Highly Preliminary and massively incomplete—please do not quote

THE DEGREE OF COMPETITION BETWEEN RETAIL AND ONLINE COMPETITORS IN THE COMPUTER INDUSTRY

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

July 1, 2000

Abstract

This paper estimates the relative price sensitivity of individuals' choice of retail venue (i.e., retail stores versus remote sellers) 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 relative price and personal characteristics. The results indicate that the decision to buy remotely is quite sensitive to the relative price of computers in retail stores. The cross price elasticity of buying remotely with respect to retail store prices is almost 2.

1. Introduction

One of the most important questions about the Internet economy is how intense the competition is that it provides with retail merchants. There has been little empirical work on the substitutability between retail and Internet commerce (see Balasubramanian, 1998). This is likely because in most sectors online merchants make up only a tiny fraction of total sales (even for books, online sales account for less than 4% of U.S. book sales). 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 competition online is not particularly intense

There is very little work estimating the degree of price sensitivity across channels, however. One exception is Goolsbee (2000) who finds that variations in retail prices caused by local sales tax rates seems has a major impact on consumers' online buying patterns suggesting the competition may be rather intense. More precise estimates of the magnitude of cross-price elasticities across online and retail stores is needed.

To be more precise about estimating the degree of competition directly requires data that is normally difficult to come by. First, there must be data on people's shopping patterns across retail and Internet channels for some type of good. Second, there must be separate retail price data for that good in every retail market. 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. One, it is one place where the data is sufficient to identify the model. Second, 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 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 location 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/direct from the manufacturer 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. The elasticity of buying remotely with respect to the retail price is almost 2.

The paper proceeds as follows: …

2. Computers & data

To do this type of estimation requires rather detailed micro data on computer purchases. I use data from a proprietary December 1998 mail survey by 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 filled-out questionnaires 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).

The data provides information on the demographics of each respondent including gender, race, income, education, age, marital status, whether they have children under 18, 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.

For anyone with a computer at the time of the survey, they also answer how many computers they currently have, how many they have ever had, when they bought their first computer, when they bought their (up to) three most recent computers, 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.

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. Table 1 give some summary statistics about computers in the sample. 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).

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 (measures of computer sophistication),

Table ** shows that about 20 percent of buyers purchased their last computer directly from the manufacturer and about 80% from a retail outlet (remember, these are residential computers, not business computers). The distribution by vendor in the sample is shown in table **.

3. Hedonic

First, using the price and computer characteristics data, we 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 the many papers they

---

[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 very 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.

cite). This literature has identified the key characteristics that influence price allows me to check the results from the Forrester data against other hedonic regressions. (*fill in here*)

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 explains computer prices 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. The regression uses only people who purchased a computer since 1996 and only those computers bought in retail stores (because the online prices are the same across markets).

The coefficients on each characteristic have the intuitive signs and plausible magnitudes. They are reported in column ** of table **. 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 hedonic regressions of the early 1990s but still sizable.

The dummy variables for each metro area are then used as an indicator of the price level in each town. 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 0.97 to 1.11 as listed in table 2.

One fear in such regressions is that unobserved characteristics that increase the price of the computer will look like higher prices when they are, in fact, higher quality. In markets where a large fraction of people buy machines that are better in the unobservable dimensions, the markets will look as if they have higher prices when in fact this is just showing the preferences

of the local computer buyers. To deal with this, I do two things. First, I add individual level

demographic information such as income dummies to the pricing regressions. The variables

should not have a direct impact on prices paid for identical machines but may be correlated with

the taste for unobserved quality. Indeed these variables are significantly correlated with price

but the impact on the other coefficients is **fill in here** as seen in column ** of table **.

A second test is to repeat the hedonic regressions but use the prices paid for computers

bought direct from the manufacturer. Since these prices are set at the national level, there should

not be any local price fixed effect (save, perhaps for the tax term). To the extent that there are,

these will be a measure of the unobservable quality premium in each city. By assuming that the

taste for unobserved quality within a city is the same for retail and for direct buyers, I can

subtract the dummy variables for each metro area here from the dummies for the same metro

areas in the retail price regression to get an alternative, unobserved quality adjusted price index

for each city. I report this hedonic regression in column ** and list the implied price index by

city in table **. **fill in here**.

4. Probability of Buying Directly versus Retail

With this price index of local computer prices, the project will 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.[2] In the work so far, we have conditioned

on those individuals who actually bought a computer. In other words, concentrating on the

cross-price effect. Overall, in places with retail prices less than 1, about 27.1% of computer

---

[2] I include all remote sales because most online computer merchants integrate their catalog and Internet sales. A customer might see an advertisement in a computer magazine, for example, that would direct them to the website for

buyers bought remotely. In places with retail prices between 1 and 1.065, about 28.2% bought

remotely, and for those with prices above 1.065 about 30.5% did so. It is important to include

individual controls, however, since they may easily be correlated with local price levels. High

price places may have more experienced users, for example, and we know that experienced users

are more inclined to buy direct from the manufacturer.

Table 4 lists the results from a logit regression of the {1,0} decision of computer buyers

of whether to buy a 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 for 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. This is, essentially, the second stage of a nested logit (see

Goldberg, 1995).

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. At the mean of the covariates, lowering the local retail price

by 1% reduces the probability of buying directly from the manufacturer by about 1.9% in column

1, and 1.5% in column 2.

5. Conclusion

---

pricing and allow them to purchase over the phone if they didn't want to use a credit card online.

** fill in **

TABLE : ESTIMATED RETAIL PRICE INDEX FOR COMPUTERS BY MARKET

CITY	Retail Computer Price Index (Providence = 1.000)
Detroit	1.103
Cleveland	1.080
Los Angeles	1.077
Philadelphia	1.076
Pittsburgh	1.075
Chicago	1.072
Seattle	1.070
New York City	1.065
Hartford	1.059
Baltimore	1.058
Sacramento	1.057
Dallas	1.056
St. Louis	1.054
Washington	1.052
Orlando	1.051
Indianapolis	1.051
Minneapolis	1.051
Sanfrancisco	1.049
Denver	1.042
Miami	1.039
Sandiego	1.039
Boston	1.036
Atlanta	1.034
Houston	1.034
Portland	1.026
Tampa	1.022
Raleigh	1.018
Phoenix	1.012
Nashville	1.000
Charlotte	0.971

LINEAR PROBABILITY OF BUYING REMOTELY AS A FUNCTION OF RELATIVE PRICE (p) AND OTHER VARIABLES.

```
. reg

Regression with robust standard errors          Number of obs =    25785
                                                F( 28, 25756) =    46.94
                                                Prob > F      =   0.0000
                                                R-squared     =   0.0475
                                                Root MSE      =   .45649

------------------------------------------------------------------------------
             |              Robust
      remote |     Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           p |   .5447854   .1285146     4.239   0.000     .2928897    .7966812
       buyon |    .073108    .007298    10.017   0.000     .0588035    .0874126
       year1 |  (dropped)
       year2 |  -.0728892   .0099456    -7.329   0.000    -.0923832   -.0533952
       year3 |  -.0360041   .0087512    -4.114   0.000     -.053157   -.0188512
       comp2 |  -.0175317   .0090869    -1.929   0.054    -.0353425    .0002791
       comp3 |   .0006364    .010293     0.062   0.951    -.0195385    .0208113
       comp4 |   .0505972   .0099896     5.065   0.000     .0310171    .0701773
      firstc1 |    .05999   .0139894     4.288   0.000       .03257    .0874101
      firstc2 |   .025133   .0154151     1.630   0.103    -.0050815    .0553474
      firstc3 |  .0144588   .0143911     1.005   0.315    -.0137487    .0426662
      firstc4 |    .002379   .0120376     0.198   0.843    -.0212153    .0259733
      firstc5 |  -.0045619   .0117919    -0.387   0.699    -.0276748    .0185509
     online2 |  -.0119661   .0114618    -1.044   0.296    -.0344318    .0104996
     online3 |   .0020116   .0109449     0.184   0.854    -.0194412    .0234643
     online4 |   .0048377   .0089409     0.541   0.588    -.0126869    .0223623
     online5 |   .0124699   .0096907     1.287   0.198    -.0065244    .0314641
     online6 |    .007974   .0123948     0.643   0.520    -.0163204    .0322685
     online7 |    .075862   .0117984     6.430   0.000     .0527364    .0989876
      laptop1 |  .0033537    .013465     0.249   0.803    -.0230384    .0297458
        autos |  -.0055274   .0026633    -2.075   0.038    -.0107476   -.0003072
        race2 |  -.0390209     .01439    -2.712   0.007    -.0672261   -.0108157
        race3 |  -.0388237   .0214577    -1.809   0.070    -.0808821    .0032346
        race4 |  -.0220547   .0171834    -1.283   0.199     -.055735    .0116256
        race5 |  -.0150402   .0234669    -0.641   0.522    -.0610367    .0309563
    hispanic |  -.0378807   .0117245    -3.231   0.001    -.0608614   -.0149001
         age |  -.0015265   .0002342    -6.518   0.000    -.0019856   -.0010674
          ed |   .0202283    .001259    16.067   0.000     .0177605     .022696
      female |   .0309678   .0058123     5.328   0.000     .0195754    .0423603
       _cons |  -.5111044   .1385215    -3.690   0.000    -.7826143   -.2395945
------------------------------------------------------------------------------


PROBIT OF BUYING REMOTELY

. probit

Note: year1 dropped due to collinearity.

Probit Estimates                                Number of obs =    25201
                                                chi2(43)      =  1286.26
                                                Prob > chi2   =   0.0000
Log Likelihood = -15206.687                     Pseudo R2     =   0.0406


------------------------------------------------------------------------------
      remote |     Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           p |   1.39849   .3806746     3.674   0.000      .652382    2.144599
    compwork |   .0702032   .0210163     3.340   0.001      .029012    .1113944
       buyon |   .2073633   .0206337    10.050   0.000     .1669219    .2478047
       year2 |  -.2120676   .0281896    -7.523   0.000    -.2673181   -.1568171
       year3 |  -.1043297   .0243501    -4.285   0.000     -.152055   -.0566043
```

```
    comp2 |   -.0491307    .0277015    -1.774    0.076    -.1034247     .0051632
    comp3 |    .0038161    .0304286     0.125    0.900    -.0558229     .0634551
    comp4 |    .1268678    .0287199     4.417    0.000     .0705779     .1831577
  firstc1 |    .1630506    .0397811     4.099    0.000     .0850812     .2410201
  firstc2 |    .0675851    .0440034     1.536    0.125    -.0186599     .1538301
  firstc3 |    .0273938    .0407601     0.672    0.502    -.0524944     .1072821
  firstc4 |   -.0018294      .03466    -0.053    0.958    -.0697619      .066103
  firstc5 |   -.0219541    .0338064    -0.649    0.516    -.0882134     .0443052
  online2 |   -.0521874    .0357463    -1.460    0.144    -.1222488      .017874
  online3 |   -.0036518    .0330298    -0.111    0.912     -.068389     .0610853
  online4 |    .0115142    .0271023     0.425    0.671    -.0416054     .0646338
  online5 |    .0301823    .0286985     1.052    0.293    -.0260657     .0864302
  online6 |     .019626    .0356869     0.550    0.582     -.050319     .0895709
  online7 |    .2002788    .0327926     6.107    0.000     .1360064     .2645513
  laptop1 |   -.0160246    .0369586    -0.434    0.665    -.0884622      .056413
    autos |   -.0405108    .0084096    -4.817    0.000    -.0569933    -.0240283
    race2 |    -.132529    .0456544    -2.903    0.004     -.22201     -.043048
    race3 |   -.1227501     .062257    -1.972    0.049    -.2447716    -.0007286
    race4 |   -.0485474    .0524065    -0.926    0.354    -.1512622     .0541674
    race5 |   -.0509372    .0741075    -0.687    0.492    -.1961853     .0943108
  hispanic |   -.1169169    .0364349    -3.209    0.001    -.1883281    -.0455057
      age |    -.004878    .0007504    -6.500    0.000    -.0063487    -.0034072
       ed |    .0462448    .0039464    11.718    0.000     .0385099     .0539796
     inc1 |   -.3442244     .086665    -3.972    0.000    -.5140846    -.1743642
     inc2 |   -.3745048    .1018862    -3.676    0.000    -.5741981    -.1748115
     inc3 |   -.2825635    .0843972    -3.348    0.001     -.447979     -.117148
     inc4 |   -.3653414    .0914586    -3.995    0.000    -.5445969    -.1860858
     inc5 |   -.2785698    .0640919    -4.346    0.000    -.4041877     -.152952
     inc6 |   -.2198265    .0508945    -4.319    0.000    -.3195779    -.1200752
     inc7 |   -.2405193    .0493718    -4.872    0.000    -.3372863    -.1437523
     inc8 |   -.2392539    .0451491    -5.299    0.000    -.3277444    -.1507634
     inc9 |   -.1787673     .044036    -4.060    0.000    -.2650762    -.0924584
    inc10 |   -.2232842    .0437306    -5.106    0.000    -.3089945    -.1375739
    inc11 |   -.2296025    .0447197    -5.134    0.000    -.3172515    -.1419535
    inc12 |    -.165716    .0310479    -5.337    0.000    -.2265687    -.1048632
    inc13 |    -.131623    .0336459    -3.912    0.000    -.1975677    -.0656783
    inc14 |    -.134741    .0375103    -3.592    0.000    -.2082598    -.0612222
    inc15 |    -.060075    .0284673    -2.110    0.035    -.1158699    -.0042802
    _cons |   -2.283104    .4134953    -5.521    0.000     -3.09354    -1.472668
----------------------------------------------------------------------------

.  log close
```