Working Paper No. 2022-02

Revisiting the Effects of Cigarette Taxes on Smoking Outcomes

By Vinish Shrestha

October 2022

© 2022 by Author. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.
Revisiting the Effects of Cigarette Taxes on Smoking Outcomes*

Vinish Shrestha†

20 October, 2022

Abstract

This study re-evaluates the efficacy of cigarette taxation in curtailing smoking. I use recent advancements in the difference-in-differences (DiD) literature to account for heterogeneous treatment effects and compare the findings to the two-way-fixed effect (TWFE) estimates. Using data from the Behavioral Risk Factor Surveillance System Selected Metropolitan/Micropolitan Area Risk Trend (BRFSS SMART) for sample periods 2004-2010 and 2015-2020, the study presents three main findings. First, the results for 2004-2010 sample show that the TWFE estimate is only about 65% of the size of the overall average treatment effect on the treated (ATT) estimate obtained using DiD framework. Second, the event-study type estimates increase gradually in magnitude following the treatment year, thus demonstrating dynamic treatment effects ignored by the TWFE estimate. Third, the ATT estimate pertaining to 2015-2020 sample is only about 63% of the ATT estimate for 2004-2010 sample. Overall, the findings point out that relying on TWFE models to obtain elasticity estimates may bias the estimates towards zero.

Keywords: Cigarette taxation, Difference-in-Differences, Treatment heterogeneity, Dynamic treatment effects, Elasticity

JEL Codes: I10, I18, D00, B23, H20

*Replication files are available on the author’s Github account. All errors are my own.
†Department of Economics, Towson University. email vshrestha@towson.edu
1 Introduction

Cigarette taxation is widely used as a policy instrument to reduce smoking as well as to increase revenue. To evaluate the efficacy of higher cigarette taxes (prices) on smoking outcomes, researchers have increasingly relied on the two-way fixed effect (TWFE) specifications.\(^1\) These studies use continuous measure of cigarette taxes (prices) and utilize within state variation (mostly increases) in cigarette taxes (prices) over the years in a multiple-treatment and multiple-control group framework to identify the target parameter.

More recent advancements in staggered difference-in-differences literature have highlighted concerns regarding the TWFE estimator (De Chaisemartin and d’Haultfoeuille (2020), Callaway and Sant’Anna (2021), Goodman-Bacon (2021)). The TWFE models pose restrictions of homogeneous effects across group and the length of exposure to the treatment. In absence of homogeneous effects, the TWFE estimators can generate effects that are biased towards zero. One of the main issues that has been highlighted is that of the negative weighting problem. If the average treatment effect varies with the length of exposure to the treatment, then early treated group forms a *bad comparison* for units treated later on in the sample period. The negative weighing problem is particularly dire if eventually a significant portion of units receive treatment. Since 38 states increased cigarette taxes at least once between 2004 and 2010, this issue can be highly pertinent when using TWFE models to evaluate the effects of cigarette taxes given that the effects are heterogenous.

This study revisits the literature evaluating the effects of cigarette taxes on smoking outcomes. I use balanced panel data from the Behavioral Risk Factor Surveillance System Selected Metropolitan/Micropolitan Area Risk Trends (BRFSS SMART) and utilize newer estimation techniques based on staggered difference-in-differences for multiple-group and multiple-time treatment framework that are less restrictive compared to the TWFE estimator to generate the average treatment effect on the treated estimates ($\hat{ATT}$). Specifically, by defining treatment as an indicator for cigarette tax increase (binary treatment), I first estimate the impacts of tax incidence using TWFE models and compare the estimates to $\hat{ATT}$ obtained from Callaway and Sant’Anna (2021).

(CS from hereon) estimator, which is based on the concept of group-time treatment effects and is robust to problems associated with the TWFE estimator. Additionally, I compare the TWFE estimates with the results from: 

i) canonical event-study design, ii) interaction-weighted estimator proposed by Sun and Abraham (2021) (SA from hereon), and iii) event-study type design following Callaway and Sant’Anna (2021). Throughout the study, separate analyses are conducted for years 2005-2010 as well as more recent period, 2015-2020. This allows inspection of cigarette taxes as a policy tool to curtail smoking in more recent years.

Although the findings from different approaches demonstrate effectiveness of tax incidence in improving smoking-related outcomes, the TWFE estimates are lower in magnitude compared to the overall $\hat{ATT}$ obtained from from CS estimator. This difference is considerably higher in the earlier sample period (2005-2010) for which the size of the TWFE estimate is only about 65% of the overall $\hat{ATT}$ from CS estimator. In other words, while the overall $\hat{ATT}$ point estimate from Callaway and Sant’Anna (2021) suggests that over a quarter of reduction in prevalence of current smoking between 2005-2010 is attributed to cigarette tax incidence, the TWFE estimate only accounts for 14% of the reduction. The results from decomposition following Goodman-Bacon (2021) suggest that the magnitude of TWFE estimate is suppressed towards zero since huge weight (31%) is placed on cases that use units treated early on in the sample as comparison for those treated later in the sample. The problem of bad comparison is not as dire in the case of later sample (2015-2020) since the majority of units remain untreated throughout the sample. Nevertheless, the overall $\hat{ATT}$ magnitude for 2015-2010 sample is only 67% of the $\hat{ATT}$ in 2004-2010 sample. In other words, the findings indicate that cigarette tax incidence has small effects that are imprecisely estimated in the recent sample period.

When estimating the dynamic effects of cigarette tax incidence, the findings from i) canonical event-study approach, ii) SA approach, and iii) event-study type method from CS all demonstrate gradual but increasing effects of tax incidence by the length of exposure to treatment. A simple comparison between the TWFE estimate and the post treatment event-study estimates show that the TWFE estimate only amounts to 60% of the average post-treatment estimates. Although the findings pertaining to the later period (2015-2020) are statistically insignificant, a break in trend
of estimates is witnessed following the treatment period. Moreover, the pre-treatment estimates in the event-study design pertaining to both samples and using different approaches fail to show any evidence of systematic differences in smoking outcomes between treated and untreated units prior to the treatment. This is consistent to the identification assumption used in the study.

The empirical methods used to evaluate the impacts of tobacco regulations can be broadly grouped into two segments: pre and post 2000 studies. Chaloupka and Warner (2000) provide a review based on pre-2000 studies and report that the price elasticity of cigarettes ranges from -0.2 to -0.35. However, many of the studies in the review use aggregated time series or cross-sectional data, which makes estimation more susceptible to omitted variable bias (Lewit and Coate (1982), Mullahy (1985), Jones (1989), Wasserman et al. (1991), Seldon and Boyd (1991), Sung, Hu, and Keeler (1994), Barnett, Keeler, and Hu (1995)). For instance, geographic units that increase cigarette taxes may have different socio-demographic characteristics as well as higher prevalence of anti-smoking sentiments even after controlling for income and other tobacco control policies, which may affect both passage of cigarette taxes as well as smoking outcomes.

The post-2000 studies to a large extent use individual level data and are methodologically based on TWFE models. These studies utilize some combination of spatial and temporal variation in cigarette taxes to identify the effects on smoking by controlling for both geographical unit (state) fixed effects and time fixed effects. While geographical unit fixed effects account for time invariant unobservables, time trends in the model absorb common trends over time across all geographical units. Given that there is sufficient within variation in cigarette taxes across units, the post-2000 studies should improve upon pre-2000 studies as unobservables such as anti-smoking sentiments are accounted by the unit fixed effects as long as the sentiments are time invariant during the span of the study. The majority of these studies find that encouraging trends in smoking outcomes can be attributed to increases in cigarette taxes (prices) (Adda and Cornaglia (2006), Tauras et

2The consensus provided by the Chaloupka and Warner (2000) study is that the range of elasticity estimates are about the same for both extensive (smoking participation) and intensive margin (conditional demand). The total elasticity ranges from -0.4 to -0.75.

The pre-2000 studies have limited ability to control for such omitted factors. Moreover, from an empirical standpoint these studies are conducted during the time period with relatively meager increases in cigarette taxes compared to the post-2000 period. Other additional problems with many of the pre-2000 include simultaneity between cigarette prices and aggregated cigarette demand and usage of annual state level tax receipts on cigarette sales as the dependent variable. We refer the reader to the Chaloupka and Warner (2000) study for more detail regarding pre-2000 studies.
This study makes three main contributions to the literature dedicated towards understanding the role of cigarette taxation in curtailing smoking. First, the study provides methodological advancements to the TWFE models previously used to estimate treatment effects. In a more recent comprehensive review discussing the role of tobacco regulations, DeCicca, Kenkel, and Lovenheim (2020) emphasize the importance of exploring methods that are robust to the problems associated with TWFE models. Specifically, this study allows for heterogeneous treatment effects and accounts for the potential problem of bad comparison group by using the Callaway and Sant’Anna (2021) estimator. Second, many studies based on TWFE models do not explicitly report suggestive evidence regarding the parallel trend assumption (DeCicca, Kenkel, and Lovenheim (2020)). Specifically, the parallel trend assumption states that in absence of the treatment the trajectory of the treated group would mimic that of the untreated group. This study conducts a detailed investigation of the parallel trend assumption using the canonical event-study approach as well as more robust methods based on Sun and Abraham (2021) and Callaway and Sant’Anna (2021). Third, the study informs policy by providing more recent estimates regarding the relationship between increases in cigarette taxation and smoking outcomes.

The study is organized as follows. Section 2 contains data and Section 3 discusses various methodologies used in this study. Section 4 provides discussion of results, while section 5 conducts additional analyses. Section 6 concludes the study.

\[\text{To our knowledge only Cotti et al. (2020) explores the parallel trend assumption. However, there are major differences between this study and the Cotti et al. (2020) study, whose main focus is to investigate whether cigarettes and e-cigarettes are economic substitutes.}\]
2 Data

2.1 BRFSS SMART

The primary data for smoking outcomes comes from the Behavioral Risk Factor Surveillance System (BRFSS) Selected Metropolitan/Micropolitan Area Risk Trends (SMART) for years 2004-2010 and 2015-2020. BRFSS data are typically used to construct state level estimates, but the SMART project was initiated to produce estimates for local areas defined as Metropolitan and Micropolitan Statistical Areas (MMSAs), delineated by the Office of Management and Budget (OMB). The respondents are associated with MMSAs by their county of residence as reported during the survey. The MMSAs are represented by the Core Based Statistical Area (CBSA) codes that falls within the Federal Information Processing Standards (FIPS).

The eligibility criteria to determine whether a particular MMSA is included in BRFSS SMART data depends on the number of observations in each weighting class. The weighting classes are based on age, race, and gender, which gives a total of 24 weighting classes. MMSAs with weighting classes comprising less than 19 observations are excluded from the SMART dataset. The SMART data allows comparison of prevalence estimates across MMSAs as the same weighting criteria are used for all MMSAs. Each MMSA includes at least 500 individuals.

The number of MMSAs vary across the BRFSS SMART survey years as MMSAs can enter or exit the survey. For example, the 2004 survey includes 134 MMSAs, while the 2010 survey includes 198 MMSAs. I focus on the status of current smoker as the main dependent variable and create a balanced panel of outcome variables collapsed at the MMSA-year level to assure that the findings are not driven due to differences in composition of MMSAs in the sample. The BRFSS SMART sample weights are using during this process. As such, the 2004-2010 and 2015-2020 balanced

---

5According to the Centers for Disease Control and Prevention, the MMSAs are defined as “a core area containing a substantial population nucleus, together with adjacent communities and all having a high degree of economic and social integration.”

6There were in total 944 MMSAs excluding Puerto Rico in 2010. The list of CBSA category (metropolitan vs. micropolitan), CBSA code, and CBSA title can be found on https://www.uspto.gov/web/offices/ac/ido/oeip/taf/clses/cbsa/cbsa_countyassoc.htm.

7Some states do not use race in post-stratification. For these states only age and gender are used to form weighing classes, which give 12 weighting classes.

8For this study, current smokers are defined as individuals who smoke cigarettes “every day” or “some days.”
panels include 108 and 95 MMSAs, respectively.

The MMSAs included in the balanced dataset for the BRFSS SMART survey years 2004-2010 and 2015-2020 are portrayed on the map using red color in panels A and B of Figure 1, while the green polygons represent MMSAs not covered by the survey. At least one MMSA is included in the balanced panel data for 46 states and there are more than two MMSAs in many states. The states that are not represented in BRFSS SMART (balanced panel) include Alaska, Hawai, North Dakota, and Rhode Island.

2.2 Cigarette taxes and tobacco control policies

The state level cigarette taxes are extracted from the consolidated files of Tax Burden of Tobacco for years 1970-2019 prepared by Orzechowski and Walker. This version of the file is obtained from the Centers for Disease Control and Prevention. Using the reported excise tax and implementation date for tax changes, I form a state-year panel data such that if the effective tax month is past July, the tax change is attributed to the following year.

Next, I create a binary variable that indicates whether a state increased cigarette tax in a given year. The indicator variable takes a value 1 in the year of tax increase and the state retains this value for the rest of the panel, while years prior to the tax change is indicated by value 0. Hence, the treatment takes a value of 0 and 1 similar to the canonical difference-in-differences framework and maintains a staggered design.

A handful of states experienced multiple tax increasements within the span of the survey. For instance, Pennsylvania increased state cigarette taxes in July 2004 and November 2009. As both episodes of tax changes fall within the survey years 2004-2010, I use the earliest episode of tax change to denote the treatment assignment. In other words, the coding of treatment assignment is dependent on the earliest tax change date in cases of states with multiple tax changes.

Additionally, I calculate the percentage of a state’s population living under a bar ban for the sample years 2004-2010 and 2015-2020 using local level information regarding smoking ban in bars.

---

9See https://chronicdata.cdc.gov/Policy/The-Tax-Burden-on-Tobacco-1970-2019/7nwe-3aj9

10Once the treatment indicator turns on, it is not switched off. In fact, none of the states during the sample periods used in the study reduced cigarette tax after increasing it.
from the American Nonsmokers’ Rights Foundation (ANRF). The ANRF data file documents the localities that implemented smoking ban in freestanding bars along with the date of implementation. I use the CBSA population estimates for years 2000 and 2010, obtained from the Census Bureau, to calculate the percentage of population living under the bar ban policy for survey years 2004-2010 and 2015-2020.

2.3 Other variables

Also, the study uses locality specific unemployment rate, the measure of anti-smoking sentiment in 1998-1999, change in the proportion of current smokers between 1998-1999 and 2001-2002 and the log of population as covariates. The locality specific unemployment rate is calculated using data from the Current Population Survey (CPS) - Merged Outgoing Rotation Group Earnings Data for years 2000 and 2010. The unemployment estimates for 2000 and 2010 are merged with survey years 2004-2010 and 2015-2020, respectively. The data for variables used to form the anti-smoking sentiment measure comes from the Current Population Survey (CPS) Supplements: Tobacco Use for years 1998-1999. The anti-smoking measure is constructed using a principal component analysis in spirit of DeCicca et al. (2008) but the measure is aggregated at the locality level instead of the state level. Not all of the MMSAs listed in the balanced BRFSS SMART data are contained in CPS tobacco supplement. In the spirit of using all observations, I use regression based imputation method to impute values for the missing observations. The change in the proportion of current smokers between 1998-1999 and 2001-2002 is constructed by using data from the CPS tobacco supplement. The sub-figures in the Appendix section, Figure 8, show the negative relationship between the measure of anti-smoking sentiment and smoking-related variables, i.e., the proportion of current smokers and those who ever smoked cigarettes. The sub-figures demonstrate a negative relationship between the intensity of anti-smoking measure and smoking related variables in 1998.

11 The list of local areas enacting smoking bans in public places (bars, restaurants, non-hospitality workplaces) are listed here.

12 A detailed discussion of the imputation can be found in the Appendix section 8.1.
3 Method

We first discuss the TWFE model and briefly summarize some problems associated with TWFE in context of cigarette taxation. Next, we thoroughly discuss the CS (Callaway and Sant’Anna (2021)) estimator. Also, we provide discussion for the Goodman-Bacon decomposition (Goodman-Bacon (2021)) and SA estimator (Sun and Abraham (2021)). Throughout the analysis we utilize the variation in increases in cigarette taxes between two periods: i) 2004-2010; and ii) 2015-2020.

3.1 Two way fixed effect (TWFE) model

The TWFE specification is given by:

\[ Y_{ist} = \alpha + \beta D_{st} + \sigma_i + \kappa_t + \eta BarBan_{ist} + \gamma X_i \times \kappa_t + \epsilon_{it} \]  

(1)

where, \( Y_{ist} \) represents smoking outcomes (i.e., percent of current smokers) collapsed at the locality-level (metropolitan or micropolitan area) \( i \) within state \( s \) at time \( t \). \( D_{st} \) is an indicator variable representing the treatment assignment of an increase in cigarette taxes. If a state experiences multiple increases in cigarette taxes within the sample period, the earliest increase in cigarette taxes is used to define the treatment.\(^\text{13}\) \( \sigma_i \) is the locality level fixed effects and captures the time invariant unobserved heterogeneity across localities, while \( \kappa_t \) absorbs common time trends in smoking outcomes.

The parsimonious treatment specification only includes the treatment indicator plus the area and time fixed effects. Additional specifications add the percent of a locality’s population living under the bar ban at time \( t \) denoted by \( BarBan_{ist} \) in equation 1. We defer from using time varying controls as much as possible due to the potential of post-treatment bias (see Rosenbaum (1984)). \( X_i \) is a vector of locality specific time invariant pre-treatment variables (the log of population, percent unemployed, a measure of anti-smoking sentiment in 1998-1999, change in the proportion

\(^\text{13}\)For example, if a state increased cigarette taxes in both 2004 and 2009, the treatment turns on in 2004 and remains on for the following years.
of current smokers between 1998-1999 and 2001-2002) interacted with the year fixed effects. However, controlling for covariates in a regression model as in equation 1 may not yield consistent estimates of ATT if treatment effects are heterogeneous across covariates, even when the conditional parallel trend assumption is satisfied; see Meyer (1995) and Roth et al. (2022) for discussion. The standard errors are clustered at the state level for all specifications.

Using the binary format of treatment is different from the majority of studies in the literature that use a continuous measure of cigarette taxes. This study uses the binary measure of treatment for two reasons. First, the binary measure complies with the staggered design of multiple-group and multiple-period difference-in-differences framework as discussed by the recent studies providing methodological advancements in the difference-in-differences literature (De Chaisemartin and d’Haultfoeuille (2020), Goodman-Bacon (2021), Callaway and Sant’Anna (2021)). Second, the binary treatment variable maps the analysis to the more standard treatment effect literature including the difference-in-differences approach.

Next, to allow for dynamic treatment we estimate a canonical event-study design. The event study specification is given as below.

\[
Y_{ist} = \alpha + \sum_{k=-K}^{L} \gamma_k D^k_{st} + \sigma_i + \kappa_t + \eta BarBan_{ist} + \gamma X_i \times \kappa_t + \epsilon_{it}
\]  

(2)

Here, \(D^k_{it} = 1(t - g_i = k)\), where \(g_i\) is the treatment (policy) year for unit \(i\) and \(t - g_i\) is the relative time around the policy year. The omitted categories include the year before the treatment and the earliest relative time, \(E\). This is to avoid problems of multicollinearity, which arises when all of the groups are eventually treated. In the case with no never-treated units, the path of a fully dynamic event-study cannot be point identified as treatment timing \(g_i\) (subsumed by the unit FE) and calendar time \(t\) can perfectly generate the relative time \(R_{it}\), where \(R_{it} = t - g_i\) (See Borusyak, 2014). The percent unemployed is measured in year 2000 and 2010 for BRFSS sample years 2005-2010 and 2015-2020, respectively. The anti-smoking variable is constructed similar to DeCicca et al. (2008) study and is measured at the locality level in 1998. To construct the proportion of change in current smokers between 1998 and 2001 (years prior to the treatment), I rely on the tobacco supplement files from Current Population Survey. The interaction of these pre-treatment variables with year indicators allow the effects of the pre-treatment variables to change over time.
Jaravel, and Spiess (2021) for an excellent discussion regarding this issue). Only a handful of states remain untreated in survey years 2004-2010. In other words, as the number of never treated groups are low, excluding the earliest period avoids concerns regarding multicollinearity.\footnote{Although the number of never treated group is quite large for the survey year 2015-2020, we still exclude the earliest time period for consistency as well as to comply with Sun and Abraham (2021) event study.}

### 3.2 Issues with TWFE Estimator in context of cigarette taxation

The main issue with the TWFE estimator originates from the model assumption that the treatment effects are homogeneous across units and by the length of exposure to the treatment. While TWFE estimate is a weighted average of all possible $2 \times 2$ DD estimators in the sample, the weights may be negative (De Chaisemartin and d’Haultfoeuille (2020), Goodman-Bacon (2021)). The negative weights can affect the TWFE estimate specifically when using the early treated units as comparison for units treated later in the sample.

To see this, we borrow explanation used in Roth et al. (2022). Using the Frisch-Lovell theorem express the TWFE estimator as:

$$\hat{\beta} = \sum_i \sum_t \frac{(D_{it} - \hat{D}_{it})(Y_{it})}{(D_{it} - \bar{D}_{it})^2}$$

where $\hat{D}_{it}$ is the fitted value from the regression $D_{it} = \tilde{\sigma}_i + \tilde{\kappa}_t + \tilde{\epsilon}_{it}$, where $\tilde{\sigma}_i$ and $\tilde{\kappa}_t$ are unit and time fixed effects. Here, $Y_{it} = Y_{it}(0) + \tau_{it}(g)$; $Y_{it}(0)$ is the potential outcome for group $g$ in absence of treatment.\footnote{$\tau_{it}(g)$ is the treatment effect for group $g$ at time $t$.} However, estimation of OLS can predict $\hat{D}_{it}$ (treatment probability) greater than 1, in which case $\tau_{it}$ gets negative weights in equation 3 ($D_{it} - \hat{D}_{it} < 0$). The negative weighting problem is more likely to affect early treated units late in the sample. This can be explained by rewriting $\hat{D}_{it}$ as: $\hat{D}_{it} = \bar{D}_i + \bar{D}_t - \bar{D}$, where, $\bar{D}_i$ is the unit specific mean $i$, $\bar{D}_t$ is the mean specific to time $t$, and $\bar{D}$ is the grand mean across time and individuals.

For the early treated units $\bar{D}_i \approx 1$ and if all units are eventually treated by the end period of the sample, $\bar{D}_t \approx 1$ for $t$ closer to the end of the sample. In this particular case, $\bar{D}_i + \bar{D}_t \approx 2$. Since, $\bar{D} < 1$ (there is a portion of untreated units or periods without treatment), $\hat{D}_{it} > 1$. This
puts negative weights on the TWFE’s decomposed 2 × 2 estimates. The negative weights are more likely to be severe towards the end of the sample when early treated units are compared to later treated units. While the negative weighting issue cancels out due to positive weights in other cases under homogeneous treatment effects, such is not the case when effects are heterogeneous.

The issue of negative weighting problem can also be seen through the lens of Goodman-Bacon (2021). Following his study, the TWFE estimate $\beta^{DD} = VWATT + VWCT - \Delta ATT$, where $VWATT$ is the variance weighted average treatment effect for the treated, $VWCT$ is the variance weighted common trends, and $\Delta ATT$ refers to the change in treatment for the early treated group over the sample period (say, later vs. early period). In case of a dynamic treatment effect, $\Delta ATT \neq 0$, which exerts a negative weight on $\beta^{DD}$.

How may this issue affect TWFE estimate in case of treatment determined by increases in cigarette taxes? Depending on the sample period selected by the researcher, eventually almost all the units may receive treatment. For instance, if the choice of sample period is 1998-2010, all states will have eventually received treatment by the end year of the sample. Between 2004-2010, 38 states increased cigarette taxes. For early treated units, $\bar{D}_i = 1$ and $\bar{D}_{t=2010} \approx 0.76$. Given that the timing of treatment varied across units, $\bar{D} < 0.76$. This means that $\hat{D}_{it}$ may be greater than 1 towards the end of the sample period, which exerts negative weight in equation 3. The negative weighting issue is less likely to be problematic when using the sample from 2015-2020 as only 18 states increased cigarette taxes during this period.

To gauge the intensity of bad comparison in TWFE setting when using early treated group in comparison to late treated group, I use the decomposition of TWFE estimate provided in Goodman-Bacon (2021) and report aggregate weights given to the following categories: i) treated units vs. untreated units, ii) treated units vs. always treated units, iii) early treated vs. late treated, and iv) late treated vs. early treated. Categories i) and iii) provide “clean” comparison, while the other two may be contaminated in cases of treatment heterogeneity.

The presence of heterogeneity not only raises questions regarding the validity of the TWFE estimate from a static model as shown in equation 1 but also affects the canonical event study estimates in equation 2. Sun and Abraham (2021) (SA) provide formal evidence to highlight problems
regarding the canonical event study as depicted in equation 2. The SA study highlights that even when assumptions of parallel trends and no anticipation effect are invoked, under heterogeneous treatment effects across cohorts (defined by the treatment timing), the event-study coefficient for a specific relative time \( l \) will be contaminated due to comparisons across other relative time periods \( (l \neq l') \) and the excluded period. As such, the pre-treatment event-study indicators can capture treatment effects from the post-treatment periods, which would contaminate the pre-treatment estimates as to validly assess whether parallel trends hold. Sun and Abraham (2021) propose an interaction-weighted estimator. First, the relative time indicators are interacted with group indicators (defined by treatment timing) using a TWFE specification to estimate dynamic average treatment effect per group.\(^{17}\) Next, the group-time estimates are aggregated using weights defined as the sample share of each group at the given relative time.

### 3.3 Callaway and Sant’Anna Estimator - An alternative to TWFE

This study uses Callaway and Sant’Anna (CS) estimator following Callaway and Sant’Anna (2021), which allows for heterogeneity across units and time. Thus, CS estimator is not susceptible to issues associated with the TWFE model. The CS estimator identifies group-time treatment effect parameter for units that are first treated at time \( g \) (hence, group \( g \)) and estimated in calendar time \( t \). The target estimand is:

\[
ATT(g, t) = E(Y_t(g) - Y_t(0)|G_g = 1)
\]

where, \( Y_t(g) \) is the outcome of group \( g \) at time \( t \), \( Y_t(0) \) is the counterfactual, and \( G \) is a set of all possible treatment timing groups.\(^{18}\) Under the following assumptions: i) unconditional parallel trend assumption which states that outcome of units treated at time \( g \) would have followed the same path as the untreated units in absence of the treatment, \( i.e., E(Y_t(0) - Y_{t-1}(0)|G_g = 1) = \)

\(^{17}\)To define units based on their treatment time, SA use the terminology “cohort”, while CS use “group.” SA’s approach is similar to CS’s approach when no additional covariates are included in the model i.e. when unconditional parallel trend assumption is invoked.

\(^{18}\)Note that this is just an extension of the \( 2 \times 2 \) DiD target estimand: \( ATT = E(Y_2(D = 1) - Y_2(D = 0)|D = 1) \), where \( D \in \{0, 1\} \) refers to the treatment status.
$E(Y_t(0) - Y_{g-1}(0)| C = 1)$, and $ii$) no anticipation assumption, i.e., $E(Y_t(g)| G_g = 1) = E(Y_t(0)| G_g = 1)$ for $t < g$, the ATT estimator is:

$$\hat{ATT}(g, t) = \frac{\sum_i (Y_{i,t} \cdot 1(G_i = g) - Y_{i,g-1} \cdot 1(G_i = g))}{\sum_i 1(G_i = g)} - \frac{\sum_i Y_{i,t} \cdot 1(G_i = C) - Y_{i,g-1} \cdot 1(G_i = C)}{\sum_i 1(G_i = C)}$$

(5)

where, $1(G_i = g)$ is a binary variable indicating whether a unit $i$ is first treated in period $g$. Similarly, $1(G_i = C)$ is an indicator for untreated group. The equation above implicitly assumes no treatment anticipation, which allows using a period prior to the reform ($g - 1$) as the reference. Additionally, following the unconditional parallel trend assumption, the estimand assumes that the path of outcome for group $g$ would have evolved similar to comparison group ($G_i = C$) from period $g - 1$ to $t$. The choice of comparison group as shown in equation 5 particularly refers to the never-treated units. Similarly, not-yet-treated units can be used as the comparison group under the parallel trend and no anticipation assumptions.

However, the unconditional parallel trend assumption is unlikely to hold when evaluating smoking outcomes as states that are more likely to increase cigarette taxes may also have relatively high anti-smoking sentiments. Hence, it is likely that smoking outcomes would have evolved differently across states with different levels of smoking sentiments, violating the unconditional parallel trend assumption. In this case, conditional parallel trend assumption may be more plausible given that one observes pre-treatment smoking sentiments across states.

An important aspect of CS method is that covariates can be flexibly incorporated in the analysis. This allows estimating conditional difference-in-differences which assumes that parallel trends hold between treated and untreated units with the same covariates. One can use inverse probability weighting (IPW) (Abadie (2005)), outcome regression (OR) (Heckman, Ichimura, and Todd (1997)), or doubly robust (DR) (Sant’Anna and Zhao (2020)) methods to recover the DiD parameters, while flexibly accounting for the pre-treatment covariates. Although all three approaches recover the same parameter under the conditional parallel trend assumption from an identification standpoint, this study relies on DR approach due to its additional robustness property. The pre-treatment locality specific covariates that are accounted for include a measure of anti-smoking
sentiment in 1998, unemployment rate (in 2000 and 2010 for the 2004-2010 and 2015-2020 samples, respectively), the change in proportion of current smokers between 1998-1999 and 2001-2002, and the log of area specific population. As discussed in detail in Sant’Anna and Zhao (2020), DR estimator is preferred from an empirical perspective (compared to other two methods) as it enjoys an additional robustness property that it only requires either propensity score or outcome regression specification to be correct for valid estimates.

The estimation of DR approach can be grouped into two steps. First it estimates the probability that a unit falls in group \( g \), denoted as \( \hat{p}_g(x) \) as well as models the outcome evolution using the approach suggested in Heckman, Ichimura, and Todd (1997). The second step uses \( \hat{p}_g(x) \) as the weights and also adjusts the predicted changes in the outcome for the treated group in absence of treatment. The latter is obtained by first estimating the conditional expectation for the untreated units using their covariates \( X_i \); the estimates from this regression are used to predict outcomes for the treated group by using \( X_i \) pertaining to the treatment group given as \( \hat{m}_{g,t}(X_i) = \hat{E}[Y_{i,t} - Y_{i,g-1} - D_i = 0, X_i] \). As such this method incorporates both IPW and OR methods. The DR estimator is given as:

\[
A\hat{TT}(g, t) = \frac{1}{N} \sum_i \left[ \left( \frac{1}{N} \sum_i 1(G_i = g) \frac{\hat{p}_g(X)}{1 - \hat{p}_g(X)} \right) - \frac{1}{N} \sum_i 1(G_i = C) \frac{\hat{p}_g(X)}{1 - \hat{p}_g(X)} \right] (Y_{i,t} - Y_{i,g-1} - \hat{m}_{g,t}(X)) \]  

Here, the units in comparison group are weighted inversely according to their probability of being untreated. This is given by weights being proportional to \( \frac{\hat{p}_g(X)}{1 - \hat{p}_g(X)} \) for untreated units. Intuitively, units in untreated group that have low probability of getting treated can be very different from the units in treatment group; hence, these units receive lower weights.

Following the recovery of disaggregated parameters, \( A\hat{TT}(g, t) \), these can be aggregated to summarize the treatment heterogeneity by using appropriate weights. One aggregation of interest for this study is to evaluate how the treatment effect varies with the length of exposure to the treatment. This portrays dynamic treatment effects similar to the canonical event-study approach from equation 2 but by relaxing the assumption that effects are homogeneous across units. To evaluate how the treatment effect varies with the length of exposure to the treatment, an aggregation
The approach given below is adapted.

\[ \theta_{es}(e) = \sum_{g \in G} w_{es}(g, t) \text{ATT}(g, g + e) \]  

(7)

where the weight, \( w_{es}(g, t) = 1(g + e \leq T).1(t - g = e).P(G = g|G + e \leq T) \). Here, \( T \) is given as the number of periods in the sample and \( e \) is the relative time period (number of periods before or following the treatment period, i.e., \( e = t - g \)). In other words, weights are just the proportion of units treated in time \( g \) when measured at the relative time \( e \). \( \theta_{es}(e) \) documents the effect of the treatment by length of exposure to the treatment, which is contextually similar to the event-study parameters from TWFE in equation 1. The standard errors are obtained using a multiplier bootstrap approach that is robust to multiple-testing problems unlike the pointwise standard error.

### 4 Results

As shown in Table 1, thirty eights states increased cigarette taxes at least once between 2004-2010, while 18 states increased taxes between 2015-2020. The table tends to suggest that cigarette tax hike (a binary treatment measure), is not systematically confined to a specific geographic region (atleast in the sample periods of the study).

Table 2 provides descriptive statistics of some important variables separated by states that increased cigarette taxes versus states that did not. The first three variables portray summary of smoking behaviors among individuals living in metropolitan/micropolitan areas (MMSAs) included in the BRFSS SMART sample. On average smoking-related variables between units exposed to tax increases versus those not exposed to tax increases are fairly similar across both sample periods. For example, around 20 percent of the 2004-2010 sample are current smokers and the magnitudes are similar across areas with and without tax increases. However, the state-specific per capita cigarette sales in the pre-treatment periods are higher on average among units exposed to tax changes in 2004-2010 sample. For example, per capita cigarette sales in 1990 amount to 107 and
100 packs in states encompassing treated vs. untreated units, respectively. But this difference vanishes when focusing on per capita cigarette sales in 2010.

Although the proportion of current smokers decreased by about 5 percentage points between 2004-2010 and 2015-2020 sample, the magnitudes are fairly similar across units with and without tax changes. While the average nominal cigarette tax is about 20 cents higher among treated units in 2004-2010 sample, the average tax amount are similar across treated and untreated units in the later sample. Also, states with tax changes on average have higher magnitude of anti-smoking measure and cover greater percentage of people living under the bar ban policy in the earlier sample period.

Table 1 also shows heterogeneity in levels of tax increases. The average tax increase for 2004 group (those increased cigarette tax in 2004) was only 24 cents, whereas the highest tax increase in 2004-2010 sample is close to a dollar for 2008 group. In the later sample, California increased cigarette taxes by $2 in 2017. Using a binary measure of treatment to identify increases in cigarette taxes (but not the levels) comes with an obvious limitation that it does not account for heterogeneity in the size of tax increments.

The identification when using a continuous measure of cigarette taxes depends on not just the parallel trends in outcome between the treated and untreated groups but also parallel trends across units receiving various dosage of treatments. For example, consider the case when unit A receives a dose of $d$ and unit B receives a dose of $d'$, where $d > d'$. In this case, the validity of identification depends on the assumption that unit A’s outcome would have evolved similarly to unit B had unit A received dosage $d'$ instead of $d$. This assumption is stronger than the parallel trend assumption based on the binary treatment measure as the potential issue of selection when using a continuous treatment measure is not just in terms of who gets treated but also the levels of treatment.

Table 3 shows that there are large differences in important pre-treatment variables across different intensity of cigarette tax increases (dosage) among units experiencing tax increases between

\[19\text{Comparison of outcome changes between units receiving the treatment of dose } d \text{ with untreated units gives the average treatment effect of dose } d. \text{ Comparison of outcomes changes between units receiving two different dosage can be considered as causal responses. See Callaway, Goodman-Bacon, and Sant’Anna (2021) for detailed discussions.}\]
2004-2010. For example, levels of tax increases are negatively correlated with pre-treatment per capita cigarette sales. Moreover, units with higher levels of anti-smoking sentiment measure are more likely to implement larger tax hikes. Hence, units facing higher vs. lower levels of taxes between 2004-2010 tend to be very different in pre-treatment variables that may also have systematically affected the evolution of smoking outcomes even in absence of tax increases. One benefit of using binary treatment in the case of cigarette taxes is that it reduces average differences in pre-treatment variables between the treated versus comparison units.

4.1 TWFE Results

The results from TWFE models are presented in Table 4 when using the percent of current smokers as the dependent variable. Columns 1-3 and 4-6 refer to 2004-2010 and 2015-2020 samples; Panel B drops units that received treatment in the very first period of the sample (always treated) as they do not contribute to the identification in TWFE models. Columns 1 and 4 provide results from the parsimonious specification that only includes a dummy for cigarette tax change along with the year and MMSA (metropolitan micropolitan statistical area) fixed effects. Columns 2 and 5 also include the percent of population living under smoking ban in bars, while Columns 3 and 6 additionally include a vector of pre-treatment characteristics including locality specific log of population, unemployment in 2000 (for 2000-2010 sample) or 2010 (2015-2020 sample), a measure of anti-smoking sentiment in 1998, and the proportion change in current smokers between 1998 and 2001. These time invariant covariates are interacted with the year dummies as shown in equation 1.

Table 4 shows that the coefficients on the treatment variable are negative across all specifications and both panels, suggesting that cigarette tax increases are negatively associated with prevalence of smoking. Column 1 suggests that an instance of cigarette tax increase on average is associated with a reduction in prevalence of current smoker by 0.6 percentage points and the coefficient is statistically significant at the 5 percent level. Next, including the percent of population living under smoking ban in bar (Column 2) and additional covariates (Column 3) do not affect the magnitude on the treatment indicator. The coefficients on the treatment indicator presented in Panel B
are similar to Panel A. The proportion of current smoker decreased by 4.25 percentage points between 2004 to 2010. If we were to allow for causal interpretation on the treatment depending on the TWFE estimate with the largest magnitude, it would suggest that on average 14 percent of reduction in current smoker can be attributed to instances of tax increase between 2004-2010. What is evident is that tax increases in relatively recent years (between 2015 and 2020) are also negatively correlated with the proportion of current smokers as shown in Columns 4-6. However, the coefficients on the treatment indicator are imprecisely estimated as well as the magnitude of the estimates are relatively smaller to the early period (2004-2010) across all specifications.

There are two main issues with the TWFE estimates presented in Table 4. First, although specifications in Columns 2, 3, 5 and 6 invoke conditional parallel trend assumption, the time invariant pre-treatment covariates enter the model specification through interaction with the year fixed effects. Given that the treatment effects vary within different values of covariates, introducing covariates linearly may not be appropriate (see Meyer (1995)). Second, issues with the TWFE comes from negative weighting problem as discussed in section 4.2.

Table 5 presents results from decomposition following Goodman-Bacon (2021) to understand how the early treated group might affect the TWFE estimates by forming bad comparison. The table summarizes the weights given for all possible $2 \times 2$ difference-in-differences estimates into four categories: 

i) early treated group (as the treated group) vs. later treated group (comparison), 

ii) later treated vs. always treated, 

iii) later vs. early treated, and 

iv) treated vs. untreated. As previously mentioned, since the effects of treatment may vary with the length of exposure, the problem with TWFE arises when comparison of early treated group forms “bad comparison” for units treated later on in the sample period. This group (category iii) contributes 32% of weight to form the TWFE estimate for the 2004-2010 sample. While it carries the highest of weight, the magnitude of estimate pertaining to this group is the lowest among the four groups. This may be explained by a large number of states (38 states) eventually being treated in the earlier sample, with a significant portion of states receiving treatment in the later years of the sample period. However, the weight given for later vs. earlier treated group (category iii) is only 4.7% for the 2015-2020 sample.
Using already treated group as comparison group is problematic in presence of treatment dynamics. In the upcoming section, we evaluate the effects on tax increases by the length of exposure using the canonical event study approach, method proposed in Sun and Abraham (2021), and CS estimator.

4.2 Results allowing for heterogeneity in treatment effects

To allow for dynamic treatment effects by the length of exposure to treatment, I first estimate the canonical event-study specification given by equation 2. In addition, I also report estimates proposed in Sun and Abraham (2021) to correct for possible contamination that might affect the canonical event-study estimates. The results from canonical event-study as well as SA design are shown in Figures 2 and 3 for sample periods 2004-2010 and 2015-2020. Panels A and B refer to results from the parsimonious specification and specification with additional covariates as controls, respectively.

The findings from both canonical event-study and SA designs presented in Panel A, Figure 2, show reductions in prevalence of being a current smoker immediately starting from the period of cigarette tax implementation. The effects are double the size in magnitude a year following the tax implementation after which the magnitudes decrease gradually in size. The average of event-study estimates in the post period shown by the horizontal black dotted line using the canonical design is slightly below -1, whereas the red dotted line refers to the TWFE estimate shown in Table 4. The TWFE estimate is around 60% as large as the average of event-study estimates in the post period. While estimates from the event-study design (that allows for heterogeneity over time) suggest that 23% of reduction in current smoker between 2004-2010 is attributed to tax increases, the TWFE estimate accounts for only 14% of the reduction. Moreover, the pre-treatment estimates from both canonical as well as SA design are statistically indistinguishable from zero, which supports the parallel trend assumption governing the difference-in-differences framework. The dynamic results presented after controlling for covariates in Panel B, Figure 2, show similar patterns as observed in Panel A.

The dynamic results referring to the later sample (2015-2020), shown in Figure 3, also demon-
strate heterogeneity over time. The coefficients gradually fall in magnitude following the year of tax implementation and the size of the coefficients are below -1 in the third and fourth relative years. This pattern is similar across both panels representing specifications with and without additional covariates. The TWFE estimates shown by the red dotted line is only about half the size in magnitude compared to the average of post treatment event-study estimates from the canonical design.

Putting the TWFE results in perspective to the findings from Goodman Bacon decomposition shown in Table 5 provide some insights as to why the TWFE estimates may be biased towards zero. When early treated units are being compared to later treated units in 2004-2010 sample period, 53.4% of the weight comes from the timing variation in treatment and relatively high weight (31.6%) is provided to the bad comparison case (when early treated units are compared to later treated units). The estimate for the bad comparison case is the lowest among four cases (-0.296). Although the problem of bad comparison group is less intense (almost non-existent) in 2015-2020 sample as a substantial portion of weight (79.4%) is placed when never-treated units are used as comparison, TWFE estimates do not account for heterogeneity over time as captured by the event study estimates.

The difference in weights for each group defined by the treatment timing when serving as treated vs control units are shown in Figure 4 using triangular markers along with the sample share for each group in solid circles. I replicate the exercise after dropping always- and never treated units to focus only on variation due to treatment timing. Units receiving treatment in 2005 and 2010 contribute disproportionately more as control units rather than treated units compared to units treated towards the middle of the sample period. In fact, units treated in 2005 serve more as control units than treated while forming the TWFE estimate. This further highlights the problem of early treated group acting as comparison for units treated later in the sample (“bad comparison group”), which will suppress the TWFE estimate towards zero in case of treatment heterogeneity by time. In the 2015-2020 sample, units treated in 2018 serve disproportionately more as treated than control units compared to any other groups. In the case of heterogenous treatment effects across units, if the 2018 group have treatment effect that is lower in magnitude compared to
other groups, then the TWFE estimate will be suppressed towards zero. This may explain why the TWFE estimate from 2015-2020 sample is about half the size in magnitude compared to the average of post treatment event-study estimates as shown in Figure 3.

Next, Figures 5 and 6 show the group-time treatment estimates aggregated to show the average treatment effects on the treated \( \hat{ATT} \)s by the length of exposure to the treatment using the aggregation scheme shown in equation 7. Figures 5 and 6 refer to years 2004-2010 and 2015-2020, respectively. Panels A and B use never treated group as the comparison group, whereas Panels C and D use not-yet-treated groups. Panels A and C show the results without additional covariates, while Panels B and C flexibly control for the pre-treatment covariates (the log of local area population, unemployment rate, a measure of anti-smoking sentiment in 1998, and the percent change in current smokers between 1998-2001) using the Doubly Robust approach. All sub-figures report the 95% simultaneous confidence band obtained from the bootstrapped standard errors.

The estimates in Figure 5, Panel A, show that the percentage of current smokers decrease following the treatment implementation. The figure does not provide any evidence of differential pre-treatment trend. The pre-treatment estimates fluctuate around zero and are statistically insignificant. The first year after the tax increase, although negative, the effects are close to zero and statistically insignificant. But the effects are pronounced starting from the second year of the treatment, where the magnitude of \( \hat{ATT} \) is below 1 and the effect is statistically significant at the 5 percent level. There is a gradual decline in the effect size as the length of treatment exposure increases. The \( \hat{ATT} \) is close to -2 following 5 year of exposure to the tax change. These \( \hat{ATT} \) effects are quite similar to the pattern as well as the size of the event-study estimates shown in Figure 2. Moreover, the dynamic effects of the treatment are similar across all panels with: i) additional covariates when using never-treated as comparison (Panel B), ii) not-yet-treated group as comparison instead of only never treated units (Panel C), and iii) accounting for covariates when using not-yet-treated group as comparison (Panel D).

Panel A, Figure 6, shows that tax increase in the later period (2015-2020) reduce the prevalence of current smoker. Consistent with the identification assumption, we find no evidence of systematic differences in trends prior to the treatment. Following the treatment, the size of \( \hat{ATT} \)
gradually increases in magnitude and the effect is around -1.5 after 5 years of exposure to treatment. Although a clear break in trend is visible in the post-treatment era, none of the $\hat{ATT}$s are statistically significant at the 5% level. This may be attributed to only a handful of states increasing cigarette taxes during 2015-2020 compared to the earlier period.

The overall summary of group-time average treatment effects are provided in Figure 7 as point estimates. Panel A refers to 2005-2010 period and Panel B presents results pertaining to 2015-2020 sample period. The 95% confidence band are plotted around the overall treatment effect. The first overall $\hat{ATT}$ (01. nt(no controls)) is generated from using never treated units as the comparison group and without additional covariates. The second $\hat{ATT}$ is generated from aggregation of group-time effects that are obtained by using DR approach with covariates. The third (03. nyt(no controls)) and fourth (04. nyt(controls)) ATT estimates are obtained similarly to the first and second estimates, but by using not-yet-treated units as comparison instead of only never treated units. The TWFE estimate with the largest magnitude from Table 4 are show by the horizontal line for reference.

Panel A shows that the overall $\hat{ATT}$ across four different estimation approaches are very similar in magnitude. More importantly, the magnitude of overall $\hat{ATT}$ is more than two times the size of the TWFE estimate, shown by the red dotted horizontal line. Specifically, the TWFE estimate suggests that close to 14% of reduction in prevalence of current smokers between 2004-2010 can be attributed to cigarette taxes, while the point estimate from group-time treatment effect approach suggests that more than a quarter of the reduction is due to increases in cigarette tax. This clearly demonstrates the difference in the treatment effect intensity generated from TWFE vs. group-time treatment approach when using 2004-2010 sample.

The magnitude of overall $\hat{ATT}$s in the 2015-2020 sample, as shown in Panel B, are similar across different approaches but are only about 65 percent of the size of $\hat{ATT}$ demonstrated in Panel A. In other words, the findings indicate that the efficacy of cigarette taxation as a tool to improve smoking outcomes has lowered in recent periods. However, the overall point estimates from the group-time treatment effects are still larger than the TWFE estimate. It has to be noted that none of the overall $\hat{ATT}$s in Panel B are statistically significant at the 5 percent level, although

---

20The prevalence of being a current smoker decreased by 4.25 percentage points between 2004 to 2010.
by small margins.

5 Robustness check

One issue when estimating the dynamic treatment effects by the length of exposure to the treatment is that the composition of units across relative time bins changes. For example, ATT for units treated in 2009 and 2010 will not be identified for relative time periods greater than 1 and 0 in 2004-2010 sample period, respectively. To address the issue of changes in composition of groups included in the relative period bins, we re-estimate CS approach by including only the units balanced in relative time period. Specifically, we include units that are atleast treated for 3 and 2 additional periods following the treatment year for 2004-2010 and 2015-2020 samples. The results from this exercise are shown in figures 9 and 10. The dynamic estimates as well as overall $\hat{ATT}$ for the sample period 2004-2010, shown in Figure 9, are similar to the main results (figures 5 and 7), suggesting that the results are not driven by compositional changes of units in relative time. For 2015-2010 sample, although the estimates increase in magnitude following the treatment year, they are statistically indistinguishable from zero. This again shows that the impacts of cigarette taxation decreased in more recent years.

6 Conclusion

This study re-evaluates the role of cigarette taxes in curtailing smoking by using data from the Behavioral Risk Factor Surveillance System (BRFSS) Selected Metropolitan/Micropolitan Area Risk Trends (SMART) for two sample periods: i) 2004-2010, and ii) 2015-2020. Although the topic of whether cigarette taxes can be used as an effective policy instrument to reduce smoking has been widely addressed, many of the studies that have come out in the past decades rely on the two-way-fixed (TWFE) approach where the treatment is rolled out in a multiple-group and multiple-timing setting. While the TWFE approach reveals the average treatment effect on the treated in absence of heterogeneity across the treated units and length of exposure to the treatment given that the parallel trend assumption is satisfied, such is not the case when the
treatment effects are heterogeneous. This study incorporates theoretical advances that have been made in the multiple-group and multiple-timing difference-in-differences framework to comment on the average treatment effect of cigarette tax incidence by comparing the newer estimates with the TWFE estimate.

The results show that the largest TWFE estimate pertaining to 2004-2010 sample is less than half the size in magnitude to the overall average treatment effect on the treated obtained by using the group-time treatment effect ($\hat{ATT}$) approach following Callaway and Sant’Anna (2021) estimator. While the TWFE point estimate suggests that 14% of the reduction in the prevalence of being a current smoker between 2004 and 2010 is explained by tax incidence, the overall group-time estimate shows that tax incidence contributes to over 25% of the reduction. Since many units were eventually treated by the end of sample period, Goodman-Bacon (2021) decomposition reveals that 31% of the weight in forming the TWFE point estimate comes from the comparison between later treated units (as treated) vs. those treated earlier in the sample (as untreated). This is precisely the comparison one would like to avoid. The canonical event-study design and the design proposed by Sun and Abraham (2021) yield dynamic estimates similar to those obtained by following Callaway and Sant’Anna (2021), indicating that the magnitude of average treatment effect on the treated increases with the length of exposure to the treatment.

Only a handful of states were treated by the end of the later sample period (2015-2020). Hence, the issue of bad comparison group is not as critical in the later period as it is in the earlier sample. However, the overall treatment effects from the group-time setting are still larger than the TWFE estimate. The comparison of $\hat{ATT}$s across two sample periods reveal that the the overall point estimate of $\hat{ATT}$ in the later sample period is only 65% of the size of $\hat{ATT}$ pertaining to the earlier sample period. This suggests that the efficacy of cigarette taxes in curtailing smoking may have decreased over the years. It has to be cautioned that such claims come with caveats as the treatment groups are different across two periods.

Several limitations are worth mentioning. First, the study uses binary form of treatment rather than different treatment intensities determined by the dosage of tax increases. While uncovering $\hat{ATT}$ based on treatment defined by dosage relies on stronger assumptions than the case when the
treatment is binary, the use of binary treatment does not allow us to evaluate the heterogeneous effects of the magnitude of tax increase. It is worth mentioning that this issue still prevails when using TWFE with continuous tax values. Second, using the BFRSS SMART data excludes some states from the analysis and only includes individuals residing in metropolitan or micropolitan areas. This may create problems of generalizing the findings to the entire population in America. Nevertheless, the results from this study can provide useful guidance to researchers as well as policy makers.
Figure 1: MMSAs in BRFSS SMART
Note: The figures show the map of Metropolitan/Micropolitan core-based statistical areas (CBSA), where the green areas are MMSAs represented in balanced panel of BRFSS SMART. There are 108 and 95 MMSAs for sample years 2005-2010 and 2015-2020.
Figure 2: Event Study Estimates (2005-2010)
Note: The figures show findings from the event-study analysis based on TWFE and SA (2020). Panel A uses the parsimonious specification, while Panel B includes controls for the percent of state population living under smoking ban in bars along with the pre-treatment variables of the locality (i. log of population measured in 2000, ii. the change in current smokers between 1998 and 2001, iii. a measure for anti-smoking sentiments in 1998, iv. unemployment rate) interacted with year dummies. The vertical bars represent the 95 percent confidence intervals constructed from standard errors clustered at the state level. The red dotted line represents the largest TWFE estimate from Table 4, while the black line is the average of post-treatment event-study estimates from the canonical model.
Figure 3: Event Study Estimates (2015-2020)
Note: The figures show findings from the event-study analysis based on TWFE and SA (2020). Panel A uses the parsimonious specification, while Panel B includes controls for the percent of state population living under smoking ban in bars along with the pre-treatment variables of the locality (i.e., log of population measured in 2010, ii. the change in current smokers between 1998 and 2001, iii. a measure for anti-smoking sentiments in 1998, iv. unemployment rate) interacted with year dummies. The vertical bars represent the 95 percent confidence intervals constructed from standard errors clustered at the state level. The red dotted line represents the largest TWFE estimate from Table 4, while the black line is the average of post-treatment event-study estimates from the canonical model.
Figure 4: Treated/Untreated Weight Difference
The figure shows the difference in weight when each group based on the treatment timing acts as treated vs. comparison group following the decomposition shown in Goodman-Bacon (2021). The square markers represent results after excluding always treated as well as never treated groups and focus solely on variation due to treatment timing. The sample size for each group are shown in solid circles.
Figure 5: CS Event-Study-Type Estimates (never treated and not-yet-treated as comparison)

Note: The figure shows the average treatment on the treated estimates by the length of exposure to the treatment using CS estimator and 2004-2010 BRFSS SMART sample. Panels A and B use only the never-treated units as comparison, while Panels C and D use the not-yet-treated-units. Panels A and C present results without the covariates when unconditional parallel trend assumption is invoked. The estimates in panels B and D use the DR estimator and include locality specific covariates: i) the log of population in 2000, ii) unemployment rate in 2000, iii) the change in proportion of current smokers between 1998 and 2001, and iv) a measure of anti-smoking sentiment in 1998. The vertical bars represent confidence intervals constructed using standard errors from multiplier bootstrapped procedure to account for multiple hypothesis testing.
Figure 6: CS Event-Study-Type Estimates (never treated and not-yet-treated as comparison)
The figure shows the average treatment on the treated estimates by the length of exposure to the
treatment using CS estimator and 2015-2020 BRFSS SMART sample. Panels A and B use only
the never-treated units as comparison, while Panels C and D use the not-yet-treated-units. Panels
A and C present results without the covariates when unconditional parallel trend assumption is
invoked. The estimates in panels B and D use the DR estimator and include locality specific
covariates: i) the log of population in 2000, ii) unemployment rate in 2000, iii) the change in
proportion of current smokers between 1998 and 2001, and iv) a measure of anti-smoking sentiment
in 1998. The vertical bars represent confidence intervals constructed using standard errors from
multiplier bootstrapped procedure to account for multiple hypothesis testing.
The figures show the overall average treatment effect on the treated using CS approach for sample years 2004-2010 in Panel A and 2015-2020 in Panel B. Different versions of CS estimator are considered. Approach 01 uses never treated units as the comparison group without covariates. Approach 02 adds covariates and still uses never-treated units for comparison. Approaches 03 and 04 follow 01 and 02 but use not-yet-treated units to form the comparison groups. The locality-specific covariates included in 02 and 04 are: i) unemployment rate, ii) the log of population (in 2000 for panel A and 2010 for panel B), iii) a measure of anti-smoking sentiments in 1998, and iv) the change in the proportion of current smokers between 1998 and 2001.
<table>
<thead>
<tr>
<th>Year</th>
<th>States</th>
<th>Count of MMSAs</th>
<th>Average tax increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>AL, HW, MI, NJ, PA, RI, VA</td>
<td>108</td>
<td>0.26</td>
</tr>
<tr>
<td>2005</td>
<td>AK, CO, KY, ME, MN, MT, NC, NH, OH, OK, WA</td>
<td>108</td>
<td>0.49</td>
</tr>
<tr>
<td>2006</td>
<td>AZ, IA, VT</td>
<td>108</td>
<td>0.67</td>
</tr>
<tr>
<td>2007</td>
<td>CT, DE, IN, SD, TN, TX</td>
<td>108</td>
<td>0.75</td>
</tr>
<tr>
<td>2008</td>
<td>DC, MA, MD, NY, WI</td>
<td>108</td>
<td>0.97</td>
</tr>
<tr>
<td>2009</td>
<td>AR, FL, MS</td>
<td>108</td>
<td>0.74</td>
</tr>
<tr>
<td>2010</td>
<td>NM, SC, UT</td>
<td>108</td>
<td>0.75</td>
</tr>
<tr>
<td>2015</td>
<td>DC, KS, LA, NV, OH, RI, VT</td>
<td>95</td>
<td>0.53</td>
</tr>
<tr>
<td>2016</td>
<td>AL, CT, PA, WV</td>
<td>95</td>
<td>0.51</td>
</tr>
<tr>
<td>2017</td>
<td>CA</td>
<td>95</td>
<td>2</td>
</tr>
<tr>
<td>2018</td>
<td>DE, KY, OK</td>
<td>95</td>
<td>0.75</td>
</tr>
<tr>
<td>2019</td>
<td>IL, NM</td>
<td>95</td>
<td>0.78</td>
</tr>
<tr>
<td>2020</td>
<td>VA</td>
<td>95</td>
<td>0.3</td>
</tr>
</tbody>
</table>

States implementing multiple tax changes within the sample period are reported by their earliest year of change.
<table>
<thead>
<tr>
<th>variables</th>
<th>tax change (04-10)</th>
<th>no change (04-10)</th>
<th>tax change (15-20)</th>
<th>no change (15-20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% ever smoked</td>
<td>0.45</td>
<td>0.44</td>
<td>0.42</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.04)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>% current smoker</td>
<td>19.57</td>
<td>21.03</td>
<td>14.74</td>
<td>13.51</td>
</tr>
<tr>
<td></td>
<td>(4.04)</td>
<td>(4.53)</td>
<td>(2.81)</td>
<td>(2.89)</td>
</tr>
<tr>
<td>% stop smoking</td>
<td>0.59</td>
<td>0.57</td>
<td>0.59</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>% current smoker (1998)</td>
<td>0.22</td>
<td>0.21</td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>nominal cig tax</td>
<td>1.36</td>
<td>0.71</td>
<td>1.9</td>
<td>1.79</td>
</tr>
<tr>
<td></td>
<td>(0.78)</td>
<td>(0.39)</td>
<td>(1.01)</td>
<td>(1.16)</td>
</tr>
<tr>
<td>population 1990</td>
<td>5662925.39</td>
<td>5101805.57</td>
<td>5930206.89</td>
<td>6420580.91</td>
</tr>
<tr>
<td></td>
<td>(4247435.7)</td>
<td>(5731874.82)</td>
<td>(5999860.9)</td>
<td>(5632039.15)</td>
</tr>
<tr>
<td>population 2000</td>
<td>6500855.7</td>
<td>5939002.2</td>
<td>6365436.98</td>
<td>7365499.56</td>
</tr>
<tr>
<td></td>
<td>(4995460.36)</td>
<td>(6692914.24)</td>
<td>(6578373.17)</td>
<td>(6476565.66)</td>
</tr>
<tr>
<td>unemployment 1990</td>
<td>5.32</td>
<td>5.43</td>
<td>5.55</td>
<td>5.34</td>
</tr>
<tr>
<td></td>
<td>(0.85)</td>
<td>(1.04)</td>
<td>(0.92)</td>
<td>(1.11)</td>
</tr>
<tr>
<td>unemployment 2000</td>
<td>3.67</td>
<td>3.9</td>
<td>4.04</td>
<td>3.68</td>
</tr>
<tr>
<td></td>
<td>(0.84)</td>
<td>(0.82)</td>
<td>(0.89)</td>
<td>(0.83)</td>
</tr>
<tr>
<td>cig sales per capita (1990)</td>
<td>106.95</td>
<td>99.91</td>
<td>105.16</td>
<td>100.05</td>
</tr>
<tr>
<td></td>
<td>(22.9)</td>
<td>(20.69)</td>
<td>(19.11)</td>
<td>(19.35)</td>
</tr>
<tr>
<td>cig sales per capita (2000)</td>
<td>86.46</td>
<td>80.59</td>
<td>90.94</td>
<td>80.33</td>
</tr>
<tr>
<td></td>
<td>(28.64)</td>
<td>(23.99)</td>
<td>(22.75)</td>
<td>(23.25)</td>
</tr>
<tr>
<td>cig sales per capita (2010)</td>
<td>51.58</td>
<td>51.25</td>
<td>58.36</td>
<td>49.61</td>
</tr>
<tr>
<td></td>
<td>(21.72)</td>
<td>(20.32)</td>
<td>(21.41)</td>
<td>(18.16)</td>
</tr>
<tr>
<td>smoking sentiment (1998)</td>
<td>-0.15</td>
<td>-0.09</td>
<td>-0.23</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.2)</td>
<td>(0.15)</td>
<td>(0.18)</td>
</tr>
<tr>
<td>% under bar ban</td>
<td>35.47</td>
<td>26.59</td>
<td>62.7</td>
<td>75.14</td>
</tr>
<tr>
<td></td>
<td>(37.11)</td>
<td>(33.99)</td>
<td>(45.9)</td>
<td>(39.55)</td>
</tr>
</tbody>
</table>

Columns 2-3 pertain to MMSAs with and without the tax change in BRFSS SMART 2004-2010 sample, while columns 4-5 refer to 2015-2020 sample. Both BRFSS SMART 2004-2010 and 2015-2020 are balanced panels with 108 and 95 MMSAs. The percent of population living under the bar ban is calculated using 2005 and 2010 population size for BRFSS SMART 2004-2010 and 2015-2020 samples, respectively.
Table 3: Cigarette tax dosage and pre-treatment variables

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>tax change dose</td>
<td>−45.76***</td>
</tr>
<tr>
<td></td>
<td>(6.22)</td>
</tr>
<tr>
<td>Observations</td>
<td>90</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01

Columns 1-4 and 5-8 refer to 2004-2010 and 2015-2020 samples. The specifications are conditional on units experiencing increases in cigarette tax within the sample period. The tax change dose represents the nominal increase in cigarette taxes. Columns 1-2 and 5-6 use per capita cigarette sales for 1990 and 2000 as the dependent variable. Columns 3 and 7 use the measure of anti-smoking sentiment in 1998, while Columns 4 and 8 use the percent of population living under bar ban in 2010.
Table 4: TWFE Estimates (BRFSS SMART 2004-2010 & 2015-2020)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A. % current smoker</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicator for tax increase</td>
<td>-0.5631**</td>
<td>-0.5612**</td>
<td>-0.6128***</td>
<td>-0.4824</td>
<td>-0.4592</td>
<td>-0.4105</td>
</tr>
<tr>
<td></td>
<td>(0.2115)</td>
<td>(0.2113)</td>
<td>(0.2117)</td>
<td>(0.3196)</td>
<td>(0.3230)</td>
<td>(0.3423)</td>
</tr>
<tr>
<td>Observations</td>
<td>756</td>
<td>756</td>
<td>756</td>
<td>570</td>
<td>570</td>
<td>570</td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>0.88960</td>
<td>0.88961</td>
<td>0.89378</td>
<td>0.86611</td>
<td>0.86629</td>
<td>0.87147</td>
</tr>
</tbody>
</table>

|                  | (1)     | (2)     | (3)     | (4)     | (5)     | (6)     |
| **Panel B. % current smoker (drop always treated)** |         |         |         |         |         |         |
| Indicator for tax increase | -0.6205*** | -0.6058*** | -0.6782*** | -0.4840 | -0.4675 | -0.3438 |
|                   | (0.2142) | (0.2148) | (0.2341) | (0.3260) | (0.3294) | (0.3471) |
| Observations      | 679     | 679     | 679     | 504     | 504     | 504     |
| **R²**            | 0.88855 | 0.88867 | 0.89270 | 0.87338 | 0.87346 | 0.87944 |

Year and MMSA FE ✓ ✓ ✓ ✓ ✓ ✓
Smoke Free air laws ✓ ✓ ✓ ✓ ✓ ✓
Additional Controls ✓ ✓ ✓ ✓ ✓ ✓
BRFSS SMART 2004-10 ✓ ✓ ✓ ✓ ✓ ✓
BRFSS SMART 2014-20 ✓ ✓ ✓ ✓ ✓ ✓

The table uses BRFSS MMSA SMART data for years: i) 2004 to 2010 in Columns (1)-(3), and ii) 2015-2020 in Columns (4)-(6). All columns include the locality level (MMSA) and year fixed effects. Columns (1) and (4) depict results from the parsimonious specification that includes a dummy for whether a state increased excise taxes on cigarettes at time $t$. Once this indicator turns on, it remains on for the rest of the years in the panel. Columns (2) and (5) control for the percent of state population living under smoking ban in bars as the Smoke Free Air Laws. Additionally, Columns (3) and (6) include a vector of pre-treatment variables (i.e. log of population measured in 2000 for 2004-2010 sample and in 2010 for 2015-2020 sample, ii. changes in the proportion of current smokers between 1998 and 2001, iii. the measure for anti-smoking sentiment in 1998, iv. unemployment rate in 2000 for 2004-2010 sample and in 2010 for 2015-2020 sample) interacted with year dummies. Panels A and B are structured in a similar way, except that Panel B drops units that fall under always treated group. The standard errors clustered at the state level are presented in parenthesis. *$p<0.1$; **$p<0.05$; ***$p<0.01$. 

37
The table provides the summary of Goodman Bacon decomposition of TWFE estimate as all possible $2 \times 2$ DiD estimates, summarized by groups mentioned in Column 1. Columns 2-3 and 4-5 refer to sample years 2004-2010 and 2015-2020. The weight column corresponds to weights given to the respective group when using estimates to form the TWFE estimate. For instance, the sum of weight (04-10)$ \times$ avg. estimate (04-10) equals to TWFE estimate in Column 1, Table 4 (Panel A).

<table>
<thead>
<tr>
<th>Type</th>
<th>weight (04-10)</th>
<th>avg. estimate (04-10)</th>
<th>weight (15-20)</th>
<th>avg. estimate (15-20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Earlier vs Later Treated</td>
<td>0.218</td>
<td>-0.898</td>
<td>0.029</td>
<td>-0.048</td>
</tr>
<tr>
<td>2 Later vs Always Treated</td>
<td>0.177</td>
<td>-0.296</td>
<td>0.13</td>
<td>-0.471</td>
</tr>
<tr>
<td>3 Later vs Earlier Treated</td>
<td>0.316</td>
<td>-0.233</td>
<td>0.047</td>
<td>0.354</td>
</tr>
<tr>
<td>4 Treated vs Untreated</td>
<td>0.29</td>
<td>-0.834</td>
<td>0.794</td>
<td>-0.549</td>
</tr>
</tbody>
</table>
8 References


Mullahy, John. 1985. *Cigarette Smoking: Habits, Health Concerns, and Heterogeneous Unob-
servables in a Microeconometric Analysis of Consumer Demand. University of Virginia.


9 Appendix


Not all of the MMSA from BRFSS SMART data are represented in CPS tobacco supplement files. In fact 39 MMSAs in 2004-2010 BRFSS SMART balanced panel data cannot be mapped to MSA/PMSA codes in CPS files. In light using all of the observations in the balanced panel, I impute the missing values on: \(i\) the measure of anti-smoking sentiments in 1998, and \(ii\) the proportion of current smokers in 1998 and 2001. First, the non-missing values on the outcomes are regressed on locality specific wage index and the log of population as well as state specific variables including the log of cigarette sales in 2000 and the measure of anti smoking sentiments in 1998. Next, the estimates from the regression are then use to construct predicted values of the outcomes. The missing values on the outcomes are replaced by the predicted values.
Figure 8: Anti-smoking sentiments and smoking
Note: The sub-figures show the relationship between the measure of anti-smoking sentiments in 1998 constructed by using the principal factor analysis and smoking-related variables. Panel A uses the proportion of individuals having reported ever smoked a cigarette (100 cigarettes or more), panel B uses the proportion of current smokers, and panel C uses the change in the proportion of current smokers between 1998 and 2001. The data comes from the Current Population Survey (Tobacco Supplement) files and is aggregated at the MMSA level.
Figure 9: CS Event-Study-Type Estimates (never treated and not-yet-treated as comparison)

Note: The figure is structured similar to Figure 5 except that the panel only consist of units that are exposed to the treatment for atleast 3 years following the treatment year. The red line represents the point estimate of the overall average treatment effect on the treated using CS approach.
Figure 10: CS Event-Study-Type Estimates (never treated and not-yet-treated as comparison)

The figure is structured similar to Figure 6 except that the panel only consist of units that are exposed to the treatment for at least 2 years following the treatment year. The red line represents the point estimate of the overall average treatment effect on the treated using CS approach.