

## MIT Open Access Articles

*Tax Sensitivity and Home State Preferences in Internet Purchasing*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Ellison, Glenn, and Sara Fisher Ellison. "Tax Sensitivity and Home State Preferences in Internet Purchasing." *American Economic Journal: Economic Policy*, 1(2)(2009): 53–71. © 2009 The American Economic Association.

**As Published:** <http://dx.doi.org/10.1257/pol.1.2.53>

**Publisher:** American Economic Association

**Persistent URL:** <http://hdl.handle.net/1721.1/73030>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of Use:** Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



## Tax Sensitivity and Home State Preferences in Internet Purchasing<sup>†</sup>

By GLENN ELLISON AND SARA FISHER ELLISON\*

*Data on memory modules sales are used to explore aspects of e-retail demand. Aggregate sales are examined in state-level regressions. Discrete choice techniques are used to examine (incomplete) hourly sales data from a price comparison site. We find a strong relationship between e-retail sales to a given state and sales tax rates that apply to purchases from offline retailers, suggesting substantial online-offline substitution and the importance of tax avoidance motives. Geography matters in two ways: consumers prefer purchasing from firms in nearby states and appear to have a separate preference for buying from in-state firms. (JEL D12, H25, H71, L81)*

The recent growth of Internet retail (e-retail) has attracted a great deal of attention in the academic literature and popular press.<sup>1</sup> The future of e-retail is of interest for intellectual and practical reasons. Intellectually, e-retail provides nice opportunities to examine consumer and firm behavior. Practically, e-retail could have significant effects on the economy. It has grown steadily at about 25 percent per year since the collapse of the dot-com “bubble.” And even a small e-retail industry could have a substantial impact on traditional retail, which employs as many Americans as all manufacturing industries combined.

In this paper, we investigate aspects of consumer behavior that will have a substantial impact on the future of Internet and traditional retail. We focus on two main issues. First, we examine the extent to which the success of e-retail is due to the *de facto* tax-free status of most e-retail purchases in the United States.<sup>2</sup> This bears

\* Glenn Ellison: Department of Economics, Massachusetts Institute of Technology, E52-380A, 50 Memorial Drive, Cambridge, MA 02142 and National Bureau of Economic Research (e-mail: [gellison@mit.edu](mailto:gellison@mit.edu)); Sara Fisher Ellison: Department of Economics, Massachusetts Institute of Technology, E52-274B, 50 Memorial Drive, Cambridge, MA 02142–1347 (e-mail: [sellison@mit.edu](mailto:sellison@mit.edu)). We would like to thank Nathan Barczi, Anna Ellison, Caroline Ellison, Eric Moos, and Andrew Sweeting for excellent research assistance, Patrick Goebel and Steve Ellison for assistance with data issues, and Austan Goolsbee, and John Morgan for helpful comments. We also thank the Center for Advanced Studies in the Behavioral Sciences, the Hoover Institute, the Institute for Advanced Study, the Toulouse Network for Information Technology, and the National Science Foundation (SES-0219205, SES-0550897) for financial support.

<sup>†</sup> To comment on this article in the online discussion forum, or to view additional materials, visit the articles page at: <http://www.aeaweb.org/articles.php?doi=10.1257/pol.1.2.53>.

<sup>1</sup> See, for example, Austan Goolsbee (2000), Judith Chevalier and Goolsbee (2003), and Ellison and Ellison (2009).

<sup>2</sup> Forty-five US states levy sales taxes on traditional retail purchases. Each of these states also has laws assessing “use taxes” on purchases that its residents make from out-of-state firms. However, the Supreme Court ruled in *Quill versus North Dakota* (1992) that absent new federal law, a state could not compel a firm without substantial physical “nexus” in that state to collect use taxes on its behalf. The 1998 Internet Tax Nondiscrimination Act makes explicit that Web presence alone does not constitute nexus. While consumers are obligated to self-report use-tax liability, few do in practice. Note that states are able to collect sales taxes on e-retailers’ in-state sales.

on the relative efficiency of e-retail, and is important to understanding what may happen if states are able to tax online sales.<sup>3</sup> Second, we examine the geography of e-retail. It is commonly supposed that geographic differentiation is an important factor allowing traditional retail stores to maintain the markups over marginal cost they need to survive. Branding, obfuscation, or other factors may allow e-retailers to survive even without geographic differentiation, but knowing whether geographic differentiation is really eliminated is also important for understanding what market structure might evolve.<sup>4</sup>

The environment we study is that examined in Ellison and Ellison (2009). We look at consumers shopping for computer memory modules using the pricewatch.com search engine, a Web site which displays sorted lists of products being sold by participating e-retailers, each of which operates their own Web site. For a period of approximately one year, we have hourly data on the 12 lowest prices listed on Pricewatch for each of several products. We know the state in which the e-retailer listing each price is located. Our quantity data is unusually good in one respect and unusually bad in another. The bad part is that we only observe purchases from two of the listed e-retailers (both located in California), so we do not know how many consumers purchased from other e-retailers (or from traditional retailers). The good part is that the data are at the individual order level and include each consumer's location.

The structure of the data provides a nice opportunity to examine consumer preferences and behavior. First, the fact that we observe the state in which each consumer is located creates an opportunity to look at the effects of geography and taxes. We can quantify the extent to which our Web sites sell more in states that levy higher sales taxes—taxes primarily affect the firm's competitive position relative to traditional retailers—and to states that are nearby. Second, there is substantial turnover in the Pricewatch lists, both in terms of which Web sites make the list of the 12 lowest-priced and in their price ranking. Hence, there are many hours in which our two Web sites are competing mostly against other California e-retailers, and others in which they are competing against e-retailers with similar prices in New Jersey, Illinois, Oregon, etc. Looking at how state-specific sales in a given hour are affected by the competitors' locations is another way to identify geography and tax effects.

The paper is organized around two analyses designed to exploit different sources of variation. In Section III, we exploit the time-invariant factors—state-level tax rates and differences in state-to-state shipping times—in the simplest way possible. We run cross-section regressions examining the total number of orders received from each state over the course of the year. These regressions provide clear evidence

<sup>3</sup> The Internet Tax Nondiscrimination Act has been extended to 2014, but there are two other ways in which the de facto tax-free status of Internet purchases in the United States might be threatened in the near future. First, the legal definition of nexus continues to be challenged in the courts on various fronts. For example, Amazon is fighting the state of New York over a law that attempts to broaden nexus to include the presence, in New York, of firms that earn referral fees for sending customers to Amazon. In addition, 18 states have joined the Streamlined Sales Tax Project in an attempt to simplify and harmonize their sales tax laws. The goal of the Project is to encourage online retailers to agree to collect use taxes for sales made in those 18 states and, eventually, to pave the way to federal legislation requiring collection of use taxes.

<sup>4</sup> See Michael D. Smith and Erik Brynjolfsson (2001); Chevalier and Goolsbee (2003); Michael R. Baye and John Morgan (2004); Ellison (2005); Ellison and Ellison (2009); and Chris Forman, Anindya Ghose, and Avi Goldfarb (2009).

that tax savings are an important motivation for online shopping. Our e-retailer's sales are substantially greater in high-tax states than they are in low-tax states. We can provide an additional piece of supporting evidence to bolster the case that the differences are due to taxes and not due to unobserved consumer heterogeneity. Our e-retailer sells *much* less in California than in comparable states. (This would be expected under the tax hypothesis because our e-retailer must charge sales tax on sales to California residents.) These cross-section regressions provide some weak evidence that geography matters for shipping time reasons.

Section IV applies standard demand estimation techniques in an unusual way to exploit the hourly variation in the data. We estimate discrete choice models that use, as their dependent variable, the number of orders of a given product from consumers in a particular state, in a particular hour.<sup>5</sup> The nonstandard part of the application is that we only have data on consumer purchases from two of the listed firms. Normally, one applies discrete choice models to datasets containing the market share of all the firms. Having data on all firms is, however, not necessary to identify the model given that we have substantial intertemporal variation in the characteristics of the competitors. It is this variation that helps us learn about substitution between e-retailers, how much attention consumers pay to geography, taxes, and so forth, simply by looking at how our firm's sales go up and down as rivals' prices and locations change.

The discrete-choice analysis provides some evidence that consumers pay attention to differences in the taxes between e-retailers. There is also evidence that geography still matters. In particular, consumers are estimated to have a preference for purchasing from e-retailers located in their own state.

Our work is related to a number of previous papers. The standard reference on Internet taxation is Goolsbee (2000). It examines a 1997 survey in which 25,000 consumers were asked whether they had ever bought products online. Consumers living in states with higher sales tax rates are found to be more likely to have bought products online. The big picture conclusion is that subjecting e-retailers to taxation could reduce online sales by 24 percent. One motivation for the tax part of our paper is to address a few potential concerns about Goolsbee's work. An elasticity derived from analyzing whether consumers ever purchase anything on the Internet could be very different from the elasticity of total quantity with respect to taxes (which will reflect more the behavior of intensive Internet shoppers), and one could also worry that some of the tax effects he finds could be due to differences in unobserved consumer characteristics across states (driven, for example, by California and Washington having high sales taxes as well as populations inclined to use the Internet).<sup>6</sup> Our tax results also relate, of course, to the literature on the effects of sales taxes on location and consumer behavior in traditional retail, e.g., William F. Fox (1986) and Michael J. Walsh and Jonathan D. Jones (1988).

<sup>5</sup> These regressions include dummy variables for each state so that the results derive from variation that is independent of the variation that identifies the cross-section regressions of Section III.

<sup>6</sup> Despite the examples of California and Washington, sales taxes in the United States are, in fact, not positively correlated with the demographic controls for computer usage we employ. For example, Louisiana, Tennessee, Oklahoma, and Alabama each have one of the eight highest average tax rates in the country and a below average fraction of households with home Internet access. Goolsbee (2000) casts doubt on the unobserved heterogeneity explanation for his results by using extensive household-level demographic controls, by including dummies for Metropolitan Statistical Area (MSA), and by showing that tax rates are not correlated with ownership of computers.

A number of other papers have used data from price search engines to examine aspects of e-retail demand. Smith and Brynjolfsson (2001) examine consumers who visited EvenBetter.com in 1999. It has a puzzling finding on taxes. Consumers are estimated to be twice as sensitive to differences in taxes as they are to differences in item prices.<sup>7</sup> It also finds strong evidence that consumers prefer branded e-retailers over lesser known firms. One limitation is that they do not actually have any quantity data. The quantity data is imputed by assuming that consumers purchased from the e-retailer they visited last. Ellison and Ellison (2009) examine the same Pricewatch data as this paper. They note that Web sites attracting customers via Pricewatch.com have extremely price-elastic demand, and investigate how it is that firms are able to maintain nontrivial markups. The primary observations on this count are that firms engage in a great deal of obfuscation, and that an adverse selection disincentive for price cutting, like that described in Ellison (2005), appears to be present. Baye et al. (2006) examine clickstream data on consumers shopping for PDAs through the Kelkoo.com search engine in 2003. They note that the lowest-priced firms get a large number of extra clicks, and address a number of interesting questions: how price-sensitivity varies with the number of listed firms; how screen- and price-rank separately influence demand; etc.<sup>8</sup>

We are not aware of any other work on spatial differentiation between e-retailers. A number of papers have examined spatial differentiation in traditional retail, including Glen E. Weisbrod, Robert J. Parcells, and Clifford Kern (1984); Lesley Chiou (2009); and Peter Davis (2006).

## I. Data

In this paper, we examine sales of four different types of memory modules, 128MB PC100, 128MB PC133, 256MB PC100, and 256MB PC133.<sup>9</sup> Our price data were obtained by downloading the first (or first and second) screens from Pricewatch's memory price lists on an hourly basis from May 2000 to May 2001 (with some gaps). Pricewatch is a price search engine where potential consumers can choose product categories, 128MB PC100 memory modules, for instance, and be given a list of participating e-retailers selling products in that category sorted by price. Consumers can then click through to one (or more) of the listed e-retailers to obtain more information or complete a transaction. Some information on retailer location, shipping terms, etc., is given to the consumer on the Pricewatch page before clickthrough. Pricewatch is still in operation, although the details of its interface, the rules it enforces, and the markets it serves have changed somewhat over the past few years.

Our data on the 128MB modules include information on the 24 lowest-priced Web sites listed on Pricewatch. The data on 256MB modules include information on

<sup>7</sup> This could be explained as an artifact of price endogeneity if higher prices are associated with higher unobserved quality whereas higher taxes are not.

<sup>8</sup> See Baye, Morgan, and Patrick Scholten (2006) for more on search and price dispersion.

<sup>9</sup> As described in Ellison and Ellison (2009), our e-retailer sells three versions of each of these types of memory modules. The three versions are clearly ranked in quality. In this paper, we restrict our attention to the lowest quality "generic" version of each type of memory module. This is the only quality level for which one can easily use Pricewatch to identify competitors' prices. Low quality memory also accounts for the majority of our firm's sales.

the 12 lowest-priced Web sites. There is a fair amount of turnover and reshuffling of the price lists from day to day (and even from hour to hour in some periods). Over the course of the year, there is a dramatic decrease in prices. For example, in the space of a year, the price of a 128MB module fell from about \$120 to about \$20.

Pricewatch does not calculate sales taxes for consumers on these pages, but it does list the home state of each e-retailer so that a consumer who knew the tax rate in his home state (and understood that sales taxes will apply if and only if he buys from an in-state firm) could take sales tax differences into account. We downloaded the state locations as well.

We obtained quantity data for these products from an Internet retailer that gets most of its traffic from Pricewatch. It operates two similar Web sites that typically have different prices for the products studied.<sup>10</sup> The quantity data, again, cover May 2000 to May 2001 (with some gaps). The raw data are at the level of the individual order. We use data on approximately 15,000 orders. The available data on each order include the Web site from which the customer made the order, details on what was ordered, and the shipping address. Our e-retailer is just one of many listing products for sale on Pricewatch. A rough estimate is that 100,000 other consumers visited Pricewatch during our sample period and purchased a corresponding product from one of the e-retailers for which we do not have quantity data.

We also use a few state-level variables. The most important of these is the state's average sales tax rate. Sales tax rates vary by county and locality in many states. Our data are averages across the various jurisdictions within a state computed by a private firm. We collected data on UPS ground shipping times by querying the UPS Web site. These data include both shipping times from our e-retailer's zip code to each state, and a state-to-state shipping time matrix.<sup>11</sup> We include a dummy variable for California because sales from our retailer into California will be subject to state sales taxes. Our other state-level variables come from US Census Bureau datasets: the fraction of households with home Internet access as reported in a 2001 survey, the population of each state in the 2000 census, and the number of computer stores and gas stations reported in the 1997 Census of Retail Industries.

## II. Analysis of Aggregate State-Level Sales

In this section, we take the most straightforward approach to examine how the time invariant variables in our dataset—sales tax rates and shipping times—affect consumer demand. We construct measures of the total number of orders received from each state, and use regressions to, for example, look at whether our e-retailer sells more in states with high sales taxes than in states with low sales taxes.

<sup>10</sup> There are several possible motivations for having multiple Web sites. The sites may be given different looks and consumers may have heterogeneous reactions. It allows the Web sites to be more specialized (which seems to be attractive to some consumers). It facilitates experimentation. It may help promote private-label branded products, and the firm may occupy multiple places on the Pricewatch screen.

<sup>11</sup> UPS provides these data on a zip-code-to-zip-code basis and there can be some within-state variation. We typically collected data using one zip code from the the largest population center in the state. In some cases, where a state did not have one dominant population center and the shipping time varied, we took an average of the times for the two largest population centers.



### A. Summary Statistics

The regressions in this section will have 51 observations, one for each state and one for the District of Columbia. We use two primary dependent variables. *Quantity128* is the number of orders for 128MB modules received over the course of the year from a given state. *Quantity256* is the corresponding number for 256MB modules.<sup>12</sup> Summary statistics for the basic regressions are presented in Table 1. Our e-retailer sells 204 128MB memory modules to the average state over the course of the year. This ranges from a low of 19 in the District of Columbia to a high of 762 in Texas. Unit sales of 256MB memory modules are about half as large. The average sales tax rate is 5.7 percent. Four states have no state or local sales taxes. The UPS ground shipping time from our retailer to the average state is about four days.<sup>13</sup> The percentage of households with home Internet access varies from a low of 40.6 percent in the District of Columbia to a high of 70.2 percent in New Hampshire. The average state has 230 computer stores. The ratio of computer stores to gas stations ranges from a low of 0.041 in West Virginia to a high of 0.184 in California.

Although prices are not used in this state-level analysis, they are relevant for the interpretation of some results. The mean price of a 128MB memory module is \$70. The mean price of a 256MB memory module is \$139. A 1 percentage point difference in tax rates, then, adds \$0.70, on average, to a 128MB module and \$1.39 to a 256MB module.

### B. Results

To analyze how the number of orders received from state  $s$  is related to the state's tax rate, we estimate a negative binomial regression model, i.e., we assume

$$Quantity_s \sim \text{Poisson}(\mu_s)$$

$$\begin{aligned} \log(\mu_s) = & \beta_0 + \beta_1 \text{OfflineSalesTaxRate}_s + \beta_2 \text{California}_s + \beta_3 \text{ShippingTime}_s \\ & + \beta_4 \frac{\text{ComputerStores}_s}{\text{GasStations}_s} + \beta_5 \text{InternetAccess}_s + \beta_6 \log(\text{Population}_s) \\ & + \varepsilon_s, \end{aligned}$$

where the  $\varepsilon_s$  are independent random variables with  $e^{\varepsilon_s} \sim \Gamma(\theta, \theta)$ , and estimate the parameters by maximum likelihood.<sup>14</sup> One can think of this as similar to estimating a linear regression with  $\log Q_s$  as the dependent variable.

<sup>12</sup> Note that in doing this we are summing both over the two Web sites for which we have data and over the two speeds of each size memory module, PC100 and PC133. We do this because there is no reason to expect that taxes or geography would have a different impact across Web sites or speeds.

<sup>13</sup> The minimum value of 1.5 days reflects that shipping times are one day for shipments to Southern California and two days for shipments to Northern California.

<sup>14</sup> The Poisson regression model is the special case of the negative binomial with  $\theta = \infty$ . In applied work, it is common to find that a specification test can reject the Poisson model in favor of other models that allow for more dispersion. The particular assumption that the errors are distributed like the logarithm of a gamma random variable (as opposed to being normally distributed, for example) is motivated by the fact that a relationship between

TABLE 1—SUMMARY STATISTICS FOR STATE-LEVEL REGRESSIONS

Variable	Mean	SD	Min	Max
Quantity128	203.5	176.0	19.0	762.0
Quantity256	85.6	84.3	5.0	391.0
OfflineSalesTaxRate	0.057	0.021	0.000	0.084
InternetAccess	0.57	0.07	0.41	0.70
ShippingTime	3.89	0.92	1.50	5.00
$\frac{\text{ComputerStores}}{\text{GasStations}}$	0.092	0.034	0.041	0.184
log (Population)	15.02	1.04	13.11	17.33
Income	41.35	6.29	29.70	55.15
FractionBA	0.15	0.03	0.09	0.22
FractionGrad	0.09	0.03	0.05	0.21
HomeComputer	0.57	0.06	0.42	0.68

Table 2 presents coefficients obtained from estimating the regression above on the total unit sales to each of the 51 states. The first column uses 128MB memory module sales as the dependent variable. The results indicate that sales taxes have a large effect on online sales. The 5.94 coefficient estimate on *OfflineSalesTaxRate* indicates that a 1 percentage point increase in a state's sales tax increases the number of orders our e-retailer receives from that state by about 6 percent. The average sales tax rate in our data is 5.7 percent. Hence, in a typical state, online purchases would be predicted to decrease by about 30 percent if the offline sales tax were eliminated. Goolsbee argues that this is a good forecast for the impact of taxing online sales—the implicit assumption is that achieving tax parity between online and offline retail should have a similar effect regardless of whether it is achieved by increasing online taxes or by decreasing offline taxes.

The coefficient on the *California* dummy provides additional support for the view that what we have estimated is a tax effect and not an artifact of unobserved state-level heterogeneity. What would we predict about our firm's sales to California if the coefficient on *OfflineSalesTaxRate* is truly a tax effect? First, since our firm has no tax advantage relative to brick and mortar stores in California—its California customers must pay sales tax—we would expect its sales to be about 35 percent lower than one would otherwise predict, given state covariates.<sup>15</sup> Second, our firm has a disadvantage relative to non-California e-retailers when selling in California. One would expect that this disadvantage would lead to an additional reduction in sales. The estimated coefficient on *California* indicates that sales to California customers are about 67 percent lower than sales to comparable states. It is implausible that an effect of this magnitude could be due to an unobserved distaste for online shopping on the part of Californians.

Poisson and gamma random variables allows the likelihood to be evaluated without a numerical integration. The distribution of  $Q_s$  turns out to be negative binomial which is what gives the model its name. Section 19.9.4 of William H. Greene (1997) provides a clear description of the model. Jerry Hausman, Brownwyn H. Hall, and Zvi Griliches (1984) discuss a number of models for count data.

<sup>15</sup> The *OfflineSalesTaxRate* variable is equal to 7.25 percent for California.



TABLE 2—STATE-LEVEL REGRESSIONS

	Dependent variable			
	Quantity128	Quantity256	Quantity128	Quantity256
OfflineSalesTaxRate	5.96 (2.16)	6.33 (2.59)	6.14 (3.04)	7.21 (2.93)
California	-1.03 (4.01)	-0.84 (3.21)	-0.99 (4.64)	-0.81 (3.43)
ShippingTime	-0.10 (2.04)	-0.07 (1.32)	-0.08 (1.40)	0.002 (0.03)
Internet Access	1.89 (2.62)	1.04 (1.25)	3.72 (2.82)	0.36 (0.23)
<u>ComputerStores</u> GasStations	1.90 (1.11)	4.39 (2.29)	2.61 (1.42)	5.48 (2.41)
log (Population)	0.85 (20.54)	0.89 (18.98)	0.90 (22.77)	0.93 (19.24)
Income			-0.04 (4.05)	-0.02 (2.01)
Fraction BA			1.53 (0.78)	1.50 (0.63)
Fraction Grad			1.49 (0.69)	-1.54 (0.56)
Home Computer			0.12 (0.08)	2.15 (1.20)
Observations	51	51	51	51

Notes: The first two columns report estimates from negative binomial regressions with *Quantity128* and *Quantity256* as dependent variables. The third and fourth columns add additional control variables. *t*-statistics are in parentheses.

The estimate on the *ShippingTime* variable provides some weak evidence that geography still matters on the Internet. Sales are estimated to be reduced by about 10 percent if UPS ground shipping to the destination state is one day longer.

The coefficients on the other control variables seem reasonable. Sales are higher in states where the fraction of residents with Internet access is higher. We cannot reject that the coefficient is one, which would correspond with sales being proportional to the number of people with home Internet access. The coefficient on the computer store–gas station ratio might be expected to have either sign. It reflects both interest in computers and the availability of computer parts at traditional retail stores. The estimated coefficient is positive but not statistically significant. Population is obviously a strong determinant of aggregate sales. Potential reasons why the coefficient might be less than one include that population is an imperfect proxy for the potential market size (which is affected by income, business activity, and other factors), and that larger population states may have better offline retail.

The second column of Table 2 presents coefficient estimates from a regression with orders for 256MB memory modules as the dependent variable. These results are very similar. Sales are substantially higher in states that levy higher sales taxes on traditional retail purchases. Sales are notably lower in California. There is weak evidence that shipping times may affect sales. The effects of the other demographic variables are similar.

As previously mentioned, a potential concern is that the estimated tax effect could be an artifact of omitted state characteristics correlated with interest in online shopping and the tax variable.<sup>16</sup> To address this concern, the third and fourth columns add a number of additional demographic characteristics that one might imagine were correlated with interest in online shopping: the state's median household income (in thousands); the fraction of households with a computer; and two measures of educational attainment of the over-25 population, the fraction with a bachelor's degree and the fraction with a graduate or professional degree. All four of the new measures are positively correlated with home Internet access and computer store-to-gas station ratio. Adding these somewhat collinear variables leads to some changes in the coefficients on the existing control variables, but only strengthens the tax-related conclusions.<sup>17</sup> The coefficient on the sales tax rate gets slightly larger and increases in significance in both regressions. The coefficient on the California dummy gets slightly smaller and increases in significance.<sup>18</sup>

### III. A Discrete-Choice Analysis

The Pricewatch environment exhibits an unusual degree of short-term variation in competitive conditions. This variation provides a nice opportunity to gain additional insight into e-retail demand and consumer behavior. In this section, we use discrete-choice models to explore substitution between e-retailers and the effects of geography and sales taxes.

#### A. Motivation

The analysis in this section is designed to exploit an important source of short-term variation in our data, turnover in the relative price rankings. The following discussion should provide some intuition first, for how that turnover arises, and second, on how it might be useful for identification.

Reshuffling in the price rankings are quite common.<sup>19</sup> There are an average of 4.1 price changes per day among the top 12 firms. Firms jumping onto the list account for about one of these. Firms raising their prices to drop off the list are about 0.8. The remaining 2.3 are smaller price changes that just reorder the list. About three-quarters of these are price cuts. They are usually \$1 or \$2, and move the price-cutting firm up by one to four places. In total, the 4.1 price changes result in 18.3 rank changes. At the firm level, there is clearly heterogeneity in position preferences. Four of the 10 Web sites we see most often have average positions of between 2 and 4 immediately following their price changes, whereas 4 others have average post-reset positions between 10 and 11.5. Activity levels are relatively similar— rates of price

<sup>16</sup> The states with the highest sales tax rates are Louisiana, Tennessee, Washington, New York, and Texas. The states with the lowest tax rates are Montana, Delaware, New Hampshire, Oregon, and Alaska.

<sup>17</sup> The *HomeComputer* and *InternetAccess* variables are particularly collinear with a correlation of 0.90.

<sup>18</sup> Additional specifications were estimated with state-level controls such as unemployment rate, racial composition, and median age, with very similar results.

<sup>19</sup> All statistics in this paragraph refer to the 128MB PC100 data.

changes among the 10 most frequently present Web sites vary from 0.2 to 0.5 price changes per day.

The retailers selling through Pricewatch are not large firms with sophisticated operations research staffs. Some are probably being run out of the back room of a retail computer store. The retailer that provided us with data is probably more typical (a proprietor working long hours, a single part-time programmer helping maintain the Web sites, someone doing accounting, a few customer service representatives, and shipping room staff putting products in boxes). The firm had dozens of products listed for sale on Pricewatch. The proprietor would manually monitor the position of the more important ones on the Pricewatch screen during breaks in his other management responsibilities.

The largest source of churning in the Pricewatch lists is a simple mechanical one. The price of a 128MB memory module declined from about \$120 early in the year to about \$25 by the end of our sample. Firms naturally cut retail prices as wholesale acquisition costs declined. There was also substantial entry and exit. One hundred and thirteen Web sites appear in the top 24 at some point during the year. In addition, we cannot rule out some demand-driven price changes.

Table 3 presents an actual example of turnover that is somewhat atypical but makes for a nice illustration. It shows the 12 e-retailers listed on the first screen of Pricewatch's 128MB PC100 memory page at 9 a.m. and 11 a.m. on August 1, 2000. Note that two e-retailers made price changes between these two times. Coast-to-Coast Memory of New Jersey, which offered the lowest price of \$112 at 9 a.m., raised its price sufficiently so as to disappear from the top 12 by 11 a.m. UpgradePlanet.com of Virginia, which was on the second page of the 9 a.m. list at \$128, reduced its price to \$111 and took over the top slot.

The first three columns show information presented on Pricewatch: the e-retailers' names, their locations, and their prices. The fourth through the sixth columns contain numbers not presented on the Pricewatch site but which consumers could compute from the given information; and the tax-inclusive prices that customers in New Jersey, Virginia, and California, respectively, would pay if they purchased from each of the e-retailers.<sup>20</sup>

We use this example to help illustrate how price turnover is useful for identification. Recall that we observe sales for two Web sites. However, we observe not just total sales but sales *into each state* at each hour. This fact, along with the turnover in relative price rankings, is crucial for our estimation strategy.

To think about how this works, suppose that our sales data were from Connect Computers.<sup>21</sup> At 9 a.m., Connect Computers' tax-inclusive price for New Jersey residents is lower than that of any other Web site. At 11 a.m., Connect Computers' tax-inclusive price for New Jersey residents is only the second lowest. Accordingly, if consumers pay attention to sales taxes, we would expect Connect Computers' sales into New Jersey to be higher at 9 a.m. than at 11 a.m. Similarly, its sales into Virginia would be higher at 11 a.m. than at 9 a.m. We can estimate tax effects controlling

<sup>20</sup> A consumer, of course, would need to know his or her local sales tax rate and the fact that sales taxes are only assessed on in-state sales to make this calculation.

<sup>21</sup> Connect Computers is, in fact, not one of the Web sites from which we have data.

TABLE 3—SAMPLE PRICEWATCH RANKINGS: 128MB PC100 MEMORY MODULES ON AUGUST 1, 2000

Web site	State	Price (USD)			
		into NJ	into VA	into CA	
<i>Pricewatch ranking at 9:01 a.m. EDT</i>					
Coast-to-Coast Memory	NJ	112	118.72	112	112
Connect Computers	CA	113	113	113	121.64
Computer Craft	FL	114	114	114	114
Advanced PCBoost	CA	115	115	115	123.80
1st Choice Memory	CA	116	116	116	124.87
Jazz Technology	CA	117	117	117	125.95
Memplus.com	CA	117	117	117	125.95
Portatech	CA	119	119	119	128.10
Augustus Technology	CA	120	120	120	129.18
EconoPC	IL	120	120	120	120
Advanced Vision	CA	121	121	121	130.26
Computer Super Sale	IL	122	122	122	122
<i>Pricewatch ranking at 11:01 a.m. EDT</i>					
UpgradePlanet.com	VA	111	111	115.99	111
Connect Computers	CA	113	113	113	121.64
Computer Craft	FL	114	114	114	114
Advanced PCBoost	CA	115	115	115	123.80
1st Choice Memory	CA	116	116	116	124.87
Jazz Technology	CA	117	117	117	125.95
Memplus.com	CA	117	117	117	125.95
Portatech	CA	119	119	119	128.10
Augustus Technology	CA	120	120	120	129.18
EconoPC	IL	120	120	120	120
Advanced Vision	CA	121	121	121	130.26
Computer Super Sale	IL	122	122	122	122

for a home-state preference by looking at how the magnitude of the 9 a.m.–11 a.m. drop in Connect's New Jersey sales compares with the 9 a.m.–11 a.m. increase in Connect's Virginia sales. A comparison of Connect's California sales at 9 a.m. and 11 a.m. will teach us about substitution between retailers. Shipping times from New Jersey and Virginia to California are the same, so the comparison should help us learn how many consumers shift from the second-lowest to the lowest-priced firm when the lowest-priced firm reduces its price by one dollar.

### B. Methodology

Let  $N_{sht}$  be the number of consumers in state  $s$  purchasing a particular type of memory module in hour  $h$  of day  $t$  from the 24 (or 12 for 256MB modules) Web sites where we observe prices. Assume that consumer  $k$ 's utility, if he purchases from Web site  $i$ , is

$$u_{iksht} = \beta_1(\text{Price}_{iht} + \beta_2\text{SalesTax}_{isht}) + \beta_3\text{ShippingTime}_{is} + \beta_4\text{HomeState}_{is} \\ + \beta_5\text{NeighborState}_{is} + \beta_6\text{SecondScreen}_{iht} + \beta_7\text{SiteB}_i + \varepsilon_{ik},$$

where *SalesTax* is the sales tax in dollars due on the purchase, *ShippingTime* is the UPS ground shipping time, *HomeState* is a dummy variable for whether Web site  $i$

is in state  $s$ , *NeighborState* is a dummy for whether Web site  $i$  is in a state bordering state  $s$ , *SecondScreen* is a dummy indicating whether Web site  $i$  only appears on the second screen of results, *SiteB* is a dummy for one of the two Web sites from which we have quantity data, and  $\varepsilon_{ik}$  is a logit random variable independent of the right-hand-side variables (and of the additional right-hand-side variables and the error  $\eta_{hst}$  introduced below).<sup>22</sup>

Writing  $\mathbf{X}_{sht}$  for the vector of attributes on the right-hand side of this expression, we have the familiar logit formula for the number of consumers in state  $s$  buying from Web site  $i$ , conditional on the total number of purchases  $N_{sht}$ :

$$E(Q_{isht} | \mathbf{X}_{sht}, N_{sht}) = N_{sht} \frac{e^{\beta \mathbf{X}_{isht}}}{\sum_{j=1}^{24} e^{\beta \mathbf{X}_{jsht}}}.$$

Our dataset only contains sales from two Web sites. It does not contain the number of consumers purchasing from other Web sites, from traditional retailers, or not at all. The total number of consumers buying through Pricewatch is affected by a number of factors. There are clear day-of-week and hour-of-day effects; Internet use is climbing over our sample period. There are substantial price declines that should increase aggregate demand. There is variation in the online-offline price gap, and there may be intertemporal price effects with the size of the potential consumer pool at a given time being affected by past prices. Our data will not allow us to separately identify all of these effects. The approach we take is simply to specify a flexible functional form for the aggregate Pricewatch demand that could reflect each of the effects. Specifically, we assume

$$(1) \quad N_{sht} = \delta_s \bar{q}_h e^{\gamma_1 \text{MinPrice}_{ht} + \gamma_2 \text{Weekend}_t + \gamma_3 \text{TimeTrend}1_t + \dots + \gamma_6 \text{TimeTrend}4_t} + \eta_{hst},$$

where  $\delta_s$  is a state fixed effect to be estimated;  $\bar{q}_h$  is an hour-of-day fixed effect,  $\text{MinPrice}_{ht}$  is the lowest price listed on Pricewatch;  $\text{Weekend}_t$  is a weekend dummy, the *TimeTrend* variables allow for linear time trends with slopes changing every 90 days; and  $\eta_{hst}$  is a random error term assumed to have mean zero, conditional on the right-hand side variables in this equation.<sup>23</sup> Given the assumptions on  $\varepsilon_{ik}$  and  $\eta_{hst}$ , and defining  $\mathbf{W}_{ht}$  as the vector of variables in the exponential term in equation (1), iterated expectations implies that

$$(2) \quad E(Q_{isht} | \mathbf{X}_{sht}, \delta_s, \bar{q}_h, \mathbf{W}_{ht}) = \delta_s \bar{q}_h e^{\gamma \mathbf{W}_{ht}} \frac{e^{\beta \mathbf{X}_{isht}}}{\sum_{j=1}^{24} e^{\beta \mathbf{X}_{jsht}}}.$$

<sup>22</sup> Note that  $\text{Price}_{iht}$  does not include shipping, but for institutional reasons, shipping is typically \$11 regardless of distance.

<sup>23</sup> Note that we do not include an “outside good” in the discrete-choice set as one might do to attempt to estimate the effect of a logit-inclusive value on aggregate demand. We are thus implicitly assuming, for example, that the total sales by Pricewatch e-retailers to state  $s$  are not affected by the states in which the e-retailers are located and the difference between the  $n^{\text{th}}$  lowest price and the lowest price. We do this because we have little data to estimate such effects, think they must be small, and prefer a more parsimonious model in which fewer coefficients are used to capture aggregate demand effects. Reasons why any inclusive-value effects would be hard to find include that prices on Pricewatch are almost always tightly bunched, and that, in any state other than California, having more than one or two e-retailers on the list from that state is extremely rare.

We estimate equation (2) via nonlinear least squares, using hour Web site destination state sales as the dependent variable. The large number of observations reflects that we have data on hourly sales into 50 states by two Web sites in up to 8,000 hours.<sup>24</sup> We carry out the estimation four times to obtain independent estimates using data on each of the four products: 128MB PC100 modules, 128MB PC133 modules, 256MB PC100 modules, and 256MB PC133 modules.

Note that we are assuming that it is not necessary to use instruments for the prices on the right-hand side of the above equations. We think endogeneity is not a big concern for two main reasons. First, we think the firms have little information about demand shocks. We say this because there is little information to be had (one would need to identify demand shocks that made moving up on the list and then drifting back down over the course of a few days better than staying put), and because our interaction with the one firm leads us to believe that firms have very little capability to uncover sophisticated demand patterns. Second, even if firms did have information about demand shocks not available to us, we do not think it could be a large part of the variation we are using to identify price and, especially, tax effects. Returning to our example, the tax effect and home-state effect are identified by the sizes of the increase/decrease in Connect Computers' sales into New Jersey and Virginia when Coast-to-Coast Memory and UpgradePlanet.com switch places. Both Coast-to-Coast Memory and UpgradePlanet.com make the vast majority of their sales outside of these states, so even if they were perfectly informed, one would expect that their decisions about where to be on the list would mostly reflect national demand shocks, not demand shocks in New Jersey and Virginia. With regard to price effects, our main estimate is an effect of relative prices. Most rank changes do not reflect active decision-making; the 4.1 price changes per day lead to 18.3 rank changes per day as each price change bumps several other firms up or down by one position. The drift in position between price changes comprises much of the variation that drives our estimates. Even if firms were jumping up when it was a relatively good time to be high on the list, we would be seeing our firm jumping into a high position when it knew demand was good in that position and being pushed out of a high position when other firms knew demand was good in that position. Hence, it is not clear if there would be a correlation between demand levels and our firm's relative price. Endogeneity could be more of a worry with regard to the price effect that is included in the number-of-consumers equation. This estimate is not a focus, however, so we are happy to regard it as a reduced-form control variable rather than as a demand elasticity.

### *C. Summary Statistics*

Table 4 reports summary statistics separately for each of the four types of memory modules. The unit of observation is an hour-state-Web site. Given that our Web sites

<sup>24</sup> We drop California from the analysis because the fact that our retailer and most other retailers are located there would make demand different under reasonable departures from our assumptions; "outside goods" could be more important because there will be hours when all of the top firms are California firms that would impose sales tax; and the impact of taxes would differ if there was a random coefficient on the tax variable rather than a coefficient that is homogeneous across consumers.



TABLE 4—SUMMARY STATISTICS FOR INDIVIDUAL-LEVEL REGRESSIONS

Variable	Mean	SD	Min	Max	Mean	SD	Min	Max
	128MB PC100				128MB PC133			
Quantity	0.007	0.085	0	4	0.006	0.077	0	4
Price	66.24	34.37	21	123	73.82	36.55	21	131
MinPrice	62.51	33.36	20	122	71.50	37.06	20	131
Rank	6.40	4.12	1	21	5.97	4.41	1	21
Observations	793,950				707,350			
	256MB PC100				256MB PC133			
Quantity	0.002	0.047	0	3	0.004	0.064	0	4
Price	130.77	65.22	43	258	146.52	79.79	39	291
MinPrice	121.13	58.39	43	215	135.99	73.50	39	269
Rank	5.79	3.15	1	12	6.20	3.08	1	12
Observations	645,200				648,150			

sell zero memory modules to a typical state in a typical hour, average sales figures at this level are quite low. For example, the average number of 128MB PC100 modules sold by a Web site in one particular hour to one particular state is 0.007.<sup>25</sup> *Price* is the price charged by our Web sites. Mean prices are about \$70 for 128MB modules and about \$140 for the 256MB modules. The dramatic price declines that occurred over the year are visible in the minimums and maximums for this variable. *MinPrice* is the lowest price listed on Pricewatch in the hour in question. Our firm's 128MB prices are about \$2 to \$4 higher than this on average. Its average rank on the Pricewatch list is sixth. The average gap between our firm's 256MB price and the lowest available price is larger. Much of this is due to a period when one firm offered these modules at a very low price. Our firm's average rank is still about sixth. We do not include California in our estimation, so all consumers in the dataset would not need to pay sales tax to buy from our Web sites. They would need to pay sales tax if they bought from an in-state firm.<sup>26</sup> The average tax that would be paid if buying from an in-state firm is about \$5 for a 128MB module and about \$9 for a 256MB module.

#### D. Results

Table 5 presents coefficient estimates obtained by performing separate nonlinear least squares estimations on the data for each of the four products: 128MB PC100, 128MB PC133, 256MB PC100, and 256MB PC133. In many ways, the four sets of results are similar.

The most basic fact about the Pricewatch environment is that it is intensely competitive (as we previously noted in Ellison and Ellison 2009). The coefficients on *Price* in the four columns range from  $-0.40$  to  $-0.81$ . The estimate for 128MB PC100 memory modules, for example, corresponds to an own-price elasticity of  $-35$

<sup>25</sup> We count a single order of multiple memory modules as having quantity one. For most of our time period, our firm limited purchases of memory modules to one per order.

<sup>26</sup> Customers living in states that do not have a sales tax are an exception.

TABLE 5—DISCRETE-CHOICE MODEL OF HOURLY SALES OF MEMORY MODULES IN TEN STATES

	Product			
	128MB PC100	128MB PC133	256MB PC100	256MB PC133
<i>Variables affecting choices between sites</i>				
Price	-0.56 (64.17)	-0.81 (53.76)	-0.43 (37.08)	-0.40 (58.46)
SalesTax	0.05 (0.59)	0.33 (3.89)	0.06 (0.78)	0.95 (2.50)
HomeState	0.47 (2.27)	1.40 (5.83)	1.06 (3.33)	0.75 (1.21)
NeighborState	-0.05 (0.38)	-0.34 (2.33)	0.73 (4.85)	0.64 (6.95)
ShippingTime	-0.03 (1.28)	-0.07 (2.00)	0.06 (1.69)	-0.05 (2.09)
SecondScreen	-1.32 (1.94)	-0.60 (2.68)		
<i>Variables affecting total Pricewatch demand</i>				
Weekend	-0.42 (20.38)	-0.41 (16.21)	-0.37 (9.42)	-0.73 (20.69)
MinPrice	-0.03 (14.09)	-0.03 (13.04)	-0.02 (11.00)	-0.03 (16.82)
TimeTrend1	0.02 (12.66)	0.02 (15.39)	0.01 (3.39)	0.03 (8.47)
TimeTrend2	-0.04 (11.54)	-0.03 (8.93)	-0.01 (2.95)	-0.05 (8.62)
TimeTrend3	0.02 (10.36)	0.00 (1.01)	-0.00 (2.30)	0.01 (4.08)
TimeTrend4	-0.00 (4.00)	-0.01 (5.60)	0.01 (5.34)	0.01 (10.86)
Observations	793,950	707,350	645,200	648,150
R <sup>2</sup>	0.03	0.02	0.01	0.03

Notes: Dependent variables are the number of distinct customers in each of the 49 states ordering from Web sites A and B approximately 7,900 hours. Regressions also contain state and Web site dummies. *t*-statistics are in parentheses.

(holding all variables fixed at their sample means). The estimates are extraordinarily significant. The decrease in demand that occurs when our firm raises its price (or is undercut) is so large as to be impossible to miss.

The coefficients on the time-trend variables illustrate the growth (and decline) of Pricewatch over our sample period. The coefficient on *TimeTrend1* in the first column indicates that overall demand was growing at about 2 percent per day (equivalent to 75 percent per month) in the first three months of our sample (May–August 2000). Growth rates for later periods are obtained by adding all of the earlier coefficients. The estimates indicate that sales decreased 50 percent per month in the fall of 2000, and then declined by a very small percentage per month over the final two quarters of our sample. Growth rates for the other three products are similar, suggesting these patterns are not just product-specific fluctuations.

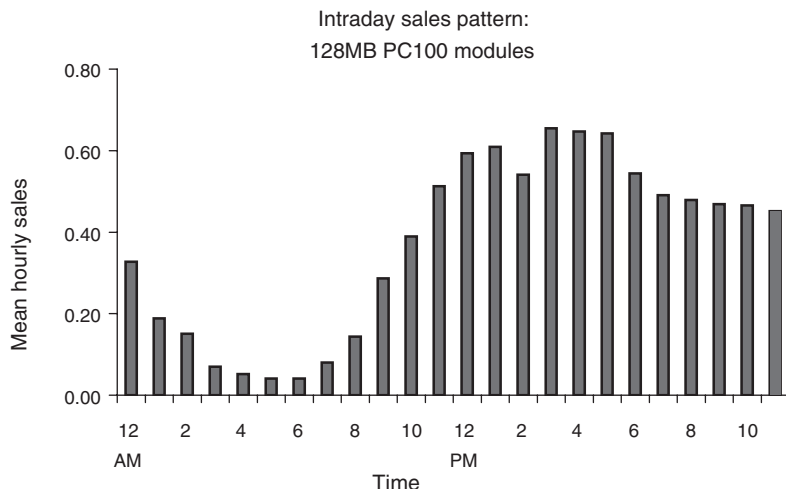


FIGURE 1. INTRADAY SALES PATTERN: 128MB PC100 MEMORY MODULES

Figure 1 presents a graph of the hour dummies.<sup>27</sup> They indicate that online shopping picks up substantially between 7 a.m. and 11 a.m., continues at approximately the 11 a.m. level past the normal workday, remains at about two-thirds of the peak value until midnight, and then drops off substantially until 6 a.m. The large number of late-night purchases suggests that greater availability may be an important factor differentiating e-retail from traditional retail.

### E. Taxes

Recall that in our demand specification consumers are assumed to evaluate products on the basis of  $Price + \beta_2 SalesTax$ , with  $SalesTax$  measured in dollars. Hence, an estimate of one on the  $SalesTax$  coefficient would correspond to the standard rational model in which consumers care only about their total expenditure, and an estimate of zero would correspond to consumers who are entirely insensitive to tax differences.<sup>28</sup>

The most general conclusion we draw from the four sets of results is that consumers pay less attention to sales taxes than the standard model predicts. The estimates in the four columns are 0.05 (standard error 0.08), 0.33 (standard error 0.08), 0.06 (standard error 0.07), and 0.95 (standard error 0.38). Note that the first three are significantly different from unity while the second and last are significantly different from zero. We interpret this as evidence that consumers are paying attention to taxes but not as much as to price differences of a similar magnitude.

<sup>27</sup> Recall that we simply set these to the sample mean quantities for each hour rather than making them part of the nonlinear least squares estimation. Sample means are computed on a time-zone adjusted basis with the times of all purchases being recorded from the consumer's perspective.

<sup>28</sup> There are clearly other "rational" models in which the coefficient would be greater than or less than one. An example of the former is if price is a signal of quality so that a high price-zero tax offer is preferable to a low-price-high-tax offer with the same total expenditure. Examples of the latter would be a model in which consumers have nonselfish preferences and value payments to local governments.

It is important to note that the fact that consumers pay less attention to tax differences than to price differences does not imply that sales taxes are not important. Our consumers are extraordinarily sensitive to price differences, so even if the coefficient on the *SalesTax* variable was 0.3, our estimates would be that a firm that must collect a 6 percent sales tax would have its sales decline by about 50 percent.

#### F. Geography

Geography enters our demand model in two ways. First, *ShippingTime* allows for the possibility that consumers may prefer to buy from e-retailers in nearby states because they will have faster delivery times with standard ground shipping. We find limited evidence of such an effect in these regressions. Two of the four estimates are negative and significant. A coefficient of 0.05 on the *ShippingTime* variable would indicate that the extra shipping time required to ship a product across the country—about 4 days more than a local shipment—reduces demand by a little less than 20 percent. This effect is, however, small relative to the price effects in our model. A 50 cent price increase will also reduce demand by 20 percent.

Second, we included the *HomeState* and *NeighborState* dummies to allow for the possibility that consumers may have an additional preference for buying from in-state firms. Here, we find stronger evidence that geography does matter. All four estimates on the *HomeState* variable are positive and three of the four are significant. Two of the four *NeighborState* estimates are positive and significant. The magnitudes of the coefficient estimates indicate that the home-state preference will roughly offset a two dollar price difference. Such preferences could exist for a variety of reasons. Consumers may simply prefer patronizing a local firm. Or local firms could enjoy a reputational advantage. Or there may be direct benefits from purchasing locally, such as the ability to make returns in person.<sup>29</sup> Regardless of their origin, however, such preferences would also favor a more geographically dispersed e-retail sector.

In light of our earlier estimates that consumers pay less attention to differences in sales taxes than to differences in prices, the home-state preference will outweigh the sales tax disadvantage on moderately priced items. For example, if the *SalesTax* coefficient is 0.33, the *Price* coefficient is  $-0.5$ , the *HomeState* coefficient is 1.0 and the tax rate is 6 percent, then the home-state preference will outweigh the tax disadvantage on items costing \$100 or less. The finding that the home-state preference is nearly strong enough to outweigh the tax disadvantage of buying from an in-state firm contrasts with what we saw in our state-level analysis. There, we saw that our firm sells much less in California than in other states. Here, we find that it fares worse in other states when competing against a local firm than when competing against other out-of-state firms. One way to reconcile these findings could be a model in which there is heterogeneity in the home-state preference. For example, it could be that 25 percent of consumers have a strong home-state preference and 75

<sup>29</sup> In practice, we doubt that the latter two effects are very important. We believe that many of these firms are pure e-retailers that are not known in their communities and that most would not accept in-person returns. The firm we visited did not even have a sign on its building giving its name, and explained that this reduced the risk of theft. Having recourse in state courts could be another benefit.

percent have none. Then, when our firm is competing for California customers, it must split the 25 percent who like California firms with several other California retailers (resulting in a low share), whereas it would sell less in New Jersey when it is competing for a New Jersey customer because it would lose almost all of the 25 percent with a home-state preference and still face 10 competitors for the other 75 percent of the New Jersey population. We cannot, however, provide any strong support for this story over many others, e.g., it could be that consumers in some states have a strong preference for buying locally whereas California consumers do not.

#### IV. Conclusion

In this paper, we have examined Internet retail demand using two different approaches: a cross-sectional analysis of demand in different states and a discrete-choice analysis of demand at an hourly frequency. The two analyses exploit separate sources of variation in the data. The state-level analysis ignores all of the variation in competitive conditions, and the discrete-choice analysis uses state fixed effects to absorb any persistent factors like tax rates.

Our most basic conclusion on sales taxes is that they are an important driver of e-retail activity. Our state-level regressions clearly show that sales are higher in states that levy higher sales taxes on traditional retail purchases. The fact that the Web sites we study sell so little in California is strong evidence that what we are picking up is a tax effect and not some artifact of unobserved heterogeneity. The environment we study is somewhat unusual in that consumers are highly savvy and price-sensitive, but in this environment at least, we would agree with Goolsbee's (2001) conclusion that applying sales taxes to e-retail sales could reduce e-retail demand by one-quarter or more. In our discrete-choice analysis, we find that consumers do not pay as much attention to differences in taxes as they do to differences in pre-tax prices when choosing between e-retailers. Taxes do matter to consumers, though, and, given how tightly distributed prices in this market are, they can have large effects on consumer behavior.

The state-level analysis indicates that geography still matters in e-retail. The Web sites we study make more sales to states that are closer to California in a shipping-time metric. In the discrete-choice analysis, we find that consumers have a preference for buying from in-state e-retailers. We think this is an interesting result on the sources of geographic differentiation. It has implications for market structure that would differ from what one would obtain from thinking about shipping times. A world where consumers care about purchasing from their home state could lead to a less concentrated e-retail sector with many small firms, whereas a world where consumers do not have a home-state preference, but do care about shipping times, could lead to a sector dominated by a few large firms that effectively use distributed warehouses to minimize shipping times and sales tax liabilities.

Taken together, we also see our results as suggesting that bounds on consumer rationality or consumer search and computation costs may be important. Taxes matter to our consumers, but we find a less than one-for-one with item prices. Smith Brynjolfsson (2001), in contrast, found that consumers react twice as strongly to tax differences as they do to item price differences. One source of the difference could

be that they study an environment in which taxes are explicitly presented to consumers in a list that is sorted on the basis of tax-inclusive prices.<sup>30</sup>

Technically, our analysis is standard. What could perhaps be more broadly useful is our suggestion that discrete-choice models may be usefully applied to datasets containing quantity data for one firm. Price data for all of the firms in a market are fairly easy to come by. Quantity data are much harder to obtain. There may, however, be many other situations, like ours, where quantity data could be obtained from one firm. (This could even be done in a field experiment.) Our example suggests that this may be a fruitful way to explore interfirm competition.

## REFERENCES

- Baye, Michael R., J. Rupert J. Gatti, Paul Kattuman, and John Morgan.** 2006. "Clicks, Discontinuities, and Firm Demand Online." <http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1075&context=iber/cpc>.
- Baye, Michael R., and John Morgan.** 2004. "Price Dispersion in the Lab and on the Internet: Theory and Evidence." *RAND Journal of Economics*, 35(3): 449–66.
- Baye, Michael R., John Morgan, and Patrick Scholten.** 2006. "Information, Search, and Price Dispersion." In *Economics and Information Systems*, Handbooks in Information Systems, Vol. 1, ed. Terrence Hendershott, 323–76. Amsterdam: Elsevier Science.
- Chevalier, Judith, and Austan Goolsbee.** 2003. "Measuring Prices and Price Competition Online: Amazon.com and Barnesandnoble.com." *Quantitative Marketing and Economics*, 1(2): 203–22.
- Chiou, Lesley.** 2009. "Empirical Analysis of Competition between Wal-Mart and Other Retail Channels." *Journal of Economics & Management Strategy*, 18(2): 285–322.
- Davis, Peter.** 2006. "Spatial Competition in Retail Markets: Movie Theaters." *RAND Journal of Economics*, 37(4): 964–82.
- Ellison, Glenn.** 2005. "A Model of Add-on Pricing." *Quarterly Journal of Economics*, 120(2): 585–637.
- Ellison, Glenn, and Sara Fisher Ellison.** 2005. "Lessons About Markets from the Internet." *Journal of Economic Perspectives*, 19(2): 139–58.
- Ellison, Glenn, and Sara Fisher Ellison.** 2009. "Search, Obfuscation, and Price Elasticities on the Internet." *Econometrica*, 77(2): 427–52.
- Forman, Chris, Anindya Ghose, and Avi Goldfarb.** 2009. "Competition between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where you Live." *Management Science*, 55(1): 47–57.
- Fox, William F.** 1986. "Tax Structure and the Location of Economic Activity Along State Borders." *National Tax Journal*, 39(4): 387–401.
- Goolsbee, Austan.** 2000. "In a World without Borders: The Impact of Taxes on Internet Commerce." *Quarterly Journal of Economics*, 115(2): 561–76.
- Goolsbee, Austan.** 2001. "Competition in the Computer Industry: Online versus Retail." *Journal of Industrial Economics*, 49(4): 487–99.
- Greene, William H.** 1997. *Econometric Analysis*. 3rd ed. Upper Saddle Ridge, NJ: Prentice-Hall.
- Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches.** 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica*, 52(4): 909–38.
- Hossain, Tanjim, and John Morgan.** 2006. "...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay." *B. E. Journal of Economic Analysis and Policy: Advances in Economic Analysis and Policy*, 6(2): Article 3.
- Smith, Michael D., and Erik Brynjolfsson.** 2001. "Consumer Decision-Making at an Internet Shopbot: Brand Still Matters." *Journal of Industrial Economics*, 49(4): 541–58.
- Walsh, Michael J., and Jonathan D. Jones.** 1988. "More Evidence on the 'Border Tax' Effect: The Case of West Virginia, 1979–84." *National Tax Journal*, 41(2): 261–65.
- Weisbrod, Glen E., Robert J. Parcels, and Clifford Kern.** 1984. "A Disaggregate Model for Predicting Shopping Area Market Attraction." *Journal of Retailing*, 60(1): 65–83.

<sup>30</sup> Tanjim Hossain and Morgan (2006) find that consumers do not fully take shipping costs into account in a neatly-designed field experiment involving selling items on eBay. A commonality between shipping costs in their experiment and tax differences on Pricewatch is that the shipping cost differences were easily available in the item descriptions, but some effort would have been required to learn the differences.



**This article has been cited by:**

1. Benjamin Edelman. 2012. Using Internet Data for Economic Research. *Journal of Economic Perspectives* **26**:2, 189-206. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
2. Avi Goldfarb, Catherine Tucker. 2011. Advertising Bans and the Substitutability of Online and Offline Advertising. *Journal of Marketing Research* **48**:2, 207-227. [[CrossRef](#)]
3. Eric T Anderson, Nathan M Fong, Duncan I Simester, Catherine E Tucker. 2010. How Sales Taxes Affect Customer and Firm Behavior: The Role of Search on the Internet. *Journal of Marketing Research* **47**:2, 229-239. [[CrossRef](#)]