COMPETITION IN THE COMPUTER INDUSTRY: ONLINE VERSUS RETAIL

Austan Goolsbee
University of Chicago, G.S.B.
American Bar Foundation and N.B.E.R.

Current Version: November, 2000

Abstract

This paper estimates the relative price sensitivity of individuals' choice of whether to buy computers online versus in retail stores using a new data source on the computer purchase behavior of almost 30,000 people. To estimate the degree of competition between the two channels, the paper uses a two step approach. First, it fits hedonic regressions for the prices paid for a computer in a retail store as a function of characteristics. The coefficients on the city fixed effects in these regressions give a measure of the retail price level The second stage then looks at whether individuals purchase their computers in stores versus online as a function of the retail price and their own personal characteristics. The results indicate that the decision to buy remotely is sensitive to the relative price of computers in retail stores. Conditional on buying a computer, the elasticity of buying remotely with respect to retail store prices is about 1.5.

I wish to thank Andrew Lee for excellent research assistance, Severin Borenstein, Judy Chevalier, Luis Garicano, Steve Levitt, and workshop participants at the NBER for helpful comments and the National Science Foundation and the Alfred P. Sloan Foundation for financial assistance.

1. Introduction

One of the most important lingering questions about Internet commerce is how much competition it provides with retail merchants. This is hard to answer, in practice, because most standard data sets do not include information about the Internet and because in most sectors, online merchants make up only a small fraction of total sales (even for books, online sales account for less than 5% of the total market). Several recent papers have emphasized the large amount of price dispersion online in individual sectors such as books and music (Brynjolfsson and Smith, 1999; Bailey, 1998; Clay et al., 2000) and seemed to suggest that price competition online may not be particularly intense and that brand and other factors are quite important.

There has been little empirical work on direct competition between retail and Internet commerce (see Balasubramanian, 1998). One exception is Goolsbee (2000) who finds that variations in retail prices caused by local sales tax rates seems to have a large impact on consumers' online buying patterns, implicitly suggesting that there is cross-channel competition. To understand the role of this cross-channel competition, though, more precise estimates of the magnitude of cross-price elasticities of demand across different venues are needed.

To do such estimation, however, requires data that is normally difficult to come by. First, one needs data on people's shopping patterns across retail and Internet channels for some type of good. Second, one needs relative price data for the good across in many different local markets if one has cross-sectional data. Unfortunately, cross-market price data on individual goods is extremely rare.

In this paper I will examine the computer industry. I choose computers for two reasons. First, it is an extremely important industry. There has been important work in industrial organization analyzing the competitive conditions in the computer industry (see the survey of

Bresnahan and Greenstein, 1999 or the work on PCs by Bresnahan, Stern and Trajtenberg, 1997). Computer goods are also the single largest category of retail goods sold online (Boston Consulting Group, 1998). In part this is an outgrowth of the well established mail-order trade in computers. Manufacturers such as Dell and Gateway have integrated their direct sales operations previously conducted through magazines and the telephone into tremendous online businesses. The second reason I look at computers is that it is one place where the data are potentially sufficient to estimate the demand system.

The approach I take will be to use a new micro data set on individual computer purchases and estimate the sensitivity of venue choice to variations in the relative price with a two step procedure. First, I will get a price index for local retail computers in each of the 50 largest metro areas by fitting a hedonic regression on purchase price data by metropolitan area for computers that were bought in retail stores. I will estimate how much the individual pays for a computer as a function of the computers characteristics, year dummies, and metro area dummies. The metro area dummies then become a local retail price index for computers. Second, using this measure of prices, I will then estimate a logit model for the discrete choice of whether an individual bought their computer in a retail store or online as a function of retail prices and of individual characteristics.

The results indicate that the variation in retail prices has a significant impact on the likelihood of buying directly from the manufacturer. Conditional on buying a computer, the elasticity of buying remotely with respect to the retail price is about 1.5.

The paper proceeds in four sections. Section 2 describes the role of the Internet in the computer industry and the data used in this paper. Section 3 presents the results from the hedonic regression and the creation of the local retail price index for computers in each market.

Section 4 explores the price sensitivity on online sales with respect to this price. Section 5 concludes.

2. Data and Industry Background

A. The Data

The estimation will rely on micro data on computer purchases from the December 1998 proprietary mail survey of Forrester research called *Technographics 99*. Forrester is a marketing research company specializing in the information economy. The fieldwork for the survey was conducted by the NPD Group. NPD Group received survey data from about 90,000 American households on their ownership patterns for computers and other electronic goods. The sampling methodology is proprietary but is meant to ensure a nationally representative sample. More details on the Technographics program can be found in Bernhoff, et al. (1998) or Goolsbee and Klenow (2000).

These data provide information on the demographics of each respondent including gender, race, income, education, age, whether they use a computer at work, whether they run a business from home, and their state and broadly defined metropolitan area of residence (specifically, what television market). They also answer whether they have a personal computer at home.

I will look at those people with a computer. For anyone with a computer at the time of the survey (conducted in December 1998), the survey asks how many computers they currently have, how many they have ever had, when they bought their first computer, and how often they use their computer. For their most recent computer, they answer where they bought it, how much they paid for it, and give a variety of characteristics of the computer such as the speed of

the chip, whether they have a modem, a laser printer, and so on. Note that these are for home computers and do not include business purchases.

I will use two different parts of the data for the two steps of the estimation procedure. In the first part the dependent variable is the log of the real price paid for the computer as a function of its characteristics. In these regressions I will look only at people who bought their computers in retail stores and I will restrict the sample to the top 50 markets (to ensure there are enough observations for the hedonic regression). The metropolitan area dummies in these regressions indicate how much more or less a computer with the same observables costs in different markets.

For the second part of the estimation, the analysis looks at all people who own a computer and the dependent variable becomes whether they bought the computer from a retail store or from a remote vendor.[1] Here the city level dummies in the price regression become the retail price index for the city and I try to explain where the customer bought their computer from as a function of individual level demographics, dummies for when they bought their first computer and for how many computers they own and other measures of computer sophistication.

B. The Computer Industry

In 1999, there were more than 36 million PCs sold in the United States (InfoTech Trends, 2000). Most of these were to businesses but given that more than half of U.S. households have a PC in their homes, residential sales are also quite important. Among computer manufacturers, there are two main methods of selling to residential customers. The first is to market computers

[1] In this category I include anyone that answers either "direct from the manufacturer" or "online" as to where they bought their computer. I do this because it is common for customers of the large direct sellers of computers such as Dell or Gateway to use the Internet to customize a computer and get a price quote and then call on the telephone to place the order. This might be reported by the customer in either category. All of the other choices are from some type of retail store such as from an electronic store, from a computer store, etc.

through distribution networks such as computer stores like CompUSA, electronics stores like Circuit City, general retailers or catalog merchants. The alternative is direct selling from the manufacturer to the consumer, usually either through an Internet site or through direct ads in computer magazines. The most well-known of these merchants are Dell and Gateway, though there are several others. Recently, Gateway has also opened computer stores of their own (Gateway Country stores) but this mainly took place after this sample.

Table 1 illustrates the point with data from Forrester on the location of the most recent purchase for members of the sample and brand of the most recent computer for members of the sample from 1996 through 1998. The data indicate that of computer owners in the top 50 metro areas, about 60 percent purchased their last computer from some kind of retail store with computer and electronics stores dominating the category, about 30 percent of computer owners purchased their last machine from a catalog, direct from the manufacturer or over the Internet, and about 10 percent received the computer as a gift. The share buying from a retail store has been slowly falling over time. In 1995, about two-thirds had bought their last computer at a store. By 1998, it was about 57%. Typically the direct sellers appeal to a more informed consumer than does the computer store. The average computer owner that bought their last computer in a store, for example, has owned 2.0 computers in their lifetime whereas the average remote buyer has owned about 2.2.

The brands represented are familiar. Compaq, Packard-Bell, and IBM make up the largest sellers. That Gateway is larger than Dell is a bit surprising since Dell's sales are larger but this may be due to the focus here being on home computers as opposed to business, government, and educational sales. More than a third of the sample bought a brand that was not in this group of well-known merchants.

3. Hedonic Regressions and A Local Price Index

First, using the price and computer characteristics data, I will estimate a hedonic regression with dummies for each metropolitan area that will provide an estimate of the local retail price level. The dummies will indicate how much more an individual in some area must pay for a computer with the same attributes. There is a large literature on the subject of hedonics in the computer industry (see Berndt, Griliches, and Rappaport, 1995 or Berndt, Dullberger and Rappaport, 2000 for results and guides to the literature). Typically these hedonic studies are based on list prices and are not quantity weighted (i.e., each model is equally weighted because there is no sales data by model). The Forrester data have the advantage of being transaction prices and being quantity weighted but have the major disadvantage that they lack the same level of detail for the characteristics of the machine. Rather than having the actual MHz of the CPU, for example, the Forrester data has only categories such as 386, 486, Pentium, Pentium II, and so on.

I estimate the hedonic price indices by looking at buyers in the 50 highest population markets (chosen because they had sufficient observations to estimate the city fixed effects rather precisely). The hedonic regression will explain log prices for computers bought in retail stores as a function of dummies for the speed of the chip, dummies for the fourteen manufacturers, year dummies, and dummies for whether the computer was bought with a modem (and the type of modem), a printer, a scanner, extra memory, an expanded hard drive, and metropolitan area dummies.

Following the results of Berndt, Dullenberger and Rappaport (2000) on the non-constancy of the parameters over time for certain characteristics (and because some of the characteristics in the Forrester depend on the year—buying a scanner in 1994 means something

quite different than buying one in 1998), I will also run a specification that includes interactions of the scanner, chip speed, modem, laptop, memory, and printer dummies with the year dummies. In doing these specifications and taking the metropolitan area dummies as a measure of the local price index, I am implicitly assuming that price differences across markets are constant across the sample and scale up the price of an identical computer in a multiplicative way. I will restrict the sample to computers bought in the period from 1996 through 1998 to keep the sample consistent.

Most of the coefficients on each characteristic have the intuitive signs and plausible magnitudes. They are reported in column 1 of table 2. Faster chip speeds, the presence of a scanner, a modem, or a laser printer are associated with higher prices, for example. The year dummies suggest that the quality adjusted prices fell almost 15% per year in the period. This is smaller than the 25%-30% declines found in the list-price based hedonic regressions of the early 1990s but still sizable. In column 2, I add the year dummy interactions with the computer characteristics. Doing this effectively eliminates the year dummies on their own but the overall trend in prices remains large and negative. Though not listed to save space, the results indicate that the price premium for each type of chip falls over time, as does the laptop premium, and so on.

The dummy variables for each metro area in these regressions are then used as an indicator of the price level in each location. Since they are in log terms, I take the exponent and then norm the price levels to be 1 in the 50th largest market (Providence, RI). The prices of the Internet/catalog computers are assumed to be the same across markets, so the local price effect is a measure of the relative price. The prices vary from 0.98 to 1.12 as listed in table 3. The correlation of the retail price index from this regression with the one from regression (1) is .99.

It is true that cities with low costs of living, such as Lancaster and Greensboro, tend to have low retail prices for computers but the index is also relatively low for cities with higher costs of living but a more information technology savvy populace such as Raleigh/Durham, Boston, and San Francisco, though Seattle is a notable exception. I will use this price index, derived from the more general model in specification (2), in the estimates below.

As this retail price index will form the core of the estimation, it is important to check that variations in it actually represent variations in local prices and not spurious factors that might also be correlated with the probability of buying online. The main fear in such regressions is that better unobserved characteristics of a person's computer that tend to increase the price will show-up as higher prices when they are, in fact, higher quality. In markets where a large fraction of people buy machines with higher MHz, conditional on the type of chip, for example, or some other measure of quality that is unobservable in the Forrester data, the price index will look higher, conditional on the observables but for the wrong reasons. If the types of places where people buy higher MHz machines, conditional on observables, are also the places where people tend to buy online and through catalogs (i.e., more sophisticated places), this will bias the results.

To deal with this potential problem, I do two things. First, I add individual level demographic information including age, income, education, and race dummies to the pricing regressions in column 3. These variables should not have a direct impact on prices paid for identical machines (unless there is discrimination) but may be correlated with the taste for unobserved quality in the computers. Indeed, as shown in the table, these variables are significantly correlated with price. Better educated and higher income people tend to have higher prices, conditional on the same observable computer characteristics (although older people do, as well). The impact of these factors on the local retail computer price index,

however, is very small. The correlation of the retail price index from this regression with that in the general model is .96 and it did not affect the results below to substitute it.

A second, more direct test is to repeat the hedonic regressions but use the prices paid for computers bought direct from the manufacturer to get a local price index for remote computers. I do this in column 4. Since these prices are national prices, however, there should not be any local variation in the price of remote computers (save, perhaps for the tax term). To the extent that there are, these may be a measure of the unobservable quality premium in each city (i.e., a city with a higher index means that people buying remotely tend to buy higher quality machines with the same observables). Later I will then include this alternative remote computer price index as a control in the decision estimates to test for the presence of spurious correlation. A finding that higher local remote prices (which can only come about because of variations in the unobservables across cities rather than actual variation in prices in those cities) are correlated with the probability of buying remotely would be rather important evidence of a flawed approach. An interesting thing to note about the remote price index, now, however, is that it is basically uncorrelated with the retail price index (a correlation of -.04).

4. Probability of Buying Remotely Versus Retail

With this price index of local computer prices, I then use information on the individual to examine their choices about whether to buy a computer remotely as a function of their observables and of relative prices in their area. Table 4 lists the results from a logit regression of the {1,0} decision of computer owners of whether they bought their computer remotely as a function of how many computers the individual has ever owned, when the person bought their first computer, how long they have had online access, whether this purchase was a laptop,

whether the respondent has ever bought a non-computer product online, the number of cars and trucks in the household (which reduces the cost of retail shopping), race, age, education, income, whether they use a computer at work, year dummies, and the price index in the city. These controls are meant to account for individual technological sophistication.

Not surprisingly, people having bought computers in the past, having previously bought online, having higher income, and so on, are significantly more likely to buy directly from the manufacturer. The price coefficient is also significant and somewhat large, suggesting direct competition between retail and the remote sales.

More importantly, however, prices are, indeed positively and significantly correlated with the likelihood of a computer owner having bought online or direct from the manufacturer. The marginal effect indicates an elasticity of 1.45 (i.e., conditional on buying a computer, an increase in retail prices of 1 percent raises the overall likelihood of buying remotely by 1.45 percent).

Next, to deal with the issue of whether the computer price index is merely picking up differences in the local housing or other costs of living, in column 2 I add the Money Magazine Cost of Living Index for the largest city in the metro area. This is also a number equal to 1 in Columbus, Ohio. The results indicate that higher local prices do make people more likely to buy computers online, separately from local computer prices, but that the effect of higher computer prices itself is still large and significant. The conditional cross-price elasticity here is slightly larger than in the base case.

Finally, in column 3, as the test of whether this results from the spurious correlation between technological sophistication, unobservable computer quality in an area, and likelihood of buying remotely, I also include the local price index for remote purchased computers, as described above. The coefficient on the remote price index is small, insignificant and of the

wrong sign while the retail price index remains almost exactly the same size (the conditional cross-price elasticity is again 1.55). This suggests against the explanation that unobserved quality is behind the variation in the price index as a counter-explanation of the cross-price sensitivity.

5. Conclusion

This paper has used micro data on individual computer purchases to estimate local retail price indices for computer equipment and to use these price data to estimate the price sensitivity of computer purchases across different channels. The results suggest that there is significant competition between online and retail sellers of computers. The conditional cross-price elasticity of buying remotely versus buying in a store with respect to the retail price is in excess of one and suggests that online and offline sales of computers are unlikely to be truly separate markets. This is one of the first estimates of direct competition between online and offline retailers.

The apparent cross-price sensitivity of demand suggests that it would be fruitful to consider the supply and pricing decisions on the part of merchants who, presumably, know what the customer elasticities are when they make such decisions. This important topic is left to future work.

TABLE 1: LOCATION AND BRAND DISTRIBUTION OF HOME COMPUTERS

Location of Last Purchase	**Percent of Total**
Computer Store	23.9
Electronics Store	23.0
Discount Warehouse/Membership Club	4.6
Office Superstore	3.8
Other Type of Retail Store	4.7
Direct, Catalog or Internet	29.6
Gift	10.3
Brand of Last Purchase	**Percent of Total**
Compaq	10.8
Packard-Bell	10.5
IBM	10.2
Gateway	9.5
HP	6.7
Dell	4.8
Acer	3.1
NEC	1.8
Toshiba	1.6
Other Brands	35.1
Unknown	5.8

Notes: Author's calculations using data from Forrester on computers purchased from 1996-1998.

TABLE 2: HEDONIC LOG PRICE REGRESSIONS

	(1) No Interactions	(2) Interactions	(3) Demographics	(4) Remote Sales
386 or less	-.065 (.027)	-.109 (.053)	-.115 (.053)	-.187 (.086)
486	-.072 (.019)	-.063 (.041)	-.076 (.041)	-.110 (.058)
Pentium	.049 (.021)	-.050 (.028)	-.058 (.028)	-.149 (.043)
Pentium II	.138 (.016)	.119 (.027)	.104 (.027)	.049 (.041)
Macintosh	-.044 (.059)	-.039 (.071)	-.045 (.070)	.003 (.073)
Modem	.071 (.029)	.008 (.050)	.004 (.049)	.240 (.054)
Scanner	.031 (.008)	.028 (.014)	.023 (.014)	.030 (.018)
Laser Printer	.114 (.030)	.073 (.039)	.058 (.038)	.050 (.045)
Laptop	.156 (.020)	.173(.028)	.152 (.028)	.312 (.035)
Sales Tax	.403 (.452)	.394 (.451)	.273 (.451)	.533 (.671)
Extra Memory	-.014 (.009)	.001 (.015)	-.005 (.015)	.001 (.021)
Year 1995	.271 (.013)	.022 (.094)	-.006 (.093)	.154 (.106)
Year 1996	.211 (.011)	-.035 (.086)	-.058 (.085)	.190 (.100)
Year 1997	.134 (.010)	-.098 (.073)	-.008 (.079)	.015 (.084)
Education			.004 (.002)	
Income(/100)			.129 (.013)	
Age(/100)			.104 (.032)	
Other Dummies:	8 other hardware 13 manufacturers Gift Metro Area	8 other hardware 13 manufacturers scanner x year chip x year modem x year memory x year laptop x year printer x year Gift Metro Area	Race 8 other hardware 13 manufacturers scanner x year chip x year modem x year memory x year laptop x year printer x year Gift Metro Area	8 other hardware 13 manufacturers scanner x year chip x year modem x year memory x year laptop x year printer x year Gift Metro Area
n	9,391	9,391	9,327	9,391
R2	.13	.14	.16	.14

TABLE 3: ESTIMATED RETAIL PRICE INDEX FOR COMPUTERS BY MARKET

Metro Area	Retail Computer Price Index	Metro Area	Retail Computer Price Index
Pittsburgh	1.118	Tampa	1.057
Norfolk	1.106	Memphis	1.056
New Orleans	1.103	Greenvillle	1.052
Grand Rapids	1.102	San Antonio	1.051
Orlando	1.101	NYC	1.048
Seattle	1.098	Sacramento	1.043
Detroit	1.097	Houston	1.040
Cincinnati	1.092	Indianapolis	1.040
Philadelphia	1.088	San Francisco	1.037
Milwaukee	1.080	Portland	1.037
Denver	1.079	Dallas	1.032
Birmingham	1.073	Atlanta	1.031
Miami	1.073	Nashville	1.026
Minneapolis	1.071	Boston	1.025
St. Louis	1.069	Columbus	1.022
Cleveland	1.069	Louisville	1.022
Baltimore	1.068	Kansas City	1.021
Albuquerque	1.067	Oklahoma City	1.019
Hartford	1.065	Charlotte	1.010
Chicago	1.065	Buffalo	1.006
San Diego	1.063	Phoenix	1.000
Washington DC	1.063	Providence	1.000
Salt Lake City	1.060	Lancaster	0.999
West Palm Beach	1.059	Raleigh	0.980
Los Angeles	1.058	Greensboro	0.976

TABLE 4: LOGIT MODELS OF THE DECISION TO BUY REMOTE VERSUS RETAIL

	(1)	(2)	(3)
Retail Computer Prices	2.129	2.281	2.265
	(.548)	(.562)	(.565)
General Cost of Living		.350	.368
		(.112)	(.122)
Remote Computer Prices			-.141
			(.401)
Ever Bought Online	.324	.333	.333
	(.038)	(.038)	(.038)
Age	-.010	-.010	-.010
	(.001)	(.001)	(.001)
Education	.077	.079	.079
	(.007)	(.008)	(.008)
Female	.168	.169	.169
	(.032)	(.032)	(.032)
Number of Automobiles	-.082	-.077	-.077
	(.016)	(.016)	(.016)
Dummies:	Year, # of comps, Yr of 1st comp, Yrs online, Race Laptop, Income.	Year, # of comps, Yr of 1st comp, Yrs online, Race Laptop, Income.	Year, # of comps, Yr of 1st comp, Yrs online, Race Laptop, Income.
n	20,724	20,326	20,326
Log Likelihood	-12433.3	-12164.8	-12164.7
Estimated Conditional Cross-Price Elasticty	1.45	1.55	1.54

Bibliography

Bailey, Joseph, *Intermediaries and Electronic Markets: Aggregation and Pricing in Internet Commerce*," Ph.D. Dissertation, Management and Policy, Massachusetts Institute of Technology [1998].

Balasubramanian, Sridhar (1998), "Mail versus Mall: A Strategic Analysis of Competition between Direct Marketers and Conventional Retailers," *Marketing Science*; 17(3): 181-95

Berndt, Ernst, Ellen Dullenberger, and Neal Rapapport (2000), "Price and Quality of Desktop and Mobile Personal Computers: A Quarter Century of History," Mimeo, MIT Sloan School of Management.

Berndt, Ernst, Zvi Griliches, and Neal Rappaport (1995), "Econometric Estimates of Price Indexes for Personal Computers in the 1990's," *Journal of Econometrics*; 68(1): 243-68.

Bernhoff, Josh, Shelley Morrisette, and Kenneth Clemmer [1998], "Technographics Service Explained," *Forrester Report*, I [1998], Issue 0.

Boston Consulting Group (1998), *The State of Internet Retailing*. Mimeo, BCG/Shop.org, November.

Bresnahan and Greenstein, "Technological Competition and the Structure of the Computer Industry," forthcoming, *Journal of Industrial Economics*

Bresnahan, Stern, and Trajtenberg, 1997), "Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s," , v28, n0: S17-44

Brynjolfsson, Erik, and Michael Smith, “Frictionless Commerce? A Comparison of Internet and Conventional Retailers,” *Management Science,* Volume 46, Number 4, April 2000

Clay, Karen, Ramayya Krishnan, and Eric Wolff, “Pricing Strategies on the Web: Evidence from the Online Book Industry,” Mimeo, Carnegie-Mellon University, 2000

Goolsbee, Austan (2000), "In a World Without Borders: The Impact of Taxes on Internet Commerce," Quarterly Journal of Economics, May, vol 115(2), pp. 561-576.

Goolsbee, Austan, and Peter Klenow (1999), "Evidence on Learning and Network Externalities in the Diffusion of Home Computers," NBER WP#7329.

InfoTech Trends (2000), “ < http://www.infotechtrends.com/cgi-bin/cif/sub_read.pl?ux=00& quar=99Q2&99239070.htm=on> accessed November, 15, 2000.

Smith, Michael, Joeseph Bailey, and Erik Brynjolfsson (1999),”Understanding Digital Markets: Review and Assessment , in Erik Brynjolfsson and Brian Kahin, eds. *Understanding The Digital Economy*, MIT Press (Cambridge, MA).