Estimating the Demand for Bandwidth

Hal R. Varian*
University of California, Berkeley

August 1999 (revised: August 29, 2000)

Abstract

One experiment in the INDEX Project offered users different bandwidths for different prices. I use the data from this experiment to estimate the demand for bandwidth and the value of waiting time for users. The parameter estimates for the demand functions for bandwidth are plausible and well-behaved. The parameter estimates for the value of time are, on average, very low, but there are some subjects with relatively high time values.

The INDEX Project is an experiment designed to estimate how much people are willing to pay for various kinds of Internet Quality of Service (QoS). The INDEX designers architected the system to provide different QoS's on demand and to record the usage of each different QoS by each user. Users can change their requested QoS instantaneously and are billed monthly for their usage. From April 1998 to December 1999 we provided approximately 70 users at UC Berkeley with residential ISDN service through the INDEX Project.

Edell and Variaya [1999] provides an overview of the project. Current information and other reports can be found on the INDEX project Web page[1].

*This work was supported by National Science Foundation Grant titled "Demand for quality-differentiated network services." I wish to thank Karyen Chu and Jorne Altmann for help with data extraction and manipulation and the entire INDEX team for helpfu discussion and critiques of this work.

[1] http://www.index.berkeley.edu

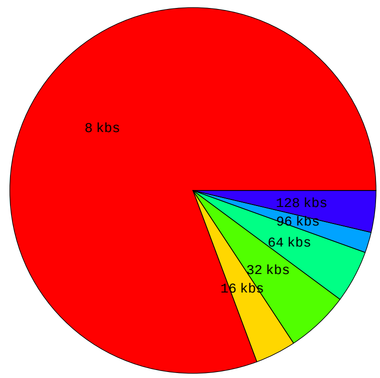

Figure 1: Bandwidth usage.

In this paper I examine one set of INDEX experiments designed to measure the willingness to pay for bandwidth. In these experiments users were offered the choice of 6 different bandwidths, ranging from 8 Kbs to 128 Kbs. Users could choose 8 Kbs service for free at any time. Each Sunday a new set of prices were chosen for the other bandwidths, ranging from .1 cents to 12 cents per minute of use. The INDEX system measured how much bandwidth subjects consumed at each different price, allowing experimenters to estimate demand for different bandwidths as a function of the price vector.

1 Reduced Form Estimates

Figure 1 depicts a pie chart of total usage. About 3/4 of the usage was 8 Kbs service. Since 8 Kbs was free, users tended to keep it on all the time. Usage was roughly equally divided among the other 5 bandwidths with positive prices.

Table 1 depicts the output of regressing the log of total minutes used on the log of the 5 different prices. Observations with zero usage were omitted. The coefficients in these log-log regressions can be interpreted as price elasticities of demand. Coefficients printed in bold are statistically significant at the 95% level.

Note that the diagonal terms (the own-price effects) are all negative and statistically significant. The subdiagonal terms are the cross-price effects for lower bandwidths. The positive numbers indicate that one-

Bandwidth	p128	p96	p64	p32	p16
128	**−2.0**	+.80	+.25	−.02	−.16
96	**+1.7**	**−3.1**	**+.43**	+.19	+.18
64	+.77	**+1.8**	**−2.9**	+.59	+.21
32	+.81	−1.0	**+1.0**	**−1.4**	+.15
16	+0.2	−.29	+.04	**+1.2**	**−1.3**

Table 1: Reduced form estimates. All own price effects are significantly negative; the cross-price effects for one-step lower bandwidths are positive.

Bandwidth	With ISE	No ISE
128	.95	.11
96	.93	.25
64	.92	.18
32	.95	.14
16	.90	.17

Table 2: Regression R^2. The R^2s with individual specific effects are large.

step lower bandwidths are substitutes for the chosen bandwidth.

This sign pattern is quite plausible. It is also worth noting that the implied elasticities are rather large. The regression for 96 Kbs service implies that a 1% increase in the price of 96 Kbs leads to a 3.1% drop in demand, and a 1% increase in 128 Kbs service leads to a 1.7% increase in the demand for 96 Kbs service.

We ran these regressions with and without dummy variables for the individual users, with little change in the estimated coefficients. Table 2 depicts the R^2s for these regressions.

Roughly speaking about 20 percent of the variance in demand is explained by price variation, about 75 percent of the variance in demand is explained by individual specific effects, and about 5 percent is unexplained. These fits are remarkably good, giving us some confidence that the subjects are behaving in accord with the traditional economic model of consumer behavior.

2 Structural estimates

The reduced form estimates given above suggest that the users are behaving in an economically sensible way. Hence it makes sense to try to model their choice behavior in more detail so we can extrapolate to other environments.

W adopt a very simple behavioral model, and assume that users get utility from the bits transferred ($u(x)$) and the time (t) it takes to transfer them. The cost of transfer time has two components: the subjective cost of time (c), which varies according to users and circumstances, and the dollar cost, which depends on the chosen bandwidth ($p(b)$). If b^* is the chosen bandwidth, optimization implies that

$$u(x) - [c + p(b^*)]t \geq u(x) - [c + p(b)]t,$$

for all bandwidths b.

Since bandwidth is by definition bits per unit time, we have $t = x/b$. Making this substitution and canceling the xs, we have

$$[c + p(b^*)]\frac{1}{b^*} \leq [c + p(b)]\frac{1}{b},$$

for all bandwidths b.[2]

It follows from simple algebra that

$$\min_{b^*<b} \frac{p(b^*)b - p(b)b^*}{b^* - b} \geq c \geq \max_{b^*>b} \frac{p(b^*)b - p(b)b^*}{b^* - b}.$$

This gives us observable upper and lower bounds for c, the user's subjective cost of time.

Figure 2 depicts these bounds graphically. Define the "total cost of time" by

$$K(c) = [c + p(b)]\frac{1}{b},$$

and plot these affine functions for each bandwidth b. A user with subjective time cost c will choose that bandwidth b with the lowest total cost. Conversely, an observed choice of b implies that the time cost must be bounded above and below as indicated. Note that a choice of the lowest available bandwidth only yields an upper bound on time cost, and a choice of the highest available bandwidth only yields a lower bound on time cost.

[2] If users waste some of their bandwidth we could write $t = ax/b$, where $a > 1$. As long as the fraction wasted is constant across bandwidths, the cancellation of ax can still be performed.

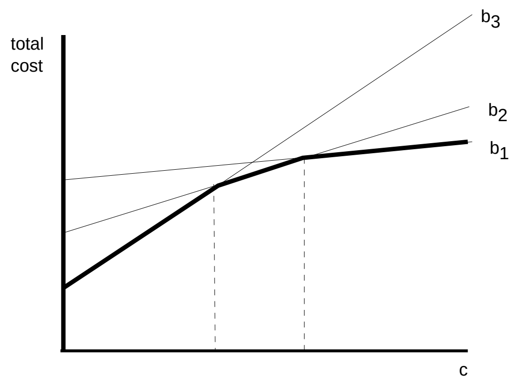

Figure 2: The straight lines are the "total cost of time" at different bandwidths. If we observe a particular bandwidth being chosen, we can calculate bounds on the subjective time cost c.

3 Estimating time cost

We assume that the user's time cost is a random parameter, drawn from a distribution $p(c)$. Sometimes the user is in a hurry, which means he or she has a high cost of time. Sometimes they are patient, which means the user has a low cost of time. This distribution of time cost is summarized by the probability distribution $p(c)$ and our objective is to estimate this distribution.

Each weekly menu of prices and bandwidths gives us a set of upper and lower bounds. Since we observe the frequency with which the user chooses bandwidth b during a week, we can construct a histogram for each user for each week illustrating the implied time costs. An example for a particular user in a particular week is given in Figure 3

4 Distribution of the time cost bounds

Table 3 shows the frequency with which the upper and lower bounds fall in a give range. For example, 39 of the users, or about 60%, have an average upper bound on the time cost of less than 1 cent a minute, 8 of the users, or about 12%, have an average upper bound greater than 1 cent a minute, but less than 2 cents a minute and so on. The last line in this table is the distribution of simple average of the upper

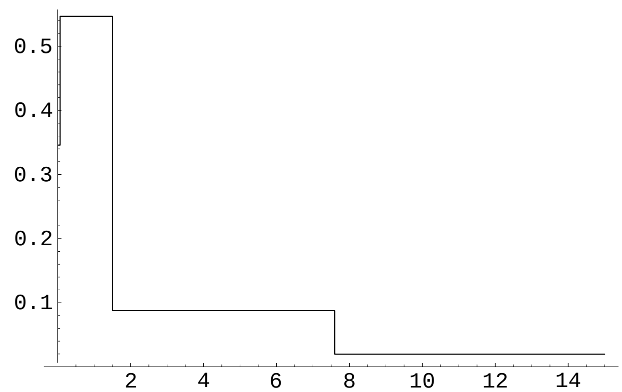

Figure 3: Histogram illustrating the fraction of the time that a particular user's time cost falls in the indicated region in a particular week.

Range	1	2	3	4	5	6	7	8	9	10	11	12
Upper bound	39	8	3	4	1	2	2	1	2	0	3	0
Lower bound	63	3	0	0	1	0	0	0	0	0	0	0
Average	47	7	2	3	3	3	1	3	1	1	0	0

Table 3: Frequency with which time cost in given range is observed.

and lower bounds which is a rough-and-ready, nonparametric estimate of the distribution of time cost across the population.

The remarkable thing about Table 3 is the low values that users place on their time. Most of the users have a time cost of less than 1 cent a minute. However there are a few users with systematically higher time costs.

The obvious question is whether we can predict which users have higher time value. Relevant variables available are occupation type, income, and whether the employer or the user pays for the service. We found that occupational dummies do a pretty good job of explaining the time costs using the following regression:

$$\text{c} = .86 \text{ professional} + 2.4 \text{ technical} + 7.02 \text{ admin} + .91 \text{ student}.$$

All coefficients are statistically significant and the R^2 for the regression is .646. Adding in both "income" and "who pays" yields an R^2 of .652, a negligible increase, suggesting that the best single predictor of

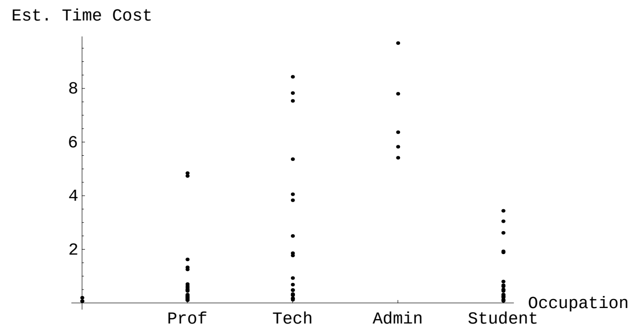

Figure 4: Time cost versus occupational category.

willingness to pay are the occupation variables.

This suggests that the time value is relatively predictable using available demographic data. For those who prefer graphs, Figure 4 shows the distribution of time values by occupational classification, which tells almost the same story as the regression. However, it must be cautioned that University of California employees may not be representative of the population as a whole, although they may be representative of early adopters of new technology.

5 Estimating a parametric distribution

In an earlier section, we showed how to derive observed bounds on the cost of time. If we observe n choices, we can construct a data set (c_U^i, c_L^i, f^i) for $i = 1, \dots, n$. If $p(c, \beta)$ is the true distribution from which the data have been drawn the probability of observing c in the region $[c_U^i, c_L^i]$ is given by

$$p^i = \int_{c_L^i}^{c_U^i} p(t, \beta)\, dt.$$

We can estimate the parameter β that makes the probability, p^i, close to the observed frequency, f^i. Various measures of closeness could be used, but a convenient one this context is the Kullbec-Leibler

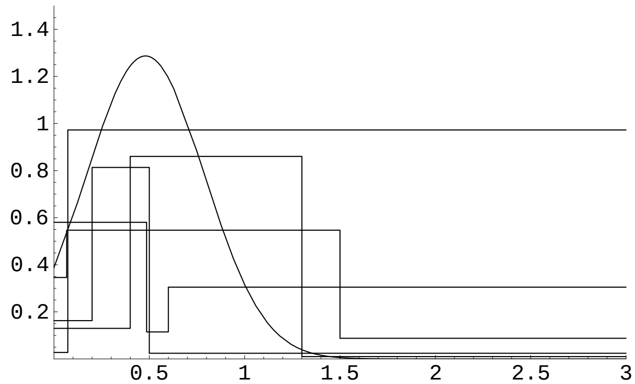

Figure 5: The histograms for a particular user superimposed on the estimated distribution.

entropy measure:

$$E = \sum_{i=1}^{n} f_i \log p_i.$$

We choose a truncated Normal distribution for the parametric form of th distribution $p(c, \beta)$. Figure 5 illustrates the fitted distribution and the five corresponding histograms. Figure 6A is a bit less messy; it depicts the empirical CDF constructed using the upper bounds, along with the theoretical CDF implied by the estimated parameters.

We could also fit the CDF directly. Figure 6B shows the theoretical CDF that minimizes the sum of squared residuals between it and the corresponding frequencies, along with the estimated parameters. Note that they are not very different from the entropy-maximizing estimates.

We applied this technique to estimate the implied parameter values for all 70 or so subjects. For about 7 subjects, the fits exhibited numerical instability. In about half of these cases, the instability was due to the fact that the user always chose the highest speed. The estimates of time value for the other cases tended to be quite low, consistent with the nonparametric results. We do not report these results since they tell essentially the same story as the nonparametric results.

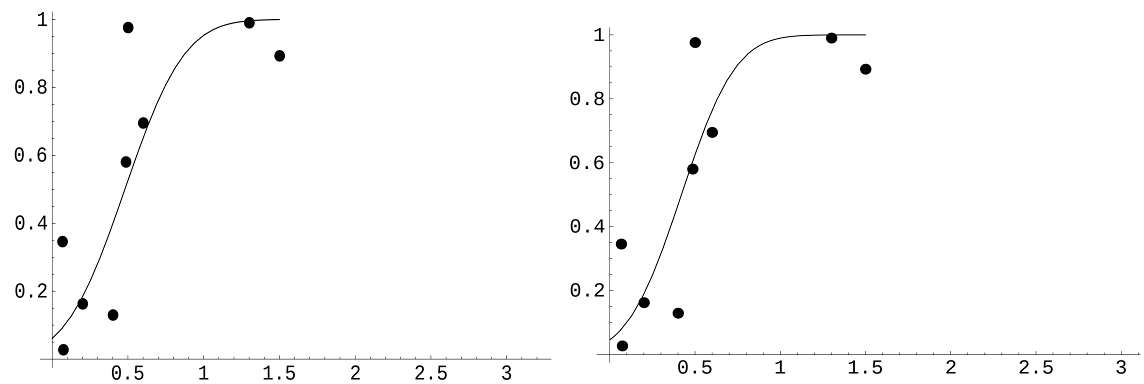

Figure 6: Data points and fitted CDFs. The entropy-maximizing fit is in panel A, the fit that minimizes the sum of squared residuals is in panel B.

6 Estimation of CDF on entire data set

Encouraged by the results described in the previous section, we estimated the CDF that minimizes the sum of squared residuals over the entire data set. The results are depicted in Figure 7. Note that the average cost of time (over the population) is very low, as would be expected from the previous results.

7 Why is the time cost so low?

These results raise the immediate question as to why the time costs are so low. A time cost of 1 cent a minute is only 60 cents an hour. Lots of student jobs at Berkeley are available for $10-$12 an hour, so this number is far below prevailing wage rates, even for students.

Several hypotheses suggest themselves.

Users are non-representative. This is likely part of the story. Our users are volunteers, many are students, and it is apparent from Figure 4 that certain occupations have much higher time values. The only way to deal with this is to try to repeat the study with a different sample, which we hope to do.

Other uses of time. Not all of the time that a user is "waiting for a

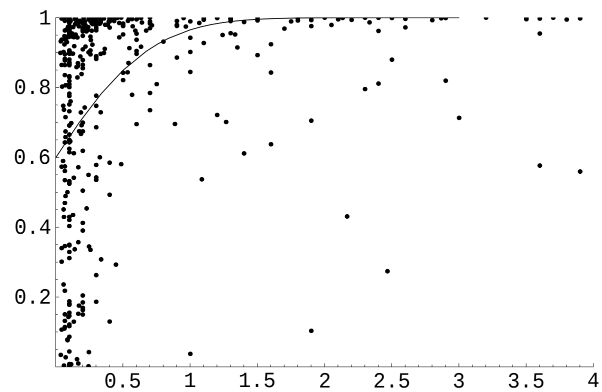

Figure 7: Fitting CDF on entire data set by minimizing the sum of squared residuals.

download" is wasted since it is common to engage in alternative activities. Indeed, we have already mentioned that users tend to leave 8 Kbs service on all the time so that email could be downloaded in the background. Because of this multitasking capability, the value of "time saved" could easily be lower than one might think. A closely related point is that certain activities, such as Web surfing, tend to involve bursts of activity, followed by a period of time spent in absorbing the acquired material. In this situation, the bandwidth per se is not necessarily the constraining factor in acquiring and absorbing information.

Service quality on rest of Internet. We can only control the quality of service on the link from the user's residence to the ISP. We have no control over bandwidth elsewhere on the Internet. If the user is accessing a site that is highly congested, an increase in speed on the residential-ISP link could have no value since it would not increase overall throughput.

Can only measure value of existing applications. We can only measure how the user values time given the existing mix of applications. If the user had access to high bandwidth at low cost, there could easily be applications that are infeasible at current bandwidths that the user could find valuable. We can not measure the value of such hypothetical applications using the methodology at our disposal.

8 Summary

The INDEX experiment is the first experiment to systematically estimate the demand for Internet bandwidth. Our estimates indicate very low willingness-to-pay for bandwidth, and very low values for time. We offer some reasons why these values may make sense, but our ultimate conclusion is that our sample of users was not willing to pay very much for bandwidth, at least given today's set of applications.

References

Richard Edell and Pravin Variaya. Providing Internet access: What we learn from the INDEX trial. *IEEE Network*, 13 (5), 1999. http://www.path.Berkeley.EDU/~varaiya/papers_ps.dir/networkpaper.pdf.