Relative tax rates, proximity and cigarette tax noncompliance:

Evidence from a national sample of littered cigarette packs

Abstract:

Relative tax rates, proximity and cigarette tax noncompliance:
Evidence from a national sample of littered cigarette packs.

We analyze data about cigarette tax compliance from the first US-based national scale littered cigarette packs collection. We code each pack based on whether an appropriate tax had been paid at the location where it was found. Noncompliance across our 132 sample communities ranges from zero to one hundred percent with an appropriately weighted mean of 21 percent. We provide evidence that noncompliance is due to both cross-border shopping and cigarette trafficking. OLS and binomial logit regressions demonstrate that the financial incentive for noncompliance is the most important explanatory variable and has a statistically and quantitatively significant impact on noncompliance. We find mixed evidence about the extent to which tax avoidance varies with distance to lower tax borders. Our simulations show that, even after accounting for increased noncompliance, virtually all areas in our study would experience increases in tax revenue if they increased cigarette tax rates.

Key words: Tax avoidance; tax evasion; cigarette tax
JEL classification numbers: H26

Introduction

Since the landmark 1964 US Surgeon General's report on smoking and health, tobacco control and specifically economic evidence about smoking has become a major element of public health policy[1]. Chaloupka and Warner (2000) provide a comprehensive review of much of the older literature about the economics of tobacco control. More recent evidence is discussed in the Economics of Tobacco and Tobacco Control (Tobacco Control Monograph Series by National Cancer Institute, 2016)[2] and suggests that significant tobacco tax increases are the single most effective policy to discourage smoking.

Tax noncompliance – either tax avoidance – i.e. legal behavioral changes designed to minimize tax payments, or tax evasion – i.e. illegal actions designed to minimize tax payments – may undermine effective cigarette tax policy. Tax noncompliance reduces government revenue and may allow some users, including youth, to obtain lower price cigarettes encouraging them to start smoking or refrain from quitting (see Institute of Medicine 2015). A long stream of economics literature has attempted to directly measure or proxy for tax avoidance and evasion (see Baltagi and Levin 1986; Becker, Grossman, Murphy 1991; Saba et. al. 1995; Gailbraith and Kaiserman 1997; Thursby and Thursby 2000; Stehr 2005; Lovenheim 2008.)

A fundamental challenge in this literature is obtaining reliable measures of tobacco consumption and even more importantly, cross-sectional and over-time variation in tobacco consumption resulting from variations in public policy, particularly variation in tobacco taxes and prices. Administrative data on tax paid cigarette sales (see Orzechowski and Walker 2008) provides evidence about the place of purchase but not necessarily the place of consumption. Data from surveys of smokers may show the place of consumption but not necessarily the place of purchase. Cigarette taxes, and hence prices, vary geographically, sometimes by a large amount in a small distance (Merriman 2010, Chernick and Merriman 2013, Hanson and Sullivan 2009). Because of this, it can be difficult to link the relevant tobacco taxes to a particular consumption location. Cigarette purchasers may buy their cigarettes in a variety of places and may pay a variety of prices. Without knowing the location of purchase, the tax that is paid often cannot be inferred. The best solution available sometimes has been, as in Lovenheim (2008), to theorize, based on relatively abstract models about human behavior and to use econometric evidence to simultaneously infer both the propensity for tax noncompliance and the price elasticity of demand for cigarettes (see also Nicholson, Turner and Alvarado 2014). Without empirical grounding for tax noncompliance estimates however, estimates of the price elasticity of demand for cigarettes are likely to be biased since econometricians cannot be certain of the saliency of home and cross-border prices (Pesko et al 2016).

In recent years there have been increasing efforts to develop data sets that allow for more direct comparison between the location of cigarette purchases and the location of consumption. Evidence about the location of purchase can often be gleaned directly from the cigarette pack because nearly all US states and many local governments require a tax stamp displayed on the cigarette pack. Several recent studies working in the unobtrusive measures tradition of Webb, Campbell, Schwartz, and Sechrest (1966) and Fullerton and Kinnaman (1996) (Chernick and Merriman 2013, Davis et. al. 2013, Fix et. al. 2014, Kurti, von Lampe and Thompkins 2013, Lakhdar 2008, and Merriman 2010) have exploited this information to determine the tax jurisdiction in which the pack was purchased. We extend this methodology by collecting the first US-based national sample of littered cigarette packs[3]. We use this sample to provide national estimates of tax noncompliance and to better understand its causes and implications. This paper makes several significant contributions to the literature. First, we demonstrate the feasibility of conducting a national scale litter data collection to measure tobacco tax avoidance based on small samples from a large variety of areas. Secondly, we empirically demonstrate that averaged across a large geographical area financial incentives are strongly correlated with tobacco tax noncompliance. But we show that the vast majority of the population faces very small incentives for noncompliance and, as a result, tobacco tax increases would increase tobacco tax revenue in almost all US jurisdictions even taking into account increases in tobacco tax avoidance and evasion.

Literature Review and Assessment of Different Data Collection Methodologies

There are three main sources of data that measure both the location of cigarette consumption and the location of cigarette purchase: (1) survey data asking smokers about their place of purchase; (2) Nielsen Homescan data that asks smokers to record both the uniform product code (UPC) on the physical cigarette pack and the location where the purchase was made and (3) examination of physical cigarette packs including both those received when smokers were invited to mail in their packs and discarded packs examined by researchers. Each source of data has inherent strengths and limitations as we discuss below.

A main source of data using approach (1) is the large and nationally representative sample provided by the Tobacco Use Supplement of the US Current Population Survey (TUS-CPS). DeCicca, Kenkel and Liu (2013) note (p.1131) that “TUS-CPS asked smokers whether their last purchase of cigarettes was in a state other than their state of residence, or over the internet or by other means.” They use data from the 2003 and 2006-2007 TUS-CPS and find that about five percent of respondents report that their last purchase was made across a state border and less than one percent made purchases by other means. Compared to non-border crossers, those who purchased in another state, faced a higher average home state tax ($1.23 compared to $0.95) and smaller distance to a state border (52 miles compared to 120 miles).

Crosby, Merriman, Wang and Chaloupka (2014) examine TUS-CPS data from January 2011 alone, as well as in combination with the May 2010 and August 2010 surveys and find avoidance rates that ranged between 4.57 and 4.58 percent for daily smokers, and between 3.00 and 3.47 percent for smokers that only smoked some days.

The main (and perhaps only) sources of data using approach (2) is the Nielsen Consumer Panel Datasets[4]. The dataset used by Harding, Leibtag and Lovenheim (2012) contains information about the approximate location of residence and the approximate location of purchases of a nationally representative panel of consumers in 2006-07. Consistent with DeCicca, Kenkel and Liu (2013) they find that about 5.5 percent of cigarette purchases are made across state borders and virtually all of these are in lower tax states.

Fix et. al. (2014) obtain data from the third source—examining physical packs mailed in by smokers. Their data were collected as part of the 2009 and 2010 waves of the nationally representative International Tobacco Control United States (ITC US) surveys. Some of those interviewed were invited to send survey administrators (p. 2) “an unopened pack of their usual brand of cigarettes, purchased from the outlet where they normally purchase their cigarettes”. Those who agreed were mailed a postage-paid return envelope and were paid twenty-five dollars for their time and effort. In 2009, 401 eligible participants returned 318 packs for a response rate of 79 percent. In 2010, 491 eligible participants returned 366 packs. A response rate of 75 percent. In 2009 20 percent of the packs collected were not taxed by the respondents’ state of residence. In 2010 an almost identical 21 percent of the packs were not taxed in the state of residence.

Barker et. al. (2016) examine a national sample of discarded packs collected in 2012—the same data we use —and, like Fix et. al. (2014), find a noncompliance rate of 21 percent. The relatively high noncompliance rates found with physical packs is consistent with many earlier, more geographically targeted, studies of littered packs. Davis et al (2013) used littered cigarette packs collected from samples of Census tracts in Boston, New York City, Philadelphia, Providence, and Washington D.C. They found that overall nearly 59 percent of packs did not have a proper local tax stamp. Noncompliance ranged from about 27 percent in Philadelphia to 81 percent in Washington, D.C. Chernick and Merriman (2013) find evidence of noncompliance rates of more than 50 percent in NYC and Merriman (2010) puts noncompliance rates at 75 percent in the City of Chicago. Aziani et. al. (2017) obtained data provided by an industry source from twenty-three collections of empty discarded packs—106,500 observations—in ten US cities from 2010 to 2014[5]. In twenty-two of the twenty-threecollections more than 5.8 percent of the packs had not paid applicable state taxes with a median rate of avoidance of 9.9 percent. Only Minneapolis' 2012 collection yielded a rate (4.4 percent) less than 5.8 percent. These findings combined with evidence from Fix et. al. (2014), Barker et. al. (2016), and many studies of littered packs from a variety of smaller areas provide motivation for additional research.

As the above review of the literature makes clear, there are a number of potential sources of data about tobacco tax noncompliance. Since tax evasion is illegal and tax avoidance is decentralized and not tracked by administrative data, it is not surprising that these phenomena are hard to measure. Each data source has some strengths and weakness so combining insights from all of them may be beneficial[6].

TUS-CPS measures of cross-border shopping have the advantage that they are based on large, representative samples collected by government agencies with well-documented and highquality data collection procedures. These data have been collected at various points in time so that over-time changes in behavior can be tracked. However, surveys about smoking have important known weaknesses. As noted in the Institute of Medicine (2015), compared with objective measures of cigarette consumption such as aggregate tax paid sales (Institute of Medicine 2015, 80) “survey participants often underreport their use behaviors”. Self-reported cigarette consumption was only about 65 percent of aggregate state tax paid sales (Institute of Medicine 2015, 97). This suggests that survey respondents either mis-remember or purposely distort their response due to a desire to report socially accepted behaviors. We do not know whether, or to what extent, this desire might bias responses about purchases of illicit tobacco. Also, consumers may not know they are purchasing illicit tobacco, particularly if they make retail purchases (Institute of Medicine 2015).

Because Nielsen Homescan respondents can record data each time they make a purchase there may be a smaller likelihood that they mis-remember where the purchase compared to other survey respondents. However, like TUS-CPS survey respondents, Nielsen respondents may wish to under-report purchases of products that are obtained through irregular channels[7]. Furthermore, although Nielsen respondents can be made to be representative of the general population in certain ways (e.g. demographically) by deciding who is accepted into the panel, they may be unrepresentative in ways that could be correlated with tax avoidance behavior. For example, respondents could have tendencies to follow rules (such as a data reporting protocol) that are different from the typical individual in their demographic and socio-economic strata.

In this paper, we use littered pack collections from 132 sample communities to better understand the reasons for, and implications of, cross-location variation in measured noncompliance. Some of our findings can be compared with earlier literature. Harding, Leibtag, and Lovenheim(2012) find that a one-cent excise tax increase raises the probability of a cross-state purchase by nearly six percent for consumers on the border. This probability declines by more than one percent for each one percent increase in distance away from the border. Merriman (2010) finds that the probability of tax noncompliance decreased about one percent with each one mile increase in distance to the lower-tax state (i.e. Indiana) border.

Chernick and Merriman (2013) collect four waves of littered cigarette pack data and examine the change in tax noncompliance in response to a 2008 New York State cigarette tax increase (from $1.50 to $2.75). They also found that distance to low tax sources of cigarettes, including NYS border, New Jersey state border, and Poospatuck Indian Reservation influenced tax noncompliance.

Our study applies the methodology used in Merriman (2010) to a national sample and therefore is able to obtain findings that may generalize to the entire nation[8]. We acknowledge that the littered pack methodology also has significant limitations since we know little about the population of those who discarded packs and therefore cannot assure that our sample is representative of that population. In particular, smokers who litter may also be disproportionately likely to engage in tax noncompliance. Unfortunately, we cannot link demographic data about smokers to any particular littered pack. Thus, we are unable to demonstrate the representativeness of our sample compared to surveys of smokers using conventional measures (such as checking the demographic characteristics of respondents)[9]. In addition, the packs in our sample may disproportionately represent smokers who spend much time outside since these smokers are more likely to finish packs on the street. However, in many areas clean indoor air laws probably mean that the vast majority of smoking now occurs outdoors[10]. Finally, we caution that some non-compliant packs are the results of tourism or normal interjurisdictional commuting patterns (See Merriman 2010 for attempts to address some of these potential weaknesses).

Conceptual Framework, Data and Methods

Conceptual Framework

Our primary objective is to collect and analyze tangible, direct and representative estimates of cigarette tax noncompliance in a large number of areas around the US. We combine this evidence with data about hypothesized determinants of tax noncompliance to estimate parameters, including the tax noncompliance gradient (i.e. the change in tax noncompliance by distance to lower tax cigarettes) that are fundamental to this literature and relevant to the policy discussion. We begin by discussing the relationship between the phenomena that we observe (i.e. the share of littered packs collected without a tax stamp or with a stamp from a lower tax jurisdiction) and the population-wide cigarette tax noncompliance rate.

Let $S_i = 1$ if the ith cigarette pack consumed in a particular location was purchased in a lower tax jurisdiction and $S_i = 0$ otherwise. Define the tax noncompliance rate as $\overline{S} = \left(\frac{\sum_{i=1}^{N} S_i}{N} \right)$ where N is the total number of cigarette packs consumed at that location. Using our littered pack collection we observe a sample of the packs consumed and calculate $\overline{S} \equiv \left(\frac{\sum_{i=1}^{N_L} S_i}{N_L} \right)$ with $N_L < N$.

If our sampled packs are an independent and identically distributed random sample of all packs then the law of large numbers implies that $E\left(\overline{S}\right) = \overline{S}$. Of course, a non-trivial concern is the extent to which the littered packs that we recovered were, in fact, a representative sample of all packs consumed in the area. Merriman (2010) and Chernick and Merriman (2012) discuss the general issue of the representativeness of littered cigarette packs and provide evidence that they are reasonably representative. In the research design used here the places that were sampled within each community were intended to be representative of the community area. The evidence provided in earlier literature and the consistency of our results with the findings in Fix et. al. (2014) gives us confidence that our sample is not biased either for, or against, areas with particularly high levels of tax noncompliance.

Data Collection Strategy

Barker et al (2016) provides detailed documentation of the protocol that was used to select collection sites and to collect and code data. We provide a brief summary of the most important points here. Our data on littered cigarette packs were collected and analyzed as a part of a larger project funded jointly by the Robert Wood Johnson Foundation and the National Cancer Institute. The project was designed to examine how recent increases in tobacco taxes affect tobacco use and related behaviors among U.S. adults and youth[11]. Data for this project came from a large variety of primary and secondary sources. The sample communities in which we attempted to collect littered packs represented school enrollment zones (SEZs) for nationally representative samples of 8th, 10th, and 12th grade public school students in the continental U.S.[12].

As a part of this primary data collection, two-person teams of data collectors were trained to follow the data collection protocol, and traveled to 160 communities to collect littered cigarette packs between May and July 2012. The collectors looked for and collected littered packs from a representative sample of streets, business entrances, and parks, and marked the location where the pack was found.

We coded the communities in which packs were found, whether a stamp was affixed, and the taxing authority reflected on the stamp. Out of a total number of 3,867 packs collected, there were 2,116 packs with cellophane[13]. Because the tax stamps are affixed to the cellophane, only these packs with cellophane were used to analyze tax noncompliance behavior. After excluding packs without cellophane, 132 communities from thirty-eight states were represented in the analysis.

In most of our analyses, we weight each community (rather than each pack) equally. Because communities that represent a larger share of the population have a higher probability of being drawn, equal weighting of communities gives estimates of population parameters (e.g. the probability of noncompliance). See Merriman (2010) for more detail.

Construction of Dependent variable: Tax Noncompliance

Since the tax stamp found on a littered pack shows the location of its purchase, a discrepancy between the jurisdiction identified by the stamp and the jurisdiction where the pack was found suggests tax noncompliance. We coded a pack as having tax noncompliance if the pack had no tax stamp or had a stamp from a lower-rate jurisdiction. At the community level, we measured tax noncompliance by the share of packs that avoided tax.

Both cross-border shopping by consumers and organized smuggling contribute to tax noncompliance. A discrepancy between the location identified by the tax stamp and the location where the pack was found indicates noncompliance, but does not reveal how noncompliance is achieved. Inappropriate tax stamps could be found because tourists or consumers in the normal course of affairs made purchases in another location; or because consumers crossed lower tax borders in order to purchase cigarettes (cross border shopping); or because someone purchased the cigarettes in a distant location and brought them to the communities in which we collected data for resale (i.e. trafficking). Merriman (2010) uses data from Chicago to demonstrate that normal commuting or tourist behavior can plausibly explain only a small fraction of such stamps.

We use the distance between the two locations to infer the extent to which noncompliance was due to trafficking. Because it would not be in their self-interest for individual consumers to travel long distances to cross-border shop it is reasonable to assume that the longer the distance, the more likely noncompliance results from trafficking. We also expect larger tax differentials on packs with longer distances between purchase and consumption locations.

We measured the distance from the centroid of the community where a pack was found to the nearest border of the state that issued the tax stamp on the pack. We assigned a zero distance if the pack had a stamp appropriate to the location where it was found. We did not assign a distance if the pack had no stamp or an unidentifiable stamp. We also calculated the tax differential between the jurisdiction where a pack was found and the issuing jurisdiction. A positive tax differential indicates monetary savings from tax noncompliance, while a negative tax differential indicates that the consumer paid higher cigarette taxes. We did not code the latter packs as being tax non-compliant. We calculated a tax differential based on the assumption that no state or local taxes are paid on packs with no tax stamp. However, it is possible that these are packs from North Carolina, South Carolina, or North Dakota, where state cigarette taxes are levied but no tax stamps are used. We could not calculate the differential if the pack had an unidentifiable stamp.

Table 1 reports a cross-tabulation of the number of packs by distances and tax differentials. Readers are cautioned that the data in table 1 weights each pack (rather than each community) equally. This is necessitated by the fact that, within a community, non-compliant packs may have inappropriate stamps from a variety of origins. Because each pack is equally weighted the analyses will not necessarily be representative of the population. Nevertheless we believe the analyses are informative.

1,680 packs out of 2,116 had appropriate tax stamps that indicate the locations where they were found and 35 packs had tax stamps indicating tax payments greater than the location where they were found. These thirty-five packs are almost certainly “incidental” cross-border shopping in the sense that they were bought in another location with little consideration given to price differentials. If people were not paying attention to price differentials, we would expect that the number of packs with higher taxes than required were roughly the same as the number with lower taxes. Our data shows that this is not the case—packs with lower than required taxes are much more prevalent. This suggests smokers are avoiding or evading cigarette taxes.

Sixty-one packs had unidentifiable stamps. 340 packs (16.5 percent, excluding the sixty-one packs that have unidentifiable stamps) indicated tax noncompliance by having no stamp or an inappropriate and low tax stamp compared to the location where they were found.

255 packs (12.4 percent) had an inappropriate low tax stamp that had been transported from locations beyond fifty miles or had no stamp. That is, cross-border shopping seems to be an unlikely explanation for non-compliant stamps on 75 percent of the non-compliant packs (255 out of 340 packs)[14]. Seventy-two of the 153 packs that had non-compliant stamps (47 percent) were transported more than 150 miles. Packs with higher tax differentials also tend to have longer distances or no stamps, indicating more incentives for tax noncompliance.

Construction of “Incentive for tax noncompliance”

Because table 1 suggests that both cross-border shopping and cigarette trafficking may be important for noncompliance our empirical analyses examines both. We hypothesize that the most important variable determining cigarette trafficking is the total tax rate because transport costs are be relatively insignificant when large quantities of cigarettes are transported.

To control for cross border shopping we hypothesize that the incentive to avoid taxes rises with the cigarette tax difference between home and nearby jurisdictions' and falls with distance to lower tax jurisdictions. Starting with census block level locations and tax rates, we constructed an index of the "incentive for tax noncompliance" (IFNC) for each community in our sample. IFNC measures the maximum per mile reduction in the tax on a pack of cigarettes purchased in an adjacent jurisdictions. For each block (i=1…N) within each community, let $BTax_i$ = cigarette tax rate in Census block I, $Atax_j$ =cigarette tax rate in state j that is adjacent to the state that houses Census block I, d_{ij} = distance (in miles) from Census block i to the border of adjacent state j. The cigarette tax rate includes the tax levied by state, county, and city governments, if applicable. Distance is the straight line as-the-crow-flies distance from the block centroid to the state border. For each Census block, the potential reduction in cigarette taxes per mile of travel (the tax savings) is:

$$TS_{ij} = \frac{\max(Btax_i - Atax_j, 0)}{d_{ij}}, \quad (1)$$

The potential reduction in cigarette taxes per mile as the result of border crossing is $TS_i = \underset{j}{Max}\left(TS_{ij}\right)$. We use the proportion of the community population in each block, $\frac{Pop_i}{Pop_c}$, to weight TS_i when we aggregate to the community level. The implication is that a block with a lower proportion of a community's population would have a lower impact on that community's incentive for tax noncompliance. IFNC is measured for each community as:

$IFNC = \sum_{i=1}^{n}\left(\frac{Pop_i}{Pop_c}\right)TSi$, where n = the number of blocks in the community.

If a smoker resides too far from the state border traveling to another state is too costly and time-consuming and therefore infeasible. Merriman (2010) found that a one mile increase in the distance from Chicago to the Indiana border reduced the probability of noncompliance by 0.026 ($p<.01$). In other words, the probability of tax noncompliance decreased from almost one at the Chicago/Indiana border to approximately zero when the Indiana border was beyond thirty-eight miles (1/0.026). To take this behavioral response into account, some of our analysis is restricted to communities within thirty-eight miles of a lower tax border.

In the vast majority of communities the average tax savings per mile traveled is less than ten cents as shown in panels 1 and 2 of figure 1. Out of the entire sample of 132 communities 115 have an IFNC of ten cents or less. Only a few have an IFNC greater than twenty cents and only two communities (located quite close to a lower tax border) have an IFNC greater than 70 cents. If we restrict the sample to communities within 38 miles of a tax border (figure 1 panel 2) fifty-one of the sixty-sevencommunities (78 percent) have an IFNC less than ten cents.

While the absolute value of these IFNCs might seem small, it is premature to conclude that they have no effect on behavior. An IFNC of ten cents implies that someone living five miles from the border would save fifty cents per pack (or five dollars per carton) by crossing the border. It is plausible that this differential might be enough to induce behavioral changes. Furthermore, IFNC will vary within a community so that some individuals close to a tax border will have a greater than average inducement to make cross border purchases. Merriman (2010) has presented evidence that small differences in proximity to lower tax cigarettes can result in relatively large differences in noncompliance. In our sampled communities, about 60 percent of the population has an IFNC of only a few cents and about 99 percent of the population has an IFNC under eighty cents per mile[15].

These descriptive statistics make it clear that the tax policy issue most relevant to the bulk of the population is the extent to which a change in IFNC from a very low level (say under five cents per mile) to a moderate level (say ten or twenty cents per mile) will change tax noncompliance. Our empirical analyses are designed to shed light on this issue.

Control variables

In some of our specifications we included covariates to control for factors other than tax differentials and distance that may affect tax noncompliance. We include median household income (natural log transformed) of the community, because smokers' motivation to avoid taxes may vary with income. We also include a measure of population density. Chernick and Merriman (2013) hypothesize that the greater density of population in New York City (NYC) may explain why poor residents of NYC seem to evade cigarette taxes at a higher rate than poor residents of Chicago. NYC's high population density brings scale economies and lowers smugglers' cost to distribute untaxed cigarettes in poor neighborhoods.

As a measure of mobility, or the cost of traveling to avoid taxes, we include the percentage of households with cars. We also include the percentage of land used in the retail or service sector as a measure of the local availability of cigarettes. In addition, we control for the geographic region in which the community is located to account for numerous factors such as climate and ease of auto travel that may affect propensity to avoid cigarette taxes[16]. Seven binary regional variables were included following the US Bureau of Economic Analysis classification scheme. The regions are Great Lakes, Midwest, New England, Plains, Rocky Mountain, Southeast, or Southwest. The reference group is the far west region.

Results

Basic Tables

Table 2 reports summary statistics, with panel 1 and panel 2 reporting the statistics for the full sample and "near the border" subset, respectively. The highest IFNC occurs in a large city with a very large population. The city government levies a cigarette tax in addition to the state tax, making the rate among the highest in the U.S. The fact that the city is close to several states facilitates noncompliance[17].

The mean of tax noncompliance is higher (.21 versus .28) in the "near the border" sample. As expected, IFNC has a higher mean (7.4 cents versus 13.8 cents) in the near the border sample. The proportion of communities in the Far West and Southwest decreases in the near the border sample, whereas the proportion of communities in New England increases.

Regression Analysis

In table 3 we report parameter estimates from the OLS regression

$$\hat{\hat{S}} = \alpha + \beta_1 IFNC + \beta X + \mu \qquad (2)$$

where, as described earlier in the paper $\hat{\hat{S}}$ is the share of littered packs that are non-compliant (i.e. have no tax stamp or a tax stamp from a lower tax jurisdiction) in the relevant community, X is a vector with a subset of the control variables in table 2 and μ is an unobserved error term with zero mean. Panel 1 presents the estimates using the full sample whereas panel 2 presents estimates using the near the border sample. Independent variables include IFNC, median household income (log-transformed), the proportion of households with cars, population density (log-transformed), proportion of land used for retail and service, and a set of regional dummies. Robust standard errors are reported for all regressions.

In the first column of panel 1 in table 3, the coefficient on IFNC is statistically significant and indicates that a ten cent per mile increase in the tax differential will increase noncompliance by 3.1 percentage points. The relatively small standard error suggests that, conditional on the model assumptions; we can be 95-percent confident that the actual increase in noncompliance will be between about 2.5 and 3.7 percentage points. In Model 2, we add a variable to measure median household income and its coefficient has the expected sign—suggesting that tax noncompliance falls as median income rises—but the coefficientis not statistically significant. The coefficient on IFNC is essentially unchanged.

Model 3 adds the share of households with cars but this variable is not a statistically significant predictor of tax noncompliance. Median household income has a statistically significant effect on noncompliance, suggesting that a ten percent increase in median household income is associated with a 1.2 percentage point decrease in noncompliance. Model 4 includes population density, which has a statistically significant negative effect on tax noncompliance. A ten percent increase in population density is associated with a 0.4 percentage point decrease in tax noncompliance. Median household income and car ownership are statistically insignificant.

In model 5 we also include the proportion of land used for retail or service. The coefficient on this variable is significant and has a large magnitude; a one-percentage point increase in the share of land used for retail or service is associated with a 3.7 percentage point decrease in tax noncompliance. None of the other control variables is associated with a statistically significant change in noncompliance. Model 6 adds a set of regional dummies. In this specification only median household income and IFNC are statistically significant predictors of noncompliance. The magnitude of the coefficient on median household income is larger compared to model 3—a ten percent increase in median household income is associated with a 1.7 percentage point decrease in tax noncompliance.

The control variables do not have robust significant effects on the proportion of packs with no tax stamp or a lower-rate tax stamp across different models. On the other hand, IFNC is a statistically significant and quantitatively important in every specification.

Panel 2 in table 3 reports analogous estimates from applying the same models to the subset of communities within thirty-eight miles of a state border. Specification 1 (column 1) reports the results using all sixty-seven such communities while specifications 2 to 7 drop two communities that are "outliers" because they have very high IFNCs[18].

In these specifications IFNC is again the only consistently statistically significant independent variable. Its magnitude is consistently two to three times as large in columns (2) to (7) of panel 2 as it is in the full sample results shown in panel 1. The explanatory power of the regressions is essentially the same in the two panels.

What do these results suggest about the impact of tax differentials on noncompliance? In the near the border sample the mean tax noncompliance rate is about 28 percent. The coefficient on IFNC in column (7) of panel 2 of table 3 suggests that a ten-cent increase in IFNC would cause noncompliance to increase by eight percentage points to a noncompliance rate of about 36 percent. While this may sound quite large, readers should understand that a fifty cent increase in the cigarette tax differential would be required to generate a ten cent increase in the IFNC of an individual living five miles from the border. An individual living ten miles from the border would have their IFNC increase by only five cents.

Our results can be compared with some earlier literature on tax avoidance gradients. In Merriman (2010) Chicago's tax was $3.105 greater than Indiana at that border. He found that the probability of tax noncompliance decreased (starting at virtually 100%) one percent for each one mile increase in distance to the border. In our framework the rate at which noncompliance falls declines with distance to the border and imply that a tax difference of $3.105 would result in noncompliance of 93 percent at one mile from the border. Moving from fourteen to fifteenmiles from the border would cause noncompliance to fall one-percent from seven to six percent.

Harding et. al. (2012) found that "a 1 ¢ tax increase raises the probability of a cross-state purchase by nearly six percent for consumers on the border". Our results are consistent with this. We find that a one cent increase in IFNC would raise the share of packs that were non-compliant by 1.2 percentage points (which is about six percent of the sample mean) at one-quarter of a mile from the boundary.

Robustness

In this section we discuss various checks on the robustness of our results. Some of our analyses are displayed in table 4 panel 1. Column (1) modifies the regression in column (6) of table 3 (panel 1) by replacing median income with the poverty rate. This causes a slight decline in explanatory power (as measured by adjusted r-squared) but has almost no impact on the magnitude or significance of the coefficient on IFNC. Columns 2 and 3 of table 4 panel 1 experiment by substituting the total tax rate and distance to a lower tax border for IFNC. This specification is, in part, motivated by the evidence presented in table 1,that suggests a substantial portion of noncompliance may result from trafficking. If this is the case, then the total tax may be more important than the tax relative to neighboring states.

Here we measure distance by the average of straight line as-the-crow-flies distance from the block centroid to the nearest state border with a lower tax weighted by population in each block that constitutes a community. This is in some ways a more general specification of incentives for noncompliance than the one used intable 3 since it allows both distance to a tax border and tax rate to separately affect noncompliance. In columns 2 and 3 we show that the total tax rate has a positive and statistically significant association with noncompliance—a one-dollar increase in the tax rate will raise the noncompliance rate by eight percentage points. Distance to the border has a negative and (in column 2) statistically significant association with noncompliance but the coefficient is surprisingly small. This suggests that, holding the total tax rate constant, a 100-mile increase in distance to the border would lower the noncompliance rate by only two percentage points. In these specifications the explanatory power of the model is reduced compared to specifications including IFNC which suggests that incentives for cross-border shopping are an important determinant of noncompliance.

In column 4 of table 4 panel 1 we include IFNC in addition to the total tax rate and distance[19]. Even with these two collinear (by construction) variables, the coefficient on IFNC is statistically and economically significant and is, quite similar to the coefficient in column 1. Column 5 is identical to column 4 except that we drop the two communities with very high values for IFNC. The magnitude of the coefficient on IFNC rises but the standard error also increases so the coefficient is not statistically significant. The coefficient on the total tax rate also becomes insignificant.

Column 6 of table 4 panel 1 reverts to the specification of table 3 panel 1 column (6) but excludes the two outliers and includes only the fifty-six communities in which we found ten or more littered packs for which we could determine noncompliance status. For this sample, in which we might have the most confidence about our measure of noncompliance, the coefficient on IFNC is statistically significant and is 2.1 times as large as the coefficient in the full sample.

Table 4 panel 2 reports regressions analogous to table 4 panel (1) (columns 1 to 4) for our close to the border sample (dropping our two outliers). In this sample, replacing median income with the poverty rate causes the coefficient on IFNC to fall and become statistically insignificant.

The coefficient on the poverty rate is very large and suggests that increases in poverty are associated with more noncompliance but is not statistically significant. In columns (2) and (3) of table 4 panel 2 we replace IFNC with separate variables for the total tax rate and distance to the border. Again, this fits the data less well than including IFNC and total tax rate has the intuitive sign although it is not statistically significant in column (3). Distance to the border is statistically insignificant and has a (non-intuitive) positive sign in both specifications (2) and (3). In column (4) we include IFNC along with total tax rate and distance to the border and find that none of the individual variables has a statistically significant effect.

Overall, the results presented in table 4 suggest caution because conclusions drawn from regressions like those presented in table 3 may be sensitive to functional form and sample. Despite this there seems to be strong evidence that noncompliance rates vary consistently with our measure of incentives for noncompliance.

Non-Linearity between Tax Noncompliance and IFNC

Although the OLS regressions provide strong evidence that financial incentives to avoid cigarette taxation are a very important determinant of behavior they have limited direct relevance for policymakers in any particular jurisdiction. Policymakers must assess the impact of a cigarette tax increase, or neighboring jurisdictions' tax increases on noncompliance, consumption and revenue in their home jurisdiction. The OLS results suggest that there will be some reaction but in any particular jurisdiction the magnitude depends on relative tax rates and proximity to lower tax jurisdictions. For example, if a jurisdiction already has much higher taxes than its neighbors, a tax increase might cause little additional noncompliance because those who can easily avoid will already have done so. The appropriate econometric treatment is to allow our estimates to vary non-linearly with the level of IFNC. We do this by hypothesizing that

$$ln\left\{\frac{E(\hat{\hat{S}})}{1-E(\hat{\hat{S}})}\right\} = \alpha + \beta_1 IFNC + \beta X \quad (3)$$

and that $\hat{\hat{S}}$ has a binomial distribution. We use Stata 14's glm (generalized linear model) command to maximize the appropriate likelihood function (given the observed data) and to recover parameters that relate our dependent and independent variables.

This specification conforms to the a priori constraint that predicted noncompliance is always between zero and one and is flexible enough to allow the marginal effect of IFNC on noncompliance to vary with its level. The parameter estimates are consistent with those in tables 3 and 4—IFNC is a statistically significant determinant of noncompliance. Because of the nonlinear specification, the magnitudes of the estimated coefficients are not very informative, and we do not report them in this paper due to space constraint[20]. Instead, we use the coefficients to produce enlightening simulations that relate tax policy to tax revenue and noncompliance. Because IFNC differs across communities there is no single revenue-maximizing tax rate. We simulate the revenue maximizing tax rate for a "typical" community at each level of IFNC taking into account changes in noncompliance.

To better understand how we do this recall that IFNC depends upon the savings per mile traveled to lower tax borders as well as the distribution of the population within the community, i.e. $IFNC = \sum_{i=1}^{n}\left(\frac{Pop_i}{Pop_c}\right)TS_i$. Taking the derivative of IFNC with respect to the local tax rate implies that $\frac{\partial IFNC_C}{\partial t_i} = \sum_{i=1}^{N}\frac{Pop_i}{Pop_c}\frac{\partial TS_i}{\partial t_i} = \sum_{i=1}^{N}\frac{Pop_i}{Pop_c}\left(\frac{1}{d_{ij}}\right)$ so the change in IFNC as a result of a tax increase will depend upon the distribution of the population as well as the distance to neighboring jurisdictions. We find that the vast majority of the population has an IFNC of less than ten cents per mile.

Using this information and GLM estimates of the parameters of equation (3) we simulate the impact of tax increases on noncompliance and tax revenue and compute the revenue maximizing tax rate[21]. Our simulated revenue-maximizing tax rates depend on a community's IFNC. Our simulations show that for the vast majority of communities the revenue-maximizing tax rate will be very high so that, any policy relevant tax increase would increase tax revenue.

Consider, for example the simulation results shown in figure 2. Remember from figure 1 that only eight of 132 communities had an IFNC greater than 20 cents per mile. Figure 2 shows that according to our simulations even with an IFNC as high as 20 cents per mile the revenue maximizing tax rate would be $4.84. Even for the very few communities with an IFNC at the very top of the IFNC distribution (50 cents or more per mile) a tax rate above two dollars per pack would maximize revenue. Thus, our empirical results suggest that virtually all communities could increase cigarette tax revenue by increasing tax rates even after taking potential increases in tax avoidance into effect.

Limitations

Limitations of this study are as follows. First, although our sample of geographic areas is national and intended to be representative of a particular group of school attendees it may not be representative of the population as a whole or the population of smokers[22]. Furthermore, the areas that were sampled within the selected communities cannot be guaranteed to be representative of those communities. Thus, we view our sample as an exploratory and experimental piece of evidence that both demonstrates the concept of national scale littered pack collections and provides data that can be usefully combined with other evidence to better understand the relationship between incentives and tax noncompliance. Given this limitation, our estimates of tax noncompliance is consistent with previous literature such as Fix et al (2014).

A second limitation of our study is that, by its nature, our data cannot link tax noncompliance to the characteristics of individual smokers. Because we observe the location in which packs were discarded (and infer the location in which they were purchased) we can link noncompliance to community characteristics and polices but not to individual characteristics. We argue, however, that controls for community level policies, rather than individual characteristics are of prime relevance to understanding the policy implications of our findings.

Third, our data cannot definitively distinguish between casual cross-border purchases by individuals (tax avoidance) and organized—probably criminal—attempts to evade taxes. We cannot demonstrate the accuracy of the rough proxy we use in this study—miles between the location of purchase and the location at which the pack was discarded.

Under what circumstances could these limitations overturn the principal findings of our study? Could the sampling frame used in our data collection (public school attendance areas) explain why we estimate avoidance that is higher than the nationally representative sample used in TUS-CPS? According to the 2012 Digest of Education Statistics about one-quarter of all families had one or more school age children and roughly 90 percent of them attended public school. Thus, about 22.5 percent of households had a child in public school in 2012 when our data was collected. If our data, which shows a noncompliance rate of about 21 percent, is representative of the areas in which these families live the total national noncompliance rate could be no lower than 4.7 percent even if families with public school children lived in completely self-contained communities and noncompliance was zero in communities without school aged children. It strikes us as highly implausible and at variance with common knowledge to suggest that households with school aged children are so unrepresentative of the general population. In fact, these households often live in communities intermingled with household that do not have school aged-children. Since TUS-CPS data yield noncompliance estimates on the order of five percent, we believe that it is highly unlikely that differences between our results and the TUS-CPS results can be explained by sampling frame alone. In any case, we can think of no a priori reason to believe that households with public school children would be less compliant than the general public.

Could assumptions inherent in our methodology exaggerate the importance of financial incentives on noncompliance? Of course, our econometric results in Tables 3 and 4 could suffer from omitted variable bias since we can include only a small selection of control variables and know nothing about the characteristics of the individual smokers who litter. In order for an omitted variable to bias the estimate of the coefficient on IFNC upward the omitted variables must be correlated with both noncompliance and IFNC. Our measure of the incentive for noncompliance depends on relative tobacco tax rates and on the geographic distribution of the population relative to tax borders. It is difficult to imagine that the geographic distribution of the population relative to tax borders is correlated with noncompliance except through the hypothesized (and modeled) impact on financial incentives.

Admittedly however, relative tax rates could reflect attitudes toward tobacco and attitudes toward tobacco might influence noncompliance independent of financial incentives. Suppose for example, that communities with strong anti-tobacco sentiments also have relatively high tobacco taxes. Could smokers in those communities be less likely to comply with tax laws simply because they enjoy violating community norms? If this were the case we might falsely attribute noncompliance to financial incentives. In this case our empirical estimates would <u>over-estimate</u> the degree to which noncompliance responded to financial incentives and the major conclusion of figure 2—that tobacco tax increases will almost surely increase tax revenue even when potential increases in noncompliance are taken into account—would be strengthened since tax increases would result in smaller increases in noncompliance. Thus, despite admitted and significant limitations in our methodology we have strong confidence in our primary policy-relevant conclusions.

Finally, we note that our methodology cannot definitively distinguish between noncompliance that is due to cross-border shopping and noncompliance that is due to organized smuggling. Several pieces of evidence are relevant, however. Table 1 shows that a significant share of non-compliant packs has no tax stamp or has been transported large distances and are reflective of much lower tax rates. This suggests that a substantial amount of noncompliance is due to trafficking. Some of our regressions (see table 4) suggest distance to a lower tax border was a much less important determinant of noncompliance than the total tax rate implying that trafficking may be important. However, these results vary somewhat with functional form and are not precise enough to draw definitive conclusions. The weight of current evidence suggests that both cross-border shopping and trafficking are important reasons for noncompliance.

Conclusion and Policy Implications

Tobacco control policies are designed to reduce tobacco use. Administrative data on tax paid cigarette sales, surveys of smokers and other measures may provide imperfect measures of cigarette consumption and cigarette tax noncompliance. Our data on a national sample of littered cigarette packs provides information about noncompliance in a large variety of areas and circumstances. The data also makes it possible to predict the effect of tax policy changes on avoidance and evasion and tax revenue as a function of both a community's relative tax rate and the distance to neighboring communities with lower tax rate. Our statistical analyses provide strong evidence that noncompliance increases with tax differentials and proximity to low tax cigarettes. Empirically our proxy for incentives for tax noncompliance—cents per mile saved by the average person in the community—explains the variance in measured noncompliance rates better than any of the other potential explanatory variables we examine.

We show that for the vast majority of people in our sample communities, potential tax savings from cross border shopping is only a few cents per mile. We also show that while increases in cigarette tax rates may result in increased noncompliance, the increase in noncompliance is low enough that tax revenues will rise in almost all cases. This evidence suggests that even in communities with quite high cigarette tax rates and close proximity to adjacent areas with relatively low cigarette taxes, increases are a viable tool for increasing revenues. Despite tax avoidance and evasion, virtually all communities will obtain increased revenues if they increase cigarette tax rates.

What are the policy implications of our study? First, evidence from our study bolsters the findings of Fix et. al. (2014) and suggests that there may be more cigarette tax noncompliance than suggested by survey data sets, like TUS-CPS. Underreporting may come about because survey respondents are reluctant to reveal illicit activities and purchase of noncompliant packs from local retailers. There are, of course, other possible explanation but to this point there are a lack of studies that investigate the discrepancies in noncompliance between different methods. Our research highlights the need for policymakers to pay more attention to the issues of tax avoidance and evasion as well as the need for diverse and rigorous research methods. However, what may matter most is the relative, rather than absolute, level of noncompliance. Our most important results explain relative levels of noncompliance across geographic areas as a function of relative incentives for noncompliance. Even if there is an upward bias in measured noncompliance using littered pack surveys there is no reason to believe that the bias would vary systematically with incentives for noncompliance. This strengthens the policy implications of our study. Second, our study illustrates why policies that promote cigarette tax harmonization or programs to track and trace cigarettes packs from point of production to the point of sale[23] may be effective means to discourage tax evasion and to promote public health. We present new evidence, consistent with earlier studies (Colman and Remler 2008; Maclean, Webber, and Marti 2014; Nicholson, Turner and Alvarado 2016; Tauras 2005), that shows that cross-region tax differentials are strongly associated with tax avoidance and evasion. We document the magnitude of this association and quantify the interaction between tax differentials and distance.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

[Description of funding]

Notes

[1] The 2014 report of the U.S. Surgeon General provided a 50 year retrospective on progress related to the health consequences of smoking and finds that (p.12) "(t)he evidence is sufficient to conclude that increases in the prices of tobacco products, including those resulting from excise tax increases, prevent initiation of tobacco use, promote cessation, and reduce the prevalence and intensity of tobacco use among youth and adults."

[2] Recent contributions in economics and public policy (rather than public health) journals include Goolsbee, Lovenheim, and Slemrod (2010), Shetty et. al (2011), Goldin and Homonoff (2013) . Maclean, Webber and Marti (2014), Ribisl, Seidenberg and Orlan (2016), and van Walbeek (2015).

[3] Wilson, Thomsons and Edwards (2009) report on a large study using similar methods in New Zealand.

[4] See https://research.chicagobooth.edu/kilts. The data is also available for direct purchase from Nielsen, see www.nielsen.com.

[5] See Institute of Medicine 2015 p.88 for cautions about industry sponsored estimates of the illicit tobacco market.

[6] See also Institute of Medicine 2015 for analysis of the strengths and weakness of various methods of estimating the size of the illicit tobacco market.

[7] See footnote 7 in Harding et. al. 2012.

[8] While our sample is national in scope it is not necessarily nationally representative. The sample was designed to be representative of communities with school children in certain age ranges. Additional detail about sampling is discussed below in the section on data collection strategy.

[9] However, note that Fix et. al. (2014) found "(r)elatively few demographic differences" between smokers who returned and did not return cigarette packs for inspection. These pack collections found avoidance rates very similar to those found in our littered pack collection.

[10] Some readers have noted that littered pack collections also are more likely to obtain data from heavy smokers since they are more likely to finish a pack while outside. We do not view this as a problem however since we wish to sample from the population of cigarette packs rather than the population of smokers. A representative sample will contain more packs per heavy smoker than packs per light smoker.

[11] See http://www.ihrp.uic.edu/study/monitoring-and-assessing-impact-tax-and-price-policies-us-tobacco-use for more information about the larger project.

[12] A community, is one in which most of the students who attend the school live. It is identified either through maps produced and available from the school district, or in a few cases (when we cannot obtain the maps and no one at the district can/will tell us) there is an algorithm that looks at location of other schools against the school district boundary area (available from census) and population data. [Bachman et al 2011]

[13] To simplify data collection and avoid arbitrary judgments in the field, collectors were told to pick up every littered pack on the route.

[14] Davis et. al. 2013 use a similar approach to estimate cigarette trafficking. In their study of littered packs in five U.S. east coast cities they find that 58.7 percent are non-compliant. They estimate that between 30.5 and 42.1 percent of all packs were the due to trafficking suggesting that between 52 and 72 percent of non-compliant packs were due to trafficking in their data.

[15] Interested readers can refer to [Redacted; a working paper online] for more details on the distribution of population by IFNC level.

[16] Regional dummies may also control for ease of cigarette trafficking since sources of and distance to low tax cigarettes differs by region.

[17] Confidentiality agreements prevent us from explicitly identifying the communities in which we collected data.

[18] The IFNCs in one of the omitted communities is \$1.13 per mile. The IFNC in the other omitted community is \$3.12 per mile. If we exclude these two outliers in OLS model using the full sample, the coefficient on IFNC is larger and about the same size as reported in columns 2 to 7 in panel 2 of table 3. Thus, the major difference between the coefficients on IFNC reported in panel 1 and 2 of table 3 is the omission of the two outliers in panel 2 columns 2 to 7.

[19] We also ran additional analysis in which we removed tax rate and distance and included only IFNC. IFNC becomes statistically significant. In the model that included the two outliers, IFNC has similar magnitude as reported in table 3, column 6 (0.002). In the model excluding the outliers, IFNC has a larger magnitude (0.008).

[20] Interested readers can consult our [redacted] working paper.

[21] More details about our procedures are available in an appendix to our [redacted] working paper.

[22] As discussed above the sampled communities represented school enrollment areas for nationally representative samples of 8th, 10th, and 12th grade public school students in the continental U.S. More details about the sample can be found at http://www.bridgingthegapresearch.org/research/community_data/. See Barker et. al. 2016 for more information about data collection procedures.

[23] See Chaloupka et. al. 2015 for more discussion of these policies.

References

Aziani, Alberto, Jonathan Kulick, Neill Norman, and James E. Prieger. 2017. Discarded Pack Data and the Prevalence of Illicit Trade in Cigarettes. Available at SSRN: https://ssrn.com/abstract=2906015 or http://dx.doi.org/10.2139/ssrn.2906015 Bachman, Jerald G., Lloyd D. Johnston, Patrick M. O'Malley, and John E. Schulenberg. 2011. The Monitoring the Future Project after thirty-seven years: Design and procedures. In *Monitoring the Future Paper Series*. Ann Arbor, MI: University of Michigan.

Baltagi, Badi H., and Dan Levin. 1986. Estimating dynamic demand for cigarettes using panel data: the effects of bootlegging, taxation and advertising reconsidered. *The Review of Economics and Statistics* 68 (1):148-155.

Barker, Dianne, Shu Wang, David Merriman, Andrew Crosby, Elissa Resnick, and Frank Chaloupka. 2016. Estimating Cigarette Tax Avoidance and Evasion: Evidence from a National Sample of Littered Packs. *Tobacco Control* 25 (suppl 1):i38-43.

Becker, Gary S., Michael Grossman, and Kevin M. Murphy. 1991. Rational Addiction and the Effect of Price on Consumption. *American Economic Review* 81 (2):237-241.

Chernick, Howard, and David Merriman. 2013. Using littered pack data to estimate cigarette tax avoidance in NYC. *National Tax Journal* 66 (3):635-668.

Colman, Gregory J., and Dahlia K. Remler. 2008. Vertical equity consequences of very high cigarette tax increases: If the poor are the ones smoking, how could cigarette tax increases be progressive? *Journal of Policy Analysis and Management* 27 (2):376-400.

Crosby, Andrew, David Merriman, Shu Wang, and Frank Chaloupka. 2014. *How much cigarette tax avoidance is there in the US? Different estimates from smoker surveys and physical pack collection methods*: University of Illinois at Chicago.

Davis, Kevin C., Victoria Grimshaw, David Merriman, Matthew C. Farrelly, Howard Chernick, Micaela H. Coady, Kelsey Campbell, and Susan M. Kansagra. 2013. Cigarette trafficking in five northeastern US cities. *Tobacco Control*:1-7.

DeCicca, Philip, Donald Kenkel, and Feng Liu. 2013. Excise tax avoidance: the case of state cigarette taxes. *Journal of Health Economics* 32 (6):1130-1141.

Fix, Brian. V., Andrew Hyland, Richard J. O'Connor, K. Michael Cummings, Geoffrey T. Fong, Frank J. Chaloupka, and Andrea. S. Licht. 2014. A novel approach to estimating the prevalence of untaxed cigarettes in the USA: findings from the 2009 and 2010 international tobacco control surveys. *Tobacco Control* 23 Suppl 1:i61-66.

Fullerton, Don. and Kinnaman, Thomas.C., 1996. Household responses to pricing garbage by the bag. *The American Economic Review*, 86(4): 971-984.

Galbraith, John W., and Murray Kaiserman. 1997. Taxation, smuggling and demand for cigarettes in Canada: evidence from time-series data. *Journal of Health Economics* 16 (3):287-301.

Hanson, Andrew and Ryan Sullivan. 2009. The incidence of tobacco taxation: Evidence from Geographic micro-level data. *National Tax Journal*, 62 (4): 677-698.

Harding, Matthew, Ephraim Leibtag, and Michael F. Lovenheim. 2012. The heterogeneous geographic and socioeconomic incidence of cigarette taxes: Evidence from nielsen homescan data. *American Economic Journal: Economic Policy* 4 (4):169-198.

Institute of Medicine and National Research Council. 2015. *Understanding the U.S. illicit tobacco market: characteristics, policy context, and lessons from international experience*. Washington DC: The National Academies Press.

Kurti, Marin K., Klaus von Lampe, and Douglas E. Thompkins. 2013. The illegal cigarette market in a socioeconomically deprived inner-city area: the case of the South Bronx. *Tobacco Control* 22 (2):138-140.

Lakhdar, Ben C. 2008. Quantitative and qualitative estimates of cross-border tobacco shopping and tobacco smuggling in France. *Tobacco Control* 17 (1):12-16.

Lovenheim, Michael F. 2008. How Far to the Border?: The extent and impact of cross–border casual cigarette smuggling." *National Tax Journal* 61 (1):7-33.

Maclean, Johanna Catherine, Douglas A. Webber, and Joachim Marti. 2014. An application of unconditional quantile regression to cigarette taxes. *Journal of Policy Analysis and Management* 33 (1):188-210.

Merriman, David. 2010. The Micro-geography of Tax Avoidance: Evidence from littered cigarette packs in Chicago. *American Economic Journal: Economic Policy* 2 (2):61-84.

National Cancer Institute. 2016. The economics of tobacco and tobacco control. In *NCI Tobacco Control Monograph Series 21*, edited by National Cancer Institute and World Health Organization. Washington DC.

Nicholson, Andrew, Tracy M. Turner, and Eduardo Alvarado. 2014. Cigarette taxes and cross-border revenue effects evidence using retail data. *Public Finance Review* 44 (3):311-343.

Pesko, Michael F., John A. Tauras, Jidong Huang, and Frank J. Chaloupka. 2016. The influence of geography and measurement in estimating cigarette price responsiveness. edited by National Bureau of Economic Research.

Ribisl, Kurt M., Andrew B. Seidenberg, and Elizabeth N. Orlan. 2016. Comprehensive e-cigarette regulation as a step toward harm reduction. *Journal of Policy Analysis and Management* 35 (2): 492-495.

Saba, Richard R., T. Randolph Beard, Robert B. Ekelund, and Rand W. Ressler. 1995. The demand for cigarette smuggling. *Economic inquiry* 33 (2):189-202.

Stehr, Mark. 2005. Cigarette tax avoidance and evasion. *Journal of health economics*, 24(2): 277-297.

Tauras, John A. 2005. Can public policy deter smoking escalation among young adults? *Journal of Policy Analysis and Management* 24 (4):771-784.

Thursby, Jerry G., and Marie C. Thursby. 2000. Interstate cigarette bootlegging: extent, revenue losses, and effects of federal intervention. *National Tax Journal* 53 (1):59-77.

van Walbeek, Corné. 2015. The economics of tobacco control (Part 2): evidence from the ITC Project. *Tobacco control* 24 (Suppl 3):iii1-iii3.

Orzechowski and Walker. 2014. *The tax burden on tobacco*. Arlington, VA: Author.

Webb, Eugene J., Donald T. Campbell, Richard D. Schwartz, and Lee Sechrest. 1966. *Unobtrusive Measures*. Chicago: Rand McNally.

Wilson, Nick, George Thomson, Richard Edwards, and Jo Peace. 2009. Estimating missed government tax revenue from foreign tobacco: survey of discarded cigarette packs. *Tobacco control* 18 (5):416-418.

Author Biography

Shu Wang is an Assistant Professor at the Department of Agricultural, Food, and Resource Economics at Michigan State University. She received her doctorate in Public Administration from University of Illinois at Chicago. Her research interests include state-local relations, fiscal federalism, and local government management.

Table 1: Number of Packs by Tax Differential and Distance between Purchase Origin and Sales Destination

Distance (d) (Miles)	Tax differential (TD)								
	TD<0	**TD=0**	**0<TD<5c**	**5c <TD <25c**	**25c <TD < 50c**	**50c < TD < $1**	**TD>$1**	**Unknown**	**Total**
d=0	2	1,680	29	0	0	0	8		**1,719**
0<d<=20	19	0	4	4	1	22	5		**55**
20<d<=50	1	0	0	0	1	4	7		**13**
50<d<=100	2	0	0	0	1	0	4		**7**
100<d<=150	0	0	1	1	0	0	4		**6**
d>150	11	0	1	6	3	4	47		**72**
No stamp(1)	0	0	0	0	31	51	101		**183**
Unidentifiable, possibly foreign(2)				Unknown				61	**61**
Total	**35**	**1,680**	**35**	**11**	**37**	**81**	**176**	**61**	**2,116**
	1.65%	79.40%	1.65%	0.52%	1.75%	3.83%	8.32%		

Notes: TD is the difference in cigarette tax rate between the jurisdiction where a pack was found and the jurisdiction that issued the stamp on the pack. TD equals to zero if the pack had the appropriate stamp that is consistent with the location it was found. A positive value of TD indicates the cents of tax saved through noncompliance. A negative tax differential indicates that the consumer paid higher cigarette taxes, and the difference in two locations is due to reasons such as travel. We do not count these packs as tax noncompliance.

The distance is measured from the centroid of the community where a pack was found the nearest border of the state that issued the tax stamp on the pack.

(1) For the packs with no stamp, distance from purchase origin to destination is unknown. We calculated tax differential based on the assumption that no state or local taxes are paid for these packs. However, it is possible that these are packs from North Carolina, South Carolina, or North Dakota, where state cigarette taxes are levied but no tax stamps are affixed on the pack.

(2) There are 61 packs that have stamps unidentifiable. These stamps can possibly be foreign stamps. Tax differential is thus unknown because we don't know where the packs were bought.

Table 2: Summary Statistics

Panel 1: all communities

	N	Mean	Median	SD	Min	Max
Tax noncompliance (share of packs)	132	0.21	0.11	0.28	0	1
IFNC (cents per mile)	132	7.43	0.63	30.38	0	312.51
Total tax rate ($s per pack)	132	1.60	1.34	1.22	0.17	5.85
Distance to tax border (miles)	132	155.42	87.16	213.99	0.53	1365.58
Number of packs per community	132	16	9	18	1	83
Median income ($s)	132	61,832	57,580	24,464	25,000	157,690
Share of households with cars	132	0.93	0.95	0.07	0.56	1
Population density (people per sq. mile)	132	2,613	1,015	4,931	21	31,145
share of retail/service land use	132	0.32	0.32	0.15	0.04	0.77
Poverty rate	132	0.12	0.11	0.08	0.03	0.41
Far West	132	0.14	0	0.35	0	1
Great Lakes	132	0.17	0	0.37	0	1
Midwest	132	0.16	0	0.37	0	1
New England	132	0.05	0	0.23	0	1
Plains	132	0.06	0	0.24	0	1
Rocky Mountain	132	0.02	0	0.15	0	1
Southeast	132	0.27	0	0.44	0	1
Southwest	132	0.13	0	0.34	0	1

Panel 2: only communities within 38 miles from the state border

	N	Mean	Median	SD	Min	Max
Tax noncompliance (share of packs)	67	0.28	0.15	0.34	0	1
IFNC (cents per mile)	67	13.78	0.86	41.81	0	312.51
Total tax rate ($s per pack)	67	1.883	1.600	1.484	0.170	5.850
Distance to tax border (miles)	67	16.24	14.21	10.65	0.20	35.01
Number of packs per community	67	17	11	17	1	76
Median income ($s)	67	61,440	53,842	27,596	25,000	157,690
share of households with cars	67	0.92	0.95	0.08	0.56	0.99
Population density (people per sq. mile)	67	2,660	665	5,901	24	31,145
Share of retail/service land use	67	0.33	0.31	0.14	0.07	0.77
Poverty rate	67	0.12	0.12	0.07	0.03	0.34
Far West	67	0.05	0	0.21	0	1
Great Lakes	67	0.16	0	0.37	0	1
Midwest	67	0.28	0	0.45	0	1
New England	67	0.10	0	0.31	0	1
Plains	67	0.06	0	0.24	0	1
Rocky Mountain	67	0.02	0	0.12	0	1
Southeast	67	0.31	0	0.47	0	1
Southwest	67	0.02	0	0.12	0	1

To construct tax noncompliance and IFNC, we collected littered packs from each SEZ, and compared the jurisdiction that issued the tax stamp on each pack with the community location. Distance to tax border is an average of straight line as-the-crow-flies distance from the block centroid to the nearest lower-rate state border weighted by population in each block that constitutes a community. Both the geographical distance and population were collected from US Census. Total tax rates were collected from government web sites. Median income, share of households with cars, population density, share of retail/service land use, and poverty rate are collected from US Census.

Table 3: Determinants for the Share of Packs with Tax Noncompliance

Panel 1: All 132 communities included						
	(1)	(2)	(3)	(4)	(5)	(6)
IFNC	0.0032***	0.0033***	0.0034***	0.0035***	0.0034***	0.0029***
	(0.0056)	(0.0062)	(0.0073)	(0.0066)	(0.0068)	(0.0074)
Ln(Median_income)		-0.1032	-0.1296*	-0.0278	-0.059	-0.1709**
		(0.0639)	(0.0732)	(0.0707)	(0.0740)	(0.0851)
% of households with cars			0.363	-0.2531	-0.5549	0.0605
			(0.3550)	(0.3537)	(0.3976)	(0.4713)
Ln(Population density)				-0.0379**	-0.0248	-0.0281
				(0.0160)	(0.0151)	(0.0175)
% of retail/service land use					-0.3707*	-0.2338
					(0.2068)	(0.1907)
Constant	0.1876***	1.3183*	1.2689*	0.9725	1.6302**	2.3153***
	(0.0235)	(0.7093)	(0.6835)	(0.6341)	(0.8008)	(0.8262)
Region dummies	No	No	No	No	No	Yes
R squared	0.107	0.118	0.118	0.148	0.159	0.184

Panel 2: Only communities within 38 miles from the state border							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IFNC	0.0027***	0.0053*	0.0062**	0.0091***	0.0089***	0.0088***	0.0082**
	(0.0004)	(0.0027)	(0.0024)	(0.0027)	(0.0024)	(0.0025)	(0.0032)
Ln(Median_income)			-0.1701*	-0.2404**	-0.0552	-0.0780	-0.1850
			(0.0968)	(0.1024)	(0.0998)	(0.1028)	(0.1492)
% of households with cars				1.091**	0.0320	-0.5039	0.0650
				(0.4173)	(0.4897)	(0.5111)	(0.7636)
Ln(Population density)					-0.0736**	-0.0537*	-0.0518
					(0.0284)	(0.0288)	(0.0369)
% of retail/service land use						-0.6588*	-0.3894
						(0.3414)	(0.3540)
Constant	0.2432***	0.2261***	2.0781*	1.8247*	1.2454	2.0765**	2.7612**
	(0.0404)	(0.0432)	(1.0708)	(1.0010)	(0.8780)	(0.9944)	(1.0625)
Region dummies	No	No	No	No	No	No	Yes
R squared	0.103	0.041	0.069	0.112	0.179	0.201	0.170
N	67	65	65	65	65	65	65

Robust standard errors in parentheses
** p<0.10; ** p<0.05; *** p<0.01*
Region dummy variables: reference group="Far West"
Two observations omitted in panel 2.
One omitted community has an IFNC of 113. The other omitted community has an IFNC of 312.

Table 4: Robustness Check with Alternative Predictors

Panel 1: All communities included & communities with number of packs >=10

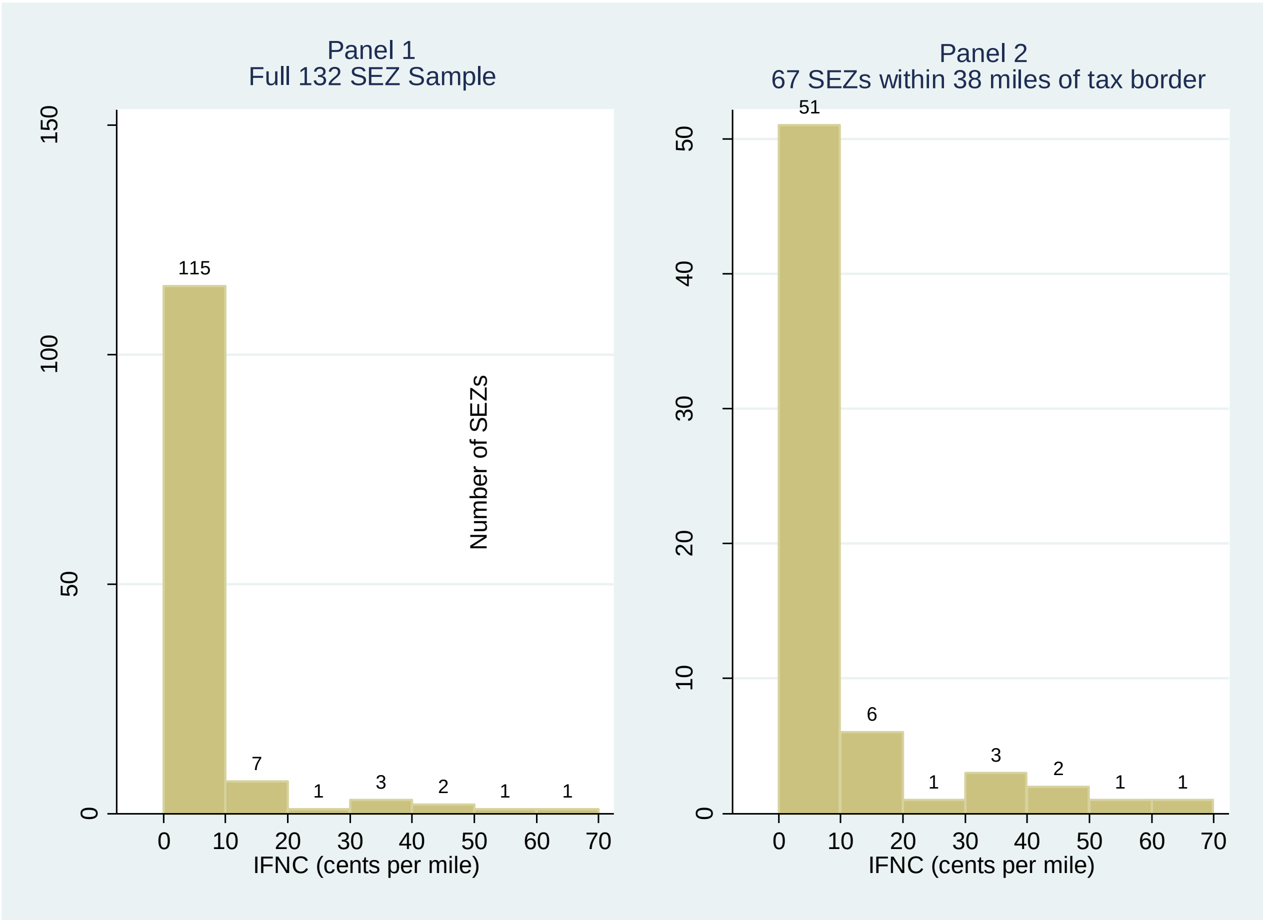
	(1)	(2)	(3)	(4)	(5)	(6)
IFNC	0.0030***			0.0025***	0.0054	0.0061**
	(0.0006)			(0.0004)	(0.0032)	(0.0020)
Total tax rate		0.0800**	0.0736*	0.0590*	0.0413	
		(0.0286)	(0.0305)	(0.0287)	(0.0286)	
Distance to border		-0.0002*	-0.0002*	-0.0002	-0.0002	
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	
Ln(Median_income)		-0.2382*				-0.1078
		(0.0931)				(0.1163)
Poverty rate	0.2533		0.5124	0.5459	0.6182	
	(0.3435)		(0.3973)	(0.3756)	(0.3883)	
% of households with cars	-0.2829	0.6480	0.2596	0.1728	0.3347	0.2772
	(0.4490)	(0.5065)	(0.5157)	(0.4464)	(0.4966)	(0.6037)
Ln(Population density)	-0.0417*	-0.0188	-0.0350*	-0.0388*	-0.0365*	-0.0155
	(0.0176)	(0.0184)	(0.0174)	(0.0168)	(0.0170)	(0.0219)
% of retail/service land use	-0.2088	-0.2577	-0.2413	-0.2412	-0.2472	0.1721
	(0.1878)	(0.1986)	(0.1968)	(0.1931)	(0.1924)	(0.2215)
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes
R squared	0.1536	0.1661	0.1324	0.1958	0.1433	0.2177
N	132	132	132	132	130	56

Panel 2: Only communities within 38 miles from the state border

	(1)	(2)	(3)	(4)
IFNC	0.0059			0.0064
	(0.0036)			(0.0038)
Total tax rate		0.0958*	0.0620	0.0199
		(0.0384)	(0.0380)	(0.0365)
Distance to border		0.0032	0.0033	0.0053
		(0.0038)	(0.0036)	(0.0037)
Ln(Median_income)		-0.2078		
		(0.1565)		
Poverty rate	-0.8401		-0.7490	-0.5836
	(0.7343)		(0.8118)	(0.7937)
% of households with cars	-1.1581	-0.0885	-1.3062	-0.9899
	(0.7915)	(0.8132)	(0.7863)	(0.8046)
Ln(Population density)	-0.0861*	-0.0504	-0.0850*	-0.0877*
	(0.0369)	(0.0392)	(0.0382)	(0.0379)
% of retail/service land use	-0.4018	-0.4610	-0.4457	-0.3538
	(0.3479)	(0.3316)	(0.3555)	(0.3461)
Region dummies	Yes	Yes	Yes	Yes
R squared	0.1635	0.159	0.145	0.1640
N	65	65	65	65

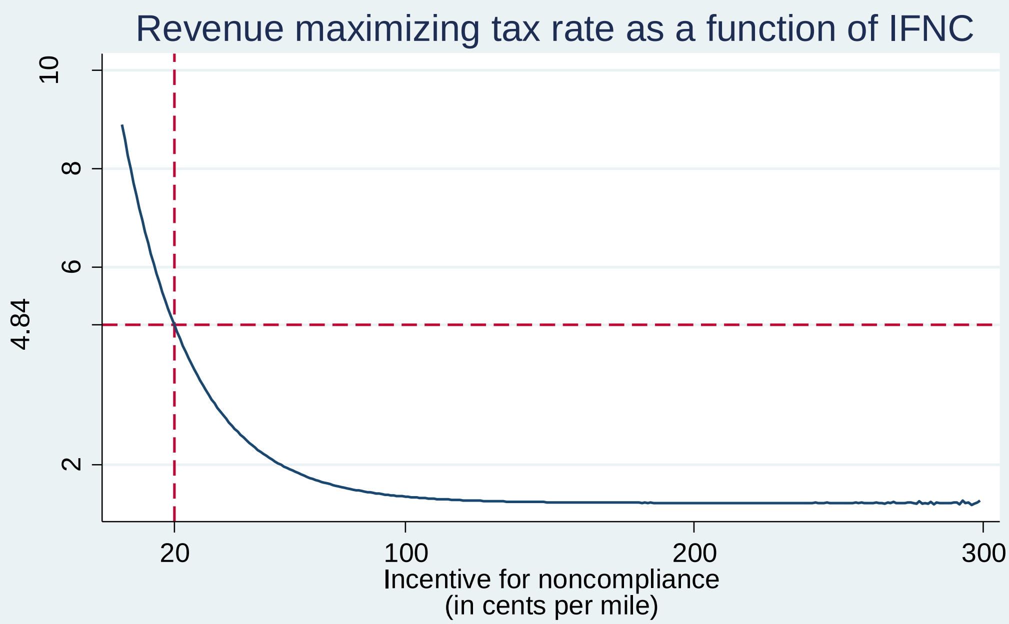
Robust standard errors in parentheses
* p<0.10; ** p<0.05; *** p<0.01
In column 5 &6 of panel 1, and all columns in panel 2, two observations omitted in panel 2. One omitted community has an IFNC of 113. The other omitted community has an IFNC of 312. Column 6 of panel 1 sample includes only communities in which 10 or more useable littered packs were found.

Figure 1: Number of Communities (SEZs) By IFNC



Notes: Table shows histogram of IFNC by number of communities, or school enrollment zones (SEZs). All SEZs have IFNC lower than 70 cents, except two SEZs that have IFNC of 113 cents and 312 cents, respectively. The graphs here exclude these two outliers.

Figure 2: Revenue Maximizing Tax Rate as a Function of IFNC



Note: Curve is simulated using observed data on independent variables and estimated parameters as described in the text. As shown in figure 1 the vast majority of communities have an IFNC<20 and thus have simulated revenue-maximizing tax rates above $4.84. See text for more detail on calculations.