Internet Purchases, Cross-Border Shopping, and Sales Taxes

Charles L. Ballard*

Jaimin Lee*

March 25, 2004

JEL Classifications: C21, D12, H26, H71, H73, R12.

*Charles L. Ballard is Professor of Economics, Michigan State University, East Lansing, Michigan 48824-1038 USA; ballard@msu.edu. Jaimin Lee is a Research Fellow at the Korea Transport Institute, Seoul, Korea; jm064@hanmail.net. We are grateful to Yoon-Sang Kim, Larry Martin, Paul Menchik, Ed Outslay, Leslie Papke, and Jeff Wooldridge for helpful comments and suggestions. Any errors are our responsibility.

Internet Purchases, Cross-Border Shopping, and Sales Taxes

Abstract

We investigate the relationship between retail sales taxes and Internet purchases, in a model that also allows for cross-border shopping. Using data from the Current Population Survey for 1997 and 2001, we estimate the probability that a consumer engages in Internet shopping, controlling for county fixed effects and a variety of demographic variables. We use variation in sales-tax rates by county to identify the effect of the sales-tax rate in the home county, as well as the effect of differences in sales-tax rates between adjacent counties. The estimates support the hypothesis that consumers in counties with higher sales-tax rates are more likely to shop on the Internet, all else equal. This evidence is consistent with the interpretation that consumers use Internet shopping as a means of evading sales taxes. In addition, consumers whose home county is adjacent to a county with a lower sales-tax rate are less likely to use the Internet for shopping, all else equal. We interpret this as reflecting the effect of cross-border shopping.

Internet Purchases, Cross-Border Shopping, and Sales Taxes

1. Introduction

Consumers can purchase private goods in any of several ways. First, they can buy goods in their own jurisdiction, paying sales taxes if the jurisdiction levies a sales tax. A second possibility is that consumers can travel across a border and purchase products in a different jurisdiction, paying the sales taxes that apply in that jurisdiction. A third possibility is that they can buy over the Internet.[1] Currently, state and local governments cannot require online firms to collect sales or use taxes unless the firms have a physical presence (*nexus*) in the taxing jurisdiction.[2] Since state and local governments cannot reliably enforce sales and use taxes on Internet purchases, evasion of these taxes is widespread.[3]

---

[1] It is also possible for consumers to engage in smuggling, in order to evade sales taes. However, the data analyzed in this paper do not shed any light on smuggling.

[2] Supreme Court rulings in 1967 and 1992 held that states do not have the power to collect taxes on sales by out-of-state mail-order companies. (See Fox and Murray, 1997.)

[3] In 2003, several large retailers announced that they would begin to collect state sales tax on their online sales. See Wedeland (2003). Also, several states are attempting to negotiate an agreement for mutual enforcement of these taxes. Nevertheless, the data analyzed in this

The empirical literature includes analyses of the effects of taxation on cross-border shopping. There is also a smaller literature on the effects of taxation on Internet sales. The contribution of our paper is to combine these two strands of literature, by analyzing Internet sales with a data set that includes the general retail sales-tax rates in the consumer's home county and in adjacent counties.[4]

Goolsbee (2000) performs an empirical test of the relation between sales-tax rates and Internet purchases, using data for 1997 from Forrester Research (a market research company). He finds that Internet purchases are highly sensitive to local taxation.[5]

Goolsbee's paper is a pioneering work in the economic analysis of electronic commerce. However, Goolsbee has only limited information regarding the residence of consumers. He knows the metropolitan area in which a consumer lives, but he does not have information about the city or county of residence. Thus, his tax-rate variable is subject to

---

paper are for 1997 and 2001, which are years for which it is believed that evasion of sales taxes on Internet purchases was very substantial.

[4] Throughout this paper, we use "county" to refer to counties as well as county equivalents, such as the parishes of Louisiana.

[5] Note, however, that Goolsbee only considers the tax-evasion aspect of Internet shopping. His estimates do not deal directly with cross-border shopping.

measurement error, because he has to assume that the tax rate is uniform throughout each metropolitan area.[6] We use a data set in which information on the county of residence is available, and we define the sales-tax rate at the county level.

The literature contains a number of empirical studies of cross-border shopping. Gordon and Nielsen (1997) estimate that 3.9 billion Danish Kroner of value added escaped taxation by Denmark in 1992 because of cross-border shopping. FitzGerald (1992) finds evidence of tax-induced cross-border shopping between the Republic of Ireland and the United Kingdom. Ferris (2000) suggests that the dramatic changes in cross-border shopping in Canada between 1989 and 1994 were caused by the substitution of the Goods and Services Tax for the manufacturer's sales tax.

In this paper, we test empirically the effect on Internet purchases of general retail sales-tax rates in the local county and neighboring counties. Our results generally indicate that Internet purchases are more likely for consumers in counties with higher tax rates, all else equal. This evidence is consistent with Goolsbee's finding that consumers use Internet shopping as a means of evading sales taxes. In addition, consumers whose home county is adjacent to a county with a lower sales-tax rate are *less* likely to use the Internet for shopping, all else equal. We interpret this as reflecting the effect of cross-border shopping.

---

[6] Goolsbee deals with this problem in a variety of ways. For example, in one specification, he restricts his sample to observations from states with uniform sales-tax rates within the state.

2. Empirical Framework

We analyze the data by county. The analysis could be carried on at the state level, rather than at the county level, but this would make it impossible to study cross-border shopping in the desired way. Virtually all residents of the United States are fairly close to a county boundary, but many are far from the nearest state border.

Counties may differ from each other in a variety of ways, even when they are in the same state, and even when they have the same sales-tax rate. For example, the counties in large metropolitan areas are likely to have better Internet infrastructure and better transportation networks. To capture the effects of these characteristics, our estimates control for county fixed effects.

2.1. Data

We use data from the Current Population Survey (CPS) for 1997 and 2001.[7] We restrict the sample to those individuals who indicate that they have access to the Internet. The CPS reports the education level and marital status for those who are 15 years old or older. Therefore, we also exclude those who are younger than 15. We also exclude consumers for whom county of residence is not given, and those for whom income information is missing.

---

[7] The CPS does not contain information on Internet use in every year. However, such information was collected in special surveys in 1997 and in 2001.

We use a binary dependent variable, which measures whether the individual engages in online purchases.[8] The explanatory variables include household income, gender, race, marital status, education level, age, and the number of computers in the household.

One key explanatory variable is *TAXRATE*, the sales-tax rate (measured in percentage points) of the consumer's home county. The states fall into four categories, with respect to the way in which they impose general retail sales taxes. First, there are no general retail sales taxes in Delaware, Montana, New Hampshire, and Oregon, either at the state level or the county level.[9] Second, 13 states impose a uniform sales-tax rate in all counties, and the District of Columbia has a uniform sales-tax rate within its borders.[10] Third, eight states have

---

[8] It would be better if we knew not just *whether* consumers use the Internet for purchases, but *how much* they spend on their online purchases. Unfortunately, in this data set, a person who spends $10 per year via the Internet is coded the same as one who spends $1000 per year.

[9] The state government in Alaska does not have a general sales tax, but its municipalities and counties impose their own sales taxes.

[10] These are Connecticut (with a uniform sales-tax rate of 6%), the District of Columbia (5.75%), Hawaii (4%), Indiana (5%), Kentucky (6%), Maine (5%), Maryland (5%), Massachusetts (5%), Michigan (6%), Mississippi (7%), Rhode Island (7%), Vermont (5%), Virginia (4.5%), and West Virginia (6%). (Mississippi has a uniform sales-tax rate of 7%,

variation in sales-tax rates across counties, but no variation within counties.[11] Fourth, in 25 states, there are variations in sales-tax rates, both across counties and within counties (*i.e.*, cities within the same county can have different sales-tax rates).[12] For these 25 states, we calculate a weighted average sales-tax rate for each county, using the populations of cities within the county as weights. In other words, the weighted average sales-tax rate for county *i* is defined as

$$TAXRATE_i = \sum_k t_{ki} \frac{pop_{ki}}{pop_i} \ , \qquad (1)$$

where $TAXRATE_i$ = weighted average sales-tax rate for county i, t_{ki} = sales-tax rate for city k in county i, pop_{ki} = population of city k in county i, and pop_i = population of county i.

We also include variables that measure the relationship between the sales-tax rates in

---

with only one exception. The city of Tupelo has its own retail sales tax of 0.25%, in addition to the state sales-tax rate.)

[11] These are Florida, Georgia, Nevada, North Carolina, Pennsylvania, South Carolina, Wisconsin, and Wyoming.

[12] These are Alabama, Alaska, Arizona, Arkansas, California, Colorado, Idaho, Illinois, Iowa, Kansas, Louisiana, Minnesota, Missouri, Nebraska, New Jersey, New Mexico, New York, North Dakota, Ohio, Oklahoma, South Dakota, Tennessee, Texas, Utah, and Washington.

the consumer's home county and adjoining counties. We do not have a strong prior belief about the best way to measure this relationship, so we use two specifications. One involves absolute differences between the sales-tax rates of adjoining counties, while the other specification involves the ratios of these sales-tax rates.

The absolute sales-tax difference, *TAXDIFF*, is equal to the sales-tax rate of the home county, minus the minimum of the sales-tax rates of the adjoining counties. This variable is measured in percentage points, and it can be positive, zero, or negative if the sales-tax rate of a consumer's own county is greater than, equal to, or less than the lowest of the sales-tax rates of neighboring counties.

The sales-tax ratio, *TAXRATIO*, is $(1+t_h)/(1+t_f)$, where t_h is the sales-tax rate of the home county and t_f is the minimum sales-tax rate among all neighboring counties.[13] For example, the residents of Washington, D.C., could buy from retail stores in the adjoining counties of Maryland and Virginia, as well as in the District of Columbia. The sales-tax rates in D.C., Maryland, and Virginia are 5.75%, 5%, and 4.5%. Thus, *TAXRATIO* for a D.C. resident is 1.0575/1.045, or about 1.012.

In any given regression, we include either *TAXDIFF* or *TAXRATIO*, but not both, because they are highly correlated. (The correlation coefficient is 0.9985.) If we include

[13] We use $(1+t_h)$ rather than t_h, and $(1+t_f)$ rather than t_f, to avoid the possibility of division by zero.

both variables in the same regression, the resulting coefficients are estimated very imprecisely, due to multicollinearity.

As mentioned above, *TAXDIFF* and *TAXRATIO* are constructed by comparing the tax rate in the consumer's home county with the lowest tax rate in any of the adjoining counties. This specification may miss certain aspects of cross-border shopping. For instance, it is possible that a consumer may engage in cross-border shopping by going two or more counties away, rather than to an adjoining county. Also, we treat everyone in a given county the same, regardless of whether they live just next to the county line, or miles away. Therefore, it should be understood that the geographic variables are not perfect. Nevertheless, we argue that these variables provide valuable information regarding cross-border shopping.

Table 1 presents summary statistics for all variables at the county level. For the sample as a whole, the average sales-tax rate at the county level is about 6.28 percent.

<Table 1 about here>

2.2. Econometric Specification

We use fixed-effects estimation to capture the unobserved group-specific effects of counties. The first model to be estimated is a linear probability model:

$$y_{ij}=\beta_0 + \beta_1 TAXRATE_{ij} + \beta_2 GEO_{ij} + \beta_3 LINCOME_{ij} + \beta_4 FEMALE_{ij} + \beta_5 WHITE_{ij} \quad (2)$$
$$+ \beta_6 MARRIED_{ij} + \beta_7 (FEMALE*MARRIED)_{ij} + \beta_8 (WHITE*MARRIED)_{ij}$$
$$+ \beta_9 (FEMALE*WHITE)_{ij} + \beta_{10} (FEMALE*WHITE*MARRIED)_{ij} + \beta_{11} HIGHGRAD_{ij}$$
$$+ \beta_{12} COLLGRAD_{ij} + \beta_{13} PROGRAD_{ij} + \beta_{14} NCH_{ij} + \beta_{15} AGE15_{ij} + \beta_{16} AGE20_{ij}$$
$$+ \beta_{17} AGE30_{ij} + \beta_{18} AGE50_{ij} + \beta_{19} AGE60_{ij} + \beta_{20} D01 + \beta_{21} PCCPI_i + a_i + u_{ij}$$

The i subscripts refer to the local county. The j subscripts refer to the individual consumer. The dependent variable, y, is equal to one when a consumer reports purchasing goods over the Internet, and zero otherwise. *TAXRATE* is the weighted average sales-tax rate in the consumer's county of residence. *GEO* is either *TAXDIFF* and *TAXRATIO*.

LINCOME is the log of the resident's household income. We expect a positive coefficient, since use of the Internet for purchases is almost certain to be a normal activity. Note that *LINCOME* may also act as a proxy for the opportunity cost of time. It would be valuable to have data on wage rates, in addition to the income data. Unfortunately, wage-rate data are not available for a substantial portion of the sample.

FEMALE, *WHITE*, and *MARRIED* are all binary variables for the individual's demographic characteristics. *HIGHGRAD*, *COLLGRAD*, and *PROGRAD* are dummy variables for the education level, indicating whether an individual has graduated from high school (*HIGHGRAD*), or from college (*COLLGRAD*), or has an M.A., Ph.D., or professional degree (*PROGRAD*). (The omitted category is the group of consumers who have less than a high-school education.) *NCH* is the number of computers in the household. *AGE15, AGE20,* *AGE30, AGE50,* and *AGE60* are age-group dummy variables. For example, *AGE15* is equal to one if a consumer's age is from 15 to 19, *AGE20* equals one for consumers aged from 20 to 29, and *AGE60* equals one for those whose age is greater than or equal to 60. (The omitted category for the dummy variables for age is the group of consumers aged 40-49.) *D01* is a yearly dummy variable, equal to one when an observation is taken in 2001, and zero when it is taken in 1997. The group-specific effect of a county is a_i, and u_{ij} is an error term.

The variable *PCCPI* is intended to control for differences in consumer prices in different parts of the country. This variable requires some explanation. Price-level data are available for certain metropolitan areas from ACCRA.[14] However, these data are normalized within each year. Therefore, they do not allow a meaningful comparison across years, either for different metropolitan areas or for the same area. Consequently, since our data come from two different years, it has not been possible to use data for the price *level*. As an alternative, we include a measure of the *rate of change* of the price level. Specifically, we include the percentage change in the Consumer Price Index (*PCCPI*) as an explanatory variable. Although we would have preferred to use a price-level variable instead of an inflation variable, the inflation rate in the local area may play a role in driving consumer awareness of the desirability of out-of-area shopping methods, such as mail-order shopping and Internet shopping. If we adopt a model with habit formation, it is possible that an increase in prices

[14] http://www.accra.org

could cause consumers to become more sensitive to economic variables than they had been previously. Therefore, we expect that inflation rates would have a positive effect on Internet shopping.

The Bureau of Labor Statistics reports the Consumer Price Index (CPI) for 26 metropolitan areas annually.[15] There are mostly for large metropolitan areas, such as Chicago, Detroit, Los Angeles, New York, Philadelphia, and San Francisco. Thus, if we include the percentage change in the Consumer Price Index (*PCCPI*) as an explanatory variable, we automatically restrict the sample to observations that are mainly from relatively large metropolitan areas. After we restrict the sample in this way, and in the other ways mentioned above, we have a sample with 16,188 observations.[16]

Even after holding constant the variables discussed above, there still may be considerable heterogeneity across different regions, in terms of effective access to the Internet, or attitudes toward it. Therefore, we also estimate the model separately for metropolitan areas of different sizes. We begin by reporting results for the entire sample for

---

[15] CPIs for metropolitan areas are available from http://www.bls.gov/cpi/home.htm.

[16] The summary statistics in Table 1 were based on a sample that was not restricted to the metropolitan areas for which *PCCPI* is available. Below, we will report estimates from a specification in which *PCCPI* is dropped from the regressions, in which case the sample size increases from 16,188 to 24,804.

which we have information on county of residence, percentage change in the Consumer Price Index, and income. After that, we restrict the sample to those who live in metropolitan areas with populations above one million, two million, and three million, and estimate the model separately for each of these subsamples.

3. Results

The first column of Table 2 shows the effects of the explanatory variables on online purchases, when we use the entire sample for which the percentage change in the CPI is available.[17] We report the robust standard errors adjusted for cluster sampling, allowing for the observations to be independent across counties, but not necessarily independent within counties.

<Table 2 about here>

[17] The results reported in these tables are for regressions in which the geographic variable is *TAXDIFF* (the absolute difference between the sales-tax rate in the local county and the minimum tax rate of the adjoining counties). When we use *TAXRATIO* instead of *TAXDIFF*, the signs, magnitudes, and significance levels of all of the coefficients are quite similar. Results for the regressions using *TAXRATIO* are available on request.

Below, we will consider the coefficients for the tax variables, *TAXRATE* and *TAXDIFF*. First, however, let us consider the results for the non-tax variables, in the left-hand column of Table 2. (i) The coefficient estimate of *LINCOME* is positive and statistically significant. The point estimate indicates that a doubling of income is associated with an increase of 0.0355 in the probability that the consumer would make online purchases. (ii) The coefficients on the education dummy variables suggest that more-educated consumers are substantially more likely to make online purchases, all else equal. For example, the coefficient on *COLLGRAD* indicates that the probability of Internet purchases is nearly 0.1 higher for a college graduate than for someone without a high-school diploma, all else equal. The coefficients for *HIGHGRAD* and *COLLGRAD* are significant at the one-percent level, and the coefficient for *PROGRAD* is significant at the ten-percent level. (iii) The coefficient estimate of *NCH* (number of computers in the household) is positive and statistically significant. When the number of computers in a consumer's household increases by one, the probability of online purchases is found to increase by about 0.08, all else equal. (iv) The coefficients for the *AGE* dummy variables indicate that the probability of Internet purchases has an inverse-U-shaped pattern by age. The estimates indicate that the probability of Internet purchases rises until consumers are in their thirties, and then declines. All else equal, a consumer in her thirties has a probability of Internet purchases that is about 0.15 greater than that for a teenager, and about 0.16 greater than that for a consumer in her sixties.

Our conjecture is that teenagers are less likely to engage in online shopping because they are less likely to have a credit card, and that the elderly may be less familiar and less comfortable with online shopping, even after controlling for other variables. (v) The coefficient on the dummy variable for the year 2001, *D01*, is large, positive, and highly significant. This is not surprising, since Internet usage increased very substantially between 1997 and 2001. (vi) The coefficient estimate of the percentage change in the Consumer Price Index (*PCCPI*) is positive and significant. This implies that, all else equal, consumers who live in metropolitan areas with higher inflation rates would be more likely to make online purchases.

Our assessment of these results is that most of the coefficients for non-tax variables in the first column of Table 2 can be given a reasonable interpretation. Many of the coefficients are highly significant, and the magnitudes are economically meaningful. For example, they suggest that, all else equal, the probability of Internet purchases by a single white male in his thirties with a Bachelor's degree is more than 0.37 greater than the probability for a single black female in her sixties without a high-school diploma.

The results in the first column of Table 2 also lend support to our hypotheses regarding the role of sales taxes in influencing Internet sales. The coefficient for the sales-tax rate in the local county (*TAXRATE*) has the expected positive sign, and it is statistically significant at the five-percent level. These results indicate that, all else equal, a resident of a county with a higher sales-tax rate is substantially more likely to use the Internet for shopping than a resident of a county with a lower tax rate. The estimated magnitude of the effect is very large. The coefficient estimate suggests that, all else equal, the probability of engaging in Internet shopping will increase by 0.208, when the sales-tax rate of the home county increases by one percentage point. These results are consistent with the notion that Internet shopping is used, in part, as a mechanism for tax evasion.

TAXDIFF has the expected negative sign, and is significant at the one-percent level. The coefficient estimate is –0.0756. This indicates that, all else equal, the probability of Internet shopping by a resident of a county with a sales-tax rate of 6% will be reduced by about 0.076 if the lowest sales-tax rate among the adjacent counties is 5%.

The negative coefficient on *TAXDIFF* indicates that a consumer who lives in a county adjacent to a low-tax county is less likely to use the Internet for shopping than he or she would be if the adjacent county had a higher tax rate, all else equal. This result does not provide any *direct* evidence regarding cross-border shopping. However, the result is certainly consistent with an interpretation that involves cross-border shopping. This is our preferred interpretation: From the consumer's perspective, shopping in the home county, shopping in an adjacent county, and shopping on the Internet are all substitutes. If a low-tax county is nearby, the tax benefits of Internet shopping are reduced for a resident of a high-tax county. He or she can reduce the sales-tax burden, simply by driving across the county line to shop.

Thus, the coefficients for *TAXRATE* and *TAXDIFF* indicate that, all else equal, the probability of using the Internet for shopping is reduced by the presence of low tax rates in *either* the home county or an adjacent county.

Next, we report the results of fixed-effects regressions in which the sample is restricted to those in metropolitan areas with populations of at least one million, two million, and three million. The results are shown in the second, third, and fourth columns of Table 2. By scanning across the rows of Table 2, it can be seen that the estimates for both the tax and non-tax variables are usually fairly consistent across the various subsamples.

Goolsbee (2000) reports elasticities of Internet shopping with respect to the tax price. He finds large elasticities, ranging from 2.3 to 4.3, depending on the specification. Based on the estimates shown in Table 2, we find tax elasticities in a similar range. In fact, the higher estimated coefficients in the third and fourth columns of Table 2 are consistent with tax elasticities that are even higher than those calculated by Goolsbee. Thus, our results are consistent with Goolsbee's results, in that they suggest that the choice of whether to shop on the Internet may be quite sensitive to taxation.

Unfortunately, however, our data only tell us about *whether* consumers use the Internet for shopping. The data do not include information on the *dollar values* of the purchases. Consequently, any attempt to estimate the effects of the Internet on sales-tax revenue would have to be fairly speculative.

In Table 3, we compare the probabilities of Internet purchases for demographic categories that differ by gender, marital status, and race. For the purpose of these comparisons, we use consumers who are unmarried, nonwhite, and male as the base group. For example, consumers who are unmarried, nonwhite, and female have a probability of Internet purchases that is lower than the base group by 0.046, all else equal, while those who are unmarried, white, and male have a probability that is higher than the base group by 0.0713.

<Table 3 about here>

Since we have a binary dependent variable, the model is amenable to Probit estimation. The Probit estimation results are fairly similar to those in Table 2, regardless of whether we use *TAXDIFF* (the absolute difference between the tax rate in the local county and the lowest rate in the adjacent counties) or *TAXRATIO* (the ratio of the tax rate in the local county to the lowest tax rate in the adjacent counties). We also use Logit estimation. The Logit results are similar to the results in the Probit case. The results of the Probit and Logit estimation are not reported here, but they are available upon request.

Earlier, we expressed some reservations about the variable *PCCPI*, which measures the percentage changes in the Consumer Price Index. Next, we consider some regressions from which *PCCPI* has been excluded. The results are shown in Table 4.

<Table 4 about here>

For many of the non-tax variables, the coefficients in Table 4 are quite similar to those in Table 2. These include the variables for income, marital status, race, gender, education, and age. In contrast to the results for the non-tax variables, however, the results for the tax variables in the first column of Table 4 are not strong. The coefficient on the sales-tax rate (*TAXRATE*) has the expected positive sign. (In other words, all else equal, a consumer who resides in a county with a high sales-tax rate is more likely to use the Internet for shopping.) However, the coefficient falls far short of statistical significance. The coefficient on *TAXDIFF* is also insignificant, and it does not even have the predicted sign. If we compare the results from the first column of Table 4 with the results from the first column of Table 2, we find that the results for the tax variables are fairly sensitive to the inclusion of the price-index variable. This suggests that, even if *PCCPI* is an imperfect measure, its absence may generate omitted-variables bias.

In Table 2, we reported estimates for the entire sample for which *PCCPI* is available, as well as for samples drawn from metropolitan areas with populations exceeding various limits. We follow a similar procedure here. In the second, third, and fourth columns of Table

4, we report the results of regressions in which the sample is restricted to those living in metropolitan areas with populations of at least one million, two million, and three million persons. In these cases, the estimates for the tax variables are more consistent with our hypotheses.

The coefficient estimates of *TAXRATE* have the expected positive signs, although they are only significant when the sample is limited to those residing in metropolitan areas with populations greater than three million. All coefficient estimates of *TAXDIFF* are negative, and they are statistically significant when the sample is restricted to metropolitan areas larger than one million or two million.

To summarize the results in Tables 2 and 4, all of the specifications give fairly strong results with respect to the demographic variables. Income and education always have a strong positive effect on the use of the Internet for purchases. The propensity to shop by the Internet rises with age until consumers are in their thirties, and then falls. When we include the percentage change in the Consumer Price Index in the regressions (see Table 2), the tax variables also give strong results: Internet shopping is stimulated in counties with high sales-tax rates, but the presence of an adjacent county with a low sales-tax rate leads to a reduction in Internet shopping, all else equal. When we exclude the price-index variable, the results for the non-tax variables are fairly robust, but the estimates for the sales-tax variables are somewhat fragile. When the price-index variable is excluded, the coefficients for the sales-tax variables are of the predicted signs when we restrict the sample to those living in the larger metropolitan areas.

4. Conclusion

In this paper, we analyze empirically the determinants of Internet shopping, using data from the Current Population Survey for 1997 and 2001. We only consider those consumers who had Internet access. We construct a data set that includes the tax rate in the consumer's local county, a measure of the tax rates in adjacent counties, and a wide range of demographic variables. We provide two sets of estimates. In one set, we control for the inflation rate in the local area; in the other, we do not. Within each of these sets of estimates, we provide separate estimates for subsamples in which we only include consumers who reside in metropolitan areas above certain population thresholds.

In all of our estimates, we find that the probability of Internet shopping is much higher for those with higher incomes, all else equal. The probability of Internet shopping increases rapidly with age until consumers are in their thirties, and then decreases. Those with a Bachelor's degree are more likely to use the Internet for shopping than are those with other levels of education, all else equal. And, not surprisingly, we find a sharp increase in the propensity to use the Internet for shopping between 1997 and 2001.

If we use the entire sample, without the inflation variable, the estimates for the sales-tax rates are not statistically significant. However, if we focus on the residents of larger metropolitan areas, the estimates for the sales-tax rates are of the expected signs, and they are statistically significant in some cases. The sales-tax rate in the consumer's own county has a positive effect on online purchases, all else equal. Our interpretation is that those who live in areas with high sales-tax rates are more likely to use the Internet for shopping, because sales-tax evasion on Internet sales is not difficult, and the benefit from evasion is greater when the sales-tax rate is higher. In addition, we find that Internet purchases are less likely for those who reside in a county that is adjacent to another county with a lower sales-tax rate, all else equal. We interpret this as evidence of cross-border shopping.

When we use the percentage change in the Consumer Price Index as an explanatory variable, the coefficients on the tax-rate variables are always of the predicted sign. The coefficients are often larger in magnitude, and they are more statistically significant in most cases. In addition, the coefficients for the tax-rate variables do not change as much when we change the size of the sample, as they did when the price-index variable was excluded. These results suggest that the specification in which we include the price-index variable is preferable.

In this paper, we have placed most of the emphasis on estimates from the linear probability model. We also use a Probit specification and a Logit specification. The Probit and Logit results are similar to those that emerge from the linear probability model. The Probit and Logit results are available on request.

Our estimates provide support for the idea that shopping in the home county, cross-border shopping in an adjacent county, and Internet shopping are substitutes, and that the consumer's choice among these forms of commerce is responsive to differences in sales-tax rates. Consumers are more likely to use the Internet for shopping if they reside in a county with a high sales-tax rate, especially if a lower tax rate cannot be found in an adjacent county.

The independent variable in our study indicates whether the consumer uses the Internet for shopping. In many cases, the coefficient estimates for the tax rate in the home county are quite large. The corresponding tax-price elasticities are comparable with those estimated by Goolsbee (2000); in some cases, the tax elasticities are as large as four, or even larger. Unfortunately, we do not have information on the *dollar value* of the Internet shopping that is undertaken by the consumers in our sample.

The coefficient estimates are sometimes stronger and more robust for the non-tax variables (such as income, age, and education level) than for the tax variables. Thus, taxes are definitely not the entire story. Moreover, our data are taken from a time period when the Internet was still in its infancy. As this new medium of commerce becomes more and more a regular feature of the economic landscape, we anticipate that the determinants of Internet shopping may change, possibly by a substantial amount. Thus, we do not consider our estimates to be the last word on the subject. Instead, we look forward to further research.

Table 1. Summary Statistics[a]

	All individuals	Internet Buyers	Non-buyers
No. of observations	24,804	10,500	14,304
TAXRATE	6.2815 (1.9361)	6.2237 (1.9742)	6.3239 (1.9066)
TAXDIFF	0.6958 (1.5142)	0.7255 (1.5532)	0.6741 (1.4845)
TAXRATIO	1.0066 (0.0148)	1.0069 (0.0152)	1.0064 (0.0145)
Household income	58,950.42 (22,930.92)	61,990.83 (21,916.30)	56,718.58 (23,398.66)
FEMALE	0.4973 (0.5000)	0.4986 (0.5000)	0.4964 (0.5000)
WHITE	0.8465 (0.3605)	0.8637 (0.3431)	0.8339 (0.3722)
MARRIED	0.5833 (0.4930)	0.6160 (0.4864)	0.5594 (0.4965)
Education	14.52 (2.8576)	15.03 (2.7626)	14.14 (2.8676)
Number of computers	1.4701 (0.6939)	1.5844 (0.7483)	1.3863 (0.6382)
Age	38.51 (14.48)	39.16 (13.11)	38.03 (15.40)

[a]The numbers in parentheses are standard deviations. The household income is measured in dollars. Education and age are measured in years. *TAXRATE* and *TAXDIFF* are measured in percentage points.

Table 2. Fixed-Effects Estimates of the Determinants of Internet Shopping[a]

Dependent Variable: Whether the consumer makes online purchases				
	Sample for which *PCCPI* available	Metro Population >= 1,000,000	Metro Population >= 2,000,000	Metro Population >= 3,000,000
TAXRATE	0.2080**	0.2268**	0.2904**	0.3481*
	(0.0946)	(0.1017)	(0.1165)	(0.1822)
TAXDIFF	-0.0756***	-0.0676**	-0.0544**	-0.0110
	(0.0269)	(0.0271)	(0.0268)	(0.0436)
PCCPI	0.0436**	0.0462**	0.0602**	0.0459
	(0.0214)	(0.0227)	(0.0280)	(0.0479)
LINCOME	0.0355***	0.0372***	0.0303***	0.0311***
	(0.0083)	(0.0089)	(0.0094)	(0.0119)
FEMALE	-0.0460***	-0.0391*	-0.0395*	-0.0438*
	(0.0178)	(0.0201)	(0.0204)	(0.0229)
WHITE	0.0713***	0.0648**	0.0599**	0.0292
	(0.0242)	(0.0256)	(0.0270)	(0.0276)
MARRIED	-0.0357	-0.0418	-0.0573**	-0.0738**
	(0.0256)	(0.0279)	(0.0282)	(0.0310)
*FEMALE*MARRIED*	0.0106	0.0096	0.0144	0.0325
	(0.0232)	(0.0278)	(0.0306)	(0.0383)
*WHITE*MARRIED*	0.0041	-0.0037	0.0056	0.0216
	(0.0272)	(0.0266)	(0.0265)	(0.0340)
*FEMALE*WHITE*	0.0246	0.0117	0.0160	0.0244
	(0.0225)	(0.0257)	(0.0265)	(0.0376)
*FEMALE*WHITE*MARRIED*	0.0237	0.0388	0.0356	0.0213
	(0.0308)	(0.0349)	(0.0381)	(0.0552)
HIGHGRAD	0.0749***	0.0708***	0.0716***	0.0659***
	(0.0165)	(0.0177)	(0.0159)	(0.0181)
COLLGRAD	0.0961***	0.1031***	0.1133***	0.1141***
	(0.0093)	(0.0099)	(0.0112)	(0.0138)
PROGRAD	0.0247*	0.0208	0.0307**	0.0345*
	(0.0132)	(0.0148)	(0.0154)	(0.0209)
NCH	0.0822***	0.0853***	0.0878***	0.0773***
	(0.0065)	(0.0072)	(0.0080)	(0.0101)
AGE15	-0.1065***	-0.1156***	-0.1164***	-0.1167***
	(0.0202)	(0.0222)	(0.0244)	(0.0332)
AGE20	0.0116	0.0037	0.0006	-0.0014
	(0.0127)	(0.0129)	(0.0156)	(0.019)
AGE30	0.0438***	0.0487***	0.0487***	0.0644***
	(0.0115)	(0.0127)	(0.0138)	(0.0175)
AGE50	-0.0417***	-0.0450***	-0.0410***	-0.0356**
	(0.0105)	(0.0116)	(0.0127)	(0.0167)
AGE60	-0.1147***	-0.1116***	-0.1155***	-0.1066***
	(0.0168)	(0.0176)	(0.0197)	(0.0275)
D01	0.3594***	0.3636***	0.3425***	0.3766***
	(0.0241)	(0.0263)	(0.0336)	(0.0597)
N	16,188	13,233	10,785	5,690
R^2	0.1819	0.1934	0.1901	0.2017

[a]The numbers in parentheses are robust standard errors, adjusted for cluster sampling. ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively.

Table 3. The Partial Effects of Variables for Race, Gender, and Marital Status

Base Group: Unmarried, Nonwhite, Male	
Unmarried, Nonwhite, Female	-0.0460
Unmarried, White, Male	0.0713
Married, Nonwhite, Male	-0.0357
Unmarried, White, Female	0.0499
Married, Nonwhite, Female	-0.0711
Married, White, Male	0.0397
Married, White, Female	0.0526

Table 4. Fixed-Effects Estimates of the Determinants of Internet Shopping, Excluding *PCCPI* From the Explanatory Variables[a]

Dependent Variable: Whether the consumer makes online purchases	Full Sample	Metro Population >= 1,000,000	Metro Population >= 2,000,000	Metro Population >= 3,000,000
TAXRATE	0.0041	0.0282	0.0581	0.2849**
	(0.0313)	(0.0400)	(0.0492)	(0.1187)
TAXDIFF	0.0059	-0.0705***	-0.0930***	-0.0243
	(0.0448)	(0.0231)	(0.0206)	(0.0394)
LINCOME	0.0326***	0.0364***	0.0291***	0.0293**
	(0.0064)	(0.0078)	(0.0091)	(0.0118)
FEMALE	-0.0496***	-0.0412**	-0.0396*	-0.0420*
	(0.0169)	(0.0189)	(0.0206)	(0.0225)
WHITE	0.0736***	0.0673***	0.0583**	0.0298
	(0.0209)	(0.0239)	(0.0266)	(0.0267)
MARRIED	-0.0320	-0.0455*	-0.0603**	-0.0767***
	(0.0235)	(0.0260)	(0.0275)	(0.0296)
*FEMALE*MARRIED*	0.0166	0.0100	0.0166	0.0329
	(0.0225)	(0.0263)	(0.0309)	(0.0381)
*WHITE*MARRIED*	0.0029	0.0008	0.0128	0.0283
	(0.0245)	(0.0248)	(0.0258)	(0.0325)
*FEMALE*WHITE*	0.0261	0.0165	0.0182	0.0235
	(0.0204)	(0.0234)	(0.0264)	(0.0363)
*FEMALE*WHITE* MARRIED*	0.0183	0.0357	0.0291	0.0163
	(0.0283)	(0.0322)	(0.0384)	(0.0548)
HIGHGRAD	0.0783***	0.0771***	0.0800***	0.0777***
	(0.0132)	(0.0163)	(0.0167)	(0.0199)
COLLGRAD	0.1007***	0.1043***	0.1136***	0.1171***
	(0.0078)	(0.0097)	(0.0106)	(0.0134)
PROGRAD	0.0127	0.0147	0.0325**	0.0408**
	(0.0117)	(0.0145)	(0.0150)	(0.0207)
NCH	0.0837***	0.0855***	0.0888***	0.0825***
	(0.0053)	(0.0065)	(0.0076)	(0.0100)
AGE15	-0.1161***	-0.1276***	-0.1099***	-0.1098***
	(0.0160)	(0.0200)	(0.0245)	(0.0331)
AGE20	0.0100	-0.0005	0.0023	0.0021
	(0.0105)	(0.0122)	(0.0149)	(0.0182)
AGE30	0.0453***	0.0483***	0.0504***	0.0638***
	(0.0089)	(0.0111)	(0.0130)	(0.0165)
AGE50	-0.0251***	-0.0407***	-0.0439***	-0.0402**
	(0.0093)	(0.0106)	(0.0125)	(0.0159)
AGE60	-0.1196***	-0.1155***	-0.1212***	-0.1196***
	(0.0137)	(0.0174)	(0.0195)	(0.0285)
D01	0.3760***	0.3918***	0.3947***	0.4268***
	(0.0119)	(0.0156)	(0.0207)	(0.0352)
N	24,804	16,014	11,341	6,087
R^2	0.1740	0.1887	0.1898	0.1973

[a]The numbers in parentheses are robust standard errors, adjusted for cluster sampling. ***, **, and * denote significance at the 1%, 5%, and 10% levels respectively.

References

Ferris, J. Stephen, "The Determinants of Cross Border Shopping: Implications for Tax Revenues and Institutional Change." *National Tax Journal* 53 No. 4 (December, 2000): 801-824.

FitzGerald, John D., "The Distortionary Effects of Taxes on Trade in Border Areas: the Republic of Ireland-UK Border." In Georg Winckler (ed.), *Tax Harmonization and Financial Liberalization in Europe*. New York: St. Martin's Press (1992).

Fox, William F., and Matthew N. Murray, "The Sales Tax and Electronic Commerce: So What's New?" *National Tax Journal* 50 No. 3 (September, 1997): 573-592.

Goolsbee, Austan, "In a World Without Borders: The Impact of Taxes on Internet Commerce." *Quarterly Journal of Economics* 115 No. 2 (May, 2000): 561-576.

Gordon, Roger H., and Soren Bo Nielsen, "Tax Evasion in an Open Economy: Value-Added vs. Income Taxation." *Journal of Public Economics* 66 No. 2 (November, 1997): 173-197.

U.S. Bureau of the Census and U.S. Bureau of Labor Statistics, "Current Population Survey". http://www.bls.census.gov/cps/cpsmain.htm.

U.S. Department of Labor, Bureau of Labor Statistics, "Consumer Price Indexes". http://www.bls.gov/cpi/.

Wedeland, Mike, “Net Shoppers’ Tax-Free Days Nearly Gone.” Detroit *News and Free Press* (February 8, 2003).