware	http://www.topshareware.com	8,770
Jumbo	http://www.jumbo.com	4,340
5 Star Shareware	http://www.5star-shareware.com	3,340
Shareware Junkies	http://www.sharewarejunkies.com	854
Paul's Pick	http://paulspicks.com	518
Netscape shareware	http://computing.netscape.com/computing/download/shareware/main.tmpl	22

Table B1: Shareware repositories ranked by visibility.

The number of links to a website (backlinks) is used as a measure of its visibility or brand awareness (see empirical section of this paper). For purpose of comparison, there were 876,000 backlinks to amazon.com at the time of collection (May 2006).

C Free and try-out software on the Internet: examples and motivations

Name	Description	Motivation	Examples
Shareware	Software given out for a limited period of time or with limited functionalities, with the option to buy later.	Quality signalling, distribution.	mIRC, WinZip, RSI Guard.
Freeware	Program distributed without its source code for free.	Complementary services, upgrade.	Wordweb, LeechFTP.
Open source	Program distributed with its source code for free.	Collaborative work, ideology.	Mozilla, Gimp, Linux, L^AT_EX.
Basic Version	Basic version of a software sold at a lower price than a fuller version.	Brand line strategy, versioning.	Microsoft Works, AVG Anti-Virus.
Adware	Free software displaying third-party advertising that may be turned off only in exchange for payment.	Advertising revenues.	Opera, Software995, Gator.
Reader version	Program that only allows one to read documents created with the full version of the program.	Network effects.	Acrobat Reader, Realplayer, MediaPlayer, iTunes.
Promotion	Software initially sold at a reduced price or given out free during a promotional period.	Lock in, market launch, visibility.	

Table C1: Free and try-out software: examples and motivations